

Trade, FDI, and the Organization of Firms

ELHANAN HELPMAN*

New developments in the world economy have triggered research designed to better understand the changes in trade and investment patterns, and the reorganization of production across national borders. Although traditional trade theory has much to offer in explaining parts of this puzzle, other parts required new approaches. Particularly acute has been the need to model alternative forms of involvement of business firms in foreign activities because organizational change has been central in the transformation of the world economy. This paper reviews the literature that has emerged from these efforts. The theoretical refinements have focused on the individual firm, studying its choices in response to its own characteristics, the nature of the industry in which it operates, and the opportunities afforded by foreign trade and investment. Important among these choices are organizational features, such as sourcing strategies. But the theory has gone beyond the individual firm, studying the implications of firm behavior for the structure of industries. It provides new explanations for trade structure and patterns of foreign direct investment, both within and across industries, and has identified new sources of comparative advantage.

1. Introduction

International trade and foreign direct investment (FDI) have been among the fastest growing economic activities around

the world. In 2003, world merchandise exports were close to 7.3 trillion dollars; world exports of commercial services were close to 1.8 trillion dollars; and world FDI inflows were close to 560 billion dollars.¹ However, between 1990 and 2001 sales by foreign affiliates of multinational corporations

* Harvard University, Tel Aviv University, and CIAR. I thank the National Science Foundation and the U.S.–Israel Binational Science Foundation for financial support. Parts of this paper were used in my European Investment Bank Lecture, the European University Institute, Florence. I am grateful to the Department of Economics at the European University Institute for its hospitality. Much of the work for this paper was done when I was Sackler Visiting Professor at Tel Aviv University. Pol Antràs, Roger Gordon, Gene Grossman, Marc Melitz, Stephen Redding, and three referees provided useful comments.

¹ FDI inflows reached a peak of 1.4 trillion dollars in 2000, but declined from 2000 to 2003; see UNCTAD (2004). According to UNCTAD (2002), foreign affiliates of multinational corporations accounted for 11 percent of world GDP and 35 percent of world trade in 2001. In the 1990s, merchandise exports grew at an annual rate of 6.4 percent in real terms while merchandise production grew at an annual rate of 2.5 percent only (see World Trade Organization 2004).

expanded much faster than exports of goods and nonfactor services.² A striking feature of this growth has been an unprecedented expansion of FDI in services; the inward stock of FDI in services increased from 950 billion dollars in 1990 to 4 trillion in 2002.³ In 2001–02, services accounted for two-thirds of FDI inflows.

These remarkable figures mask equally remarkable changes in the nature of trade and FDI flows. The fast expansion of trade in services has been accompanied by fast-growing trade in intermediate inputs.⁴ Moreover, the growth of input trade has taken place both within and across the boundaries of the firm, i.e., as intrafirm and arm's-length trade.⁵ In the United States, the latter has grown particularly fast. Many studies have documented the growth of international vertical specialization, as reflected in the flows of inputs across national borders for further processing and final assembly.⁶ These trends are closely related to the growing fragmentation of production, in which multinational corporations play a central role. Technological change, such as computer-aided design and computer-aided manufacturing, contributed to this process. The

same technological changes also contributed to growing outsourcing within and across national borders.⁷

In addition to these broad trends, new data sets enable researchers to uncover previously unobserved patterns of trade and FDI flows. Especially important is the finding that a systematic relationship exists between the characteristics of business firms and their participation in foreign trade and investment. Exporting firms are not a random sample of the population of firms in an industry, and neither are firms engaged in FDI. Only a small fraction of firms export, they are larger and more productive than firms that serve only the domestic market, and more firms export to larger markets.⁸ A small fraction of firms engage in FDI, and these firms are larger and more productive than exporting firms. A lot of within-industry heterogeneity exists, and the distribution of firms by size or productivity varies substantially across industries.⁹

Sourcing strategies of business firms have become more complex than ever before, and so have the integration strategies of multinational corporations.¹⁰ As a result, the traditional classification of FDI into vertical and horizontal forms has become less meaningful in practice. Large multinationals invest in low-cost countries to create export platforms from which they serve other countries around the world, and the large flows of FDI across industrial

² According to UNCTAD (2002), by almost 7 percent per year.

³ See UNCTAD (2004).

⁴ See Alexander J. Yeats (2001).

⁵ See Robert C. Feenstra (1998) and Maria Borgia and William J. Zeile (2004). According to Borgia and Zeile (2004), exports of U.S. parent companies to their foreign affiliates for further processing have increased from 8.5 percent of total U.S. exports of goods in 1966 to 14.7 percent in 1999, and from 39.3 percent of total exports of goods by U.S. parents to their foreign affiliates in 1966 to 64.7 percent in 1999. These shares vary substantially across industries; they are particularly large in electronic and other electric equipment as well as in transportation equipment, and particularly small in petroleum manufacturing as well as in food and kindred products.

⁶ See, for example, Jose Campa and Linda S. Goldberg (1997) for the United States, United Kingdom, and Canada; Vanessa Strauss-Kahn (2003) for France; and Hummels, Rappoport, and Yi (1998) and Hummels, Jun Ishii, and Yi (2001) for other OECD countries.

⁷ See Katharine G. Abraham and Susan K. Taylor (1996) and Ann P. Bartel, Saul Lach, and Nachum Sicherman (2005) on outsourcing trends in the United States.

⁸ See Jonathan Eaton, Samuel Kortum, and Francis Kramarz (2004). They report that only 17.4 percent of French firms in manufacturing industries export, and they export 21.6 percent of the aggregate manufacturing output. These numbers hide large variations across industries, however. In food and tobacco industries, for example, only 5.5 percent of the firms export, while in chemicals 55.4 percent of the firms export.

⁹ See Andrew B. Bernard, J. Bradford Jensen, and Peter K. Schott (2005) for a portrait of U.S. firms.

¹⁰ See UNCTAD (1998).

countries cannot be satisfactorily classified as horizontal FDI.¹¹

New theories have been developed to explain these changes. While the new theories do not replace comparative advantage explanations of intersectoral trade and FDI flows or replace imperfect competition explanations of intraindustry trade, they do bring to trade theory a new focus: the organizational choices of individual firms. By focusing on the characteristics of individual firms, the theory can address new questions: Which firms serve foreign markets? And how do they serve them, i.e., which choose to export and which choose to serve foreign markets via FDI? How do they choose to organize production, do they outsource or integrate? Under what circumstances do they outsource in a foreign country rather than at home? And if they choose integration, under what circumstances do they choose to integrate in a foreign country, via FDI, rather than to integrate at home?¹²

I discuss this literature in two sections. Section 2 examines insights from models of heterogeneous firms in which the internalization decision, i.e., outsourcing versus integration, is put aside. This proves to be a useful simplification because the resulting predictions go a long way toward explaining why firms sort into exclusive domestic producers, exporters, or foreign direct investors, and the structure of complex integration strategies. Naturally, these models cannot explain why some firms outsource while others integrate. This issue is taken up in

section 3, which examines the implications of the theory of incomplete contracts for internalization and offshoring decisions. The result is a trade theory with rich sourcing patterns.¹³

Various studies emphasize different trade-offs in the decision to internalize or offshore, and no model integrates all considerations into a single framework. But the studies discussed in section 3 all build on a common assumption, namely that some inputs are highly specific to a final product and that their supply is not fully contractible. This assumption is enough to study (1) the impact of variations across industries in the intensity of inputs that suffer from agency problems; (2) Ricardian-type comparative advantage that arises when legal systems of different quality interact with sectoral differences in contract dependency; (3) the impact of different degrees of contract incompleteness, which may vary across countries; (4) the role of matching between buyers and sellers of intermediate inputs, and the resulting “thick market” effect; and (5) the interaction between within-industry heterogeneity with incomplete contracts, which yields joint predictions about internalization and offshoring. In particular, it predicts the relative prevalence of the four main organizational forms: integration at home, outsourcing at home, integration abroad, and outsourcing abroad.

While the main purpose of this article is to review the theoretical literature, I report empirical evidence wherever possible. The interplay between theory and empirics is

¹¹ See Karolina Ekholm, Rikard Forslid, and James R. Markusen (2004) and Susan E. Feinberg and Michael P. Keane (2003). See also section 2.5 for more details.

¹² I attach traditional meanings to the terms “outsourcing” and “integration.” That is, outsourcing means the acquisition of an intermediate input or service from an unaffiliated supplier, while integration means production of the intermediate input or service within the boundary of the firm. These choices are distinct from the choice of country in which to engage in these activities, because outsourcing can be carried out in the home country of the firm, or in any number of foreign countries, and similarly for integration.

¹³ Some of the issues examined in section 3 are discussed in Barbara J. Spencer (2005). I have chosen to focus on incomplete contracts, thereby not covering the work on managerial incentives, such as Gene M. Grossman and Elhanan Helpman (2004) and Dalia Marin and Thierry Verdier (2005). The reason for this choice is that there is a lot of common ground in the approaches reviewed in section 3, while the papers on managerial incentives are somewhat idiosyncratic. I also do not review earlier work on incomplete contracts, such as Spencer and Larry D. Qiu (2001) and Qiu and Spencer (2002), which have a narrow focus, such as Keiretsu-type organizations, and have no obvious implications for the broader issues discussed in the introduction.

particularly important here because many of these theoretical studies have been motivated by evidence. As one would expect, the theoretical models deliver new empirical implications that can be confronted with data. I report empirical studies that do that, but other empirical implications have not yet been tested. Some will undoubtedly be tested in the near future, while others will have to wait because they require data that are not yet available. These issues are discussed in the closing section of the paper.

2. *Heterogeneous Productivity*

In the 1980s, trade theory introduced within-industry heterogeneity resulting from product differentiation and monopolistic competition. Heterogeneity in these studies was not designed, however, to explain asymmetries across firms in productivity or size. Not because it was not known at the time that firms differ along these dimensions, but rather because the aim was to explain large volumes of trade between countries with similar factor compositions and large volumes of intraindustry trade. For this purpose, differences in productivity or size were not considered to be important. As a result, the models assumed (for the most part) symmetry across firms within an industry in terms of the available technology, which implied in turn similar productivity levels and similar participation in foreign trade. The monopolistic competition models implied that all firms export to all countries unless there is pressure for the formation of multinational corporations.¹⁴

¹⁴ See Helpman and Paul R. Krugman (1985, chapters 7 and 12). Differential incipient pressure on factor prices across countries can lead to the formation of multinational corporations despite the prevalence of factor price equalization. Under these circumstances, some firms become multinationals while others do not. This produces asymmetries in the organizational forms of different firms in the same industry, and different trading patterns, but these firms do not differ in productivity or size.

Detailed empirical studies of exporting firms have led to a recognition of the limitations of the symmetry assumption. As new firm-level data became available, it became clear that not all firms within an industry export, nor are exporting firms a random sample of the population of firms in an industry. This evidence accumulated in the 1990s and showed that only a small fraction of firms export and that exporters are larger and more productive than nonexporters.¹⁵ Eaton, Kortum, and Kramarz (2004) find, for example, that in the mid-1980s only 17.4 percent of French firms in manufacturing industries exported, that they exported only 21.6 percent of their output, and that both averages hide wide variations across industries. And Helpman, Marc J. Melitz, and Stephen Ross Yeaple (2004) report that, in a large 1996 sample of U.S. firms, exporters had a 39 percent labor productivity advantage over nonexporters.¹⁶ Finally, there appear to exist large sunk costs of exporting in developed and developing countries alike.¹⁷

In view of these findings, Melitz (2003) developed a theoretical model of monopolistic competition with heterogeneous firms that

¹⁵ See Sofronis K. Clerides, Lach, and James R. Tybout (1998) for Colombia, Mexico, and Morocco; Bernard and Jensen (1999) for the United States; Bee-Yan Aw, Sukkyun Chung, and Mark J. Roberts (2000) for Taiwan; Miguel A. Delgado, Jose C. Fariñas, and Sonia Ruano (2002) for Spain; and John R. Baldwin and Wulong Gu (2003) for Canada.

¹⁶ A detailed account of the characteristics of U.S. firms that trade in goods is provided by Bernard, Jensen, and Schott (2005). In their data too only a small fraction of firms export and they export a small fraction of their own output.

¹⁷ See Roberts and Tybout (1997) for Colombia and Bernard and Jensen (2004) for the United States. While these studies only report large persistence in exporting status, Sanghamitra Das, Roberts, and Tybout (2005) use a structural model to estimate the size of sunk exporting costs for Colombian firms in three industries. They find that the sunk costs for small producers are between \$412,000 and \$430,000, and for large producers between \$344,000 and \$402,000. Moreover, they find that fixed exporting costs are important for at least some of these firms.

was designed to explain these features of the data.¹⁸ His model has become the cornerstone of a growing literature that examines the role of heterogeneity in international trade and foreign direct investment.¹⁹ The success of Melitz's model derives from the fact that, when combined with old and new approaches to trade theory, it yields rich predictions that can be confronted with data, and so far the model has performed admirably well.

The main insights from Melitz's model are derived from an interaction between productivity differences across firms and fixed costs of exporting. The fixed export costs are interpreted as distribution and servicing costs in foreign markets, and a firm has to bear them in every country to which it exports. As a result, the total fixed export costs are larger the more foreign countries the firm chooses to serve.²⁰

To illustrate the nature of these interactions, consider an industry supplying a differentiated product, in which each of a continuum of firms manufactures a different brand. The demand function for firm j 's brand is $x(j) = Ap(j)^{-\varepsilon}$, where x is the quantity and p is the price, A is a measure of the demand level, and $\varepsilon \equiv 1/(1 - \alpha)$ is the demand elasticity. The demand elasticity is

¹⁸ Other, related models of this type are developed in Catia Montagna (2001) and Sébastien Jean (2002). In addition, Bernard, Eaton, Jensen, and Kortum (2003) propose a model of heterogeneous firms with a different market structure (i.e., Bertrand competition instead of monopolistic competition) in order to address similar questions. I focus on Melitz (2003) because his model has proved to be most adaptable to a wide range of applications, including integration with the literature on incomplete contracts and the international organization of production (see section 3). Richard E. Baldwin (2005) provides an alternative discussion of this model.

¹⁹ Melitz (2003) builds on the work of Hugo A. Hopenhayn (1992), who studied the entry and exit dynamics of firms in an industry.

²⁰ Earlier studies, including Baldwin (1988), Baldwin and Krugman (1989), Dixit (1989), and Roberts and Tybout (1997), used sunk costs of exporting, yet only Roberts and Tybout touch upon some of the issues addressed by Melitz. Their model, which is designed to estimate the impact of sunk costs on export decisions, is not as useful, however, for dealing with the wide range of issues to which Melitz's model has been applied.

assumed to be constant, with $0 < \alpha < 1$, which implies $\varepsilon > 1$.²¹ Although the demand level A is endogenous to the industry, it is treated as exogenous by producers because every producer is of negligible size relative to the size of the industry.

Firm j discovers its productivity $\theta(j)$ only after it enters the industry. Let $c/\theta(j)$ be its variable production cost per unit of output and let cf_D be its fixed cost, where c measures the cost of resources (e.g., the wage rate when there is only labor input); and f_D is a measure of fixed production costs in terms of resources. Then, if the firm chooses to sell the product, its profit-maximizing strategy is to charge $p(j) = c/\alpha\theta(j)$, which yields the operating profits $\pi(j) = \theta(j)^{\varepsilon-1}B - cf_D$, where $B \equiv (1 - \alpha)A(c/\alpha)^{1-\varepsilon}$.

Figure 1 depicts these profits as a function of the productivity measure $\Theta \equiv \theta^{\varepsilon-1}$. The firm index j is dropped because profits do not depend on the identity of the firm, only on its productivity level; firms with higher productivity have higher profits. The profit function in the figure is:

$$(1) \quad \pi_D(\Theta) = \Theta B - cf_D.$$

As is evident from the figure, firms with productivity levels below Θ_D choose not to produce because, for these firms, variable profits do not cover their fixed cost, while firms with higher productivity supply their brands to the market. Given a productivity distribution $G(\Theta)$, we can calculate the fraction of firms that serve the domestic market as the fraction of firms with productivity above the cutoff Θ_D .

2.1 Export

Now interpret the profit function $\pi_D(\Theta)$ as applying to sales in the domestic market, so that A is the demand level in the domestic

²¹ As is well known, this form of demand function can be derived from a constant elasticity of substitution (CES) utility or production function. In this event $A = E/\int_{j \in J} p(j)^{1-\varepsilon} dj$, where E is total spending on these products and J is the set of available brands.

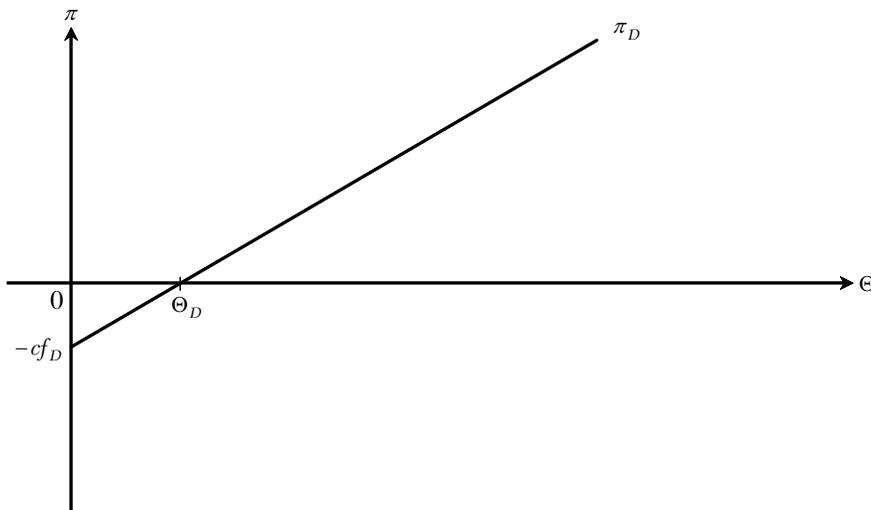


Figure 1. Producing and Nonproducing Firms

market. And assume that firms can sell their products in country ℓ as well, which has the demand function $x(j) = A^\ell p(j)^{-\varepsilon}$. That is, the demand elasticity is the same in the two markets but the demand level is not necessarily the same at home as in country ℓ . In addition, there are melting iceberg trading costs for the shipment of every brand of the product from home to ℓ , such that $\tau > 1$ units have to be shipped for one unit to arrive, and there are fixed export costs cf_X . The variable trading costs typically include transport costs, insurance, fees, duties, and other impediments that may stem from language barriers, differences in the legal systems, and the like.²² Under these circumstances, a firm that chooses to sell in the domestic market, i.e., one with productivity $\Theta > \Theta_D$, can make additional profits

$$(2) \quad \pi_X^\ell(\Theta) = \tau^{1-\varepsilon} \Theta B^\ell - cf_X$$

from export sales, where $B^\ell \equiv (1 - \alpha)A^\ell(c/\alpha)^{1-\varepsilon}$.

Figure 2 depicts both $\pi_D(\Theta)$ and $\pi_X^\ell(\Theta)$, for the case in which $A^\ell = A$ (thus $B^\ell = B$) and $\tau^{\varepsilon-1}f_X > f_D$. When the two demand levels are the same, π_D is steeper than π_X^ℓ as a result of the trading costs, and the assumption on the relative size of the fixed costs then ensures $\Theta_X^\ell > \Theta_D$. It follows that low-productivity firms, with $\Theta < \Theta_D$, still choose to close down, because they lose money from domestic sales as well as from exporting, while firms with productivity above Θ_D make money from serving the domestic market. Now, however, high-productivity firms, with $\Theta > \Theta_X^\ell$, also make money from exporting. Such firms choose, therefore, to serve the domestic market as well as the market in ℓ . Firms with intermediate productivity levels, between Θ_D and Θ_X^ℓ , attain the highest profits by serving the domestic market only, i.e., they

²² See James E. Anderson and Eric van Wincoop (2004) for estimates of the size of these costs.

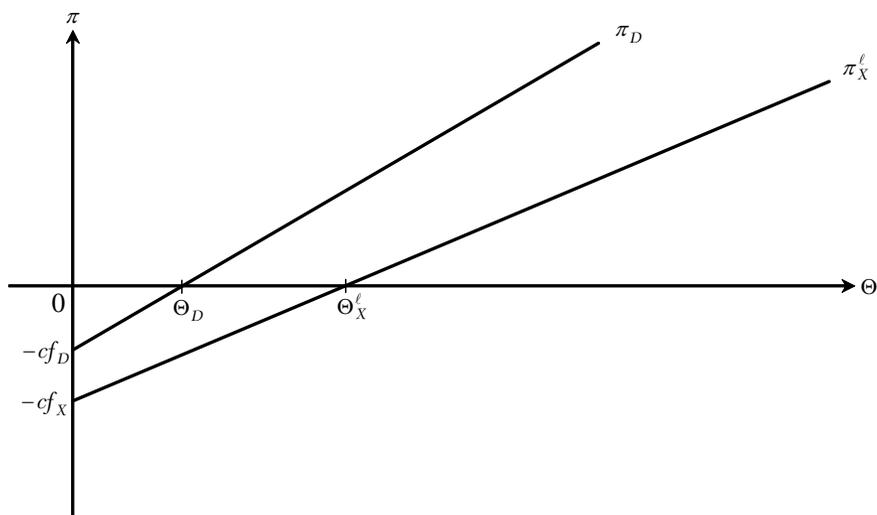


Figure 2. Exporting and Nonexporting Firms

choose not to export. The sorting pattern depicted in this figure implies that exporting firms are more productive than nonexporters and that they are bigger. The last implication follows from the fact that more-productive firms sell more in the domestic market and they sell in the foreign country as well. Evidently, this model's predictions are consistent with the data, in which exporters are larger and more productive than nonexporters.

Next observe that we can add as many profit functions from exporting as there are foreign countries ℓ . Assuming that the foreign countries differ only in market size, A^ℓ , would then imply a negative correlation between market size and the export cutoff Θ_X^ℓ . That is, the smaller the foreign country ℓ the larger its cutoff Θ_X^ℓ . For simplicity, suppose that $\min_\ell \Theta_X^\ell > \Theta_D$.²³ In this event, all

exporting firms sell in the domestic market too, and there exist firms, with productivity between Θ_D and $\min_\ell \Theta_X^\ell$, which serve the domestic market but do not export. All firms with productivity levels above $\min_\ell \Theta_X^\ell$ export. In this multicountry world, the positive correlation between productivity and export status is preserved. In addition, we obtain a new prediction which is consistent with the data: there exists a positive correlation between the size of an export market and the number of firms that export to it.²⁴ Naturally, this correlation may not hold when the trading cost τ is not the same with every foreign country. Nevertheless, it should still hold once we control for the cross-country variation in trading costs.

2.2 Turnover

I described a static version of Melitz's (2003) model. This is sufficient for the issues

²³ This should be true in big countries but may not be true in small countries. In any case, the analysis can be carried out without this assumption.

²⁴ See the evidence in Eaton, Kortum, and Kramarz (2004) for French firms.

discussed above as well as for a number of other issues to be discussed below. Yet, the original formulation of the model is dynamic, shedding light on entry, exit, and turnover of firms. In the dynamic version of the model, the fixed production and export costs f_D and f_X have to be borne every period. There also exists an entry cost f_E that is a capital cost; it has to be borne only once, at entry. Moreover, there is a constant probability of death δ_e of every firm, irrespective of its productivity. In this setting, free entry requires the expected present value of profits to equal the entry cost. In a steady state, firms constantly leave the industry, as a fraction δ_e die every period. At the same time, there is a constant inflow of new firms, and a fraction of these firms—those whose productivity is above the cutoff Θ_D —remain in the industry. In the steady state equilibrium, the inflow equals the outflow, so that the number of firms remains constant in every productivity category. As a result, the ratio of new entrants per period to the stock of active firms, a measure of turnover, equals $\delta_e/[1 - G(\Theta_D)]$, where $G(\cdot)$ is the cumulative distribution of Θ .²⁵ This setup can be used to study the determinants of turnover, which I illustrate in the next section with a discussion of trade liberalization.

2.3 Trade Liberalization

Consider multilateral trade liberalization, which leads to a proportional reduction of trading costs τ in all countries. On impact, this reduction in trading costs raises the profits of exporters and reduces the cutoff Θ_X^ℓ . As a result, a larger proportion of firms choose to export. But the presence of a larger number of exporters in a market reduces the demand facing every supplier, which cuts into the profits of exporters and nonexporters alike.²⁶ After allowing the general

equilibrium effects to work themselves out, the final outcome is a lower export cutoff Θ_X^ℓ (although not as low as one would predict from the impact effect) and a higher domestic cutoff Θ_D . It follows that trade liberalization leads to higher average productivity, since only the more-productive firms survive entry, and output is reallocated toward more productive firms. These are interesting implications, which illustrate important issues that this model can address, and which could not be addressed by earlier models of international trade. Moreover, Daniel Trefler (2004) finds that both of these predictions are consistent with the impact of the Canada–U.S. Free Trade Agreement on Canadian industries.²⁷

2.4 Horizontal FDI

Melitz's (2003) model can be generalized to handle horizontal foreign direct investment. The traditional classification of FDI has been into horizontal and vertical FDI, where the former concerns subsidiaries that serve the local market in the host country while the latter concerns subsidiaries that add value to products that are not destined (necessarily) for the host country market (more on this in the next section).²⁸ Following Helpman, Melitz, and Yeaple (2004), suppose that a home-country firm can build a (second) production facility in country ℓ , at cost cf_I , that will enable it to produce its brand of the product in country ℓ at unit cost c^ℓ/θ , where θ is the firm's productivity. Then if the firm exports to country ℓ , its profits from exporting are given by (2), while if it chooses to serve the foreign market via FDI, the firm's profits from FDI are

$$(3) \quad \pi_I^\ell(\Theta) = \Theta B_I^\ell - cf_I,$$

²⁷ Tybout and M. Daniel Westbrook (1995) also find important market share reallocations from low to high productivity plants in response to trade liberalization in Mexico.

²⁸ According to the BEA data, the destination of sales of U.S. subsidiaries are distributed as follows: 65 percent in the host country markets, 11 percent in the United States, and 24 percent in other countries (see J. Steven Landefeld and Raymond Mataloni 2004).

²⁵ Let N be the stock of active firms and let n_E be the flow of new entrants per period. Then $[1 - G(\Theta_D)]n_E$ is the inflow of active firms and δN is the outflow. In steady state the two are equal. Therefore $n_E/N = \delta/[1 - G(\Theta_D)]$.

²⁶ See the determinants of the demand level A in footnote 21.

where $B_I^\ell \equiv (1 - \alpha)A^\ell(c^\ell/\alpha)^{1-\varepsilon}$. Comparing (2) with (3) we note that, as long as $f_I > f_X$ and $c^\ell < c\tau$, the firm faces a proximity-concentration trade-off, for which Brainard (1997) provides empirical evidence. Namely, by choosing FDI instead of exporting the firm gives up concentration of production, which raises its fixed costs, but saves on variable unit costs by avoiding trade costs (and possibly on unit production costs). Figure 3 describes this trade-off for the case in which $c^\ell = c$, $B_I^\ell = B^\ell$, (i.e., the demand level is the same in the two countries), and $f_I > \tau^{\varepsilon-1}f_X > f_D$. Under these circumstances, $\Theta_I^\ell > \Theta_X^\ell > \Theta_D$. It follows that the most productive firms, with $\Theta > \Theta_I^\ell$, serve the foreign market via subsidiary sales; lower productivity firms, with $\Theta_X^\ell < \Theta < \Theta_I^\ell$, serve the foreign market via export; and still lower productivity firms, with $\Theta_D < \Theta < \Theta_X^\ell$, serve only the domestic market. Evidently, this sorting pattern is consistent with the empirical evidence that multinational corporations are more productive than exporters who are not multinationals, and exporters who are not multinationals are more productive than firms who serve only the domestic market.²⁹ Since more productive firms produce more output, this sorting pattern also implies that multinational firms are larger than exporters, and exporters are larger than firms who serve only the domestic market.

Helpman, Melitz, and Yeaple (2004) also show that when the distribution of productivity θ is characterized by a Pareto distribution, the size distribution of firms also is Pareto, and the model then predicts more

subsidiary sales relative to export sales in sectors with greater productivity (and therefore size) dispersion. This is a particularly interesting implication because it suggests that heterogeneity can be a source of comparative advantage. The use of a Pareto distribution is compelling in this case because the actual size distribution of firms is well approximated by such a distribution (see Robert L. Axtell 2001). Helpman, Melitz, and Yeaple also show that the shape parameter of a Pareto distribution can be precisely estimated in almost every one of fifty-two sectors for which they have data, and these estimates exhibit large variations in the degree of dispersion across sectors. Using these measures of dispersion, as well as non-parametric measures, they estimate the impact of heterogeneity on the ratio of subsidiary sales to export sales of U.S. firms in a sample of twenty-seven countries, and a broader sample of thirty-eight countries, both in 1994. Their estimates, which control for the variation in fixed costs and other relevant variables, are precise and consistent with the theory. Moreover, the estimates are large economically; they compare in size to the impact of freight, tariffs, and measures of fixed costs on the ratio of export to subsidiary sales, which have been routinely used in studies of the proximity-concentration trade-off.³⁰

2.5 Technology Adoption

Paula Bustos (2005) introduces a technology choice into Melitz's (2003) model in order to study the impact of trade liberalization on technology upgrading in Argentina. For this purpose, suppose that a firm located in Argentina can serve the domestic market or it can serve the domestic market and also export to a foreign market (we disregard the FDI option). But unlike the Melitz model, now, upon entry, and after learning its productivity θ , the firm can choose to use an advanced

²⁹ Helpman, Melitz, and Yeaple (2004) find that, in 1996, U.S. firms that engaged in FDI had a 15 percent labor productivity advantage over exporters who did not engage in FDI, and the latter had a 39 percent labor productivity advantage over firms who engaged in neither export nor FDI. See also Keith Head and John Ries (2003) for evidence from Japan; Sourafel Girma, Holger Görg, and Eric Strobl (2004) for evidence from Ireland; and Girma, Richard Kneller, and Mauro Pisu (2005) for evidence from the United Kingdom.

³⁰ The comparability in size is of beta, or standardized, coefficients.

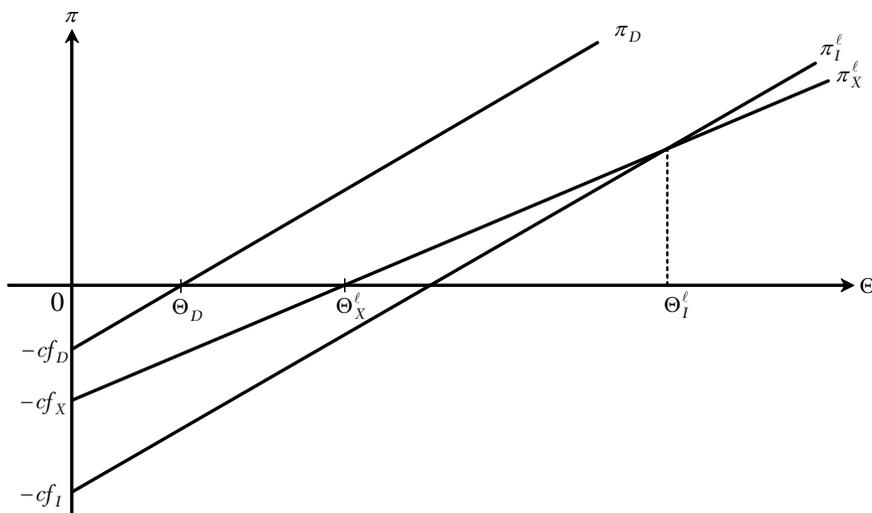


Figure 3. Multinationals, Exporting, and Nonexporting Firms

technology H or a traditional technology L , as in Yeaple (2005). The advanced technology requires higher fixed costs, so that $f_{DH} > f_{DL}$ and $f_{XH} > f_{XL}$, but its advantage is that it has lower variable costs, so that variable unit costs are ca_H/θ if the firm uses technology H and ca_L/θ if the firm uses technology L , $a_H < a_L$. With suitable restrictions on these parameters, a firm with productivity below a cutoff Θ_D exits the industry because all choices afford it negative operating profits, a firm with productivity between Θ_D and Θ_{XL} uses technology L to serve the domestic market only, a firm with productivity between Θ_{XL} and Θ_{XH} uses technology L to serve the domestic market and to export, and a firm with productivity above Θ_{XH} uses technology H to serve the domestic market and to export. Naturally, $\Theta_{XH} > \Theta_{XL} > \Theta_D$. In other words, more productive firms use the more advanced technology but some low-productivity exporters use the traditional technology. This sorting pattern is consistent with Bustos's data.

Now consider multilateral trade liberalization, which reduces trading costs to

Argentinian firms. This raises the operating profits of all exporters, but proportionately more so from the use of the advanced technology if an exporter's productivity is close to Θ_{XH} . As a result, Θ_{XH} declines, and some exporters who used technology L switch to H , while exporters who used the better technology have no incentive to switch to technology L . Firms that serve only the domestic market also have no incentive to switch technologies, and they keep using technology L . The model therefore predicts that only firms with intermediate productivity levels upgrade their technology in response to trade liberalization. And indeed, Bustos finds an inverted U shaped relationship between productivity and technology upgrading in Argentina.³¹

³¹ Bustos (2005) also examines skill upgrading, which is positively correlated with the upgrading of technology. She finds that of the 17 percent rise in the demand for skilled workers after trade liberalization, 15 percent took place within firms in each of three skill categories: production, nonproduction, and R&D workers. Since she has data on the education level of workers within each one of these skill categories, she measures skill upgrading as the rise in average years of schooling.

2.6 Complex Integration Strategies

Although horizontal FDI of the type described in the previous section is prevalent, the evidence points to a growing importance of more complex integration strategies by multinational corporations. Feinberg and Keane (2003) find, for example, that among U.S. multinationals with affiliates in Canada, only 12 percent are of the purely horizontal type (i.e., they have negligible intrafirm flows of intermediate inputs) and only 19 percent are of the purely vertical type (i.e., they have negligible intrafirm flows of intermediate inputs in one direction only). The remaining 69 percent of the firms pursue more complex integration strategies.³² Yeaple (2003) provides the first analysis of such complex strategies, identifying an important complementarity between the two types of FDI. In what follows, I briefly discuss insights from Grossman, Helpman, and Adam Szeidl (forthcoming) who combine heterogeneity features from Melitz (2003) with the modelling of the two types of FDI from Yeaple (2003) in order to explore patterns of FDI in an environment that offers a rich choice of integration strategies.

The model has a simple structure. There are two symmetric countries in the North and one country in the South. Every Northern country has a population of firms who know how to produce varieties of a differentiated product. A typical firm has a production function $\theta F(m, a)$, where θ is (as before) a firm-specific productivity level and $F(\cdot)$ is a concave constant-returns-to-scale production function, common to all firms; m represents intermediate inputs and a represents assembly. That is, every final good is produced with a combination of intermediate inputs and assembly. The elasticity of substitution between m and a is smaller than one. And in this model there are no fixed manufacturing costs f_D nor fixed exporting costs f_X .

³² See also UNCTAD (1998), where the term “complex integration strategies” was coined.

Intermediate inputs and assembly are produced from a bundle of primary inputs at cost c per unit, where c is higher in the North than in the South. As a result, there is a cost advantage to locating these activities in South, unless other costs enter the calculus. To introduce a tradeoff in the location decision, it is assumed that no fixed costs are borne by a firm that locates both activities in the Northern country in which it is headquartered, but that such a firm has to bear a fixed cost g if it locates the production of intermediates in a different country and a fixed cost f if it locates assembly in a different country. The firm may also incur transport costs for either intermediate inputs or final goods. In combination, this cost structure induces a nontrivial decision problem in which the optimal integration strategy depends on these cost parameters as well as on the demand levels in the three countries. The demand function is $Ap(j)^{-\varepsilon}$ (as before), and A is higher in a Northern country than in South.

First consider the case in which there are no transport costs. Then, given the fixed cost f of FDI in assembly, there are four integration strategies that may be chosen by a firm in equilibrium, depending on the fixed cost of FDI in intermediates g and the firm’s productivity. They are depicted in figure 4. Region $\{S, H\}$ describes a strategy whereby the firm manufactures intermediates in South and assembles final goods in the home country, i.e., the country in which the firm is headquartered. The other regions have similar interpretations; the first letter denotes the location of intermediate inputs while the second letter denotes the location of assembly. The fixed cost of FDI in intermediate inputs varies along the vertical axis while the productivity measure $\Theta \equiv \theta^{\varepsilon-1}$ varies along the horizontal axis.

We see that for low fixed costs g the least-productive firms perform both activities at home, intermediate-productivity firms produce intermediates in South and assemble final goods at home, and high-productivity

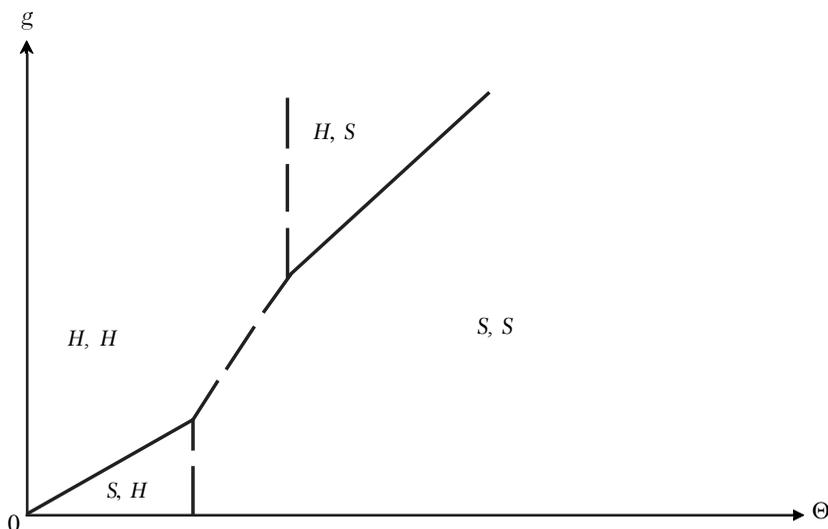


Figure 4. Optimal Integration Strategies

firms perform both activities in South. That is, the least-productive firms do not engage in FDI; they produce intermediates and perform assembly in the home country and export the final product to the other Northern country and to South. Firms with intermediate productivity engage in partial FDI; they produce intermediate inputs in South, import them to the home country, assemble them there into a final product, and then export the final product to the other Northern country and to South. Finally, the most-productive firms engage in FDI to the greatest possible extent; they produce intermediate inputs in South and assemble there the final product. The final product is then exported to the two Northern countries, i.e., the South serves as an export platform to the rest of the world.³³

The figure also shows that for an intermediate range of FDI costs g there are only two optimal integration strategies; low productivity firms do everything at home while high

productivity firms do everything in South. Finally, for high values of g , low productivity firms do everything at home, the highest productivity firms do everything in South, and firms in between produce intermediates in the home country and assemble final goods in South.

It is also clear from the figure that, given a distribution of Θ , the fraction of firms that do both activities at home is rising with g while the fraction of firms that do both activities in South is declining with g . Moreover, as shown by the broken lines, the fraction of firms that assemble final goods in South declines with g . That is, FDI in intermediates and in assembly are complementary; as the fixed cost of FDI in intermediate goods increases, the fraction of firms assembling in South declines.³⁴

³³ See also Ekholm, Forslid, and Markusen (2004) on export-platform FDI.

³⁴ This is what Grossman, Helpman, and Szeidl (forthcoming) call unit-cost complementarity, which has its origin in Yeaple (2003). It arises from the fact that when intermediates are produced in South at lower unit cost, it becomes more attractive to assemble final goods there because the larger final good sales make it easier to cover the fixed cost of FDI in assembly.

In the absence of trading costs, horizontal FDI has no economic justification. And indeed, figure 4 shows no instance in which a firm in one Northern country chooses to perform assembly in the other Northern country. At most there is vertical FDI (region $\{S, H\}$) and complex integration (region $\{S, S\}$). But horizontal FDI becomes a viable option when trade in final goods is costly. So consider a modified version of this model with melting iceberg transport costs of final goods (but still free trade in intermediate inputs). For low transport costs, the equilibrium integration strategies are the same as in figure 4. But for intermediate levels of such transport costs and relatively low demand in South, the multinationals pursue different integration strategies for high values of g . The least-productive firms perform both activities in the home country while the most-productive firms perform both activities in South. However, firms with productivity between these extremes produce intermediate inputs in the home country but choose different strategies for serving foreign markets depending on how productive they are within this range; the less-productive firms choose subsidiary sales in the other Northern country and export to South, while the more-productive firms choose subsidiary sales in both foreign countries. As a result, all these firms engage in horizontal FDI except that the more-productive firms do not export at all; they serve every market with local subsidiary sales. In this case too there is complementarity between the two forms of FDI; as g increases, a smaller fraction of firms engage in subsidiary sales in foreign countries.³⁵ As

³⁵ The composition of FDI in assembly is now driven by an additional source of complementarity, what Grossman, Helpman, and Szeidl (forthcoming) call source-of-components complementarity, which stems from the fact that, for moderate transport costs of final goods, a Northern market is cheaper to serve from assembly lines in South if and only if intermediate inputs are also produced in South.

Grossman, Helpman, and Szeidl (forthcoming) show, this type of complementarity is robust in the sense that it holds also for high transport costs of final goods and for high transport costs of intermediate inputs.³⁶

2.7 Variable Markups

The constant-elasticity demand function that was used above has been the workhorse of monopolistic competition studies in economics, including international trade. It is a convenient tool in many applications and it is easily derived from either CES preferences or a CES production function. It has one particularly undesirable feature, however: it implies that markups depend neither on cost nor on demand levels.³⁷ As a result, the distribution of prices is a scaled version of the distribution of marginal costs, with no impact of market size or the number of competitors on the shape of the price distribution. Yet empirical evidence on regional markets in the United States suggests that higher demand, as measured by market density, reduces markups and price dispersion.³⁸ Moreover, with this type of demand, free entry implies that total spending on the industry's products has no effect on firm size because higher spending raises the demand level A but entry of new firms then reduces this demand level, so that at the end of the

³⁶ When transport costs of final goods are high and the demand level in South is not very high, firms do not assemble final goods in South for export to North; they either export from the home country or they serve foreign markets through subsidiary sales.

³⁷ A firm with marginal cost c/θ that faces the demand function $x(j) = Ap(j)^{-\epsilon}$, where $\epsilon = 1/(1 - \alpha) > 0$, maximizes profits by charging price $p(j) = c/\alpha\theta$. Under these circumstances, the ratio of price to marginal cost—which is a measure of the markup—equals $1/\alpha$, and it does not depend on marginal cost. Moreover, it does not depend on the demand level A .

³⁸ See Chad Syverson (2005) for a study of ready-mixed concrete plants.

process A does not change.³⁹ This too is inconsistent with the evidence; market size actually is positively correlated with firm size.⁴⁰ In order to accommodate these features of the data, it is necessary to find an alternative specification of demand in which markups are endogenous. The theory will be more consistent with the evidence when the model implies that a larger market size reduces a firm's markup because, in this case, the firm also raises sales at constant cost and productivity.

Although comparable evidence on variation across countries does not exist, it is quite likely that markups, prices, and firm size vary across countries in similar fashion. To address these issues, Melitz and Gianmarco I. P. Ottaviano (2005) combine supply-side features from Melitz (2003) with demand side features from Ottaviano, Takatoshi Tabuchi, and Jacques-François Thisse (2002) to construct a model of international trade with variable markups in which market size affects average prices, price dispersion, and firm size. The model yields interesting predictions concerning trade and the impact of trade liberalization on productivity and price distributions.

Ottaviano, Tabuchi, and Thisse (2002) use the following quadratic, quasilinear utility function:

$$u = x_0 + \zeta \int_{j \in J} x(j) dj - \frac{1}{2} \gamma \int_{j \in J} x(j)^2 dj - \frac{1}{2} \eta \left[\int_{j \in J} x(j) dj \right]^2,$$

³⁹ For simplicity, consider a closed economy with no export opportunities. Using the optimal pricing strategy $p(j) = c/\alpha\theta$, a firm with productivity θ earns operating profits that equal either $\Theta B - cf_D$ or zero, whichever is larger, where $B = (1 - \alpha)A(c/\alpha)^{1-\varepsilon}$ (see (1)). Then free entry implies that the expected present value of these operating profits equals the entry cost. This free entry condition depends on the cutoff Θ_D and on the demand level A . Together with the equation for the cutoff, i.e., $\Theta_D B = cf_D$, the two equations uniquely determine the cutoff Θ_D and the value of A . It follows that larger spending on these products is precisely offset by a larger number of entrants (brands) so that A is not affected.

⁴⁰ See Jeffrey R. Campbell and Hopenhayn (2005) for evidence on retail trade industries across U.S. cities.

where x_0 is consumption of an outside good that yields constant marginal utility, J is the set of brands available in the market, and ζ and η are positive parameters. When $\gamma = 0$ all brands are perfect substitutes, and the brands are less substitutable for each other the larger is γ .⁴¹ Let $\gamma > 0$. Then, assuming that the consumer has enough income to justify positive consumption of the outside good, his demand for brand j is:

$$(4) \quad x(j) = \frac{1}{\gamma} \left[\frac{\zeta\gamma + \eta N \bar{p}}{\gamma + \eta N} - p(j) \right],$$

where N is the number of products he consumes and \bar{p} is the average price of these products.⁴² This is a linear demand function in which the demand level is decreasing in the own price, increasing in the average price \bar{p} (i.e., the competitors' prices), and declining in the number of products N .⁴³ That is, as the competitive pressure intensifies, either because prices of competing products decline or the number of competing products increases, the manufacturer of brand j faces lower demand.⁴⁴ In an economy populated by Q such consumers, aggregate demand for the brand equals $Qx(j)$. Facing production unit cost c/θ , this manufacturer maximizes profits by charging price

$$(5) \quad p(j) = \frac{1}{2} \left(\frac{\zeta\gamma + \eta N \bar{p}}{\gamma + \eta N} + \frac{c}{\theta} \right).$$

⁴¹ This utility function represents preferences over two goods: a homogeneous outside good x_0 and a differentiated product with consumption $x(j)$ of brand j . It can be generalized to many goods.

⁴² The consumer chooses to purchase all products with

$$p(j) \leq \frac{\zeta\gamma + \eta N \bar{p}}{\gamma + \eta N},$$

and in an equilibrium with these types of consumers the set J consists only of such products.

⁴³ Note that the inequality in the previous footnote together with (4) imply $\bar{p} \geq \zeta$.

⁴⁴ J. Peter Neary (2003) uses a related demand structure in his model of general oligopolistic competition, in which preferences are not quasi linear and $\eta = 0$. Nevertheless, in his case too the demand level of a single product depends on the average price of other products, because it affects the marginal utility of consumption.

Under these circumstances, the markup—defined as the ratio of price to marginal cost—is increasing in the average price and declining in the number of products. Moreover, the markup is increasing in γ , implying that it is higher the less substitutable are brands for each other.

In this model, firms with very low productivity do not produce even if they bear no fixed production costs f_D , because demand drops to zero at a finite price.⁴⁵ It is then possible to solve the entire model of a closed economy by adding a free entry condition, with or without a positive fixed cost f_D . The solution for θ_D then implies that the cutoff for positive production is declining in γ and in Q . That is, less-productive firms survive entry the less substitutable the products are for each other and the larger the market size is. It follows that in sectors with less substitutability there is more productivity and size dispersion and average productivity is lower.⁴⁶

Melitz and Ottaviano (2005) assume that countries differ only in their numbers of consumers Q , that there are neither fixed production nor fixed export costs, but that there are variable trade costs $\tau > 0$. The trade costs introduce a degree of market segmentation that produces cross-country variation in the number of consumed products that is positively correlated with market size, as measured by the number of consumers. Moreover, average productivity is higher in larger markets, because low-productivity

firms find it harder to compete in larger markets. This is similar to the result in Melitz (2003) but for different reasons. In Melitz (2003), trade raises the profits of high productivity firms—which export—and these exporters raise the demand for domestic inputs. As a result, domestic producers who do not export are hit by competition from foreign exporters on the one hand and by higher input prices on the other, which forces the least productive of them to leave the business. The cutoff θ_D increases, and so does average productivity. In contrast, in Melitz and Ottaviano (2005), trade does not change input prices, but it reduces markups as a result of the increase in competitive pressure from foreign exporters, and this raises the cutoff θ_D and average productivity. The implication is that not only do consumers in larger markets have access to more products, they also pay lower prices.⁴⁷

Multilateral liberalization raises the number of products in all markets, which raises competition and cuts into markups. Only more-productive firms survive this pressure, resulting in higher productivity and lower prices. Evidently, this sort of trade liberalization is beneficial to all countries concerned. In contrast, when only a subset of countries liberalize trade amongst themselves, the impact on the liberalizing countries differs markedly from the impact on the excluded countries. In the former countries average

⁴⁵ With $f_D = 0$, demand is not negative for

$$p(j) \leq \frac{\zeta\gamma + \eta N\bar{p}}{\gamma + \eta N}.$$

Together with (5) it implies

$$\frac{c}{\theta} \leq \frac{\zeta\gamma + \eta N\bar{p}}{\gamma + \eta N}.$$

Therefore the cutoff productivity level θ_D satisfies

$$\theta_D = c \frac{\gamma + \eta N}{\zeta\gamma + \eta N\bar{p}}.$$

⁴⁶ Syverson (2004) provides evidence of these effects for a cross section of U.S. manufacturing industries.

⁴⁷ The differences between Melitz (2003) and Melitz and Ottaviano (2005) stem from two sources: they use different shapes of demand functions (constant elasticity in one case and linear in the other) and they make different assumptions about the outside good (no outside good with constant marginal utility in one case and the presence of such a good in the other). The absence of an outside good in Melitz (2003) generates impacts on input costs that are absent in Melitz and Ottaviano (2005), where quasi-linearity fixes the unit cost c . But this is distinct from the impacts of the shapes of the demand functions for final goods, which are isoelastic in one case and linear in the other. In the isoelastic case, the demand level has no effect on the cutoff because any shift in demand is offset by entry, in contrast to the case of linear demand. And while the markup is constant in the isoelastic case, it responds to demand and entry in the linear case.

productivity rises, markups and prices decline, and the number of products increases. The opposite takes place in the excluded countries. Under these circumstances, the liberalizing countries gain while the other countries lose.

2.8 *Factor Proportions*

Although Melitz (2003) places the firm at the center of analysis, his approach has implications for trade flows at the sectoral level. This is apparent from the fact that sectoral average productivity levels are endogenous, and they depend on the determinants of the sectoral cutoffs, θ_D (or Θ_D). These endogenous productivity levels generate Ricardian-type comparative advantage that affects the sectoral patterns of trade flows.

Bernard, Stephen Redding, and Schott (forthcoming) have extended the Melitz (2003) model to accommodate variable factor proportions, producing a richer model of trade in differentiated products than the standard Helpman and Krugman (1985) version. They consider a two-sector, two-factor world with constant expenditure shares on each sector's output, CES preferences for varieties in every sector, and Cobb–Douglas production functions for activities that generate either fixed or variable costs. And they achieve great simplicity by assuming that the Cobb–Douglas production functions have the same exponents in all activities within a given sector, while they vary across sectors. In a world with no trading frictions, i.e., neither fixed nor variable costs of exporting, the analysis proceeds along the now familiar lines of the integrated equilibrium approach, with results similar to Helpman–Krugman. The sectoral cutoffs θ_D are not affected by trade, and therefore neither are sectoral productivity levels. The intersectoral pattern of trade is of the Heckscher–Ohlin type: every country is a net exporter of goods that use relatively more intensively

the input with which the country is better endowed.⁴⁸

Next they introduce melting iceberg variable trade costs and fixed export costs, where the sectoral fixed export cost, arising from a Cobb–Douglas production function, has the same factor intensity as the other sectoral activities. These costs segment markets across countries. Now trade has an influence on the cutoffs θ_D ; they rise in every country and every industry. This means that trade raises average productivity everywhere in the world. Importantly, however, in every country it raises average productivity proportionately more in the comparatively advantaged industry, i.e., the sector that is relatively intensive in the input with which the country is relatively well endowed. Under the circumstances, the Heckscher–Ohlin-type comparative advantage, which emanates from factor composition, also produces Ricardian comparative advantage; and the two forms of comparative advantage are positively correlated. This is an important result, because the empirical evidence suggests that it is necessary to control for TFP differences across countries in order to estimate the impact of factor proportions on trade flows.⁴⁹ In addition, trade increases firm size, and relatively more so in sectors having comparative advantage. Finally, trade raises the rate of gross job destruction and gross job creation, thereby raising turnover. But net job creation rises in comparatively advantaged industries and declines in the other sectors. These are very interesting predictions that will undoubtedly influence empirical analysis.

2.9 *Gravity Equation of Trade Flows*

The gravity equation is a major tool for the empirical analysis of trade flows. It has been

⁴⁸ This result does not hold in a world in which different activities within a sector have different factor proportions.

⁴⁹ See Treffer (1995) and Donald R. Davis and David E. Weinstein (2001).

used to study the impact on trade flows of international borders, currency unions, membership in the WTO, and other variables. And it has been used outside trade for instrumental variable estimation of the impact of variables such as social infrastructure or political institutions on measures of economic success.⁵⁰ In all these applications, the standard procedure is to estimate a gravity equation of bilateral trade flows on a sample of countries that export to each other. This selected sample of countries represents, however, only about half of the country pairs in large samples of countries; in the majority of the other half of country pairs, the countries do not trade with each other; and in the remaining pairs, one country exports to the other but not vice versa.⁵¹ These facts raise two questions: First, what accounts for the absence of trade among so many pairs of countries? And second, to what extent are estimates of trade flows that disregard the nontrading countries reliable?

Helpman, Melitz, and Rubinstein (2006) show that a modified version of Melitz (2003) can account for the lack of trade between potential trading partners and that the modified theoretical framework provides guidance for an estimation procedure that exploits the information contained in the zero trade flows. In particular, they argue that lack of trade is not random, but rather arises from economic conditions, and that therefore we should simultaneously explain which countries trade bilaterally and, amongst those that do, how much is traded. The model suggests that the standard estimation procedure introduces two types of biases: a sample-selection bias and an omitted-variable bias. The sample-selection bias problem is well known and it can be corrected for with standard methods. The omitted-variable problem is novel, however. It stems from the fact that, in addition to the

intensive margin of trade, i.e., the response of a firm's export to changing conditions, there is an extensive margin, which consists of the response of the number of exporting firms to changing conditions. If determinants of the number of exporting firms are not accounted for in the estimation of trade flows, the resulting estimates suffer from an omitted-variable bias. Helpman, Melitz, and Rubinstein propose a method for decomposing the impact of covariates, such as distance between countries, on trade flows into an intensive and extensive margin, and they find that the extensive margin is empirically quite important.

The main ingredient of the modified model is a cumulative distribution function of productivity θ that has finite support $[\theta_L, \theta_H]$, where θ_L is the lowest productivity level and $\theta_H < \infty$ is the highest. It is evident from figure 2 that if $\Theta_H \equiv (\theta_H)^{\varepsilon-1}$ falls between the domestic cutoff Θ_D and the export cutoff Θ_X^ℓ , then home firms produce for the home market but none of them finds it profitable to export to country ℓ . Moreover, Θ_H can be below the export cutoff of some countries and above the export cutoff of other countries, so that domestic firms may find it profitable to export to some countries but not to others. The export cutoff Θ_X^ℓ is smaller the larger is the market in ℓ and the lower are the fixed export or trading costs with ℓ . The variables that affect the cross-country variation in Θ_X^ℓ therefore explain to which foreign countries the home country should export.

Using the firms' optimal pricing and sales strategies together with the free-entry condition, the model implies two equations for every export flow, say from country j to country i . One equation describes the log of exports from country j to country i , m_{ij} , when exports are positive, as a function of standard covariates, such as distance between the countries and whether they share a common language, as well as exporter and importer fixed effects. But in addition, it includes a

⁵⁰ See the discussion in Helpman, Melitz, and Yona Rubinstein (2006).

⁵¹ See Helpman, Melitz, and Rubinstein (2006).

variable w_{ij} which is an increasing function of the fraction of country j firms that export to country i :

$$m_{ij} = \beta_0 + \lambda_j + \chi_i - \gamma d_{ij} + w_{ij} + u_{ij}.$$

In this equation, λ_j is country j 's fixed effect as an exporter, χ_i is country i 's fixed effect as an importer, d_{ij} is the distance between the two countries, and u_{ij} is an error term that describes the unobserved variation across country pairs in variable trade costs.⁵² Covariates other than distance are accommodated in similar fashion. The second equation describes a latent variable that is positive if and only if j exports to i . It is defined as the log of the ratio of variable export profits for the most productive firm to the fixed export cost, where the latter is the same for all firms in j . When this latent variable, z_{ij} , exceeds zero, some firms from country j export to country i because the most productive firm makes profits large enough to cover the fixed export cost. The resulting equation is

$$z_{ij} = \gamma_0 + \xi_j + \zeta_i - \gamma d_{ij} - \kappa \phi_{ij} + \eta_{ij},$$

where ξ_j is an exporter fixed effect, ζ_i is an importer fixed effect, ϕ_{ij} is an observed variable that impacts the fixed cost of exporting from j to i , and η_{ij} is an error term that combines the unobserved variation across countries in variable trade costs u_{ij} (which also appears in the trade flow equation) and a variable v_{ij} that represents unobserved variation across countries in the fixed export costs. Evidently, η_{ij} is correlated with u_{ij} .

It is common to estimate the equation of trade flows m_{ij} for country pairs with positive trade flows, without controlling for the impact of the fraction of exporting firms through w_{ij} . These are the sources of the selection and omitted-variable biases.

Helpman, Melitz and Rubinstein (2006) show how to use standard methods to correct for the sample selection bias, by estimating the equation for z_{ij} and the equation for m_{ij} . This bias turns out to be rather small in their data. In addition, they show how to account for the impact of w_{ij} on the trade flows. Their procedure recognizes the fact that no data on w_{ij} are available, nor are there data on the fraction of exporting firms, a variable that impacts w_{ij} .⁵³ In particular, they show that the estimated (selection) equation for z_{ij} can be used to construct estimates of the w_{ij} 's, which can then be used in the estimation of the trade flow equation. In this way one can separately identify the impact of a variable, such as distance d_{ij} , on the intensive margin (via its direct impact on m_{ij}) and the extensive margin (via its impact on m_{ij} through w_{ij}). The theoretical underpinning of this procedure derives from the fact that fixed costs affect the latent variable z_{ij} directly, but affect the trade flow m_{ij} only indirectly via their impact on the fraction of exporting firms.

It should be evident from this section that productivity differences across firms in an industry have important implications for trade, trade policies, and FDI flows. Not only does this new way of thinking shed light on a host of substantive issues, it also helps in formulating better empirical strategies for estimating trade flows. To illustrate, this approach suggests that countries seeking to integrate into the world trading system or to join a free trade area should expect substantial reallocations across firms within industries, which will impact sectoral productivity levels. In addition to raising the profits of exporting firms, lower protection raises the competitive pressure from foreign enterprises, which cuts into the domestic firms' markups, raises domestic

⁵² The need for separate importer and exporter fixed effects has been well known; see, for example, Feenstra (2004). The theoretical model provides, however, a clear interpretation of the determinants of these fixed effects.

⁵³ That is, there are no data on the fraction of exporting firms for the large samples of countries used to estimate gravity equations. Note also that even if data on w_{ij} were available, it could not be used directly because w_{ij} is an endogenous variable, and one would therefore need to instrument it.

factor costs, and drives the country's least-productivity firms out of business.

3. *Incomplete Contracts*

My discussion of trade and FDI has so far focused on final products.⁵⁴ Importantly, none of the studies reviewed in the previous section, where firms make FDI choices, explicitly analyzes the internalization decision. That is, it is assumed that foreign operations are organized in foreign affiliates, be it for the purpose of manufacturing final products designated for the foreign market, manufacturing components in a foreign country to be assembled at home or in a foreign country, or assembling final goods to be sold in the home or foreign market. Yet the choice of whether to manufacture components inhouse or acquire them from an unaffiliated firm is a key decision about organizational form, as is the decision of whether to source such components at home or in a foreign country. The same applies to assembly, which is just another activity in the chain of tasks that need to be performed in order to deliver a product to a final user. A better understanding of these choices is needed in order to explain the trends in trade and FDI and their relation to the evolving organization of production and distribution.

Two facts stand out that triggered a major research effort into the international organization of production. First, with the advent of computer-aided design, computer-aided manufacturing, and institutional changes in labor markets, outsourcing has rapidly expanded.⁵⁵ This is true about both

domestic and international outsourcing, where rising domestic outsourcing means an increase in the purchase of intermediate goods and services from domestic unaffiliated firms, and rising foreign outsourcing means an increase in the purchase of intermediate goods and services from foreign unaffiliated firms. These trends have been widespread across different sectors and different inputs.⁵⁶ Second, the sourcing of inputs from foreign countries has increased at a rapid pace, both via arm's-length trade (outsourcing) and via intrafirm trade (FDI), a phenomenon known as offshoring.⁵⁷ In order to understand these trends, we need to understand the two-dimensional decision problem of business firms: whether to outsource or insource (i.e., integrate), and whether to offshore or not.⁵⁸ This choice yields four possibilities: insourcing at home, outsourcing at home, insourcing abroad (FDI), and outsourcing abroad. The first two organizational forms do not involve foreign trade, while the latter two do: intrafirm trade in the case of

⁵⁶ See Edward J. Bardi and Michael Tracey (1991), Elizabeth Gardner (1991), Susan Helper (1991), James Bamford (1994), Abraham and Taylor (1996) and Bartel, Lach, and Sichernman (2005) for evidence on the growth of outsourcing in various industries. *The Economist* (1991) provides an early overview.

⁵⁷ Feenstra and Gordon H. Hanson (1996) find more than a doubling of the share of imports in total purchases of intermediates from 1972 to 1990 in the United States (from 5.0 percent to 11.6 percent), while Campa and Goldberg (1997) find similar trends in Canada and the United Kingdom. And Hummels, Ishii, and Yi (2001) and Yeats (2001) find that foreign trade in components has grown faster than foreign trade in final goods. Finally, Hanson, Mataloni, and Matthew J. Slaughter (2005) find that intrafirm trade within U.S. multinationals has grown very fast, although somewhat less than international outsourcing by U.S. firms, and Feinberg and Keane (2005) report that sales of U.S. parent firms to their Canadian affiliates as a fraction of the affiliates' total sales, as well as sales of the Canadian affiliates to their U.S. parents as a fraction of the parents' total sales, have almost doubled between 1984 and 1995.

⁵⁸ Two additional decisions, which are equally important but received only scant attention, concern the types of inputs that should be acquired by means of each one of these organizational forms, and if an input is to be sourced abroad, to which country it should be offshored.

⁵⁴ The only exception being the discussion of the interrelationship between FDI in components and FDI in assembly in section 2.6.

⁵⁵ I use "outsourcing" to mean the acquisition of an input or service from an unaffiliated company. This is the standard terminology used in industrial organization. A narrower definition is used in some of the literature, e.g., Mary Amiti and Shang-Jin Wei (2005). A notable example of a very narrow definition is Jagdish Bhagwati, Arvind Panagariya, and T. N. Srinivasan (2004), who restrict the term to outsourcing of services from foreign unaffiliated companies. I find it preferable to use the traditional definition.

FDI, and arm's-length trade in the case of outsourcing.⁵⁹

An analysis of these issues by means of the incomplete contracts approach to the theory of the firm helps in understanding why some companies source inputs abroad primarily via FDI, while other companies source them abroad primarily via outsourcing. It also helps in understanding why intrafirm trade as a fraction of total trade is positively correlated with capital-labor ratios across U.S. industries and capital-labor ratios across countries from which the U.S. imports (see Pol Antràs 2003). Moreover, this approach helps in understanding why differences across countries in the quality of legal systems generate comparative advantage, and thereby impact the patterns of trade (see Andrei A. Levchenko 2004 and Nathan Nunn forthcoming). Finally, when combined with productivity variation across firms within industries, this approach helps in predicting the relative prevalence of alternative forms of the international organization of production as a function of sectoral characteristics and differences in features of the trading partners.

3.1 *The Incomplete Contracts Approach*

To illustrate the incomplete contracts approach to the theory of the firm, consider the following example.⁶⁰ A final good producer makes profits $\pi_0 \geq 0$ if she does not use a specialized intermediate input. If, however, she uses one unit of the specialized intermediate input, her profits become $\pi_1 > \pi_0$, where π_1 does not include the cost of the

input to the final good producer. For simplicity, assume that the final good producer can use only one unit of this input.

In order to acquire the input, the final good producer needs to engage a supplier. The supplier can produce the required input at cost c . Importantly, however, the final good producer and the supplier cannot sign an enforceable contract that specifies the nature of the specialized intermediate input, but the final good producer can recognize ex post, after the input is delivered, whether the input has the requisite features. For this reason the supplier can choose the characteristics of the input, and when delivering it to the final good producer he can bargain with the final good producer for payment. At the bargaining stage, the cost c of the intermediate input is sunk, and it therefore plays no role in determining the bargaining outcome. But it does play an important role in determining whether the supplier chooses to manufacture the requisite intermediate input in the first place.

There are two stages to the game. In the first stage, the supplier decides whether to manufacture the intermediate input and, if he does, whether to endow the input with the special characteristics requested by the final good producer. In stage two, the supplier delivers the input and bargains for payment. As usual, the game is solved backwards, starting from stage two.

Assume that whenever the final good producer and the supplier bargain they reach an agreement according to the Nash bargaining solution, with the bargaining weight $\beta \in (0, 1)$ for the final good producer and $1 - \beta$ for the supplier. In this event, their payoffs are derived as follows. In case of a breakup of the negotiation, the final good producer has the outside option π_0 while the supplier has the outside option $\sigma_0 \geq 0$. The size of σ_0 depends on how specialized the intermediate input is. If, for example, it is so highly specialized that no one else can use it, i.e., it has no value outside the relationship, then

⁵⁹ The segmentation of production across different countries has become so large that it prompted the WTO to describe in its 1998 annual report the detailed acquisition of inputs by U.S. car manufacturers in different countries, concluding that only 37 percent of a car's value was generated in the United States. Rone Tempest (1996) describes an equally global sourcing strategy of Mattel in the manufacturing of Barbie dolls (see Feenstra 1998).

⁶⁰ See Oliver D. Hart (1995) for a detailed discussion of this approach, and Patrick Bolton and Mathias Dewatripont (2005) for a textbook treatment.

$\sigma_0 = 0$. If, however, the intermediate input can be used by other manufacturers, then $\sigma_0 > 0$. In either case, the payoff of the supplier equals his outside option plus the share $1 - \beta$ of the surplus from the relationship, while the payoff of the final good producer equals her outside option π_0 plus the share β of the surplus from the relationship. The size of the surplus depends on whether the intermediate input satisfies the specifications needed by the final good. If it does, the surplus equals π_1 minus the outside options of the two players; that is, it equals $\pi_1 - \pi_0 - \sigma_0$. If it does not, it equals $-\pi_0 - \sigma_0$, because the input adds no value to the relationship. Naturally, the latter case does not arise in equilibrium. Therefore in an equilibrium in which the supplier delivers the requisite input the payoffs from the bargaining game are

$$P_f = \pi_0 + \beta(\pi_1 - \pi_0 - \sigma_0)$$

for the final good producer and

$$P_s = \sigma_0 + (1 - \beta)(\pi_1 - \pi_0 - \sigma_0)$$

for the supplier. Note that $P_f + P_s = \pi_1$, so that they split the profits π_1 . We can interpret P_s as the payment of the final good producer to the supplier for the intermediate input, so that her profits net of the input cost, $\pi_1 - P_s$, equal her payoff P_f .

This solution to the bargaining game determines the incentives of the supplier to engage in a business relationship with the final good producer (in stage one of the game). If $P_s \geq c$, this relationship generates a profitable deal; otherwise it does not. That is, in case $P_s < c$ the supplier will not produce the specialized intermediate input.

In this example, there is a one-sided holdup problem. The supplier is held up by the final good producer because he makes a relationship-specific investment. More generally, however, the final good producer may

also be required to make a relationship-specific investment, in which case there will be a two-sided holdup problem. Moreover, the outside options of the two parties may depend on the organizational form of the business firm, e.g., whether the intermediate input is produced in-house or outsourced. In the former case, the supplier is an employee of the final good producer while in the latter case he is not (see Sanford J. Grossman and Hart 1986 and Hart 1995). Finally, when these interactions are placed in a general equilibrium setup, the outside options become endogenous, and they depend on the nature of the technology and the organization of the industry (see Grossman and Helpman 2002).

I now turn to the application of this approach to the main issues discussed in the preamble to this section. I have ordered the topics in a way that eases the exposition, and not in the order in which they appeared in the literature.

3.2 Contractual Input Intensity

Traditionally, input (or factor) intensity refers to the relative requirement of various inputs (or factors of production) in the manufacturing of a good. But the theory of incomplete contracts identifies another important measure of input intensity: the relative requirement of intermediate inputs that are under direct control of the final good producer, and intermediate inputs that require the engagement of suppliers. The importance of the second intensity measure, which I term the *contractual input intensity*, stems from the fact that intermediate inputs under the direct control of the final good producer suffer less from agency problems than intermediate inputs that require the engagement of suppliers. Naturally, the two intensity measures can be correlated. And when they are, the theory yields interesting predictions about the structure of trade.

The contractual input intensity impacts the power of incentives that a final good producer wants to give a supplier. In particular, the more intensive the production process is in intermediate inputs that are controlled by suppliers, the more powerful incentives she wants to give the suppliers. Yet her most desired incentives are not extreme. That is, she never wants to give them the strongest possible or the weakest possible incentives. Under the circumstances her choice of organizational form, such as outsourcing versus integration, is determined to an important degree by its effect on the incentives of suppliers.

To understand the role played by contractual input intensity, consider an industry of a differentiated product in which the demand function is, as before, $x(j) = Ap(j)^{-\varepsilon}$, $\varepsilon = 1/(1 - \alpha) > 0$. Now, however, the production of brand j requires two customized inputs, headquarter services $h(j)$ and components $m(j)$. These intermediate inputs are combined via a Cobb–Douglas production function to produce either brand j of the differentiated product, $x(j)$, or another intermediate input of type j , say $y(j)$, which is used to assemble $x(j)$. In the latter case $x(j) = y(j)$.⁶¹ I express this production relationship as

$$(6) \quad z(j) = \theta \left[\frac{h(j)}{\eta} \right]^\eta \left[\frac{m(j)}{1 - \eta} \right]^{1 - \eta}, \quad 0 < \eta < 1,$$

where z is either x or y , θ represents productivity, which for the time being is the same for all firms in the industry, and η measures contractual input intensity. The larger η is the more intensive the sector is in headquarter services (but η does not vary across firms in the industry). The critical assumption is that h has to be supplied by the final good producer while m

requires the engagement of a supplier, which can take place either inside or outside the firm. But in either case the supplier controls m . In this event, the internalization decision is only about the intermediate input m , not about h . In a simple version of the model, there is only labor and both h and m are produced with a fixed amount of labor per unit output. More generally, there can be many factors of production (primary inputs), and h and m may be produced with different factor proportions. It then follows that the overall factor intensity of an industry is jointly determined by its contractual input intensity and by the factor intensities of headquarter services and components.

Using the demand function $x(j) = Ap(j)^{-\varepsilon}$ and the production function (6), we can calculate revenue as a function of the inputs h and m , say $R[h(j), m(j)]$. The assumption is that the final good producer bears directly the cost of headquarter services and decides the level of h , while the supplier, who may be working for the final good producer or be independent, chooses m . Great simplification is attained by assuming that the final good producer can obtain as many suppliers as she wants by offering a reward structure consisting of an upfront payment and a share of the profits at the bargaining stage. In this event competition among suppliers leaves them with no rents, and a supplier's total net income (net of input cost) equals his opportunity cost. At the bargaining stage, the distribution of revenue $R[h(j), m(j)]$ depends on the bargaining weights, which are β for the final good producer and $1 - \beta$ for the intermediate good producer, and on organizational form, which determines every party's outside option.

Consider outsourcing. Under this organizational form, the outside options at the bargaining stage are zero for both parties because one party owns h and the other owns m , and both inputs have been customized for product j to a degree that they

⁶¹ These two possibilities may seem to be an unnecessary complication at this stage, but they provide a unified treatment of distinct papers in the literature, as will become clear below.

have no value outside the relationship.⁶² As a result, the final good producer receives the fraction β of the revenue while the supplier receives the fraction $1 - \beta$.

Next consider integration. Now both h and m belong to the final good producer, because the supplier is her employee. But, following Grossman and Hart (1986), assume that if the bargaining fails and the supplier does not cooperate, then the final good producer cannot deploy the inputs as effectively as she can when the supplier cooperates. In particular, without the cooperation of the supplier she is able to produce only a fraction δ of the output in (6). Under the circumstances the outside option of the supplier at the bargaining stage is zero, while the outside option of the final good producer is fraction δ^α of the revenue $R[h(j), m(j)]$.⁶³ As a result, in the bargaining stage the final good producer receives a fraction $\beta_V = \delta^\alpha + \beta(1 - \delta^\alpha)$ of the revenue $R[h(j), m(j)]$, and the supplier receives a fraction $1 - \beta_V$.

An important trade-off in the choice of organizational form by the final good producer is derived from a comparison of the optimal distribution shares of revenue, $R[h(j), m(j)]$, with the shares that arise under outsourcing and integration. Let β^* be the final good producer's most preferred share, which maximizes her profits. First note that it cannot be zero, because if it were zero she would have no incentive to provide headquarter services, and in the absence of h ,

revenue equals zero. Second, note that it cannot be one, because if it were one the supplier would have no incentive to provide components, and in the absence of m , revenue would equal zero. Evidently, β^* is strictly positive and strictly smaller than one. Moreover, it can be shown that β^* is an increasing function of the intensity of headquarter services, as measured by η . The shape of the relationship between β^* and η is depicted in figure 5. β^* equals zero when $\eta = 0$, it equals one when $\eta = 1$, and it rises in between. Moreover, it is concave for low values of η and convex for high values.

The figure also shows the distribution of revenue shares under outsourcing and integration, β and β_V , respectively; they are above the optimal share β^* when an industry is component-intensive, so that η is small (such as η_M), and they are below β^* when an industry is headquarter-intensive, so that η is large (such as η_H). The arrows show the direction of rising profits; that is, profits rise when the final good producer's share shifts vertically toward β^* . This characterization implies that there exists a cutoff η_c —not drawn in the figure—with $\eta_M < \eta_c < \eta_H$, such that the final good producer has higher profits from outsourcing when η is below η_c and higher profits from integration when η is above η_c . It follows that, based on the power of incentives consideration alone, final good producers prefer outsourcing in component intensive industries and integration in headquarter intensive industries. However, the

⁶² The outside options need not be zero in this case. For example, the final good producer may have the option of using a generic intermediate input m instead of the specialized variety, in which case her outside option will not be zero. Similarly, the intermediate good producer may have the option of selling the input $m(j)$ to another firm, which will provide it with a positive outside option. Grossman and Helpman (2002) provide an analysis in which the suitability of a component m to a final good producer is measured by the distance between the supplier and the final good producer in technology space, and suppliers have the option to choose the location of their intermediate inputs in this space. When bargaining breaks down, final good producers and component suppliers enter a secondary market. The secondary market equilib-

rium then determines the outside options in the primary market. In the Grossman and Helpman technology space a generic input is defined as the input that is equal-distant from all final goods. Feenstra and Spencer (2005) also develop a model of contracting in the presence of generic inputs, and they use it to analyze the organization of Chinese suppliers. The model discussed in the text focuses on the simpler case of zero outside options.

⁶³ The finding that the outside option of the final good producer under integration is the fraction δ^α of revenue rather than the fraction δ stems from the concavity of the revenue function in the quantity sold. That is, revenue as a function of the quantity x is proportional to x^α , where (recall) α determines the demand elasticity $\varepsilon = 1/(1 - \alpha)$ and $0 < \alpha < 1$.

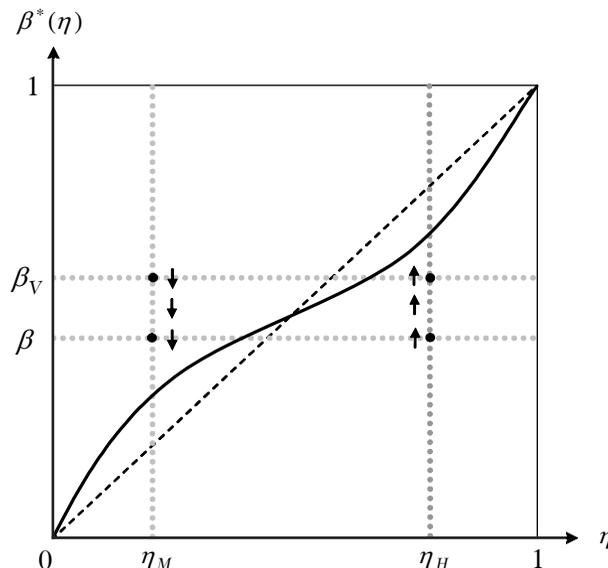


Figure 5. Optimal Bargaining Share

final verdict on whether to outsource or integrate does not depend on these considerations alone if there also exist cost differences in running firms with different organizational forms.⁶⁴

Antràs (2003) uses a variant of this model in which there are no organization-specific costs; there are fixed entry costs, but these costs are independent of whether the firm chooses to outsource or integrate. In this event the power of incentives dominates the integration decision. That is, given the headquarter intensity measure η a firm prefers integration if $\eta > \eta_c$ and outsourcing if $\eta < \eta_c$. He assumes that h is capital intensive and m is labor intensive, and that h and m are not tradeable across borders. In this event every final good producer has to deploy h and m in the same country.

⁶⁴ Feenstra and Hanson (2005) estimate a related model of the firm from Chinese export-processing data. A plant that processes imported inputs for sales to a foreign firm can be owned either by a foreign firm or by a Chinese entity. Similarly, imported inputs for further processing can be owned by a foreign firm or by the processing plant. In the latter case the inputs are controlled by the plant's manager. The organizational form, which consists of the ownership of the plant and the ownership of the imported

inputs, are used to manufacture an intermediate input y that can be freely shipped across borders. The production function is given by (6), where $z = y$. The final good x is produced from y in the destination country with one unit of y per unit x . Finally, consumers spend fixed budget shares on goods in every sector and they have CES preferences across brands.

In a two-country, two-sector version of this model, trade structure can be derived from the integrated equilibrium, similarly to Helpman and Krugman's (1985) analysis of trade in differentiated products. Assuming that one sector has η above the cutoff η_c and another sector has η below this cutoff implies that firms are integrated in the former sector and firms outsource in

inputs, is determined according to the property rights approach, as in the model described in the text. Feenstra and Hanson find that the prevalence of alternative organizational forms varies across Chinese regions in accordance with the model's prediction. This is a good case study, because 55.6 percent of Chinese exports during the sample period, 1997 to 2002, are of this nature (i.e., export-processing), and the distribution of export-processing exports into the four organizational forms has a nonnegligible fraction in every regime.

the latter sector. Moreover, since h is capital intensive while m is labor intensive, the sector with integrated firms is capital intensive and the sector with outsourcing firms is labor intensive. As a result, there is intrafirm trade in the tradeable intermediate inputs in the capital-intensive sector and arm's-length trade in tradeable intermediate inputs in the labor-intensive sector. This implies a positive correlation between capital intensity and the share of intrafirm trade. A multicountry version of this model also implies a positive correlation between the share of intrafirm imports and the capital abundance of the exporting country. Antràs (2003) provides evidence supporting these predictions. In U.S. data, intrafirm imports as a fraction of total imports are positively correlated with the capital intensity across twenty-three manufacturing industries, and intrafirm imports as a fraction of total imports are positively correlated with capital abundance across twenty-eight exporting countries.

Antràs (2005) applies a one-factor variant of this model to product cycles. The two countries are North and South. He assumes that both headquarter services and final goods can be produced only in North. In addition to whether to integrate or outsource, however, a final good producer has to decide in which country to source the component m , i.e., whether to offshore m or not. Integration or outsourcing in North imply no trade in components, integration in South implies intrafirm trade, and outsourcing in South implies arm's-length trade in components. Contracts are complete in North but incomplete in South. That is, the two countries differ in the degree of contract incompleteness.

The main result is that there exist two cutoff values of the contractual input intensity measure η , η_c and $\eta_n > \eta_c$, which determine the desired organizational form. When headquarter intensity is above the upper threshold η_n , final good producers source m in North (the model is silent on whether they outsource or integrate there, because

contracts are complete in North). For values between η_c and η_n , final good producers invest in subsidiaries in South and source m from their affiliates in South. And when headquarter intensity is below the lower cutoff η_c final good producers outsource in South. Interpreting η as a feature of technology that changes over time—so that η is high for a new product and it declines over time as experience in production is gained—these results imply a product cycle of the Raymond Vernon (1966) type: all parts of the value chain of a new product are produced in North, over time the production of components is shifted to subsidiaries in South, and as the product matures, the components are outsourced to Southern manufacturers.

3.3 Contractual Input Intensity and Productivity Heterogeneity

A combination of variation in contractual input intensity across sectors and variation in productivity across firms within industries generates equilibria in which all four organizational forms—insourcing at home, outsourcing at home, insourcing abroad, and outsourcing abroad—coexist in an industry and their relative prevalence varies across industries as a function of sectoral characteristics. Note that these four organizational forms do not coexist in the previous models.

Following Antràs and Helpman (2004), assume that the production function (6) applies to a typical industry, but that the productivity level θ varies across firms. As in Melitz (2003), an entrant into the industry obtains a productivity draw θ after sinking the entry cost. After entry, and knowing her productivity, the final good producer has to decide on her organizational form.

There are two countries, North and South, with the wage rate in North exceeding the wage rate in South. Labor is the only primary input. All final good producers are located in North, where they also produce headquarter services h . The intermediate inputs m can be produced either in North or South with the

same labor input per unit output. This makes the variable costs of m lower in South. But there are different fixed costs of sourcing in North and South, and these fixed costs also differ for outsourcing and integration. In particular, Antràs and Helpman (2004) focus on the case $f_V^S > f_O^S > f_V^N > f_O^N$, where f_V^S is the fixed cost of integration in South (FDI), f_O^S is the fixed cost of outsourcing in South, f_V^N is the fixed cost of integration in North, and f_O^N is the fixed cost of outsourcing in North, all measured in terms of Northern labor.⁶⁵

Under these circumstances outsourcing dominates integration in component-intensive industries, because (1) outsourcing has lower fixed cost; and (2) for low values of η outsourcing provides better incentives to suppliers of intermediate input m (see figure 5). It follows that in component-intensive industries all firms outsource, and the only remaining trade-off is between domestic and foreign outsourcing. In the offshoring decision, the trade-off is between lower variable cost in South and lower fixed cost in North. This trade-off is depicted in figure 6, where π_O^S represents profits from outsourcing in South and π_O^N represents profits from outsourcing in North as a function of the productivity measure $\Theta \equiv \theta^{\varepsilon-1}$. The profit line π_O^S is steeper because variable costs are lower in South. Evidently, firms with productivity below Θ_D exit the industry, high-productivity firms—with Θ above Θ_M —import components from unaffiliated producers in South, and firms with productivity between Θ_D and Θ_M acquire components from unaffiliated firms in North. That is, among the active firms low-productivity firms outsource at home and high-productivity firms outsource abroad.

A similar analysis of a headquarter-intensive sector shows that all four organizational forms can coexist. The trade-off

between outsourcing and integration in North is depicted in figure 7, where π_V^N represents the profits of an integrated producer and π_O^N represents the profits of an outsourcing producer. The profit line π_V^N is steeper because integration in a headquarter-intensive sector provides better incentives to suppliers of parts (see figure 5). In this case, low-productivity firms—with Θ below Θ_D —exit the industry; high-productivity firms—with Θ above Θ_O —integrate; and firms with intermediate productivity levels outsource. Combining this analysis with a similar analysis of the trade-off between outsourcing and integration in South, and accounting for the fact that offshoring has an advantage in terms of variable costs but a disadvantage in terms of fixed costs, we obtain the sorting pattern depicted in figure 8. That is, the least-productive firms exit the industry while the most-productive firms use FDI to produce intermediate inputs in South. In between, the less-productive firms outsource in North, the more-productive firms outsource in South, and firms with intermediate productivity levels integrate in North.

Three interesting results emerge from a comparative statics analysis of this model. First, offshoring declines with headquarter intensity η . Second, more productivity dispersion leads to more offshoring; in component-intensive sectors it leads to more outsourcing in South while in headquarter-intensive sectors it leads to more integration plus outsourcing in South.⁶⁶ In addition, in headquarter intensive sectors, where there is both outsourcing and integration, more productivity dispersion leads to more integration and less outsourcing. These predictions apply to variations across industries; e.g., the model predicts more offshoring in sectors with higher component intensity and sectors with more productivity dispersion. Third, an improvement in the competitive advantage of South,

⁶⁵ They also provide a brief discussion of the implications of other orderings of fixed costs. See also Grossman, Helpman, and Szeidl (2005).

⁶⁶ Productivity dispersion is measured by the shape parameter of the Pareto distribution.

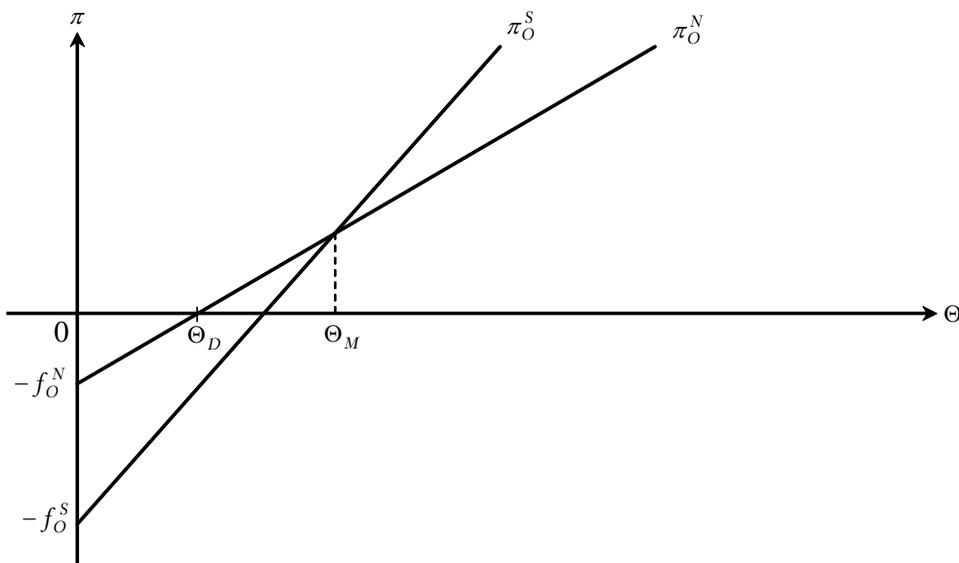


Figure 6. Importing and Nonimporting Firms in a Component-Intensive Sector

be it as a result of declining relative wages in South or declining protection in North, raises offshoring in all sectors; and in headquarter-intensive sectors, outsourcing of components from foreign suppliers rises proportionately more than purchases of intermediate inputs from foreign affiliates.⁶⁷

3.4 Matching and Thick Market Effects

In the previous models, final good producers could attract suitable suppliers at will as long as the suppliers could gain from the relationship with a final good producer as much as they expected to gain from

alternative activities. In other words, there was an infinitely elastic supply of component producers. This, of course, is not entirely realistic, because matching between buyers and sellers is a complex process that involves risks on both sides. In particular, the quality of a match with a supplier that a final good producer can expect when she outsources depends on the number of potential suppliers in the market and on their expertise, i.e., whether their knowledge and experience are suitable for the manufacturing of the type of intermediate inputs required by her brand.

One simple approach, which places matching between buyers and sellers of intermediate inputs at the heart of the analysis, has been developed by John McLaren (2000) and Grossman and Helpman (2002). In this approach, potential buyers of an intermediate input find it more attractive to outsource the “thicker” the market for the input is, in the sense

⁶⁷ Grossman, Helpman, and Szeidl (2005) use a variant of this model to examine the complementarity between outsourcing and offshoring. In their model, the fraction of offshoring firms is larger the smaller the fixed cost of outsourcing. This is the sense in which offshoring is complementary to outsourcing; as the fixed cost of outsourcing changes, it generates a positive correlation between the fraction of firms that outsource and the fraction of firms that offshore.

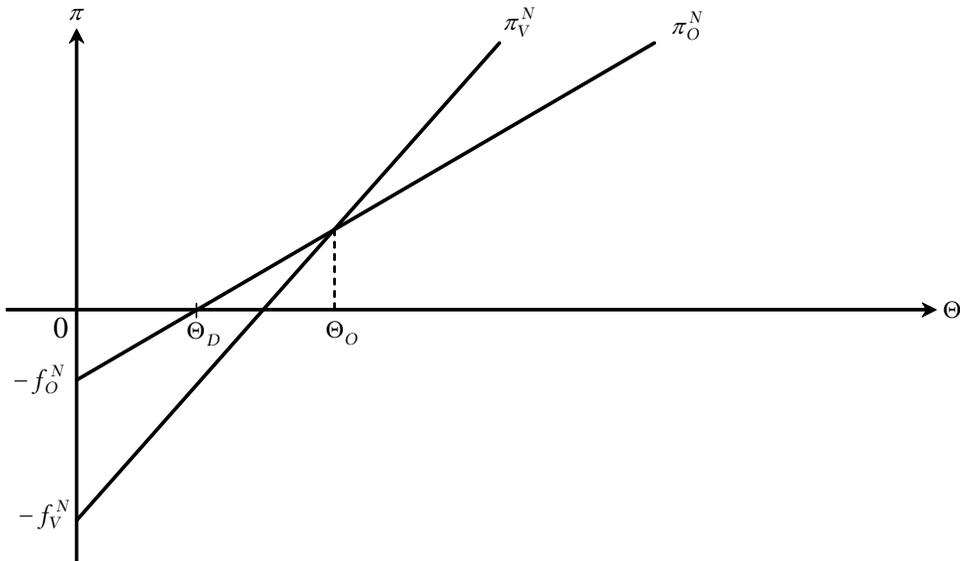


Figure 7. Insourcing and Outsourcing Firms in North in a Headquarter-Intensive Sector

that there exist more sellers to serve the buyers' needs. And similarly, sellers of an intermediate input find it more attractive to operate the larger the number of potential buyers is. Although there can be more than one reason for this type of market externality, both papers use an endogenous probability of successful matching between buyers and sellers as the main driving force of this process. In this type of environment, international trade (or "globalization," using McLaren's terminology) affects the trade-off between outsourcing and integration. In particular, in the presence of economies of scale to matching, trade encourages outsourcing.

Since Grossman and Helpman's (2002) analysis is closer in form to what we have seen in previous sections, and they also show how to deal with these issues in general equilibrium, I will use their framework to illustrate this approach. To this end, consider an industry supplying a differentiated product, in which the demand for variety j is,

as before, $x(j) = Ap(j)^{-\varepsilon}$, where $\varepsilon = 1/(1 - \alpha) > 1$ (i.e., $0 < \alpha < 1$). Now assume that, in order to produce brand j , the manufacturer of the final good needs to acquire an input that is highly specific to this brand. As before, assume that the input has to be tailor-made for brand j , and once it has been tailor-made for j it cannot be used for any other brand, nor can it be put to any other use. For simplicity, assume that one unit of the intermediate input is needed per unit of final good and that no other inputs are required.⁶⁸

First consider a closed economy in which the producer of brand j has two organizational options: she can produce the intermediate good inhouse or outsource. If she

⁶⁸ This is a special case of the production function (6), with $\eta = 0$. It implies that all else equal, the final good producer would like to give the supplier the most powerful incentives possible. As noted in footnote 62, Grossman and Helpman (2002) also develop a richer analysis in which the quality of a match depends on the distance between the final good producer and the supplier in technology space.

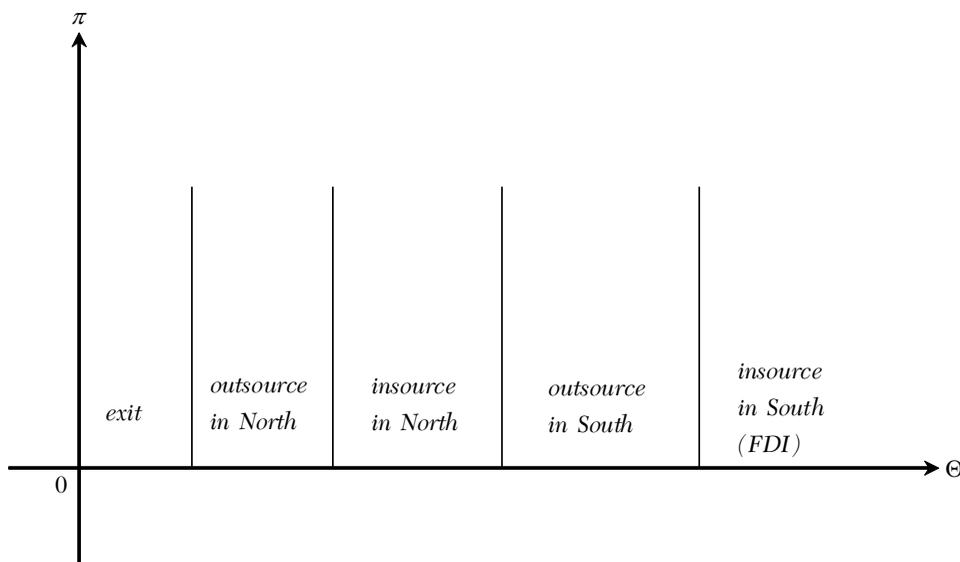


Figure 8. Sorting Pattern in a Headquarter-Intensive Sector

produces inhouse, she needs $1/\theta > 1$ units of labor for every unit of the tailor-made intermediate input, where θ is a measure of productivity, common to all firms. In addition, she has to bear a fixed labor cost f_v , which includes her entry cost (the entry cost covers the acquisition of the technology, the cost of setting up shop, and the like). After entry, her optimal pricing strategy generates a profit level $\pi_v(A)$ which is increasing in the demand level A . It follows that integration is viable in a free entry equilibrium if and only if the demand level equals A_v , at which the integrated firm breaks even; that is, $\pi_v(A_v) = 0$. Obviously, the demand level cannot be higher than A_v , because this would induce entry of additional integrated final good producers, and if it were lower than A_v no final good producer would choose to integrate.

Next consider a final good producer who chooses to outsource. For this she needs to be matched with a supplier of the intermediate input, because inputs with her specialized

needs are not readily available in the market. It is assumed that once she is matched with a supplier, they cannot sign a contract for the delivery of the brand-specific intermediate input. In this event there exists a holdup problem; the supplier can choose how much of the input to produce, but then he has to bargain with the final good producer for payment. A specialized supplier of inputs can produce them with one unit of labor per unit output, which gives him a cost advantage over the integrated firm (which needs $1/\theta > 1$ units of labor per unit output). In the ensuing Nash bargaining, both parties have zero outside options and the final good producer gets a fraction β of the surplus. Evidently, the distribution of the bargaining power between the two parties affects payoffs. Using these payoffs it is then possible to calculate the expected profits of a final good entrant who plans to outsource and the expected profits of an intermediate good producer. These expected profits depend on the probabilities of being

matched and on entry costs, in addition to the payoffs at the bargaining stage.

Let $\mu(N, M)$ be the matching function, which describes the number of matches that take place in a market with N outsourcing final good producers and M producers of intermediate inputs. This function is increasing in both arguments and $\mu(N, M) \leq \min\{N, M\}$. Then the probability of a final good producer being matched is $\mu(N, M)/N$ and the probability of an intermediate good producer being matched is $\mu(N, M)/M$. Once a supplier and a final good producer have been matched, the supplier manufactures an intermediate input tailored to the specific needs of the final good producer. In this model, all suppliers are equally capable of manufacturing every such input. I discuss asymmetries in the qualifications of suppliers below.

When the matching function exhibits constant returns to scale $\mu(N, M)/N = \mu(1, M/N)$ and $\mu(N, M)/M = \mu(N/M, 1)$. We can then use the entry costs of final and intermediate good producers to formulate two free entry conditions: the expected profits (before entry) of an outsourcing final good producer equal her entry cost as an outsourcing enterprise, and the expected profits (before entry) of an intermediate good producer equal his entry cost. These expected profits are functions of the demand level A and the ratio of entrants M/N , that is, $\pi_N(A, M/N)$ and $\pi_M(A, M/N)$. Both expected profits are rising in A , but the final good producer's profits $\pi_N(A, M/N)$ are rising in M/N while the intermediate good producer's profits $\pi_M(A, M/N)$ are declining in M/N . Hence, there is complementarity between entry of intermediate good producers and entry of outsourcing-oriented final good producers. Other things equal, an increase in M raises the expected profits of final good producers while an increase in N raises the expected profits of intermediate good producers. It follows that more entry of one type encourages

more entry of the other type.

Viability of outsourcing in the resulting equilibrium requires zero expected profits for both final and intermediate good producers; that is, $\pi_N(A, M/N) = 0$ and $\pi_M(A, M/N) = 0$. The two free entry conditions are satisfied for unique values of the demand level and the ratio of entrants, say A_O and $r_O = M_O/N_O$.

Grossman and Helpman (2002) show that an equilibrium with integrated firms only always exists, but that it is not stable unless $A_V < A_O$. Namely, stability requires the demand level that ensures zero profits of integrated firms to be lower than the demand level that ensures zero profits of outsourcing final good producers and their suppliers of parts. The reason that an equilibrium with integrated firms only always exists is that, in the absence of suppliers of intermediate inputs, the final good producers' optimal strategy is to enter as integrated manufacturers; and in the absence of outsourcing final good producers the optimal strategy of intermediate good producers is not to enter. This is one consequence of the above-discussed entry complementarity. And Grossman and Helpman also show that there is no mixed equilibrium in which some final good producers insource while others outsource.⁶⁹ Finally, they show that a unique stable equilibrium exists, in which final good producers integrate when $A_V < A_O$ and outsource when $A_V > A_O$.⁷⁰ It follows that structural features determine whether integration or outsourcing prevails.

The analysis so far has focused on the equilibrium organizational form, which does not depend on the size of the economy. Together with a resource constraint, our equilibrium conditions determine the number of

⁶⁹ An exception is a special case in which the parameters of the economy are such that $A_V = A_O$.

⁷⁰ Recall the definitions of the demand levels A_V and A_O ; at A_V an integrated firm just breaks even, while at A_O and a ratio $M/N = r_O$ of entrants an outsourcing firm just breaks even.

entrants. The main implication is that the number of entrants rises proportionately with the size of the economy. Namely, if, say, the labor force doubles, so does the number of entrants in an equilibrium with integrated firms only, as well as in an outsourcing equilibrium. Under these circumstances, the opening of free trade between two countries that differ only in size does not change the equilibrium organizational form and the number of entrants in the world economy. This is a direct consequence of the assumption that the matching function exhibits constant returns to scale, because with this sort of matching technology the probabilities of finding suppliers or buyers of intermediate inputs do not depend on the number of entrants, only on their ratio N/M , and therefore the critical demand levels A_V and A_O do not depend on market size.

In the absence of constant returns to scale in matching the probabilities of a match, $\mu(N, M)/N$ and $\mu(N, M)/M$, depend not only on the ratio of entrants M/N but also on the absolute number of entrants. As a result, there is feedback from country size to organizational form. When $\mu(N, M)$ exhibits increasing returns to scale two stable equilibria can coexist: one with integration, the other with outsourcing. Moreover, in this case outsourcing is more likely the larger the country is, because larger market size raises the probability of a match for both buyers and sellers of intermediate inputs for every ratio of entrants N/M . This implies that opening trade between two countries that differ only in size makes outsourcing more likely. In particular, it is possible to have a situation in which every country in isolation is too small to support an outsourcing equilibrium, yet by opening to trade, the world economy sustains an outsourcing equilibrium (see also McLaren 2000). If the outsourcing equilibrium is unique, it implies that trade changes the organization of production from integration to outsourcing. More generally, increasing returns to

matching imply that market integration encourages outsourcing through the thick-market effect.⁷¹

One drawback of this approach is that, in an outsourcing equilibrium, international trade in intermediate inputs results from the random matches of buyers and sellers from different countries. Although the volume of trade in intermediate inputs is well determined in both directions, it is not related to a deliberate effort of final good producers in one country to seek out suppliers of parts in a different country. In other words, in this model offshoring is not a strategic choice of business firms. The approach described in the next section makes explicit the offshoring decision and introduces a role for different degrees of contract incompleteness. It also introduces natural asymmetries into the matching of final good producers and suppliers of intermediate inputs.

Grossman and Helpman (2003, 2005) develop a different variant of organizational choice under incomplete contracts, in which technological proximity between final good producers and suppliers of intermediate inputs plays a key role. In this model firms choose in which country to search for an outsourcing partner, and countries may differ in their degrees of contract incompleteness. These modifications introduce separate roles for variations across countries in market thickness, legal systems, and other institutional features, as

⁷¹ The above described model is special in many ways. It clarifies, however, the role of market thickness in the link between trade and the organization of production. One of its stark implications is that all firms choose the same organizational form. To avoid this outcome, one can introduce heterogeneity. Thus, for example, one could divide explicitly the fixed costs into entry and operating fixed costs. By paying the fixed entry cost a final good producer would find out its productivity θ , drawn from a known distribution, as in Melitz (2003). After that the final good producer would decide whether to outsource, integrate, or leave the industry. Under these circumstances outsourcing could coexist with integration, whereby low-productivity firms outsource while high-productivity firms integrate.

determinants of the sourcing strategies of business firms.

To understand the basic mechanism of outsourcing in Grossman and Helpman (2003, 2005), consider a simplified version of a closed economy in which integration is not an option. An industry supplies a differentiated product with an isoelastic demand function for every brand $x(j) = Ap(j)^{-\varepsilon}$, $\varepsilon = 1/(1 - \alpha) > 1$. A unit of $x(j)$ is produced with one unit of a tailor-made intermediate input that has no other uses, and it takes one unit of labor to manufacture one unit of the intermediate input by specialized suppliers of parts.

There are N final good producers, each one specializing in a different brand, and M producers of intermediate inputs. Unlike the previous model, however, in which N and M were finite numbers, now M is a finite number while N is a mass. In this formulation each supplier serves many downstream firms. The final good producers are all located on the circumference of a circle of length one. This circumference represents a technology space; a point in this space represents the expertise of an intermediate good producer or the expertise needed by a final good producer for her intermediate input. The finite number of intermediate good producers is symmetrically spaced at distance $1/M$ from each other, while the mass N of final good producers is uniformly distributed with density N at each point on the circumference. I will shortly discuss how these firms found themselves spaced in this way. For now, take these locations as given.⁷²

A final good producer cannot manufacture her product without outsourcing its tailor-made input to a supplier. The cost of

manufacturing an intermediate input has two parts: a variable cost of one unit of labor per unit output plus a fixed cost of customization to the special needs of the final good producer. The cost of customization is proportional to the distance d in technology space between the seller and buyer of the input, say wvd , where w is the wage rate and v is a cost parameter. That is, it is more costly to customize the input when the seller and buyer are far away from each other than when they are close to each other. Under the circumstances every final good producer chooses to source her input from the closest supplier, with the distance d varying between zero and $1/2M$.

It is assumed that the investment in customization has to be made by the producer of the intermediate input, and that this investment is not contractible.⁷³ Moreover, once a final good producer and an intermediate input supplier form a relationship, the final good producer is bound to acquire her input from this partner.⁷⁴

After the investment in customization, the two parties sign an order contract, which stipulates the production of intermediates, assembly of final goods, and the distribution of profits from sales. At this stage both parties seek to maximize joint profits.⁷⁵ This generates a profit π_o that is distributed according to the Nash bargaining weights, which are taken to be $1/2$ for each party.⁷⁶ The profit π_o determines the incentive of the intermediate good producer to customize the input. If $\pi_o/2 \geq wvd$, the

⁷³ I will introduce partial contractibility shortly.

⁷⁴ Here again other options are possible, such as the existence of a secondary market or the use of generic inputs.

⁷⁵ Grossman and Helpman (2005) assume that at this stage, after the customization costs are sunk, there is no further agency problem, i.e., the only agency problem exists at the customization stage. Naturally, one could introduce at this stage too an agency problem in which there is a role for incomplete contracts, similarly to the formulation in sections 3.1 and 3.2.

⁷⁶ It is not difficult to allow more general weights β and $1 - \beta$.

⁷² The circumference of a circle is used in many applications as a space for bilateral matching. In international trade it has been used, for example, by Helpman (1981) for matching buyers and sellers in the presence of monopolistic competition, and by James E. Rauch and Vitor Trindade (2003) for matching buyers and sellers in the presence of informational frictions.

intermediate good producer is willing to customize the product, otherwise he is not, because the expected payoff does not cover investment costs.⁷⁷

In view of these considerations, we can calculate the aggregate postentry profits of an intermediate input producer as the sum (integral) of profits across all final good producers who purchase from him intermediate inputs, call it Π_M . And we can calculate the post-entry profits of a final good producer who finds herself at distance d from the nearest producer of intermediate inputs, call it $\Pi_N(d)$. Then the expected value of $\Pi_N(d)$, where the expectation is taken over the distance d , determines the expected pre-entry profits of a final good producer, call it Π_N . To calculate this expectation, assume that when a final good producer enters the industry she is equally likely to be located at any point on the circumference of the circle. And indeed, ex post, the final good producers are uniformly distributed in this technology space. As for the intermediate good producers, they can each choose their location in the technology space. But in the Nash equilibrium of the entry game they choose equal distances from each other. Moreover, entry of intermediate input producers proceeds until the expected profits Π_M equal the entry cost wf_M , and entry of final good producers proceeds until the expected profits Π_N equal the entry cost wf_N . These conditions together with the resource constraint then determine the equilibrium number of entrants, M and N .

Note that in this model too there is complementarity between entry of the final good and the intermediate input producers; the more entry there is of one type the more profitable is entry of the other type. This is the thick market effect. And the

more suppliers of parts enter the industry the smaller the average distance between buyers and sellers of parts is.

To introduce trade, Grossman and Helpman (2005) consider a two-country (North and South) world, in which final good producers enter only in North and intermediate input producers enter in both the North and the South. As in the closed economy described above, final good producers have to outsource intermediate inputs. But now they have to pay a fee for finding the location of input suppliers in the technology space, and this fee is separate for each country. Therefore, when the search costs for component suppliers are large enough, a final good producer searches in one country only, either in North or in South. This generates segmentation of input markets across countries, and introduces a deliberate decision of where to search for a supplier. This decision involves two considerations, in addition to search costs. First, wages differ across countries, making it attractive to search in the low-wage South, where higher profits can be shared. Second, the number of suppliers of parts differs across countries, making it attractive to search in the country with a larger number of suppliers, where the probability of finding a good match is higher. It follows that if search costs are the same in North and South, the outsourcing of intermediate inputs in both countries can take place only if the number of suppliers is smaller in South.

Grossman and Helpman (2005) characterize a general equilibrium of a trading world of this type and analyze its determinants. They find multiple equilibria. The positive feedback that produces multiple equilibria is the following. As more input suppliers enter a particular country, the country becomes more attractive to final good producers searching for suppliers of parts, because the suppliers are more closely packed there in technology space, making it more likely for a final good producer to find a supplier who will undertake the requisite investment in

⁷⁷ It follows that if $1/M > \pi_0/2wv$, then there exist final good producers who cannot find suppliers for their specialized intermediate inputs, and they exit the industry. This is similar to the presence of an exit cutoff in the models discussed in section 2. The discussion below proceeds under the assumption that this is indeed the case.

customization. Moreover, the larger the number of final good producers searching in a country, the more attractive it is for intermediate input suppliers to set up shop there. For this reason there can be one equilibrium with intermediate inputs produced in both countries, and another equilibrium with intermediate inputs produced only in North.⁷⁸

Focusing on an equilibrium with suppliers of parts in both countries, Grossman and Helpman (2005) derive comparative statics results. An increase in the size of South raises the number of suppliers of intermediate inputs in South and lowers their number in North; the volume of outsourcing rises in South and declines in North; the volume of trade rises relative to income; and the wage rises in South relative to North. That is, unlike in a neoclassical world, here growth in labor of one type does not reduce its relative factor reward. In addition, uniform improvements across countries in the customization technology have no effect on the numbers of input suppliers, the volumes of outsourcing, the relative wage, or the volume of trade relative to income. But improvements in customization that are biased toward South increase the entry of parts suppliers in South, reduce their entry in North, and shift outsourcing from North to South. Moreover, such improvements in technology raise the relative wage of South and the volume of trade relative to income.

One may argue that computer-aided manufacturing and computer-aided design have reduced the cost of customization. If so, then this analysis suggests that the observed patterns of outsourcing and trade expansion cannot be explained by this technological improvement alone, unless we have reason

to believe that it has been particularly effective in reducing customization costs in South.

To discuss the impact of different degrees of contract incompleteness, Grossman and Helpman (2005) extend the model at the customization stage. Instead of assuming that the investment in customization is not contractible, they assume that a fraction of this investment is contractible, and that the supplier of an intermediate input and its potential buyer negotiate an investment contract, which specifies an upfront payment for the contractible part of the investment. As a result, there exists a range of distances d in which customization did not take place before (in the absence of contractibility), but takes place now, and this range is larger the larger the fraction of contractible investment is. It follows that contractibility enlarges the set of active matches.

This generalization has a number of implications. First, starting with no contractibility, the introduction of a positive fraction of contractible customization costs in North increases the number of suppliers of parts in North, reduces their number in South, and raises the relative wage of North. As a result, the volume of outsourcing rises in North, declines in South, and trade relative to income shrinks. Second, an improvement in contracting institutions in South, which raises the contractible fraction of customization costs there, may not expand outsourcing in South. When a significant fraction of customization costs are contractible in North but not in South, initial improvements in contractibility in South raise outsourcing in both South and North, raise South's relative wage, and raise the trade volume relative to income. But once the fraction of contractible costs crosses a threshold, further improvements in contractibility in South reduce outsourcing there, further raise outsourcing in North, reduce the relative wage of South, and diminish the ratio of trade to income. In other words, the response to better contracting institutions in South is not monotonic,

⁷⁸ In a two-country world, the positive feedback is limited by a relative wage response, which stems from the fact that expanding economic activity in a country raises the demand for its labor, which raises in turn its wage relative to the wage in the other country. This general equilibrium effect limits to some extent the concentration of economic activity in one country only, despite the presence of a thick market effect.

and it depends on how far the South's institutions lag behind those of North. This nonmonotonicity has no simple intuitive explanation; it results from a complex interaction between direct effects of changes in the degree of contractibility and indirect effects that operate through labor and product markets in the general equilibrium.

The analysis has so far dealt with outsourcing, where the choice is between the acquisition of intermediate inputs at home (in North) or abroad (in South). Grossman and Helpman (2003) discuss a variant of the same model in which a final good producer can either outsource or integrate, but this trade-off is analyzed at the expense of abandoning the endogeneity of wages and the trade-off between locating the activity in North or South. In particular, they assume that the production of intermediates takes place in South, so that intermediates are offshored, and a firm has to decide only whether to produce its intermediates in a subsidiary or acquire them at arm's-length from an unaffiliated supplier. And they assume constant wages in every country.

The trade-off is the following. As in Grossman and Helpman (2002), an integrated firm has a cost disadvantage in producing intermediates. Therefore, while a specialized supplier of parts needs only one unit of labor per unit of intermediate input, a final good producer needs $1/\theta > 1$ units of labor per unit of intermediate input, where θ is common to all firms. But, the final good producer has a cost advantage in customization; his customization costs are zero while a specialized supplier of parts bears customization costs wvd , which are (as before) proportional to the distance in technology space between him and the producer of the final good. As a result, a final good producer who chooses integration makes profits Π_V , which can be calculated in the usual way.

A final good producer who chooses outsourcing seeks out the closest supplier of parts in technology space, and negotiates with him an investment contract (to be followed by

an order contract after customization takes place). The largest distance d that makes such a relationship viable depends on the degree of contract incompleteness: the larger the contractible fraction of the investment in customization, the larger this distance. It follows that the profits of an outsourcing final good producer depend on how far she is from the closest supplier of parts, $\Pi_N(d)$.

Figure 9 depicts the profits Π_V and $\Pi_N(d)$ as functions of the distance d . Naturally, Π_V is flat, because it does not depend on this distance. But $\Pi_N(d)$ is flat up to d_s , and declines gradually after a downward drop at d_s , where d_s is defined as the largest distance at which the supplier has an incentive to customize the input without an investment contract. The flat part results from the fact that up to distance d_s the supplier's payoff from the order contract, which is independent of the distance d , exceeds the customization costs, in which case no investment contract is signed and the supplier of parts invests in customization nevertheless.⁷⁹ Just slightly above d_s , however, the customization costs wvd exceed the supplier's payoff from the order contract, in which case the supplier of parts does not invest in customization unless an investment contract is signed, and the equilibrium investment contract allocates the customization costs equally between the supplier and the buyer of intermediate inputs. The larger the distance between the two parties the larger the contribution of the final good producer to the customization costs and the smaller her profits.

Under these circumstances there exists a critical distance d_o , which satisfies $\Pi_N(d_o) = \Pi_V$, such that all final good producers with $d < d_o$ prefer to outsource and all final good producers with $d > d_o$ prefer to integrate (or exit if $\Pi_N(d) < 0$). Since d is

⁷⁹ Let P_S be the supplier's payoff from the order contract, the same for all d . Then at d_s this payoff just equals the customization costs wvd_s , i.e., $wvd_s = P_S$.

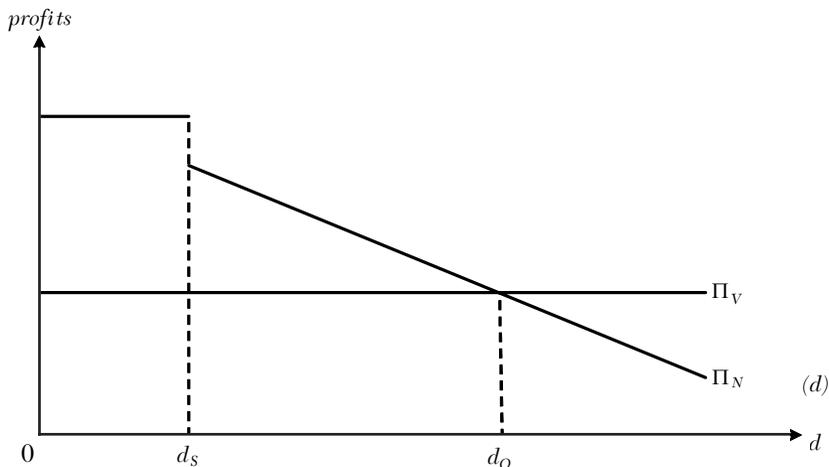


Figure 9. Choice of Integration Versus Outsourcing Depends on Technological Distance from Input Suppliers

random before entry, we can use the uniform distribution of location on the circumference of the circle together with the number of intermediate input producers to calculate the expected profits of a final good producer who enters the industry. Entry proceeds until these expected profits, net of entry cost, equal zero. We can similarly calculate the free entry condition for intermediate good producers.

Grossman and Helpman (2003) find that outsourcing is more prevalent in larger markets, and that the thick market effect is responsible for the positive correlation between market size and the fraction of outsourcing firms and their market share. They also find that better contracting institutions in South, which render larger fractions of the customization costs contractible, increase the prevalence of outsourcing.

Analyzing the trade-off between outsourcing at home or abroad and the trade-off between outsourcing or integration, provides useful insights but it gives only a partial

view of organizational choices. A complete analysis requires simultaneously allowing firms to choose between outsourcing and integration in every country, thereby admitting an interaction between the offshoring decision and the internalization decision. No such analysis exists for the class of models discussed in this section, only for the model discussed in section 3.3.

3.5 Ricardian Comparative Advantage

Differences across countries in legal systems and institutions that shape the enforcement of contracts, and thereby the degree of contract incompleteness, have the potential for influencing patterns of comparative costs across countries. Ricardian comparative advantage, as reflected in the cross sectoral variation in productivity levels, can arise as a result of institutional variation across countries when the relative requirement of contract-dependent inputs varies across sectors, because institutions impact costs in sectors with a larger need for contract-

dependent inputs relatively more than in sectors with less need for contract-dependent inputs.

Nunn (forthcoming) provides a detailed empirical analysis of the impact of the degree of contract incompleteness on international trade flows.⁸⁰ As the main representative of the degree of contract incompleteness, he uses a measure of the rule of law, which consists of a weighted average of a number of variables that gauge the effectiveness of the judiciary, its predictability, and its enforcement of contracts.⁸¹ He finds that the results do not change much when this variable is replaced with other, more objective measures of the efficacy of courts.⁸² To compute an index of contract dependence for every final good sector, Nunn uses U.S. input–output tables to compute the proportion of intermediate inputs used in every final good, and he classifies intermediates into those that are traded on an organized exchange, those that have a reference price, and those that have none of these.⁸³ He assumes that a good is more contract dependent the larger is the fraction of its intermediate inputs that have no organized exchange nor a reference price, or alternatively, the larger is the fraction of its intermediate inputs that have no organized exchange only. The main empirical finding is that countries with better legal systems export relatively more in sectors that are more intensive in contract-dependent inputs. This finding is robust to controls

of other determinants of trade flows, alternative specifications of the estimated equation, and alternative estimation methods. Moreover, not only has the quality of the legal system a statistically significant impact on trade flows, it also has a large economic impact. In particular, its impact, as measured by the beta coefficient, is of similar magnitude to that of human capital and physical capital combined. In other words, contracting institutions are an important source of comparative advantage.

Daron Acemoglu, Antràs, and Helpman (2006) propose a model in which Ricardian comparative advantage emerges from the interaction of contract incompleteness with the deliberate choice of technology by final good producers. In their model, a final good producer can choose how to divide the production process, so as to have many or few intermediate inputs. Every intermediate input has to engage a supplier. The supplier of the input, who can be a worker in the firm or an outside supplier, has to execute a set of activities in order to produce it. A subset of these activities are contractible, while the others are not. The fraction of noncontractible activities provides a measure of contract incompleteness.

On the one hand, more sophisticated technologies—that allow more intermediate inputs in the production process—are more costly to acquire, and they may involve larger organizational costs. On the other hand, more sophisticated technologies are more productive. Using this trade-off, a final good producer makes an optimal technology choice, and this choice depends on features of the industry and the degree of contract incompleteness. Acemoglu, Antràs, and Helpman (2006) find that better contracting institutions lead to the choice of more sophisticated technologies, and that the impact of contracting institutions on technology choice is relatively larger in sectors with lower elasticities of substitution across intermediate inputs, because low substitutability makes the sector more sensitive to

⁸⁰ Levchenko (2004), who preceded Nunn (forthcoming), makes related arguments. I focus on Nunn (forthcoming), however, because he provides the sharper empirical analysis. Both Levchenko and Nunn develop simple theoretical models to guide their empirical work.

⁸¹ These variables are estimated from subjective perceptions; see Daniel Kaufmann, Aart Kraay, and Massimo Mastruzzi (2003).

⁸² These objective measures are from Simeon Djankov, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2003) and they are available for a smaller sample of countries.

⁸³ The classification of inputs into those that have an organized exchange, those that have a reference price, and those that have no organized exchange and no reference price is from Rauch (1999).

contractual frictions. As a result, countries with better contracting institutions have a relative productivity advantage, and therefore comparative advantage, in sectors with less substitutable inputs.

Arnaud Costinot (2005) proposes a different model in which contracting institutions interact with technological features to create Ricardian comparative advantage. In his model, every industry is characterized by a set of tasks that have to be performed and these sets are exogenous. Industries are ordered by the complexity of their technology, which is measured by the number of tasks in their set. Workers are assigned to tasks. A worker has to spend a fixed amount of time to learn a particular task. As a result, there are increasing returns to scale in the performance of tasks. But a worker can shirk and not perform his task. In the event of shirking, no output is produced because every task is essential. The degree of contract incompleteness is measured by the probability that a worker shirks, which is exogenous and independent across workers.

When a team of workers produces a product, it is efficient to assign every worker the same number of tasks. Given the size of the team, it is then possible to compute expected output per worker. The resulting optimal team size, which maximizes output per worker, is larger in more complex industries and in countries with better legal institutions, in which contracts are enforced with higher probabilities. In a competitive economy, better institutions raise output per worker proportionately more in more complex industries. As a result, a country with better contract-enforcing institutions gains a comparative advantage in more complex industries.

4. *Concluding Remarks*

New developments in the world economy have called for new developments in the theory of international trade and foreign direct investment, designed to better understand the shifts in trade and investment

patterns and the reorganization of production across national borders. Although traditional trade theory has much to offer in explaining parts of this puzzle, such as the international fragmentation of production, the theory had to be generalized in order to provide a better understanding of the trends in the data.⁸⁴ Particularly acute has been the need to model different forms and degrees of involvement of business firms in foreign activities because organizational change has been a central element in the transformation of the world economy. As a result, theoretical refinements have focused on the individual firm, studying its choices in response to its own characteristics, the nature of the industry in which it operates, and the opportunities afforded by foreign trade and investment. Important among these choices are modes of serving foreign markets and sourcing strategies.

But the theory went beyond the individual firm, studying the implications of firm behavior for the structure of an industry and, by implication, structural differences across industries. These variations deliver new explanations for trade structure and patterns of FDI, both within and across industries. For example, they identify new sources of comparative advantage, such as the degree of heterogeneity within industries and the quality of contracting institutions.

Heterogeneity plays a key role in this literature in two ways. First, there is heterogeneity as a result of productivity differences across firms within industries, because some firms happen to be luckier than others.⁸⁵ Second, there is heterogeneity in organizational form. The two are related, however, because differences in productivity induce different choices for the organization of production and distribution. In this theory, trade and FDI patterns are jointly deter-

⁸⁴ See Ronald W. Jones (2000).

⁸⁵ The empirical literature supports the view that causality goes from productivity to, say, exports, rather than the other way around; see, for example, Bernard and Jensen (1999) and Clerides, Lach, and Tybout (1998).

mined with organizational structures, such as sourcing and integration strategies.

Some implications of these models have been tested empirically. Examples include the sorting patterns of firms into exporters and foreign direct investors. Other implications have not been tested. These include patterns of sorting into outsourcing at home, integration at home, outsourcing abroad, and integration abroad, because this cannot be done with the available data sets. More generally, hypotheses that require detailed firm-level data about trade in different types of products, such as intermediate inputs versus final goods, and whether this trade takes place within the boundary of the firm or at arm's-length, cannot be examined. The theoretical models point out, however, what additional data need to be collected in order to improve the empirical analysis. Since this is still a lively area of research, we can expect to see much more theoretical and empirical work on these topics.

REFERENCES

- ▶ Abraham, Katharine G., and Susan K. Taylor. 1996. "Firms' Use of Outside Contractors: Theory and Evidence." *Journal of Labor Economics*, 14(3): 394–424.
- Acemoglu, Daron, Pol Antràs, and Elhanan Helpman. 2006. "Contracts and Technology Adoption." Harvard University. Mimeo
- Amiti, Mary, and Shang-Jin Wei. 2005. "Fear of Service Outsourcing: Is It Justified?" *Economic Policy*, 42: 307–39.
- ▶ Anderson, James E., and Eric van Wincoop. 2004. "Trade Costs." *Journal of Economic Literature*, 42(3): 691–751.
- ▶ Antràs, Pol. 2003. "Firms, Contracts, and Trade Structure." *Quarterly Journal of Economics*, 118(4): 1375–1418.
- ▶ Antràs, Pol. 2005. "Incomplete Contracts and the Product Cycle." *American Economic Review*, 95(4): 1054–73.
- ▶ Antràs, Pol, and Elhanan Helpman. 2004. "Global Sourcing." *Journal of Political Economy*, 112(3): 552–80.
- Aw, Bee Yan, Sukkyun Chung, and Mark J. Roberts. 2000. "Productivity and Turnover in the Export Market: Micro-Level Evidence from the Republic of Korea and Taiwan (China)." *World Bank Economic Review*, 14(1): 65–90.
- ▶ Axtell, Robert L. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science*, 293(5536): 1818–20.
- ▶ Baldwin, John R., and Wulong Gu. 2003. "Export-Market Participation and Productivity Performance in Canadian Manufacturing." *Canadian Journal of Economics*, 36(3): 634–57.
- Baldwin, Richard E., 1988. "Hysteresis in Import Prices: The Beachhead Effect." *American Economic Review*, 78(4): 773–85.
- Baldwin, Richard E., 2005. "Heterogeneous Firms and Trade: Testable and Untestable Properties of the Melitz Model." NBER Working Papers, no. 11471.
- ▶ Baldwin, Richard E., and Paul R. Krugman. 1989. "Persistent Trade Effects of Large Exchange Rate Shocks." *Quarterly Journal of Economics*, 104(4): 635–54.
- Bamford, James (1994), "Driving America to Tiers," *Financial World*, November 8, pp. 24–27.
- Bardi, Edward J., and Michael Tracey. 1991. "Transportation Outsourcing: A Survey of U.S. Practices." *International Journal of Physical Distribution and Logistics Management*, 21(3): 15–21.
- Bartel, Ann P., Saul Lach, and Nachum Sicherman. 2005. "Outsourcing and Technological Change." NBER Working Papers, no. 11158.
- ▶ Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum. 2003. "Plants and Productivity in International Trade." *American Economic Review*, 93(4): 1268–90.
- ▶ Bernard, Andrew B., and J. Bradford Jensen. 1999. "Exceptional Exporter Performance: Cause, Effect, or Both?" *Journal of International Economics*, 47(1): 1–25.
- ▶ Bernard, Andrew B., and J. Bradford Jensen. 2004. "Why Some Firms Export." *Review of Economics and Statistics*, 86(2): 561–69.
- Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. 2005. "Importers, Exporters, and Multinationals: A Portrait of Firms in the U.S. That Trade Goods." NBER Working Papers, no. 11404.
- Bernard, Andrew B., Stephen Redding, and Peter K. Schott. Forthcoming. "Comparative Advantage and Heterogeneous Firms." *Review of Economic Studies*.
- ▶ Bhagwati, Jagdish, Arvind Panagariya, and T. N. Srinivasan. 2004. "The Muddles over Outsourcing." *Journal of Economic Perspectives*, 18(4): 93–114.
- Bolton, Patrick, and Mathias Dewatripont. 2005. *Contract Theory*. Cambridge and London: MIT Press.
- Borga, Maria, and William J. Zeile. 2004. "International Fragmentation of Production and the Intrafirm Trade of U.S. Multinational Companies." Bureau of Economic Analysis Working Paper WP2004-02.
- Bustos, Paula. 2005. "Rising Wage Inequality in the Argentinean Manufacturing Sector: The Impact of Trade and Foreign Direct Investment on Technology and Skill Upgrading." Harvard University. Mimeo.
- Campa, Jose, and Linda S. Goldberg. 1997. "The Evolving External Orientation of Manufacturing: A Profile of Four Countries." *Federal Reserve Bank of New York Economic Policy Review*, 3(2): 53–81.
- ▶ Campbell, Jeffrey R., and Hugo A. Hopenhayn. 2005. "Market Size Matters." *Journal of Industrial Economics*, 53(1): 1–25.
- ▶ Clerides, Sofronis K., Saul Lach, and James R. Tybout. 1998. "Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco." *Quarterly Journal of Economics*, 113(3):

- 903–47.
- Costinot, Arnaud. 2005. "Contract Enforcement, Division of Labor, and the Pattern of Trade." Department of Economics, Princeton University. Mimeo.
- Das, Sanghamitra, Mark J. Roberts, and James R. Tybout. 2001. "Market Entry Costs, Producer Heterogeneity, and Export Dynamics." NBER Working Papers, no. 8629.
- Davis, Donald R., and David E. Weinstein. 2001. "An Account of Global Factor Trade." *American Economic Review*, 91(5): 1423–53.
- ▶ Delgado, Miguel A., Jose C. Fariñas, and Sonia Ruano. 2002. "Firm Productivity and Export Markets: A Non-Parametric Approach." *Journal of International Economics*, 57(2): 397–422.
- ▶ Dixit, Avinash K. 1989. "Hysteresis, Import Penetration, and Exchange Rate Pass-Through." *Quarterly Journal of Economics*, 104(2): 205–28.
- ▶ Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2003. "Courts." *Quarterly Journal of Economics*, 118(2): 453–517.
- ▶ Eaton, Jonathan, Samuel Kortum, and Francis Kramarz. 2004. "Dissecting Trade: Firms, Industries, and Export Destinations." *American Economic Review*, 94(2): 150–54.
- Ekhholm, Karolina, Rikard Forslid, and James R. Markusen. 2004. "Export-Platform Foreign Direct Investment." University of Colorado. Mimeo.
- ▶ Fan, Joseph P. H., and Larry H. P. Lang. 2000. "The Measurement of Relatedness: An Application to Corporate Diversification." *Journal of Business*, 73(4): 629–60.
- Feenstra, Robert C. 1998. "Integration of Trade and Disintegration of Production in the Global Economy." *Journal of Economic Perspectives*, 12(4): 31–50.
- Feenstra, Robert C. 2004. *Advanced International Trade: Theory and Evidence*. Princeton and Oxford: Princeton University Press.
- Feenstra, Robert C., and Gordon H. Hanson. 1996. "Globalization, Outsourcing, and Wage Inequality." *American Economic Review*, 86(2): 240–45.
- ▶ Feenstra, Robert C., and Gordon H. Hanson. 2005. "Ownership and Control in Outsourcing to China: Estimating the Property-Rights Theory of the Firm." *Quarterly Journal of Economics*, 120(2): 729–61.
- Feenstra, Robert C., and Barbara J. Spencer. 2005. "Contractual versus Generic Outsourcing: The Role of Proximity." University of California, Davis. Mimeo.
- Feinberg, Susan E., and Michael P. Keane. 2003. "Accounting for the Growth of MNC-Based Trade Using a Structural Model of U.S. MNCs." Yale University. Mimeo.
- Feinberg, Susan E., and Michael P. Keane. 2005. "Tariff Effects on MNC Decisions to Engage in Intra-firm and Arms-Length Trade." Yale University. Mimeo.
- Gardner, Elizabeth (1991), "Going on Line with Outsiders," *Modern Healthcare* 21 pp. 35–47.
- ▶ Girma, Sourafel, Holger Görg, and Eric Strobl. 2004. "Exports, International Investment, and Plant Performance: Evidence from a Non-Parametric Test." *Economics Letters*, 83(3): 317–24.
- ▶ Girma, Sourafel, Richard Kneller, and Mauro Pisu. 2005. "Exports versus FDI: An Empirical Test." *Review of World Economics/Weltwirtschaftliches Archiv*, 141(2): 193–218.
- ▶ 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. 2003. "Outsourcing versus FDI in Industry Equilibrium." *Journal of the European Economic Association*, 1(2–3): 317–27.
- ▶ Grossman, Gene M., and Elhanan Helpman. 2004. "Managerial Incentives and the International Organization of Production." *Journal of International Economics*, 63(2): 237–62.
- ▶ Grossman, Gene M., and Elhanan Helpman. 2005. "Outsourcing in a Global Economy." *Review of Economic Studies*, 72(1): 135–59.
- ▶ Grossman, Gene M., Elhanan Helpman, and Adam Szeidl. 2005a. "Complementarities between Outsourcing and Foreign Sourcing." *American Economic Review*, 95(2): 19–24.
- Grossman, Gene M., Elhanan Helpman, and Adam Szeidl. Forthcoming. "Optimal Integration Strategies for the Multinational Firm." *Journal of International Economics*.
- ▶ Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 94(4): 691–719.
- Hanson, Gordon H., Raymond J. Mataloni Jr., and Matthew J. Slaughter. 2001. "Expansion Strategies of U.S. Multinational Firms." In *Brookings Trade Forum: 2001*, ed. Susan M. Collins and Dani Rodrik. Washington, D.C.: Brookings Institution, 245–94.
- ▶ Hanson, Gordon H., Raymond J. Mataloni Jr., and Matthew J. Slaughter. 2005. "Vertical Production Networks in Multinational Firms." *Review of Economics and Statistics*, 87(4): 664–78.
- Hart, Oliver D., 1995. *Firms, Contracts, and Financial Structure*. Clarendon Lectures in Economics. Oxford and New York: Oxford University Press, Clarendon Press.
- ▶ Hart, Oliver D., and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy*, 98(6): 1119–58.
- ▶ Head, Keith, and John Ries. 2003. "Heterogeneity and the FDI versus Export Decision of Japanese Manufacturers." *Journal of the Japanese and International Economics*, 17(4): 448–67.
- ▶ Helper, Susan. 1991. "Strategy and Irreversibility in Supplier Relations: The Case of the U.S. Automobile Industry." *Business History Review*, 65(4): 781–824.
- ▶ Helpman, Elhanan. 1981. "International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition: A Chamberlin–Heckscher–Ohlin Approach." *Journal of International Economics*, 11(3): 305–40.
- Helpman, Elhanan, and Paul R. Krugman. 1985. *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*. Cambridge: MIT Press.
- Helpman, Elhanan, Marc J. Melitz, and Yona Rubinstein. 2006. "Trading Partners and Trading Volumes." Harvard University. Mimeo.

- ▶ Helpman, Elhanan, Marc J. Melitz, and Stephen Ross Yeaple. 2004. "Export versus FDI with Heterogeneous Firms." *American Economic Review*, 94(1): 300–316.
- ▶ Holmes, Thomas J. 1999. "Localization of Industry and Vertical Disintegration." *Review of Economics and Statistics*, 81(2): 314–25.
- Holmstrom, Bengt, and Paul Milgrom. 1994. "The Firm as an Incentive System." *American Economic Review*, 84(4): 972–91.
- ▶ Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, 60(5): 1127–50.
- ▶ Hubbard, Thomas N. 2001. "Contractual Form and Market Thickness in Trucking." *RAND Journal of Economics*, 32(2): 369–86.
- ▶ Hummels, David, Jun Ishii, and Kei-Mu Yi. 2001. "The Nature and Growth of Vertical Specialization in World Trade." *Journal of International Economics*, 54(1): 75–96.
- ▶ Jean, Sébastien. 2002. "International Trade and Firms' Heterogeneity under Monopolistic Competition." *Open Economies Review*, 13(3): 291–311.
- Jones, Ronald W. 2000. *Globalization and the Theory of Input Trade*. Ohlin Lectures, Vol. 8. Cambridge and London: MIT Press.
- Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2003. "Government Matters III: Governance Indicators for 1996–2002." World Bank. Mimeo.
- Landefeld, J. Steven, and Raymond J. Mataloni Jr. "Offshore Outsourcing and Multinational Companies," presentation at the Brookings Institution, June 22, 2004.
- Levchenko, Andrei A. 2004. "Institutional Quality and International Trade." IMF Working Paper 04/231.
- Marin, Dalia, and Thierry Verdier. 2005. "Corporate Hierarchies and International Trade: Theory and Evidence." University of Munich. Mimeo.
- McLaren, John. 2000. "'Globalization' And Vertical Structure." *American Economic Review*, 90(5): 1239–54.
- ▶ Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.
- Melitz, Marc J., and Gianmarco I. P. Ottaviano. 2005. "Market Size, Trade, and Productivity." NBER Working Papers, no. 11393.
- ▶ Montagna, Catia. 2001. "Efficiency Gaps, Love of Variety and International Trade." *Economica*, 68(269): 27–44.
- ▶ Neary, J. Peter. 2003. "Presidential Address: Globalization and Market Structure." *Journal of the European Economic Association*, 1(2–3): 245–71.
- Nunn, Nathan. Forthcoming. "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade." *Quarterly Journal of Economics*.
- ▶ Ottaviano, Gianmarco I. P., Takatoshi Tabuchi, and Jacques-François Thisse. 2002. "Agglomeration and Trade Revisited." *International Economic Review*, 43(2): 409–35.
- Puga, Diego, and Daniel Treffer. 2002. "Knowledge Creation and Control in Organizations." NBER Working Papers, no. 9121.
- ▶ Qiu, Larry D., and Barbara J. Spencer. 2002. "Keiretsu and Relationship-Specific Investment: Implications for Market-Opening Trade Policy." *Journal of International Economics*, 58(1): 49–79.
- ▶ Rauch, James E. 1999. "Networks versus Markets in International Trade." *Journal of International Economics*, 48(1): 7–35.
- ▶ Rauch, James E., and Vitor Trindade. 2003. "Information, International Substitutability, and Globalization." *American Economic Review*, 93(3): 775–91.
- Roberts, Mark J., and James R. Tybout. 1997. "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs." *American Economic Review*, 87(4): 545–64.
- ▶ Spencer, Barbara J. 2005. "International Outsourcing and Incomplete Contracts." *Canadian Journal of Economics*, 38(4): 1107–35.
- ▶ Spencer, Barbara J., and Larry D. Qiu. 2001. "Keiretsu and Relationship-Specific Investment: A Barrier to Trade?" *International Economic Review*, 42(4): 871–901.
- Strauss-Kahn, Vanessa. 2003. "The Role of Globalization in the Within-Industry Shift Away from Unskilled Workers in France." NBER Working Papers, no. 9716.
- ▶ Syverson, Chad. 2004. "Product Substitutability and Productivity Dispersion." *Review of Economics and Statistics*, 86(2): 534–50.
- Syverson, Chad. 2005. "Prices, Spatial Competition, and Heterogeneous Producers: An Empirical Test." University of Chicago. Mimeo.
- Tempest, Rone. "Barbie and the World Economy," *Los Angeles Times*, September 22, 1996, A1 and A12.
- The Economist* (1991), "The Ins and Outs of Outsourcing," August 31, pp. 54–56.
- Trefler, Daniel. 1995. "The Case of the Missing Trade and Other Mysteries." *American Economic Review*, 85(5): 1029–46.
- ▶ Trefler, Daniel. 2004. "The Long and Short of the Canada–U.S. Free Trade Agreement." *American Economic Review*, 94(4): 870–95.
- ▶ Tybout, James R., and M. Daniel Westbrook. 1995. "Trade Liberalization and the Dimensions of Efficiency Change in Mexican Manufacturing Industries." *Journal of International Economics*, 39(1–2): 53–78.
- United Nations Conference on Trade and Development. 1998. *World Investment Report: Trends and Determinants*. New York and Geneva: United Nations Publications.
- United Nations Conference on Trade and Development. 2002. *World Investment Report: Transnational Corporations and Export Competitiveness*. New York and Geneva: United Nations Publications.
- United Nations Conference on Trade and Development. 2004. *World Investment Report: The Shift towards Services*. New York and Geneva: United Nations Publications.
- ▶ Vernon, Raymond. 1966. "International Investment and International Trade in the Product Cycle." *Quarterly Journal of Economics*, 80(2): 190–207.
- World Trade Organization. 1998. *Annual Report 1998*. Geneva: World Trade Organization.
- World Trade Organization. 2004. *Annual Report 2004*.

- Geneva: World Trade Organization.
- ▶ Yeaple, Stephen Ross. 2003. "The Complex Integration Strategies of Multinationals and Cross Country Dependencies in the Structure of Foreign Direct Investment." *Journal of International Economics*, 60(2): 293–314.
 - ▶ Yeaple, Stephen Ross. 2005. "A Simple Model of Firm Heterogeneity, International Trade, and Wages." *Journal of International Economics*, 65(1): 1–20.
 - Yeats, Alexander J. 2001. "Just How Big Is Global Production Sharing?" In *Fragmentation: New Production Patterns in the World Economy*, ed. Sven W. Arndt and Henryk Kierzkowski. Oxford and New York: Oxford University Press.