Chapter 2

Workhorse Models

The field of International Trade has witnessed a true revolution in recent years. Firms rather than countries or industries are now the central unit of analysis. The workhorse trade models used by most researchers both in theoretical work as well as in guiding empirical studies were published in the 2000s. The purpose of this Chapter is to provide a succinct account of the rich intellectual history of the field and to offer an overview of these modern workhorse models. While these benchmark frameworks ignore contractual aspects, they constitute the backbone of the models developed later in the book, so it is important to gain an understanding of their key features.

Two Centuries of Trade Theory

The recent revolution in the International Trade field would perhaps not be apparent when browsing the leading undergraduate textbooks covering the basics of international trade and investment. Neoclassical trade theory still constitutes the core of what we teach to college students. This should not be surprising: the concept of comparative advantage is as relevant today as it was almost two hundred years ago when David Ricardo initiated the formal modeling of foreign trade in his *Principles of Political Economy and Taxation* (1817). The first hundred and seventy years of the International Trade field were largely devoted to refining Ricardo's rudimentary description of the gains from specialization. The benchmark two-good, two-country Ricardian model found in most introductory textbooks is the culmination of an intellectual endeavor to which John Stuart Mill, Frank Graham, and Lionel McKenzie contributed key advances.

Starting with the work of Eli Heckscher and his disciple Bertil Ohlin, another branch of the neoclassical theory studied models in which comparative advantage is endogenously shaped by the interaction of differences in relative factor abundance across countries and differences in relative factor intensities across sectors. The formalization of the so-called Heckscher-Ohlin model is often associated with the great Paul Samuelson, but Abba Lerner, Ronald Jones and Alan Deardorff should also be credited for particularly significant contributions.[1]

The core theorems of neoclassical trade theory – the Heckscher-Ohlin, the Stolper-Samuelson and the Rybczinski theorems – are the product of these intellectual efforts. These beautiful and incredibly sharp results still shape to date the way that most economists think about the determinants and consequences of international trade flows. Why is China the single largest exporter to the U.S.? How does trade with China affect the relative pay of skilled and unskilled workers in the U.S.? How does immigration affect sectoral employment in the U.S.? You would be hard-pressed to answer these questions without appealing to the insights of neoclassical theory.

Neoclassical trade models deliver sharp results but also make strong assumptions. The benchmark models assume a very low number of goods and factors, often only two of each. In higher-dimensional environments, the classical theorems become much less beautiful and much less sharp.[2] More importantly, in neoclassical models, technology is typically assumed to feature constant returns to scale and market structure is characterized by perfect competition, thus making these frameworks of limited use for firm-level studies of international trade. Indeed, in neoclassical trade theory it is not firms but rather countries that trade with each other.

Trade theory witnessed a first revolution in the late 1970s and early 1980s when a group of young trade economists, led by Paul Krugman and Elhanan Helpman, developed new models attempting to account for some empirical patterns that were hard to reconcile with neoclassical theory. Most notably,

[1] A lucid exposition of neoclassical trade theory with extensive references can be found in Jones and Neary (1984).

[2] It is important to emphasize, however, that the implications of the theory for the net factor content of trade – the so-called Vanek (1968) equations – have been shown to be robust to variation in the number of goods and factors. It is no surprise then that beginning with the seminal work of Leamer (1984), empirical testing of the Heckscher-Ohlin model has largely focused on these factor content predictions (see Trefler, 1993*a*, 1995, Davis and Weinstein, 2001 and Trefler and Zhu, 2010).

traditional theory rationalized the existence of mutually beneficial intersectoral trade flows stemming from cross-country differences in technology or endowments. In the real world, however, the bulk of trade flows occurs between countries with similar levels of technological development and similar relative factor endowments, and a significant share of world trade is accounted for by two-way flows within fairly narrowly-defined sectors (i.e., 'intraindustry' instead of intersectoral trade).

This new wave of research, which was dubbed 'new trade theory', emphasized the importance of increasing returns to scale, imperfect competition, and product differentiation in accounting for these salient features of the data. Intuitively, even two completely identical countries will find it mutually beneficial to trade with each other as long as specializing in particular *differentiated* varieties of a sector's goods allows producers to expand their sales and operate at lower average costs, as would naturally be the case whenever technology features *economies of scale*. The relevance of *imperfect competition* for these theories stems from the simple fact that (internal) economies of scale are inconsistent with perfect competition.

A key hurdle facing the pioneers of new trade theory was the absence of a generally accepted modeling of product differentiation and imperfect competition. While there is only one way in which goods can be perfectly homogeneous, there are many ways in which products can be differentiated. Differentiation can arise because individual consumers enjoy spreading their income across different varieties of particular goods (as in the case of cultural goods), or because different consumers prefer to consume different varieties or qualities of the same good (as with tablets or cars). Even when focusing on one of these modeling approaches, there remains the issue of how to mathematically characterize product differentiation in preferences. Similarly, there is only one way in which markets can be perfectly competitive, while there are various possible approaches to modeling imperfect competition.

There are two main reasons why 'new trade theory' was able to overcome these difficulties and become mainstream in a relatively short period of time. First, researchers quickly converged in the use of *a particular* modeling of product differentiation and market structure associated with Krugman (1979, 1980), who in turn borrowed from Dixit and Stiglitz (1977). This served the important role of providing a common language for researchers in the field to communicate among themselves. Still, the heavy use of specific functional forms in representing preferences and technology was viewed with some reservations by the old guard in the field.[3]

The second key factor in the success of new trade theory was the publication of a landmark treatise by Helpman and Krugman (1985). This concise book established the generality of most of the insights from Krugman's work and also illustrated how the new features of new trade theory could be embedded into neoclassical trade theory. As a result, these new hybrid models could explain the features of the data that motivated the new models, while at the same time preserving the validity of some of the salient results from neoclassical theory, such as the Vanek (1968) equations characterizing the factor content of trade. With the publication of this manuscript, the walls of resistance came tumbling down, new trade theory became the new paradigm, and Krugman's modeling choices gained a prominent spot in the toolbox of trade theorists (and of applied theorists in other fields).

In recent years, international trade theory has witnessed a second revolution which in many respects parallels the one witnessed thirty years ago. As in the case of new trade theory, and consistently with Kuhn's (1996) description of the structure of scientific revolutions, the need for a new paradigm was fueled by the discovery of a series of new empirical facts that were inconsistent with new trade theory models. To understand these inconsistencies, it is important to note that in Krugman-style models, all firms within a sector are treated symmetrically. Although firms produce differentiated products, they do so under a common cost function, and all varieties enter symmetrically into demand with an elasticity of substitution between any pair of varieties that is constant and common for any pair. As a result, firm behavior within an industry is 'homogeneous'. Furthermore, under the common assumption of iceberg (or ad valorem) trade costs, new trade theory models deliver the stark implication that all firms within a differentiated-good sector will export their output to every single country in the world.

In the 1990s, a wave of empirical papers using newly-available longitudinal plant and firm-level data from various countries demonstrated the existence of significant levels of heterogeneity in revenue, productivity, factor inputs, and trade behavior across firms within sectors. In fact, in some cases, heterogeneity in performance was shown to be almost as large within sectors

[3] As an illustration of this resistance, Krugman's 1979 seminal article was rejected by the *Quarterly Journal of Economics* in 1978 and was subsequently salvaged by Jagdish Bhagwati at the *Journal of International Economics* despite two negative referee reports (see Gans and Shepherd, 1994; note however that Ethier, 2001, offers a slightly less glorifying account of Bhagwati's role in rescuing the paper at the JIE).

than across sectors (see, for instance, Bernard et al., 2003). With regards to export behavior, studies found that only a small fraction of firms engage in exporting, and that most exporting firms sell only to a few markets. This so-called extensive margin of trade has been shown to be important in order to understand variation in aggregate exports across destination markets. Several studies have also documented that exporters appear to be systematically different from non-exporters: they are larger, more productive, and operate at higher capital and skill intensities. In addition, firm heterogeneity has been shown to be of relevance for assessing the effects of trade liberalization, as those episodes appear to lead to market share reallocations towards more productive firms, thereby fostering aggregate productivity via new channels.

Access to micro-level data has also served to confirm the importance of multinational firms in world trade. For instance, according to 2009 data from the Bureau of Economic Analysis, 75 percent of the sales by U.S. firms in foreign markets is carried out by foreign affiliates of U.S. multinational enterprises (MNEs), and only 25 percent by exports from the U.S. (Antràs and Yeaple, 2013). Furthermore, not only do intrafirm trade flows constitute a very significant share of world trade flows (as mentioned in Chapter 1), but an important share of the volume of arm's-length international trade is accounted for by transactions involving multinational firms as buyers or sellers. For instance, data from the U.S. Census Bureau indicates that roughly 90 percent of U.S. exports and imports flow through multinational firms (Bernard et al., 2009). New trade theory did not ignore the importance of multinational firms or intrafirm trade in the world economy (see Helpman, 1984, or Helpman and Krugman, 1985, Chapter 12 and 13), but by focusing on complete-contracting, homogenous-firm models, it was unable to account for central aspects of multinational activity, such as the rationale for internalizing foreign transactions and the existence of heterogeneous participation of firms in FDI (or affiliate) sales and in global sourcing.[4]

Motivated by these new empirical findings, recent trade theory has been developed in frameworks that incorporate intraindustry firm heterogeneity. The seminal paper in the literature is that of Melitz (2003), which follows closely the structure of Krugman (1980). Although Melitz's framework fea-

[4]The fact that firms engaged in FDI sales and in importing appear to be distinct from other firms has been documented, among others, by Helpman, Melitz and Yeaple (2004) and Bernard, Jensen and Schott (2009). In addition, Ramondo, Rappoport and Ruhl (2013) have recently documented that U.S. intrafirm trade appears to be highly concentrated among a small number of large foreign affiliates.

tures no multinational activity, no global sourcing and no contractual frictions, it is natural to begin our incursion into theoretical territory with a variant of his model.

A Multi-Sector Melitz Model

Consider a world consisting of J countries that produce goods in $S+1$ sectors using a unique (composite) factor of production, labor, which is inelastically supplied and freely mobile across sectors. One sector produces a homogenous good z, while the remaining S sectors produce a continuum of differentiated products. Preferences are identical everywhere in the world and given by:

$$U = \beta_z \log z + \sum_{s=1}^{S} \beta_s \log Q_s, \tag{2.1}$$

with $\beta_z + \sum_{s=1}^{S} \beta_s = 1$ and

$$Q_s = \left(\int_{\omega \in \Omega_s} q_s(\omega)^{(\sigma_s - 1)/\sigma_s} \, d\omega \right)^{(\sigma_s/\sigma_s - 1)}, \quad \sigma_s > 1. \tag{2.2}$$

It is worth pausing to discuss the specific assumptions we have already built into the model. The preferences in (2.1) feature a unit elasticity of substitution across sectors, so industry spending shares are constant. Within differentiated-good sectors, the preferences in (2.2) are of the Dixit-Stiglitz type: there is a continuum of varieties available to consumers and these enter preferences symmetrically and with a constant, higher-than-one-elasticity of substitution between any pair of varieties. These assumptions are special, but they are standard in the International Trade field. In particular, the preferences in (2.1) and (2.2) are a strict generalization of those in Krugman (1980) and Melitz (2003), which correspond to the case $\beta_z = 0$ and $S = 1$.[5] I incorporate multiple differentiated-good sectors because this will facilitate the derivation of cross-sectional predictions, while the presence of a homogeneous-good sector will simplify the general equilibrium aspects of the

[5] To be precise, the last section of Krugman (1980) develops a two-industry model featuring cross-country differences in demand patterns.

model. I will however consider the Krugman-Melitz, one-sector version of the model at times in the book. It would be valuable to follow the approach of Helpman and Krugman (1985) and work out the robustness of the results below to more general preference structures, but I will not attempt to do so in this book.[6]

Given (2.1), consumers in country j will optimally allocate a share β_z of their spending E_j to good z and a fraction β_s to differentiated-good sector s. I will use the subscripts i and j to refer to countries, with i denoting producing/exporting countries and j denoting consuming/importing countries. In order to keep the notation as neat as possible, I will drop the subscript s associated with differentiated-good sectors and their sector-specific parameters. Similarly, and although the model is dynamic (time runs indefinitely), I will omit time subscripts throughout since I will focus on describing stationary equilibria.

Within a representative differentiated-good sector then consumers allocate spending across varieties to maximize Q in (2.2), which gives rise to following demand for variety ω in country j:

$$q_j(\omega) = \beta E_j P_j^{\sigma-1} p_j(\omega)^{-\sigma}, \tag{2.3}$$

where $p_j(\omega)$ is the price of variety ω, P_j is the ideal price index associated with (2.2),

$$P_j = \left[\int_{\omega \in \Omega_j} p_j(\omega)^{1-\sigma} d\omega \right]^{1/(1-\sigma)}, \tag{2.4}$$

and Ω_j is the set of varieties available to consumers in j.

Consider next the supply side of the model. The homogenous good is produced with labor under conditions of perfect competition, and according to a constant-returns-to-scale technology which is allowed to vary across countries. In particular, output is equal to

$$z_i = L_{zi}/a_{zi}, \tag{2.5}$$

[6] As it will become apparent, however, the Cobb-Douglas assumption in (2.1) is of little relevance for the main results derived in future chapters of the book. Also, the literature has developed versions of the Melitz (2003) model with alternative, specific functional forms for the aggregate industry index Q_m (see, for instance, Melitz and Ottaviano, 2008, or Novy, 2013). Relaxing the assumption of a continuum of varieties would severely complicate the analysis by introducing strategic pricing interactions across firms within an industry.

where L_{zi} is the amount of labor in country i allocated to the production of good z, and a_{zi} is country i's unit labor requirement in that sector. The homogeneous good z is freely tradable across countries and will serve as the numéraire in the model.

The differentiated-good industries are instead monopolistically competitive. Each variety is produced by a single firm under a technology featuring increasing returns to scale, and there is free entry into each industry. The existence of internal economies of scale stems from the presence of three types of fixed costs. First, the process of entry and differentiation of a variety entails a fixed cost of f_{ei} units of labor in country i. Second, production of final-good varieties in country i entails an overhead cost equal to f_{ii} units of country i's labor. Finally, firms in country i need to incur an additional fixed 'market access' cost equal to f_{ij} units of labor in order to export in country $j \neq i$. These fixed export costs capture costs associated with marketing and distributing goods in foreign markets that need to be incurred regardless of the volume exported. I will specify these costs in terms of the exporting country's labor, but not much would change if they were specified in terms of the importing country's labor. Notice, that we do not assume that $f_{ij} > f_{ii}$, but the latter type of fixed costs of production need to be incurred before the firm can sell in any market.

The fixed cost parameters f_{ei}, f_{ii} and f_{ij} are common for all firms within an industry. Intraindustry heterogeneity stems from differences in the marginal cost of production faced by firms. In particular, after incurring the fixed cost of entry f_{ei}, firms learn their productivity level φ, which determines their marginal cost of production, $1/\varphi$, in terms of labor. These productivity levels are drawn independently from a cumulative distribution function $G_i(\varphi)$ which is assumed Pareto with shape parameter $\kappa > \sigma - 1$, so

$$G_i(\varphi) = 1 - \left(\frac{\underline{\varphi}_i}{\varphi}\right)^{\kappa}, \quad \text{for } \varphi \geq \underline{\varphi}_i > 0. \tag{2.6}$$

The marginal cost of servicing foreign markets is further magnified by 'iceberg' trade costs such that $\tau_{ij} > 1$ units of output need to be shipped from country i for 1 unit to make it to country j. The firm productivity parameter φ is time invariant, but firms face a common, exogenous probability $\delta \in (0, 1)$ of being subject to a (really) bad shock that would force them to exit, which keeps the value of the firm bounded for any φ.

When selling to local consumers, firms need not incur variable trade costs ($\tau_{ii} = 1$) nor market access costs in excess of the production fixed cost f_{ii}.

Under the mild assumption that any firm with positive production sells some amount of output in their domestic market, we can then succinctly express the cost for a firm with productivity φ of producing q units of output in country i and selling them in country j as

$$C_{ij}(q) = \left(f_{ij} + \frac{\tau_{ij}}{\varphi} q \right) w_i, \tag{2.7}$$

where note that the formula applies both for foreign $(i \neq j)$ as well as for domestic sales $(i = j)$.

This completes the description of the model. Before discussing some features of the equilibrium, it is worth briefly relating the model above to other ones in the literature. The structure of the model is most closely related to that of the multi-sector Melitz models in Arkolakis, Demidova, Klenow and Rodríguez-Clare (2008) and Helpman, Melitz and Rubinstein (2008).[7] The original model in Melitz (2003) corresponds to the particular case in which $\beta_z = 0$ and $S = 1$, and parameters are fully symmetric across countries, so $f_{ei} = f_e$, $f_{ii} = f$, $f_{ij} = f_X$, $\tau_{ij} = \tau$ and $L_i = L$, where L_i is the stock of labor in country i.[8] As hinted above, the seminal paper of Krugman (1980) – except for its last section – is also a special case of the framework above, in which on top of the assumptions in Melitz (2003), there are no fixed marketing costs $f_X = 0$ and the distribution of productivity $G_i(\varphi)$ is degenerate, so firms are homogenous.[9]

Selection into Exporting

I next illustrate how this simple model is able to explain some of the firm-level exporting facts discussed above. Given the isoleastic demand in (2.3), firms will charge a price in each market in which they sell equal to a constant markup $\sigma/(\sigma-1)$ over the marginal cost of servicing that market. As a

[7] Melitz and Redding (2013*a*) have recently used a model with a very similar structure to navigate the literature on heterogeneous firms and trade. Chaney (2008) also develops a multi-sector Melitz framework but does not allow for free entry.

[8] The above model is less general than Melitz (2003) in that I impose that $G(\varphi)$ is Pareto, while he considers a general cumulative probability distribution.

[9] A hybrid model in the spirit of Helpman and Krugman (1985) could also be derived from our model (whenever $\beta_z > 0$) if we allowed sectors to use two factors of production (say capital and labor) under different factor intensities. Also, our benchmark model could easily be turned into the standard neoclassical Ricardian and Heckscher-Ohlin models by setting $\sigma \to \infty$ and all fixed costs to 0.

result, the potential operating profits for a firm from i with productivity φ considering servicing a particular market j can be concisely written as

$$\pi_{ij}(\varphi) = (\tau_{ij} w_i)^{1-\sigma} B_j \varphi^{\sigma-1} - w_i f_{ij} \tag{2.8}$$

where

$$B_j = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} P_j^{\sigma-1} \beta E_j. \tag{2.9}$$

Notice that $\pi_{ij}(\varphi)$ increases linearly with the transformation of productivity $\varphi^{\sigma-1}$ and that for a sufficiently low φ, $\pi_{ij}(\varphi)$ is necessarily negative. More formally, only the subset of firms from i with productivity $\varphi \geq \tilde{\varphi}_{ij}$, where

$$\tilde{\varphi}_{ij} \equiv \tau_{ij} w_i \left(\frac{w_i f_{ij}}{B_j}\right)^{1/(\sigma-1)}, \tag{2.10}$$

will find it optimal to export to country j. Other things equal, the higher are trade barriers between i and j (τ_{ij} and f_{ij}), the lower will be the share of firms in i choosing to service j. This contrasts with homogeneous firm models, in which all firms from i would sell to all possible markets j.

The model also sheds light on the fact that exporters typically appear to be more productive than non-exporters. In particular, provided that the profit shifter B_j does not vary too much across countries, and given that $f_{ij} > 0$ for all $j \neq i$, firms will find it relatively harder to profitably sell in foreign markets than in their local market. Furthermore, provided that $\tau^{\sigma-1} f_{ij} > f_{ii}$ for all $j \neq i$, it will necessarily be the case that a positive measure of firms sells domestically but does not export. These intuitive results regarding selection into exporting and productivity differences between exporters and non-exporters are depicted in Figure 2.1. The figure exploits the fact that firm profits $\pi_{ij}(\varphi)$ in (2.8) increase linearly with the transformation of productivity $\varphi^{\sigma-1}$. The lower slope of the export profit line reflects the large variable transport costs $\tau_{ij} > 1$ (remember that we are assuming small differences in the profit shifter B_j across countries). In the figure, it is also assumed that $f_{ij} > f_{ii}$, although we have noted above that $\tau^{\sigma-1} f_{ij} > f_{ii}$ is sufficient to obtain selection into exporting.[10]

[10]Note that no matter how low f_{ij} is, whenever B_j is identical across countries, it will never be the case that a firm from i produces *only* for a particular export market $j \neq i$. This is because in such a case, these *pure* exporters would need to incur the overhead costs f_{ii} on top of the marketing costs f_{ij}, and thus we would have $f_{ij} + f_{ii} > f_{ii}$. Lu

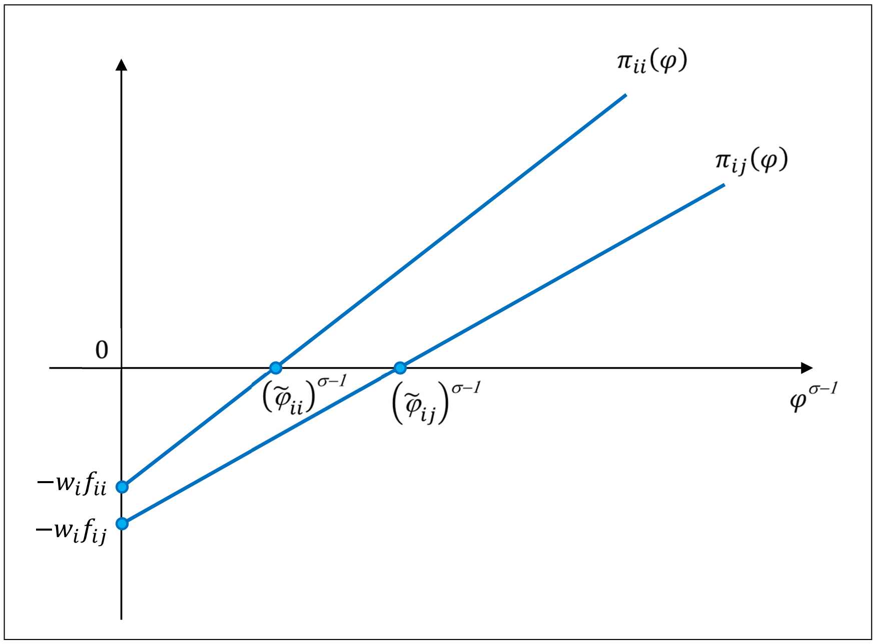

Figure 2.1: Selection into Exporting with Heterogeneous Firms

The Extensive Margin, Gravity and Reallocation Effects

The logic behind the fact that a model with heterogeneous firms and fixed export costs can deliver selection into exporting based on productivity is hardly earth-shattering. The beauty of the Melitz (2003) model resides in the fact that, despite its simple structure, it can account for several additional features documented in empirical studies. These additional results from the model are less central for the set of results emphasized in this book, but nonetheless it is worth discussing them briefly.

Consider first the implications of the model for aggregate exports at the sectoral level. Letting X_{ij} denote aggregate exports from i to j in a representative differentiated-good sector, and denoting by N_i the measure of potential producers from i in that sector (i.e., the set of firms that have paid

(2011) shows that a 'reverse' sorting is observed among Chinese manufacturing firms, a fact that she attributes to a particularly low value of B_j in labor-intensive industries in China relative to other countries.

the fixed cost of entry $w_i f_{ei}$), we have

$$X_{ij} = N_i \int_{\tilde{\varphi}_{ij}}^{\infty} \sigma \left(\tau_{ij} w_i\right)^{1-\sigma} B_j \varphi^{\sigma-1} dG_i\left(\varphi\right), \tag{2.11}$$

where we have used (2.8) and the fact that export revenues are a multiple σ of $\pi_{ij}\left(\varphi\right) + w_i f_{ij}$.

A first point to notice is that variation in exporting across destination markets j is composed of an extensive margin and an intensive margin. In particular, we can write

$$X_{ij} = N_{ij} \cdot \bar{x}_{ij},$$

where $N_{ij} = \left(1 - G_i\left(\tilde{\varphi}_{ij}\right)\right) N_i$ is the actual measure of firms from i selling in j (the extensive margin) and

$$\bar{x}_{ij} = \frac{1}{1 - G_i\left(\tilde{\varphi}_{ij}\right)} \int_{\tilde{\varphi}_{ij}}^{\infty} \sigma \left(\tau_{ij} w_i\right)^{1-\sigma} B_j \varphi^{\sigma-1} dG_i\left(\varphi\right), \tag{2.12}$$

are average firm-level exports (the intensive margin). As first worked out by Arkolakis, Demidova, Klenow and Rodríguez-Clare (2008), when productivity is distributed Pareto as in (2.6), integrating (2.12) and using (2.10) to simplify, delivers

$$\bar{x}_{ij} = \frac{\kappa}{\kappa - \sigma + 1} \sigma w_i f_{ij}, \tag{2.13}$$

and thus the intensive margin is independent of variable trade costs and of market size of the destination country. In other words, the model is consistent with export volumes from i to j being lower for smaller and more distant markets, but the reason for this is very different than in homogeneous firm models à la Krugman. It is not because firms export on average lower volumes to those markets but rather because a smaller set of firms export to those markets. As shown by Chaney (2008), this is not an immaterial distinction, since it critically affects, for instance, how the elasticity of trade flows to trade frictions depends on the elasticity of substitution σ.

Another remarkable feature of the model is that it delivers a modified sectoral version of the gravity equation for trade flows, which has been shown to fit the data rather well. As shown by Melitz and Redding (2013*a*) (see also Chapter 3 for a related derivation), in the Pareto case, aggregate exports in (2.11) can alternatively be expressed as

$$X_{ij} = \frac{Y_i}{\Theta_i} \left(\frac{\beta E_j}{P_j^{1-\sigma}}\right)^{\frac{\kappa}{(\sigma-1)}} \tau_{ij}^{-\kappa} f_{ij}^{-\frac{\kappa-(\sigma-1)}{\sigma-1}}, \tag{2.14}$$

where Y_i is the aggregate industry output in i (i.e., $Y_i \equiv \sum_j X_{ij}$) and Θ_i is a structural measure of country i's market potential in that industry.[11] Notice that equation (2.14) structurally justifies the use of empirical log-linear specifications for sectoral trade flows with importer-sector and exporter-sector asymmetric fixed effects and measures of bilateral trade frictions. In the one-sector models of Krugman (1980) and Melitz (2003) (i.e., $\beta_z = 0$ and $S = 1$), the model predicts that the gravity equation will hold for aggregate bilateral trade flows across countries, and as shown by Helpman, Melitz and Rubinstein (2008), for estimation purposes, the model serves a very useful role in structurally correcting for the large number of bilateral zero trade flows in the data (we will cover their contribution in more detail in Chapter 3).

In this same one-sector version of the model, Arkolakis, Costinot and Rodríguez-Clare (2012) have derived a neat formula for the welfare effects of trade in terms of two sufficient statistics: the import penetration ratio and the elasticity of imports with respect to variable trade costs. Arkolakis, Costinot and Rodríguez-Clare (2012) have also shown that, remarkably, this formula is identical to the one obtained in the Anderson and van Wincoop (2003), Eaton and Kortum (2002), and Krugman (1980) models.[12]

One final aspect of the model that is worth discussing is its ability to rationalize the reallocation effects following trade liberalization documented by the empirical literature. This is most elegantly derived in the symmetric, one-sector model of Melitz (2003) in which no parametric assumptions on $G(\varphi)$ are imposed. Essentially, what Melitz shows is that reductions in trade costs will not only expand the number and revenues of exporting firms, but will also (via competition effects) reduce the scale of non-exporting firms and will also lead to the exit of a set of producers that were marginally profitable before the reduction in trade costs.[13] Formally, in terms of the

[11] In particular, $\Theta_i \equiv \sum_j \left(\frac{\beta E_j}{P_j^{1-\sigma}}\right)^{\frac{\kappa}{(\sigma-1)}} \tau_{ij}^{-\kappa} f_{ij}^{-\frac{\kappa-(\sigma-1)}{\sigma-1}}$.

[12] Because the import penetration ratio and the 'trade elasticity' respond to trade opening in distinct manners in these different frameworks, their results do not necessarily imply, however, that information on the microstructure of these models is irrelevant for assessing the welfare consequences of trade liberalization (see Melitz and Redding, 2013*b*, for more on this).

[13] Even though the size of continuing exporters increases and that of continuing non-exporters shrinks, in the case in which $G(\varphi)$ is Pareto, the extensive margin responses ensure that the average size of exporters and non-exporters will remain unaffected by

notation above, Melitz (2003) shows that reductions in trade costs will not only reduce $\tilde{\varphi}_{ij}$, but will also increase $\tilde{\varphi}_{ii}$ thus forcing firms with productivity marginally above $\tilde{\varphi}_{ii}$ to shut down. As discussed by Baldwin and Forslid (2010) and Arkolakis, Demidova, Klenow and Rodríguez-Clare (2008), this in turn leads to 'anti-variety' effects by which the measure of varieties available to consumers will, under plausible conditions, decrease following trade liberalization.

The Melitz (2003) model has been extended in a variety of fruitful ways, ranging from the exploration of alternative demand systems, the introduction of Heckscher-Ohlin features into the model, the modeling of multi-product firms, and many others. Several applications and extensions of the model are reviewed in Melitz and Redding (2013*a*). I will next focus on an extension of the model that is particularly relevant for the study of the global organization of production, which is the central topic of this book.

Global Sourcing with Heterogeneous Firms

In the Melitz (2003) model the only involvement of firms with foreign markets is via the exportation of final goods produced with local labor. As documented in Chapter 1, the recent process of globalization has led to a disintegration of the production process across borders in which international trade in intermediate inputs has been a dominant feature in the world economy. I next develop a simple variant of the Melitz framework in which firms not only export, but also make global sourcing decisions related to the location and quantity of inputs to buy from different countries.

In order to meaningfully study offshoring, one needs to consider multi-stage production processes, and a natural starting point is a two-stage model. With that in mind, assume that the production of varieties in the differentiated-good sectors now involves two stages, which we will refer throughout the book as *headquarter services* and *manufacturing production.* Headquarter services may include a variety of activities such as R&D expenditures, brand development, accounting, and finance operations, but may also involve high-tech manufacturing or assembly. The important characteristic of this stage in terms of the model is that these activities need to be produced in the same country in which the entry cost f_{ei} was incurred. Manufacturing production can instead be thought of as entailing low-tech manufacturing or assembly

changes in variable trade barriers, as shown in (2.13).

of inputs into a final product. Crucially, we will depart from Melitz (2003) in allowing manufacturing production to be geographically separated from the location of entry and headquarter services provision. This is a highly simplified characterization of the process of offshoring, but we will work to enrich the model later in the book.

Relative to the multi-sector Melitz (2003) framework developed above, the key new decision facing firms is thus whether to maintain plant production in the same country in which entry and headquarter service provision takes place, or whether to offshore that stage. In order to simplify the model and isolate the new insights arising from the modelling of offshoring, we shall assume that there are no costs, fixed or variable, associated with exporting final goods so that the exporting decision is trivial and all firms producing final goods export them worldwide. Conversely, the decision of whether to source locally or engage in offshoring will be nontrivial: offshoring will be associated with a reduction in production costs but will also entail additional fixed and variable transportation costs that might lead some firms to opt out of that strategy.

More formally, the overall costs of producing q units of a final-good variety incurred by a firm with headquarters in country i and manufacturing production in country j (with possibly $j = i$) are given by

$$C_{ij}\left(q, \varphi\right) = f_{ij} w_i + \frac{q}{\varphi} \left(a_{hi} w_i\right)^{\eta} \left(\tau_{ij} a_{mj} w_j\right)^{1-\eta}. \tag{2.15}$$

As before, φ is a firm-specific productivity parameter. The parameters f_{ij}, τ_{ij}, η, a_{hi} and a_{mj} are instead sector specific but common across firms within a sector s, while the wage rates w_i and w_j vary only across countries. The parameters f_{ij} and τ_{ij} appeared already in the Melitz (2003) model (see equation (2.7)) but their interpretation is somewhat different in the present context. In particular, f_{ij} and τ_{ij} now reflect the fixed and variable trade costs associated with a particular sourcing strategy. Although, we will often associate f_{ij} and τ_{ij} with the costs of transporting intermediate inputs across countries, these parameters can be interpreted more broadly to reflect other technological barriers associated with international fragmentation, such as communication costs, language barriers or search costs. For these reasons, it is now natural to assume not only that $\tau_{ij} > \tau_{ii}$, but also that $f_{ij} > f_{ii}$ whenever $j \neq i$.

Relative to the specification of technology in (2.7), the new parameters are η, a_{hi} and a_{mj}. The first of these captures the headquarter services intensity (or *headquarter intensity* for short) of the production process, and the associated primal representation of technology (leaving aside fixed costs and trade costs) is a Cobb-Douglas technology in headquarter services h and manufacturing production m:

$$q(\varphi) = \varphi \left(\frac{h(\varphi)}{\eta} \right)^{\eta} \left(\frac{m(\varphi)}{1-\eta} \right)^{1-\eta}, \quad 0 < \eta < 1. \tag{2.16}$$

Finally, the parameters a_{hi} and a_{mj} capture the unit labor requirements associated with headquarter service provision and manufacturing production and these are allowed to vary across sector and countries reflecting comparative advantage considerations.

Although the benchmark model of offshoring we have developed is quite stylized it is a generalization of a complete-contracting variant of the heterogeneous firm model in Antràs and Helpman (2004). In particular, in Antràs and Helpman (2004) it is further assumed that:

- The world consists of only two countries, North and South.
- The homogenous good z is always produced in both countries but with a higher labor productivity in the North, thus implying that wages rates are fixed at $w^N = 1/a_{zN} > 1/a_{zS} = w^S$.
- The South features either very low productivity in producing headquarter service or very high fixed costs of entry, so that all entry and headquarter service provision occurs in the North, where $a_{hN} = 1$.
- Plant production can be done with the same physical productivity – in particular, $a_{mN} = a_{mS} = 1$ – in both North and South, so offshoring to South offers a production cost advantage.

With these additional assumptions, and simplifying further the notation by denoting $f_{NN} = f_D$, $f_{NS} = f_O$, $\tau_{NS} = \tau$, the total cost of production associated with **D**omestic sourcing and **O**ffshoring can be written, respectively, as

$$C_D(q,\varphi) = \left(f_D + \frac{q}{\varphi} \right) w_N, \tag{2.17}$$

and

$$C_O(q,\varphi) = f_O w_N + \frac{q}{\varphi} (w_N)^{\eta} (\tau w_S)^{1-\eta}. \tag{2.18}$$

Selection into Offshoring

We can now study the implications of the above framework for the selection of firms into offshoring. For now I will focus on the simplified two-country framework in Antràs and Helpman (2004) since it has featured prominently in the literature, but at the end of the chapter I will discuss how the results can be extended to a multi-country environment.

In light of the cost functions in (2.17) and (2.18), and given that firms charge a price for the final good equal to a constant markup $\sigma/(\sigma-1)$ over the marginal cost of production, the potential operating profits for a Northern firm with productivity φ associated with Domestic sourcing and Offshoring can be expressed as

$$\pi_D(\varphi) = (w_N)^{1-\sigma} B\varphi^{\sigma-1} - f_D w_N \tag{2.19}$$

and

$$\pi_O(\varphi) = \left((w_N)^{\eta}(\tau w_S)^{1-\eta}\right)^{1-\sigma} B\varphi^{\sigma-1} - f_O w_N, \tag{2.20}$$

respectively, where

$$B = \frac{1}{\sigma}\left(\frac{\sigma}{(\sigma-1)P}\right)^{1-\sigma}\beta(w_N L_N + w_S L_S)$$

and P is the common price index in (2.4) in each country, given costless international trade in final goods.

As in Melitz (2003), the profit functions $\pi_D(\varphi)$ and $\pi_O(\varphi)$ are linearly increasing in the transformation of productivity $\varphi^{\sigma-1}$ and for a sufficiently low φ, both of these profit levels necessarily take negative values. Hence, upon observing their productivity, the least productive firms in an industry will optimally decide not to produce. Furthermore, the fact that $f_O > f_D$ ensures that for sufficiently low levels of productivity, we have $\pi_D(\varphi) > \pi_O(\varphi)$, and offshoring is not a viable option in situations in which domestic sourcing might be profitable. In fact, whenever

$$f_O > \left(\frac{w_N}{\tau w_S}\right)^{(1-\eta)(\sigma-1)} f_D, \tag{2.21}$$

there always exists a subset of firms in the industry that find it optimal to opt out of offshoring and decide instead to source locally in the North. In order for some firms within the industry to find it optimal to offshore it is

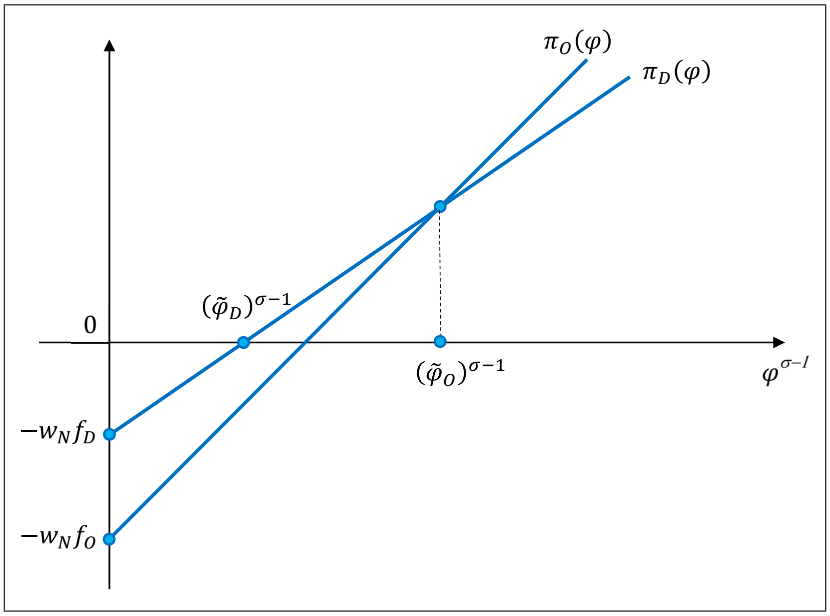

Figure 2.2: Equilibrium Offshoring Sorting with High Wage Differences

necessary to assume that offshoring trade costs τ are low enough to ensure that $w_N > \tau w_S$. This case is depicted in Figure 2.2, which also drawn under the implicit assumption that condition (2.21) holds.

As shown by the Figure, the model features selection into offshoring by which only the most productive firms within an industry find it worthwhile to pay the fixed costs of fragmentation to benefit from the lower production costs associated with manufacturing in the South.[14] In particular, offshoring is the preferred option only for firms with productivity $\varphi \geq \tilde{\varphi}_O$, where

$$\tilde{\varphi}_O \equiv \left(\frac{f_O - f_D}{B} \frac{w_N}{\left((w_N)^{\eta} (\tau w_S)^{1-\eta} \right)^{1-\sigma} - (w_N)^{1-\sigma}} \right)^{1/(\sigma-1)}.$$

The sorting pattern in Figure 2.2 is consistent with the evidence on selection into importing in Bernard, Jensen, Redding and Schott (2007), who show that not only U.S. exporting firms but also U.S. importing firms appear

[14]In the much less interesting case in which $w_N < \tau w_S$, no firm in the industry finds it optimal to offshore and if this condition holds for all sectors of the economy, then the South is fully specialized in the production of the homogenous good z.

to be more productive than purely domestic producers. Their results are reproduced in Table 2.1 below, which shows that U.S. manufacturing plants that import employ more workers, sell more, are more productive, pay higher wages and are more capital and skill intensive than plants that do not source abroad.

Table 2.1. Trading Premia in U.S. Manufacturing, 1997

	Exporter Premia	Importer Premia
Log Employment	1.50	1.40
Log Shipments	0.29	0.26
Log Value-Added per Worker	0.23	0.23
Log TFP	0.07	0.12
Log Wage	0.29	0.23
Log Capital per Worker	0.17	0.13
Log Skill per Worker	0.04	0.06

Source: Bernard, Jensen and Schott (2009), Table 8.

More specifically, firms that import appear to be 12% more productive than firms that do not, while the productivity advantage of exporting plants is only of 7%. Furthermore, Bernard, Jensen, Redding and Schott (2007) report that only 14 percent of U.S. manufacturing plants report positive imports (versus 27 percent of plants reporting positive exports), which is again suggestive of the existence of significant fixed costs of importing.

One might be concerned that the patterns observed by Bernard, Jensen, Redding and Schott (2007) do not necessarily support the sorting pattern implied by the theory because it is unclear whether the goods that firms are importing are intermediate inputs, rather than finished goods. In the latter case, one might worry that Table 2.1 is simply picking the role of large intermediaries (wholesalers or retailers) in bringing consumer goods into the United States. The fact that the importer premia reported in Table 2.1 correspond to the operations of U.S. *manufacturing* firms should however dispel that concern. Furthermore, recent work by Fort (2013) using U.S. Census data demonstrates that a similar sorting pattern is observed when focusing on imports of contract manufacturing services, which cover exclusively offshoring of inputs that are customized to specific U.S. firm's production needs. More specifically, Fort (2013) finds that U.S. firms that offshore contract manufacturing services feature on average 13% higher valued-added labor productivity than U.S. firms in the same six-digit NAICS that purchase those services only domestically.

Figure 2.3 provides further confirmation of the superior performance of offshoring firms with 2007 data from the Spanish Encuesta sobre Estrategias Empresariales (ESEE). The dataset distinguishes between firms that purchase inputs only from other Spanish producers and firms that purchase inputs from abroad. As is clear from the picture, the distribution of productivity of firms that engage in foreign sourcing is a shift to the right of that of firms that only source locally.[15]

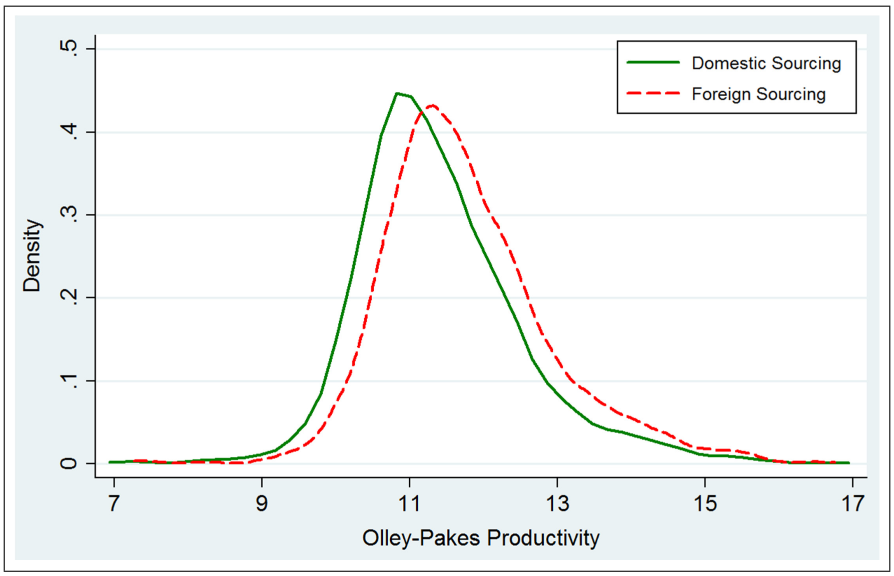

Figure 2.3: Selection into Offshoring in Spain

Determinants of the Prevalence of Offshoring

We can next use this simple model of global sourcing to study the determinants of the relative prevalence of offshoring in an industry. For instance, consider computing the share of spending on *imported* manufacturing inputs over total manufacturing input purchases in a particular industry. Given the Cobb-Douglas technology in (2.16) and the CES preferences in (2.1), manufacturing input purchases will constitute a share $(\sigma - 1)(1 - \eta)/\sigma$ of revenue

[15]More details on this dataset and on the measurement of productivity according to the Olley-Pakes method will be provided in Chapter **??**.

for all firms, while revenue itself will be a multiple σ of firm operating profits (defined as revenue minus variable costs). Using the profit functions (2.19) and (2.20) and cancelling common terms, we can thus express the share of imported manufacturing input purchases in a given industry as

$$\Upsilon_O = \frac{\left(\frac{w_N}{\tau w_S}\right)^{(1-\eta)(\sigma-1)} \int_{\tilde{\varphi}_O}^{\infty} \varphi^{\sigma-1} dG_i(\varphi)(\varphi)}{\int_{\tilde{\varphi}_D}^{\tilde{\varphi}_O} \varphi^{\sigma-1} dG(\varphi) + \left(\frac{w_N}{\tau w_S}\right)^{(1-\eta)(\sigma-1)} \int_{\tilde{\varphi}_O}^{\infty} \varphi^{\sigma-1} dG_i(\varphi)}.$$

Particularly sharp results can be obtained when assuming that the distribution of firm productivity is Pareto as in equation (2.6) in which case we obtain

$$\Upsilon_O = \frac{\left(\frac{w_N}{\tau w_S}\right)^{(1-\eta)(\sigma-1)}}{\left(\frac{\tilde{\varphi}_O}{\tilde{\varphi}_D}\right)^{\kappa-(\sigma-1)} - 1 + \left(\frac{w_N}{\tau w_S}\right)^{(1-\eta)(\sigma-1)}}. \tag{2.22}$$

where

$$\frac{\tilde{\varphi}_O}{\tilde{\varphi}_D} = \left[\frac{f_O/f_D - 1}{\left(\frac{w_N}{\tau w_S}\right)^{(1-\eta)(\sigma-1)} - 1}\right]^{1/(\sigma-1)}. \tag{2.23}$$

As indicated by equation (2.22) and (2.23), the prevalence of offshoring is naturally increasing in the wage gap (w_N/w_S) and decreasing in fragmentation barriers (f_O/f_D, τ). These comparative statics are quite intuitive. Note that the elasticity of substitution σ and the parameter κ governing the thickness of the right tail of the Pareto distribution also have an impact on the prevalence of offshoring in an industry. The intuition for these effects is analogous to that in Helpman, Melitz and Yeaple (2004). In particular, the Pareto parameterization of productivity combined with CES preferences imply that the distribution of sales of all active firms is also Pareto with shape parameter $\kappa/(\sigma-1)$. As a result, a decrease in κ raises the sales and input purchases of firms with productivity $\varphi > \tilde{\varphi}_O$ – i.e., firms that find offshoring optimal – relative to the sales and input purchases of firms with productivity $\varphi \in (\tilde{\varphi}_D, \tilde{\varphi}_O)$ – i.e., firms that source domestically. Because the standard deviation of the logarithm of sales by all active firms in the industry is equal to $(\sigma-1)/\kappa$, this result can be interpreted as indicating that the prevalence of offshoring should be higher in industries with a larger dispersion in firm size.[16]

Note that the elasticity of substitution σ affects positively the share of imported inputs for an additional reason – see the exponent of $w_N/(\tau w_S)$ in (2.22) and (2.23). The intuition for this effect is simpler: the more substitutable final-good varieties are, the more elastic will demand be and the higher will be the incentive of firms to engage in a costly investment (in this case offshoring) to reduce the marginal cost of input provision from w_N down to τw_S.

Back to the Multi-Country Model

Having worked with a simplified two-country model to build intuition, we can now go back to the multi-country environment in which the overall costs of producing q units of a final-good variety faced by a firm with headquarters in country i and manufacturing production in country j are given by equation (2.15). Given CES preferences over final-good varieties, it is then straightforward to show that the operating profits associated with that sourcing strategy are given by

$$\pi_{ij}(\varphi) = \left((a_{hi}w_i)^{\eta}(\tau_{ij}a_{mj}w_j)^{1-\eta}\right)^{1-\sigma} B\varphi^{\sigma-1} - f_{ij}w_i, \qquad (2.24)$$

where B is now given by

$$B = \frac{1}{\sigma}\left(\frac{\sigma}{(\sigma-1)P}\right)^{1-\sigma}\beta\sum\nolimits_j w_j L_j$$

and P is the common price index (2.4) for final-good varieties in each country.

Equation (2.24) illustrates again that the profit levels associated with different sourcing strategies are all linear in $\varphi^{\sigma-1}$ and thus the sourcing decision of firms can be analyzed with graphs analogous to that in Figure 2.2. Of course, with multiple countries the range of possible sorting patterns is much more complex, but we can still derive some general results.

For instance, as long as $f_{ij} > f_{ii}$ for all $j \neq i$, so domestic sourcing is the sourcing strategy associated with the lowest fixed costs, the model can only deliver a positive amount of offshoring in an industry whenever $\tau_{ij}a_{mj}w_j < a_{mi}w_i$ for some country $j \neq i$. Importantly, in such a case,

[16] Other measures of industry firm size dispersion, such as the Theil index also vary monotonically with $(\sigma-1)/\kappa$.

if firms sourcing domestically and abroad coexist within an industry, then firms that offshore are necessarily larger and more productive than firms that source domestically. In sum, under the plausible condition $f_{ij} > f_{ii}$ for all $j \neq i$, the model continues to predict selection into offshoring based on productivity in a manner consistent with the U.S. import premia in Table 2.1 and with the evidence from Spain depicted in Figure 2.3.

It is also noteworthy that in contrast to the simple two-country model above, this multi-country extension of the model can easily generate two-way intermediate input trade flows across countries. For instance, a given country i can feature a high manufacturing productivity level $1/a_{mj}$ in some industries and a very low one in others. In the former type of industries, this country i may well export inputs to firms with headquarters located in other countries (particularly when country i's productivity in headquarter provision is low in that industry), while it may well import manufacturing inputs in the latter type of industries.[17]

With multiple countries, firms not only decide on whether to offshore manufacturing production or not but also choose the optimal location of production among all possible ones. It is evident that, other things equal, firms based in country i will be drawn to locations j entailing low fixed costs of sourcing f_{ij} and low variable costs of manufacturing, as summarized by $\tau_{ij} a_{mj} w_j$. Some highly productive firms might however be drawn to locations with high sourcing fixed costs as long as those locations offer a particularly favorable marginal cost of input manufacturing.

Figure 2.4 depicts a possible equilibrium in a world of four countries, a 'Home' country i and three 'Foreign' countries j, k and l. Domestic sourcing is the least fixed cost sourcing strategy, and as argued above, this is the preferred option for the least productive among the active firms in the industry. Offshoring to country l entails high fixed costs and also high variable costs (perhaps due to high transportation costs τ_{il} or high productivity-adjusted manufacturing wages $a_{ml} w_l$), and thus no firm finds it optimal to import inputs from l. Country j offers the largest marginal cost savings when offshoring there, but the fixed costs of fragmentation are high there, so only the most productive firms within an industry find it optimal to import inputs from j. Finally, country k is associated with moderate fixed costs of

[17] From this discussion, it should be obvious that the two-country model developed above failed to deliver two-way input trade flows because of its assumptions on technology (e.g., ruling out headquarter services provision in the South), and not because it only featured two countries.

offshoring and offers a cost advantage relative to domestic sourcing, so a subset of middle-productivity firms chooses it as their optimal location of manufacturing input production.

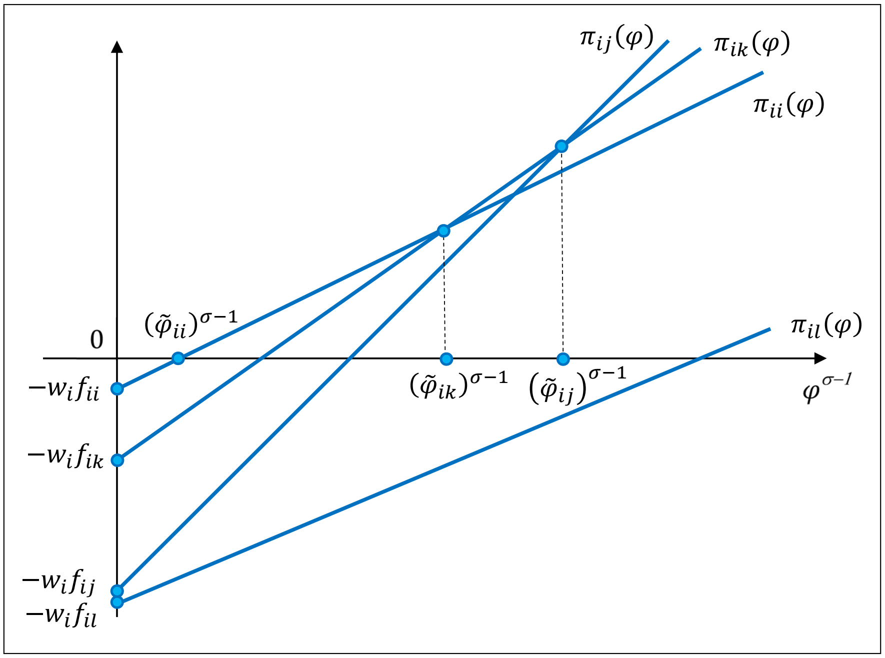

Figure 2.4: Selection into Offshoring with Multiple Countries

Naturally, the example illustrated in Figure 2.4 is rather special and, more worryingly, very different sorting patterns could emerge with mild changes in the key productivity and cost parameters. To illustrate this sensitivity, consider the case in which all foreign countries share the same level of offshoring fixed costs or $f_{ij} = f_{iO}$ for all $j \neq i$. It is then clear, that conditional on finding it optimal to offshore, firms headquartered in country i will offshore manufacturing to the location j that minimizes marginal costs, or $j^* = \arg\min_j \{\tau_{ij} a_{mj} w_j\}$. Small changes in any of these parameters could thus lead to discontinuous jumps in the prevalence of imports of inputs from particular countries.

Another limitation of this multi-country model is that it is not well designed to aggregate all firm decisions within an industry in order to guide empirical analyses of the determinants of the relative prevalence of offshoring to particular countries depending on some fundamental parameters of those countries. For similar reasons, the model is not a particularly useful tool for quantitative analysis, particularly when envisioning a more realistic world with multiple inputs.

Fortunately, below we will be able to make some progress on these limitations by borrowing some neat modeling techniques from a recent paper by Tintelnot (2013), who in turn builds on the seminal work of Eaton and Kortum (2002).[18]

Bringing Eaton and Kortum (2002) Inside the Firm

Imagine now that the manufacturing stage of production entails the procurement of a continuum of measure one of inputs indexed by v, rather than just one input as assumed so far. I let these inputs be imperfectly substitutable with each other with a constant and symmetric elasticity of substitution equal to $1/(1-\rho)$. Very little will depend on the particular value of ρ. The cost function associated with producing q units of a final-good variety faced by a firm with headquarters in country i is now given by

$$C_{i\{j(v)\}_{v=0}^{1}}(q,\varphi) = w_i \sum_{j\in\mathcal{J}_i(\varphi)} f_{ij} + \frac{q}{\varphi}(a_{hi}w_i)^{\eta}\left(\int_0^1 \left(\tau_{ij(v)}a_{mj(v)}(v)\,w_{j(v)}\right)^{1-\rho}dv\right)^{(1-\eta)/(1-\rho)}, \tag{2.25}$$

where $j(v)$ corresponds to the country in which input v is produced and $\mathcal{J}_i(\varphi) = \{\hat{\jmath} : j(v) = \hat{\jmath} \text{ for some } v\}$ is the set of locations from which this firm with productivity φ sources inputs.

I will depart from the previous model in allowing the manufacturing productivity parameters $a_{mj(v)}$ to be *firm-specific*, and following Eaton and Kortum (2002), in treating them as the realization of random variables rather than as being deterministic. More formally, by paying the fixed cost f_{ij} of offshoring to country j, a firm headquartered in country i gains the ability of having *any* input v produced in that country j under an input-specific unit labor requirement drawn (independently from other inputs) from the Fréchet distribution

$$Pr(a_{mj}(v) \leq a) = e^{-T_j a^{-\theta}}, \quad \text{with } T_j > 0 \text{ and } \theta > \rho/(1-\rho).$$

[18]Tintelnot's (2013) framework does not feature trade in intermediate inputs, but the same tricks he develops can be fruitfully adapted to the current setting. Garetto (2013) also applies the Eaton and Kortum (2002) framework to a global sourcing environment, but does so in a two-country model and with other goals in mind.

As in Eaton and Kortum's (2002) model, T_j governs the industry-level state of technology in country j, while θ determines the variability of productivity draws across inputs, with a lower θ fostering the emergence of comparative advantage *within* input subsectors across countries.

In order to simplify matters, it is assumed that firms only learn their particular realization of $a_{mj}(v)$ after they have incurred all sunk costs of offshoring. Hence, regardless of the different amounts that firms paid to have the ability to source from particular countries, the choice of location of production of any input v will simply solve $j^*(v) = \arg\min_{j(v)\in\mathcal{J}_i(\varphi)} \{\tau_{ij} a_{mj}(v) w_j\}$. Remember that the set $\mathcal{J}_i(\varphi)$ from which $j^*(v)$ is chosen corresponds to the set of countries in which a firm from country i with productivity φ paid the associated fixed costs of offshoring f_{ij}. I will refer to $\mathcal{J}_i(\varphi)$ as the *sourcing strategy* of a firm headquartered in i with productivity φ.

Because this model has many moving parts, it is worth pausing to review the timing of events. Firms in a given sector s (subscripts omitted) initially pay a fixed cost of entry $f_{ei} w_i$ to enter country i and gain the ability to later produce headquarter services there at a unit labor cost equal to $a_{hi} w_i$. After paying this entry cost, firms learn their core productivity φ which affects firm productivity in a Hicks-neutral manner. Firms next select a set of countries $\mathcal{J}_i(\varphi)$ from which to be able to import inputs and pay all fixed offshoring costs $w_i \sum_{j\in\mathcal{J}_i(\varphi)} f_{ij}$. Once those countries have been selected, the firm observes the vector of input-location-specific productivity draws $\{a_{mj}(v)\}_{v\in[0,1]}$ for each $j \in \mathcal{J}_i(\varphi)$. The firm then decides from which country to buy a particular input v, after which headquarter services and manufacturing inputs are produced, and the final good is assembled and sold in world markets.

We have obviously made the model significantly more complicated than it originally was. Some readers might then be wondering: to what effect? To understand the purpose of this added structure, consider first the choice of location of manufacturing inputs, once all offshoring fixed costs have been paid. As argued above, at that point, a firm headquartered in i with productivity φ simply solves $j^*(v) = \arg\min_{j(v)\in\mathcal{J}_i(\varphi)} \{\tau_{ij} a_{mj}(v) w_j\}$. The beauty of the Fréchet distribution (see Eaton and Kortum, 2002) is that the probability that a given location j is chosen for any input v is simply given by

$$\chi_{ij}(\varphi) = \frac{T_j (\tau_{ij} w_j)^{-\theta}}{\Theta_i(\varphi)}. \tag{2.26}$$

where

$$\Theta_i(\varphi) \equiv \sum_{k \in \mathcal{J}_i(\varphi)} T_k (\tau_{ik} w_k)^{-\theta} \tag{2.27}$$

summarizes the *sourcing potential* of firm φ from i. With a continuum of inputs, $\chi_{ij}(\varphi)$ corresponds to the fraction of inputs sourced from j conditional on the sourcing strategy $\mathcal{J}_i(\varphi)$. Even more remarkably, the distribution of the actual price paid for any input v turns out to be independent of the actual source j of those inputs (again, see Eaton and Kortum, 2002, for details), which implies that $\chi_{ij}(\varphi)$ in (2.26) also corresponds to country j's share of all manufacturing input purchases by a firm with sourcing strategy $\mathcal{J}_i(\varphi)$.

Hopefully, the reader is beginning to appreciate that the extra machinery is starting to pay off. According to expression (2.26), and conditional on the set of active locations $\mathcal{J}_i(\varphi)$, sourcing decisions at the level of the firm now vary smoothly with the key parameters of the model. Furthermore, each country's market share in a firm's purchases of intermediates corresponds to this country's contribution to the sourcing potential $\Theta_i(\varphi)$ in (2.27). Countries in the set $\mathcal{J}_i(\varphi)$ with lower wages w_j, more advanced technologies T_j, or lower distance from country i are predicted to have higher market shares in the intermediate input purchases of firms based in country i.

Although it might seem that the core productivity parameter φ no longer plays a relevant role in the model, it is important to stress that the set of 'activated' offshoring locations $\mathcal{J}_i(\varphi)$ is endogenous and will naturally be a function of that core productivity level. To see this, let us then turn to studying the determination of the set $\mathcal{J}_i(\varphi)$.

After choosing the least cost source of supply for each input v, the overall cost function associated with producing q units of a final-good variety can be written, after some nontrivial derivations, as

$$C_i(q, \varphi, \mathcal{J}_i(\varphi)) = w_i \sum_{j \in \mathcal{J}_i(\varphi)} f_{ij} + \frac{q}{\varphi} (a_{hi} w_i)^{\eta} (\gamma \Theta_i(\varphi))^{-(1-\eta)/\theta}, \tag{2.28}$$

where $\gamma = \left[\Gamma\left(\frac{\theta+1-\rho}{\theta}\right)\right]^{\theta/(1-\rho)}$ and Γ is the gamma function.[19] Note that the addition of a new location to any potential set of active locations necessarily lowers the marginal cost faced by firms. Intuitively, an extra location grants the firm an extra cost draw for all varieties $v \in [0, 1]$. It is thus natural

[19] These derivations are analogous to those performed by Eaton and Kortum (2002) to solve for the aggregate price index in their model of final-good trade.

that greater competition among suppliers will reduce the expected minimum sourcing cost $\tau_{ij^*}a_{mj^*}(v)\,w_{j^*} = \min_{j(v)\in\mathcal{J}_i(\varphi)}\{\tau_{ij}a_{mj}(v)\,w_j\}$ per intermediate. In fact, the addition of a country to $\mathcal{J}_i(\varphi)$ lowers the expected price paid for *all* varieties v, and not just for those that are ultimately sourced from the country being added to $\mathcal{J}_i(\varphi)$.[20]

Following analogous steps as in the previous models to solve for the profit function associated with the cost function in (2.25), we can express the profits associated with the optimal sourcing strategy of a firm from country i with productivity φ as the solution to the following problem:

$$\pi_i(\varphi) = \max_{\mathcal{J}_i(\varphi)}\left\{ (a_{hi}w_i)^{-\eta(\sigma-1)} (\gamma\Theta_i(\varphi))^{(\sigma-1)(1-\eta)/\theta} B\varphi^{\sigma-1} - w_i \sum\nolimits_{k\in\mathcal{J}_i(\varphi)} f_{ik}\right\}. \tag{2.29}$$

As is clear from equation (2.29), when deciding whether to add a new country l to the set $\mathcal{J}_i(\varphi)$, the firm trades off the reduction in costs associated with the inclusion of that country in the set $\mathcal{J}_i(\varphi)$ against the payment of the additional fixed cost $w_i f_{il}$.

The problem in (2.29) is not straightforward to solve because the decision to include a country j in the set $\mathcal{J}_i(\varphi)$ naturally interacts with the decision to add any other country j'. For this reason, although the larger is the core productivity level φ, the higher will the marginal benefit of adding a location to any given set $\mathcal{J}_i(\varphi)$, it is not necessarily the case that the set $\mathcal{J}_i(\varphi)$ is 'increasing' in φ. Or, more precisely, the choice of locations $\mathcal{J}_i(\varphi_0)$ of a firm with productivity φ_0 is not necessarily a strict subset of the set of locations $\mathcal{J}_i(\varphi_1)$ chosen by a firm with a higher productivity level $\varphi_1 > \varphi_0$. For example, a highly productive firm from i might pay a large fixed cost to be able to offshore to a country l with a particularly high value of $T_l(\tau_{il}w_l)^{-\theta}$, after which the marginal incentive to add further locations might be greatly diminished whenever $(\sigma-1)(1-\eta) < \theta$.[21]

As we show in the Theoretical Appendix, however, these complications do not arise whenever $(\sigma-1)(1-\eta) \geq \theta$, in which case the addition of a

[20] Hence, the addition of an input location decreases costs and increases revenue-based productivity for reasons quite distinct than in the love-for-variety frameworks in Halpern, Koren and Szeidl (2011), Goldberg, Khandelwal, Pavcnik and Topalova (2010), and Gopinath and Neiman (2013).

[21] The difficulties in solving for $\mathcal{J}_i(\varphi)$ are nicely discussed in Blaum, Lelarge and Peters (2013) in a model of input trade with very different features. It is worth pointing out, however, that one can easily show that the endogenous sourcing potential $\Theta_i(\varphi)$ is necessarily increasing in φ regardless of parameter values.

location to the set of active locations does not decrease the marginal benefit of adding further locations.[22] As a result, one can show that the number of locations to which a firm offshores is a monotonically increasing function of productivity φ, and even more strongly, that $\mathcal{J}_i(\varphi_0) \subseteq \mathcal{J}_i(\varphi_1)$ for $\varphi_1 \geq \varphi_0$. The model thus delivers a 'pecking order' in the extensive margin of offshoring that is reminiscent to the one typically obtained in models of exporting with heterogeneous firms, such as in Eaton, Kortum and Kramarz (2011). Furthermore, for a sufficiently low value of core productivity φ, the only profitable location of input production might be one associated with a low fixed cost of sourcing. Under the maintained assumption that $f_{ij} > f_{ii}$ for all $j \neq i$ – so domestic sourcing is the sourcing strategy associated with the lowest fixed costs –, the model thus continues to deliver selection into offshoring based on firm core productivity.

The more tractable case with $(\sigma - 1)(1 - \eta) \geq \theta$ is more likely to apply whenever demand is elastic and thus profits are particular responsive to variable cost reductions (high σ), and whenever input efficiency levels are relatively heterogeneous across markets (low θ), so that the expected reduction in costs achieved by adding an extra country into the set of active locations is relatively high. Naturally, this scenario is also more likely whenever headquarter intensity η is low, and thus changes in the cost of the input bundle cost have a relatively high impact on profits.

We can obtain sharper characterizations of the solution to the sourcing strategy problem in (2.29) by making further specific assumptions. For instance, when the fixed cost of offshoring is common for all foreign countries, so $f_{ij} = f_{iO}$ for all $j \neq i$, then regardless of the value of $(\sigma - 1)(1 - \eta)/\theta$, it is clear that locations j associated with a high value of $T_j(\tau_{ij}w_j)^{-\theta}$ will necessarily be more attractive than locations associated with low values of this term. In such a case, and regardless of the value of $(\sigma - 1)(1 - \eta)/\theta$, one could then rank foreign locations $j \neq i$ according to their value of $T_j(\tau_{ij}w_j)^{-\theta}$, and denote by $i_r = \{i_1, i_2, ..., i_{J-1}\}$ the country with the r-th

[22] It is not evident which of these cases is the most relevant empirically. The share of intermediate inputs in gross output of tradeable goods is roughly 50% (see, for instance, Alvarez and Lucas, 2007), indicating $\eta \approx 0.5$. Many estimates for the elasticity of substitution exist, but the consensus is that σ is somewhere between 3 and 6 (see, for instance, Broda and Weinstein, 2006). Eaton and Kortum (2002) estimated a value of $\theta = 8.28$, which would clearly suggest $(\sigma - 1)(1 - \eta) < \theta$. Nevertheless, Simonovska and Waugh (2011) have recently estimated lower values of θ, somewhere between 2.5 and 4.5, which could imply $(\sigma - 1)(1 - \eta) > \theta$ for a value of σ in the high range of typical estimates.

highest value of $T_j(\tau_{ij}w_j)^{-\theta}$. Having constructed i_r, it then follows that for any firm with productivity φ from i that offshores to at least one country, $i_1 \in \mathcal{J}_i(\varphi)$; for any firm that offshores to at least two countries, we have $i_2 \in \mathcal{J}_i(\varphi)$, and so on. In other words, not only does the extensive margin increase monotonically with firm productivity, but it does so in a manner uniquely determined by the ranking of the $T_j(\tau_{ij}w_j)^{-\theta}$ terms.

Even with variation of fixed costs of offshoring, a similar sharp result emerges in the knife-edge case in which $(\sigma-1)(1-\eta)=\theta$. In that case, the addition of an element to the set $\mathcal{J}_i(\varphi)$ has no effect on the decision to add any other element to the set, and the same pecking order pattern described in the previous paragraph applies, but when one ranks foreign locations according to the ratio $T_j(\tau_{ij}w_j)^{-\theta}/f_{ij}$. This result is analogous to the one obtained in standard models of selection into exporting featuring constant marginal costs, in which the decision to service a given market is independent of that same decision in other markets.

After having solved the sourcing strategy problem in (2.29), it is straightforward to compute the aggregate volume of intermediate inputs from any country j in the industry under consideration. These imports are given by

$$M_{ij}=(\sigma-1)(1-\eta)\widetilde{B}N_i\int_{\tilde{\varphi}_{ij}}^{\infty}\chi_{ij}(\varphi)\,\Theta_i(\varphi)^{(\sigma-1)(1-\eta)/\theta}\varphi^{\sigma-1}dG(\varphi), \quad (2.30)$$

where N_i is the measure of final-good entrants in country i, $\chi_{ij}(\varphi)$ is given in (2.26), $\widetilde{B}=(a_{hi}w_i)^{-\eta(\sigma-1)}\gamma^{(\sigma-1)(1-\eta)}B$, and $\tilde{\varphi}_{ij}$ is the productivity of the least productive firm from i offshoring to j. As long as a higher value of $T_j(\tau_{ij}w_j)^{-\theta}$ is associated with a (weakly) higher probability that country j belongs to the set $\mathcal{J}_i(\varphi)$, it is then clear from (2.30) that a high value of $T_j(\tau_{ij}w_j)^{-\theta}$ leads to a large volume of imports from that country j on account of both the intensive and extensive margins of trade.

Interestingly, in the special case in which the fixed costs of offshoring are low enough to ensure that all firms acquire the capability to source inputs from *all* countries, equation (2.30) reduces to a modified version of the gravity equation, analogous to that in Eaton and Kortum (2002). To see this, note that whenever $\mathcal{J}_i(\varphi)=\{1,2,....,J\}$ for all φ and i, (2.30) can be written as

$$M_{ij}=(\sigma-1)(1-\eta)\widetilde{B}N_i\left(\bar{\Theta}_i\right)^{(\sigma-1)(1-\eta)/\theta}\chi_{ij}\int_{\underline{\varphi}_i}^{\infty}\varphi^{\sigma-1}dG(\varphi), \quad (2.31)$$

where

$$\bar{\Theta}_i \equiv \sum_{k=1}^{J} T_k \left(\tau_{ik} w_k\right)^{-\theta}$$

and

$$\chi_{ij} = \frac{T_j \left(\tau_{ij} w_j\right)^{-\theta}}{\bar{\Theta}_i}.$$

Defining $A_i = \sum_j M_{ij}$ as the total absorption of intermediate inputs by firms in i, and by $Q_j = \sum_l M_{lj}$ as the total production of intermediates in country j, it is straightforward to verify that (2.31) in fact reduces to

$$M_{ij} = \frac{\left(\tau_{ij}\right)^{-\theta} \frac{A_i}{\bar{\Theta}_i}}{\sum_l \left(\tau_{lj}\right)^{-\theta} \frac{A_l}{\bar{\Theta}_l}} Q_j,$$

which is analogous to equation (11) in Eaton and Kortum (2002).

When the extensive margin (or sourcing potential $\Theta_i(\varphi)$) varies across firms, such a neat expression no longer applies. This suggests that one might be able to infer the importance of the firm-level extensive margin (and of cross-country variation in the fixed costs of offshoring) from observed deviations from the traditional gravity equation. This is one of the approaches currently being explored by Antràs, Fort and Tintelnot (2014) in their study of the extensive margin of offshoring of U.S. firms.

Further Reading

This concludes my overview of the key benchmark models of international trade I will be building on in future chapters. Although most of the papers I have discussed are quite recent, there exist already a number of useful reviews of this literature. For instance, three of the chapters in the forthcoming fourth volume of the *Handbook of International Economics*, namely Melitz and Redding (2013*a*), Antràs and Yeaple (2013), and Costinot and Rodríguez-Clare (2013), cover these models in significant detail. The multi-country model of global sourcing is novel to this book, but it is being further developed in Antràs, Fort and Tintelnot (2014), where it is used to interpret and structurally estimate the extensive margin of U.S. intermediate input imports.

With this machinery at hand, we are now ready to begin our theoretical exploration of the implications of contractual imperfections for the global organization of production.