Firms, Contracts and Global Production

Pol Antràs

September 2013

Contents

Preface

This book has grown out of the *CREI Lectures in Macroeconomics* that I delivered in Barcelona in June of 2012. I am grateful to the Series Editor, Hans-Joachim Voth, for inviting me to deliver these lectures and for encouraging me to accept the invitation. Part of my initial reticence, I can now admit, was related to the fact that I viewed the invitation to give these prestigious lectures as a flagrant instance of 'home bias'. At the time, the CREI Lecture series committee was composed of a coauthor of mine (Hans-Joachim Voth), one of my Ph.D. advisors (Jaume Ventura) and two of my favorite teachers during my undergraduate studies at Universitat Pompeu Fabra (Antonio Ciccone and Jordi Galí). Regardless of their motivations, it was an honor to have been selected as the 2012 CREI Lectures speaker. I am grateful for the comments and feedback I received during my lectures and also for CREI's hospitality during my many visits there.

The book is largely aimed at graduate students and researchers interested in learning about recent developments in the field of International Trade. I have attempted, however, to make the style of the book a bit less terse than is standard in professional journals and graduate-level textbooks. This may alienate some technically-oriented readers, but will hopefully encourage some advanced undergraduate students and trade practitioners to venture into the material. Chapter 1, in particular, provides an overview of the topics covered in later chapters at a highly accessible level. Although it would be hard to sell this book as being a set of *Lectures in Macroeconomics*, I hope that some of the material in this lecture will appeal to researchers in that field, as well as readers interested in Organizational Economics, and Applied Contract Theory.

Although the bulk of the contents of this book has appeared in some form or other in academic journals, many chapters of this book include new and original work. For instance, the multi-country global sourcing model introduced in Chapter 2 and further developed in Chapters 4, 5 and 6 is a novel contribution of this book. The same is true about the multi-country model of limited commitment in Chapter 3. Many of the empirical parts of the book are original as well, although they build heavily on previous work in terms of methodology and data sources.

I have taught most of the material in these lectures at Harvard but also at Study Center Gerzensee, the London School of Economics, Penn State University, the University of Zurich and Northwestern University. I have found that between four and five 90-minute lectures are generally sufficient to cover the contents of this book. I am grateful to all these institutions for their hospitality and to the lecture participants for many useful comments.

Although I have attempted to provide a broad overview of the topics in this lecture, the spirit of the CREI Lectures dictated me to have my own work feature prominently in this book. For this reason, my greatest debt is to my coauthors on the papers overviewed in this book, including Daron Acemoglu, Davin Chor, Fritz Foley, Esteban Rossi-Hansberg, Bob Staiger, Steve Yeaple, and especially Elhanan Helpman. I am also particularly grateful to my colleagues Elhanan Helpman and Marc Melitz for many stimulating discussions that have shaped my thinking on the material of this book. My interest in the contracting aspects of global production dates back to my Ph.D. years at MIT and I am indebted to Daron Acemoglu, Gene Grossman, Bengt Holmstrom and Jaume Ventura for their encouragement during those initial phases of this intellectual adventure.

Turning my lecture slides into a book manuscript has proved to be much harder than I first anticipated. Mireia Artigot, Teresa Fort, and Felix Tintelnot read different parts of the first draft of this book and provided very useful feedback and corrections. I am also grateful to Eric Unverzagt for his careful editorial assistance.

Several colleagues have kindly shared their data for some of the empirical material in this book. These include Andrew Bernard, Davin Chor, Robert Johnson, Nathan Nunn, and Greg Wright. I have also benefited from the outstanding research assistance of Ruiqing Cao, Yang Du, and Boo-Kang Seol during various periods over which this book was written. They are of course not responsible for any mistakes left in the manuscript. Finally, I am forever grateful to my wife Lucia and my daughters Daniela and Martina for their patience during the many hours I have spent mulling over and writing this book.

Part I

Introduction

Chapter 1

Made in The World

It does not feel like a long time ago when I began my undergraduate studies in Economics at Universitat Pompeu Fabra (UPF), the same institution that hosts the Centre de Recerca en Economia Internacional (CREI), where these Lectures were delivered. It was 1994 and I felt like a lived in a truly global economy. The music I listened to and the movies I watched were mostly British or American. Most of the clothes I wore were manufactured abroad, some of them in rather exotic places such as Morocco or Taiwan. My favorite beer was Dutch. At UPF, about half of my teachers were foreign, a third of the classes were taught in English and most of the textbooks were the same ones used in universities around the globe.

In hindsight, it seems pretty clear, however, that the world had not yet witnessed the full advent of globalization. What has changed since 1994? *First* and foremost, the last two decades have brought a genuine information and communication technology (ICT) revolution that has led to a profound socioeconomic transformation of the world in which we live. The processing power and memory capacity of computers have doubled approximately every two years (as implied by Moore's law), while the cost of transmitting a bit of information over an optical network has decreased by half roughly every nine months (a phenomenon often referred to as Butter's law). The number of internet users has increased by a factor of 100, growing from around 20 million users in 1994 to more than 2,000 million users in 2010 (see World Development Indicators). As a result of these technological developments, the cost of processing and transmitting information at long distances has dramatically fallen in recent times. Consider the following example: the 3.3GB file containing my favorite movie of 1994, *Pulp Fiction*, can be downloaded today from Amazon.com in about 11 minutes and 16 seconds using a standard broadband connection with a download speed of 5 megabits per second. In 1994, downloading that same file using a dial-up connection and the state-of-the-arts modem, which allowed for a maximum speed of 28.8 kilobits per second, would have kept your phone line busy for at least 33 hours and 23 minutes![1]

Second, during the same period, governments have continued (and arguably intensified) their efforts to gradually dismantle all man-made trade barriers. This process dates back to the initial signing of the General Agreement on Tariffs and Trade (GATT) in 1947, but it has experienced a revival in the 1990s and 2000s with the gradual expansion of the European Union, the formation of the North American, Mercosur, and ASEAN free trade agreements, the signing of a multitude of smaller preferential trade agreements under the umbrella of GATT's Article XXIV, and China's accession to the WTO, just to name a few. As a consequence, the world's weighted average tariff applied on traded manufactured goods fell from 5.14% in 1996 to 3.03% in 2010 (see World Development Indicators).[2]

Third, political developments in the world have brought about a remarkable increase in the share of world population actively participating in the process of globalization. These changes largely stemmed from the fall of communism in Eastern Europe and the former Soviet Union, but also from an ensuing ideological shift to the right in large parts of the globe. Thus, not only did former communist countries embrace mainstream capitalist policies, but these policies themselves became more friendly towards globalization, as exemplified by the deepening of trade liberalization mentioned in the last paragraph, but also by a notable relaxation of currency convertibility and balance of payments restrictions in several low and middle-income countries.[3]

[1] Paraphrasing a memorable quote from Samuel L. Jackson's character in *Pulp Fiction*, download speeds today and in 1994 "ain't the same [*freaking*] ballpark. They ain't the same league. They ain't even the same [*freaking*] sport."

[2] Technological developments since 1994 have also reduced the quality- (or time-) adjusted costs of transporting goods across countries (see Hummels, 2007), while investments in infrastructure in less developed economies have also contributed to spreading the effects of globalization across regions in those countries.

[3] The late 1990s also saw the emergence of a left-leaning anti-globalization movement, which drew particular attention during the 1999 WTO meetings in Seattle. There is little evidence, however, of this movement having led to any significant slow down in the process of globalization (see, for instance, Harrison and Scorse, 2010).

The Slicing of the Value Chain

One of the manifestations of these three developments in the world economy has been a gradual disintegration of production processes across borders. More and more firms now organize production on a global scale and choose to offshore parts, components or services to producers in foreign and often distant countries. The typical "Made in" labels in manufactured goods have become archaic symbols of an old era. These days, most goods are "Made in the World."

A variety of terms have been used to refer to this phenomenon: the "slicing of the value chain", "fragmentation of the production process", "disintegration of production", "delocalization", "vertical specialization", "global production sharing", "unbundling", "offshoring" and many more (see Feenstra, 1998). I shall use these terms interchangeably throughout the book.

The case of Apple's iPad 3 tablet nicely illustrates the magnitude of this new form of globalization. The slim and sleek exterior of the tablet hides a complex manufacturing process combining designs and components provided by multiple suppliers with operations in various countries. Although Apple does not disclose detailed information on its input providers, a clear picture of the global nature of the iPad 3 production process emerges when combining information from tear-down reports (such as those published by isuppli.com and ifixit.com) with various press releases.[4] For instance, it is well-known that the tablet itself is assembled in China (and since 2012 also in Brazil) by two Taiwan-based companies, Foxconn and Pegatron. The revolutionary retina display is believed to be manufactured by Samsung of South Korea in its production plant in Wujiang City, China. The distinctive touch panel is produced (at least, in part) by Wintek, a Taiwan-based company that also owns plants in China, India and Vietnam, while the case is provided by another Taiwanese company, Catcher Technologies, with operations in Taiwan and China. A third important component, the battery pack, also originates in Taiwan and is sold by Simplo Technologies and Dynapack Inter-

[4]Facing strong criticism over the working conditions in its suppliers' factories, Apple released a full list of its 156 global suppliers early in 2012 (see http://images.apple.com/supplierresponsibility/pdf/Apple_Supplier_List_2011.pdf). Teardown reports further faciliate a mapping between the iPad parts and their respective producers. Press releases sometimes also identify particular suppliers with specific iPad 3 components (see, for instance, Forbes' "Batteries Required?" available at http://www.forbes.com/global/2010/0607/best-under-billion-10-raymond-sung-simplo-technology-batteries-requried.html).

national. Apart from these easily identifiable parts, the iPad 3 incorporates a variety of chips and other small technical components provided by various firms with headquarters and R&D centers in developed economies and manufacturing plants scattered around the world. A non-exhaustive list includes (again) Korea's Samsung, which is believed to manufacture the main processor (designed by Apple), U.S.-based Qualcomm supplying 4G modules, and Italo-French STMicroelectronics contributing key sensors.[5]

Apple's sourcing strategies are far from being an isolated example of a global approach to the organization of production. In fact, the increasing international disintegration of production processes has been large enough to be salient in aggregate statistics. During the 1990s and early 2000s, when this phenomenon was still in its infancy, researchers devised several approaches to measuring the quantitative importance of global production sharing.[6] Feenstra and Hanson (1996*b*), for instance, used U.S. Input-Output tables to infer the share of imported inputs in the overall intermediate input purchases of U.S. firms; they found that this share had already increased from 5.3% in 1972 to 11.6% in 1990. Campa and Goldberg (1997) found similar evidence for Canada and U.K., but surprisingly not for Japan, where the reliance on foreign inputs appeared to have declined between 1974 and 1993. Hummels, Ishii and Yi (2001) instead constructed a measure of vertical specialization capturing the value of imported intermediate inputs (goods and services) embodied in a country's exported goods and found that it already accounted for up to 30% of world exports in 1995, having grown by as much as 40% since 1970.

The work of Johnson and Noguera (2012*a*, 2012*b*) constitutes the state of the arts in the use of Input-Output tables to quantify the importance of global production sharing and its evolution in recent years.[7] The main innovation of their methodology is in the attempt to compute a *global* Input-Output table from which one can back out the value-added and intermediate input contents of gross international trade flows. In particular, their VAX ratio (the value-added to gross-value ratio of exports) is an appealing inverse measure of the importance of vertical specialization in the world production: the

[5] A more extensive list can be found at: http://www.chipworks.com/en/technical-competitive-analysis/resources/blog/the-new-ipad-a-closer-look-inside/.

[6] The task is complicated by the fact that data on trade flows of goods is collected on a gross output basis, without regard to the particular sources of the value added embodied in these goods.

[7] See also Koopman, Wang and Wei (2013).

lower is this measure is, the larger is the value of imported inputs embodied in exports. As is clear from Figure 1.1, their VAX ratio has declined rather significantly since 1970 with about two-thirds of the decline occurring after 1990. Johnson and Noguera (2012*b*) show that this decline is explained solely by increased fragmentation within manufacturing. Furthermore, they also find that fragmentation has grown disproportionately in emerging economies and also appears to increase following the signing of regional trade agreements.

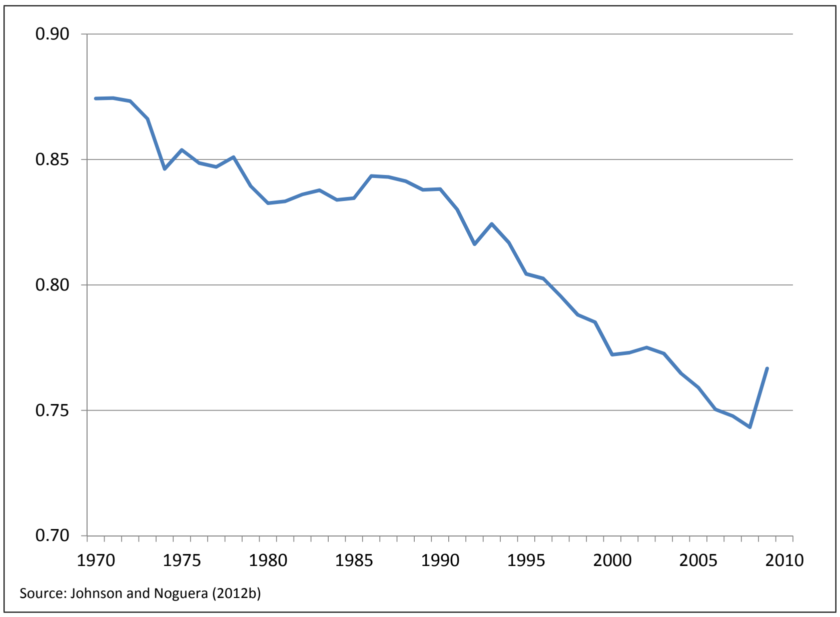

Figure 1.1: Ratio of Value Added to Gross Exports (VAX), 1970-2009

Two limitations of the fragmentation measures discussed so far are that they rely on fairly aggregated Input-Output data and that they impose strong proportionality assumptions to back out the intermediate input component of trade. A different approach to measuring the degree to which production processes are fragmented across countries was first suggested by Yeats (2001), and consists in computing the share of trade flows accounted for by SITC Rev.2 industry categories that can be safely assumed to contain *only* intermediate inputs (as reflected by the use of the word "Parts of" at the beginning of the category description). It turns out that all these industries are in the "Machinery and Transport Equipment" industrial group (or SITC 7). Yeats (2001) found that intermediate input categories accounted for about 30% of OECD merchandise exports of machinery and transport equipment in

1995, and that this share had steadily increased from its 26.1% value in 1978. A limitation of Yeats' measure is that, by focusing on industries composed *exclusively* of inputs, it naturally understates the importance of input trade. This might explain why when updating this methodology to present times, one finds little evidence of a further increase in this share.[8]

An alternative to categorizing trade flows as *either* final goods or intermediate inputs is to attempt to calculate a more continuous measure of the "upstreamness" of the goods being traded. This is the approach in Antràs, Chor, Fally and Hillberry (2012), who use Input-Output data to construct a weighted index of the average position in the value chain at which an industry's output is used (i.e., as final consumption, as direct input to other industries, as direct input to industries serving as direct inputs to other industries, and so on), with the weights being given by the ratio of the use of that industry's output in that position relative to the total output of that industry. Intuitively, the higher is this measure, the more removed from final good use (and thus the more upstream) is that industry's output.[9] Antràs, Chor, Fally and Hillberry (2012) use the measure to characterize the average upstreamness of exports of different countries in 2002, but it can also be employed to illustrate how the upstreamness of world exports has evolved in recent years. As shown in Figure 1.2, world exports became significantly more upstream in recent years, particularly in the period 2002-08. The patterns are in line with those illustrated in Figure 1.1, and also suggest an increase predominance of input trade in world trade. Although a significant share of the observed increase in upstreamness is related to an increase the relative weight of petroleum-related industries, even when netting those out, one observes a significant upward trend in upstreamness (see Figure 1.2). Interestingly, both Figures 1.1 and 1.2 identify a disproportionate fall in global production sharing relative to the overall fall in world trade during the early years of the recent 'great recession'.

[8] Other authors attempting to compute the share of intermediate inputs in world trade using alternative methodologies have also found little evidence of a trend in the series (see, for instance, Chen, Kondratowicz and Yi, 2005, or Miroudot, Lanz and Ragoussis, 2009). I have obtained similar results when computing the relative growth of overall trade and input trade using the classification of goods developed by Wright (2012). As argued by Johnson and Noguera (2012*b*), even when taking this finding at face value, it is not necessarily inconsistent with the observed rise in indices of vertical specialization, which better capture the use of imported inputs in producing goods *that are exported.*

[9] The measure was independently developed by Antràs and Chor (2013) and Fally (2012), and its properties were further studied in Antràs, Chor, Fally and Hillberry (2012).

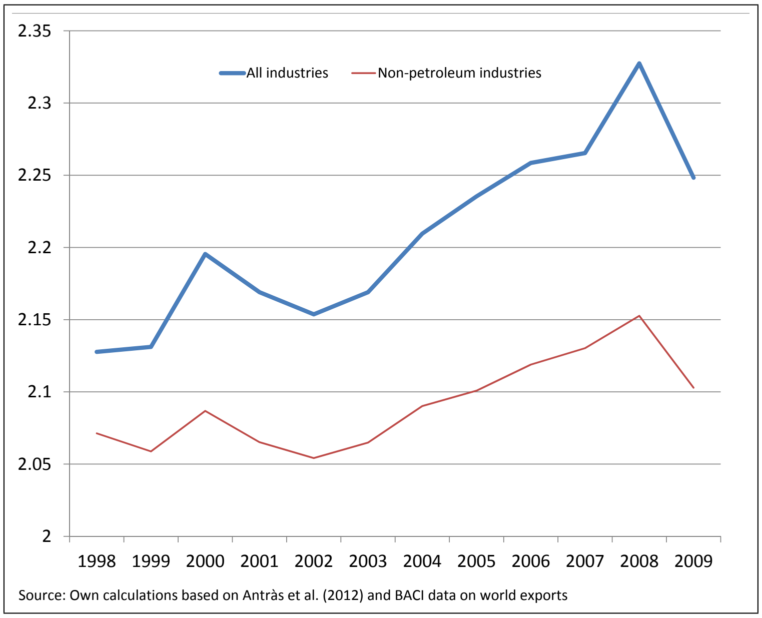

Figure 1.2: Average Upstreamness of World Exports, 1998-2009

Old and New Theories

The noticeable expansion in input trade has also captured the attention of international trade theorists eager to bridge the apparent gap between the new characteristics of international trade in the data and the standard representation of these trade flows in terms of final goods in traditional and new trade theory.

One branch of this new literature has focused on incorporating the notion of fragmentation in otherwise neoclassical models with homogeneous goods, perfectly competitive markets and frictionless contracting. Key contributions include Feenstra and Hanson (1996*a*), Jones (2000), Deardorff (2001), and Grossman and Rossi-Hansberg (2008). The main idea in these contributions is that the production process (as represented by an abstract mapping between factors of production and final output) can be decomposed into smaller parts or stages that are themselves (partly) tradable. Different authors assign different labels to these parts: some refer to them as intermediate inputs, others call them vertical production stages, while others view them as tasks. Regardless of the interpretation of the process under study, a common lesson from this body of work is that the possibility of fragmentation generates nontrivial effects on productivity, and that these endogenous changes in productivity in turn deliver novel predictions for the effects of reductions in trade costs on patterns of specialization and factor prices. Antràs and Rossi-Hansberg (2009) elaborate on this broad interpretation of this branch of the literature and also offer more details on the specific results of each of these contributions.[10]

As insightful as this body of work has proven to be, it seems clear that modeling global production sharing as simply an increase in the tradability of homogeneous inputs across countries misses important characteristics of intermediate input trade. Prominent among these features is the fact that parts and components are frequently customized to the needs of their intended buyers (remember our example above with the iPad 3). In other words, the disintegration of the production process is more suitably associated with the growth of trade in *differentiated* (rather than homogeneous) intermediate inputs.[11]

Another important characteristic of global production networks is that they necessarily entail intensive *contracting* between parties located in different countries and thus subject to distinct legal systems. In a world with perfect (or complete) contracting across borders, this of course would be of little relevance. Unfortunately, this is not the world we (or at least, I) live in. Real-world commercial contracts are incomplete in the sense that they cannot possibly specify a course of action for *any* contingency that could arise during the course of a business relationship. Of course, the same can be said about domestic commercial transactions, but the cross-border exchange

[10] Another common feature of the theoretical frameworks developed in these papers is that the number of primitive factors of production is assumed small, and normally equal to two. Another branch of the literature has developed perfectly-competitive, frictionless models in which offshoring results from the assignment of a population of a large number of heterogeneous agents into international hierarchical teams (see Kremer and Maskin, 2006, or Antràs, Garicano and Rossi-Hansberg, 2006).

[11] Admittedly, there does not exist much evidence to substantiate this claim. Antràs and Staiger (2012) offer a back-of-the envelope quantification applying the methodology suggested by Schott (2004) to identify international trade in intermediate goods and using the "liberal" classification of Rauch (1999) to distinguish between differentiated and homogeneous goods. They find that the share of differentiated inputs in world trade more than doubled between 1962 and 2000, increasing from 10.56% to 24.85% of world trade. Behar and Freund (2011) show that during the late 1990s and 2000s, intermediate inputs traded within the EU became more sophisticated and involved more relationship-specific investments (in the sense of Nunn, 2007).

of goods cannot generally be governed by the same contractual safeguards that typically accompany similar exchanges occurring within borders.

Given the subject of this book, it is worth pausing to describe in more detail some of the factors that make international contract enforcement particularly problematic.

Contracts in International Trade

A first natural difficulty in contractual disputes involving international transactions is determining which country's laws are applicable to the contract being signed. In principle, the parties can include a *choice-of-law* clause specifying that any dispute arising under the contract is to be determined in accordance with the law of a particular jurisdiction, regardless of where that dispute is litigated. Nevertheless, many international contracts do not include that clause and, in any case, it is up to the court of law adjudicating a dispute to decide whether it will uphold the expressed desire of the parties. If the court is not familiar with the law specified in the contract, as may often occur in international transactions, the court might decide to rule on the basis of its own law, or they may inadvertently apply the desired foreign law incorrectly.

A second difficulty relates to the fact that even when local courts are competent (in a legal sense), judges may be reluctant to rule with regard to a contract dispute involving residents of foreign countries, especially if such a ruling would entail an unfavorable outcome for local residents. The evidence on the home bias of local courts is mixed, but even those authors advocating that a formal analysis of case law does not support the hypothesis of biases against foreigners readily admit a widespread belief of the existence of such xenophobic biases (see Clermont and Eisenberg, 2007).

A third complication with international contracts relates to the enforcement of remedies stipulated in the court's verdict. For instance, the court might rule in favor of a local importer that was unsatisfied with the quality of certain components obtained from an exporter, and the verdict might require the exporter to compensate the importer for any amount already paid for the components, as well as for any court or even attorney fees incurred. An issue arises, however, if the exporter does not have any assets (say bank accounts or fixed assets) in the importer's country. In such case, it is not clear that the exporter will feel compelled to accept the verdict and pay the importer.

In recent years, there have been several coordinated attempts to reduce the contractual uncertainties and ambiguities associated with international transactions. A particularly noteworthy example is the United Nations Convention on Contracts for the International Sale of Goods (or CISG), or Vienna Convention, which attempts to provide a set of uniform rules to govern contracts for the international sale of goods. The idea is that even when an international contract does not include a choice-of-law clause, parties whose places of business are in different signing countries can rely on the Convention to protect their interests in courts. As ambitious as the CISG initiative is, it has arguably fallen short of its objectives. For instance, several countries or regions (most notably, Brazil, Hong Kong, India, South Africa, Taiwan, and the United Kingdom) have yet to sign the agreement. Furthermore, a few of the signing countries have expressed reservations and choose not to apply certain parts of the agreement. Finally, it is not uncommon for private parties to explicitly opt out of the application of the Convention, as allowed by its Article 6. The reluctance to unreservedly embrace the Convention has been associated with the somewhat vague language of the text, which might foster the natural inclination of judges to interpret the Convention through the lens of the laws of their own State.[12]

Another attempt to ameliorate the perceived contractibility of international transactions consists in resorting to international arbitration. More specifically, an international trade contract can include a (so-called) *forum-of-law* clause establishing that a particular arbitrator, such as the International Chamber of Commerce (ICC) in Paris, will resolve any contractual dispute that may arise between the parties. International arbitration is appealing because it avoids the aforementioned uncertainties associated with litigation in national courts. It is also relatively quick and parties benefit from the fact that arbitrators tend to have more commercial expertise than a typical judge. Furthermore, arbitration rulings are confidential and are generally perceived to be more enforceable than those of national courts because they are protected by the Convention on the Recognition and Enforcement of Foreign Arbitral Awards, also known as New York Convention. Despite its attractive features, international arbitration is rarely used in practice because its cost

[12] The Institute of International Commercial Law at Pace Law School maintains a website (http://www.cisg.law.pace.edu/) with comprehensive information on the CISG, including a database of thousands of legal cases in which the Convention was invoked. The details of these cases offer a vivid account of the nature of contractual disagreements in international trade.

is too high for most firms to bear.[13]

One might argue that even when explicit contracts are incomplete and perceived to be unenforceable, parties in international transactions can still resort to implicit contracting to sustain 'cooperation'. Nevertheless, it is particularly hard to render international commercial relationships self-enforcing. On the one hand, international parties are less likely to meet face-to-face and to transact on a repeated basis than domestic parties, in part due to distance and trade costs, but also due to shocks (such as exchange rate movements) that can quickly turn efficient relationships into inefficient ones. On the other hand, the possibility of collective or community enforcement is hampered again by distance but also by the fact that parties might have different cultural and societal values. In sum, and in the words of Rodrik (2000), "ultimately, [international] contracts are often neither explicit nor implicit; they simply remain incomplete."

Although contractual risks are also of relevance for the exchange of final goods (see Chapter 3), the detrimental effects of imperfect international contract enforcement are likely to be particularly acute for transactions involving intermediate inputs. This is so for at least two reasons. First, input transactions are often associated with relatively long time lags between the time an order is placed (and the contract is signed) and the time the goods or services are delivered (and the contract is executed). Second, parts and components often entail significant relationship-specific investments and other sources of lock-in on the part of both buyers and suppliers, which make contractual breaches particularly costly. As argued above, suppliers often customize their output to the needs of particular buyers and would find it hard to sell those goods to alternative buyers, should the intended buyer decide not to abide by the terms of the contract. Similarly, buyers often undertake significant investments whose return can be severely diminished by incompatibilities, production line delays or quality debasements associated with suppliers not

[13] It may be instructive to illustrate this claim with some figures. Using the arbitration cost calculator available from the ICC website, the estimated cost of arbitration (involving a single arbitrator) would be $5,401 for a $10,000 dispute (or a 54% cost-to-dispute-amount ratio), $15,425 for a $100,000 dispute, $61,094 for a $1 million dispute, and $170,799 for a $10 million dispute (or a mere 1.7% cost-to-dispute-amount ratio). It is thus little surprise that there were only 796 ICC arbitrations requests in 2011 and that the amount in dispute was under one million U.S. dollars in only 22.7% of these cases (see http://www.iccwbo.org/products-and-services/arbitration-and-adr/arbitration/cost-and-payment/cost-calculator/).

going through with their contractual obligations.[14]

Firm Responses to Contractual Insecurity

When designing their global sourcing strategies, firms face two key decisions. The first one concerns the *location* of the different stages in the value chain and involves deciding in which country or region will firms conduct R&D and product development, where parts and components should be produced, what is the best place to assemble the finished good, and so on. The second key decision relates to the extent of *control* that firms exert over these different production stages. For instance, firms may decide to keep these production stages within firm boundaries, thus engaging in foreign direct investment (FDI) when the integrated entity is in a foreign country. Other firms may be less inclined to keep a tight control over certain stages and thus choose to contract with suppliers or assemblers at arm's-length.

Neoclassical models of fragmentation focus exclusively on the first of these decisions and emphasize that fragmentation will emerge as part of a competitive equilibrium whenever firms find it cost-minimizing to break up production process across countries. The source of the cost-advantage associated with fragmentation varies by model, sometimes it stems from differences in relative factor endowments across countries (which, for instance, naturally confer comparative advantage in labor-intensive stages to relatively labor-abundant countries), while other times they are motivated by technological differences across countries.

Neoclassical models are silent on the issue of control. This is not because these models assume perfect competition, increasing returns to scale, or homogeneous goods. Instead, the key assumption that renders those models (and just about *any* model in the field of International Trade) vacuous when tackling the notion of control is the assumption of perfect or complete contracting. Indeed, if firms could foresee all possible future contingencies, and if they could costlessly write contracts that specify in an enforceable manner the course of action to be taken in all of these possible contingencies, then

[14] A third more specific reason for which input trade might be perceived to be less contractually secure relates to the fact that Article 3 of CISG explicitly excludes from the applicability of the Convention situations in which "the party who orders the goods undertakes to supply a substantial part of the materials necessary for such manufacture or production," thus making the Convention less relevant for sustaining cooperation in global production sharing networks.

firms would no longer need to worry about "controlling" the workers, the internal divisions or the supplying firms with whom they interact in production. The complete contract would in fact confer *full* control to the firm regardless of the ownership structure that governs the transactions between all these producers. In other words, and as Coase (1937) anticipated more than seventy-five years ago, firm boundaries are indeterminate in a world of complete contracts.[15]

In the real world, however, contracts are very much incomplete and especially so in international transactions, where as argued above, the enforceability of contracts is particularly questionable. In response to this perceived contractual insecurity, firms spend a substantial amount of time and resources figuring out the best possible way to *organize* production in the global economy. In some cases, foreseeing that producers located in a particular country might not feel compelled to follow through with their contractual obligations, firms contemplating doing business in that country might decide to do so within their firm boundaries, either by setting up a new, wholly- or partially-owned affiliate or by acquiring a controlling stake in an existing firm in that country. In some circumstances, however, the lack of contract enforceability might precisely turn firms to independent suppliers for the procurement of parts because such an arrangement might elicit the best performance from foreign producers. In other words, it is important to keep in mind that internalization is a double-edged sword: it may partly protect the integrating party from the vagaries of international contracting, but it might dilute the incentives to produce efficiently of the integrated party, which is now more tightly controlled and has less power in the relationship (cf., Grossman and Hart, 1986).

The boundaries of firms in the world economy are thus the result of the (constrained) optimal decisions of firms attempting to organize production in the most profitable way possible. A recurring theme of this book is that much can be learned from a theoretical and empirical study of the *fundamental* forces that appear to shape whether international transactions are internalized or not, independently of the firm or sector one is studying.

Practitioners (and perhaps some academics too) might react skeptically

[15] It is worth stressing that even in the presence of product differentation and market power, firm boundaries remain indeterminate when contracts are complete. For example, the often-cited double-marginalization rationale for vertical integration rests on the assumption that firms and suppliers cannot sign simple two-part tariff contracts, and as such, it also constitutes an incomplete-contracting theory of firm boundaries.

to the idea that low-dimensional models can possibly capture the reasoning behind the complex and idiosyncratic decisions of firms in the world economy. Business school cases often highlight the peculiarities of particular organizational decisions, making it hard to envision that much can be gained from extrapolating from those particular cases. The fact that comprehensive datasets on the integration decisions of firms are not readily available might have only compounded this belief, as most empirical studies of integration decisions rely on data from specific industries or firms.[16]

A Comparative Advantage of Trade Statistics

An advantage of studying the global integration decisions of firms is that data on international transactions are particularly accessible due to the widespread existence of official records of goods and services crossing borders. For instance, it is well-known that researchers can easily access data on U.S. imports from any country of the world at the remarkably detailed ten-digit Harmonized Tariff Schedule classification system, which consists of nearly 17,000 categories.[17] A less well-known fact is that, in some countries, these same detailed country- and product-level data contain information on the extent to which trade flows involve related parties or non-related parties. Most notably, the "U.S. Related Party Trade" data collected by the U.S. Bureau of Customs and Border Protection and managed by the U.S. Census Bureau provides data on related and non-related party U.S. imports and exports at the six-digit Harmonized System (HS) classification (which consists of over 5,000 categories) and at the origin/destination country level. This amounts to hundreds of thousands of observations *per year* on the relative prevalence of integration versus nonintegration across products and countries.[18]

What do these data tell us about the global sourcing strategies of firms? The first thing that one notices when using U.S. related party trade data is

[16] See Baker and Hubbard (2003) for a particularly careful study using data from the trucking industry and Lafontaine and Slade (2007) for a broad survey of the empirical literature on vertical integration.

[17] Downloading this data from the NBER website, one can readily verify that in 2001 France exported $15,747 worth of frozen potatoes to the United States (HTS code 2004.10), yet none of those were French fries (HTS code 2004.10.8020).

[18] The U.S. Related Party Trade data are publicly available at: http://sasweb.ssd.census.gov/relatedparty/. This website permits downloading the data at the six-digit NAICS level. The finer six-digit Harmonized System (HS) data are available from the U.S. Census for a fee.

how predominant intrafirm transactions are in U.S. trade. In 2011, intrafirm imports of goods totaled $1056.2 billion and constituted a remarkable 48.3 percent of total U.S. imports of goods ($2,186.9 billion). In fact, the share of intrafirm trade has been higher than 46.5 percent in every year since 2000. On the export side, related-party exports are also pervasive, with their share in total U.S. exports ranging from 28 to 31 percent in recent years. These figures illustrate the importance of multinational firms for U.S. trade.[19]

A second evident feature of the data is that the share of U.S. intrafirm imports varies widely across countries. On the one hand, in 2011 intrafirm imports equalled 0 for 10 countries and territories (including Cuba), all exporting very low volumes to the U.S. On the other hand, in that same year the share of intrafirm trade reached a record 89.6 percent for U.S. imports from Western Sahara. Leaving aside communist dictatorships and disputed territories, and focusing on the 50 largest exporters to the U.S., Figure 1.3 illustrates that the share of intrafirm trade still varies significantly across countries, ranging from a mere 2.4 percent for Bangladesh to an astonishing 88.5 percent for Ireland.

Similarly, the share of intrafirm trade varies widely depending on the type of product being imported. Again, the raw data contain infrequently traded goods with shares equal to 0 and 100, but even when focusing on the top 20 six-digit HS manufacturing industries by importing volume, in Figure 1.4 one observes significant variation in the share of intrafirm trade, which ranges from a share of 11.4 percent for U.S. imports of sweaters, pullovers and sweatshirts made of cotton (HS 611020) to 98.8 percent for imports of automobiles with engines of more than 3000 cc (HS 870324). This variation persists even when focusing on much narrower sectors. As shown in Figure 1.5, when analyzing imports across subcategories of the four-digit Harmonized System sector 8708 ('Parts and accessories of motor vehicles'), the share of intrafirm trade still ranges from 19.8 percent for drive axles (HS 870850) to 71.2 percent for steering wheels (HS 870894). It is thus clear that

[19]In contrast, Atalay, Hortacsu and Syverson (2013) study intrafirm shipments across U.S. multiplant firms and find that these constitute a very small share of total shipments, a finding that they interpret as indicating that firm boundaries are shaped by issues related to the the transfer of intangible inputs, rather than of physical goods. However, as argued above, contractual insecurity in the exchange of physical inputs is much more significant in international transactions than in domestic ones (especially, in the U.S. case), and thus firm boundaries might well be shaped by different factors in cross-border relationships than in domestic ones.

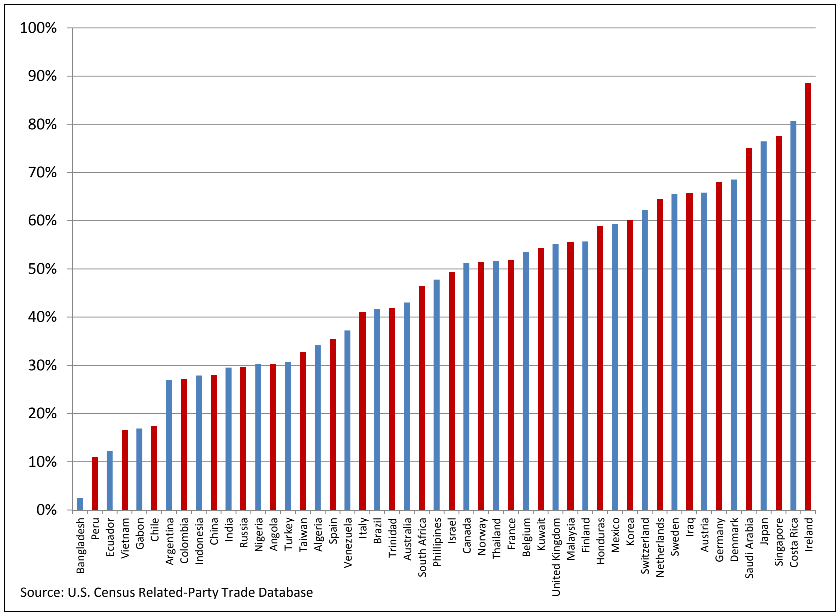

Figure 1.3: Share of U.S. Intrafirm Imports for Largest 50 U.S. Exporters in 2011

U.S.-based firms appear to source different auto parts under quite different ownership structures.

As a final illustration of the richness and variation in the data, consider the six-digit HS industry with the largest share of intrafirm imports in Figure 1.5, namely HS 870894 (Steering Wheels, columns and boxes for motor vehicles). Figure 1.6 reports the share of intrafirm trade for all 56 countries with positive exports to the U.S. in that sector. As is clear from the graph, even when focusing on a narrowly-defined component, a similar pattern to that in Figure 1.3 emerges, with U.S.-based producers appearing to source particular inputs quite differently depending on the location from which these products are bought. Imports from 17 of the 56 countries are exclusively transacted at arm's-length, while one country (Liechtenstein) sell steering wheels to the U.S. almost exclusively within multinational firm boundaries. The remaining 38 countries feature shares of intrafirm trade fairly uniformly distributed between 0 and 100 percent.

The large variation in the relative importance of intrafirm transactions across types of goods and countries might seem to validate the skeptics' view that the decision to integrate or outsource foreign production processes is

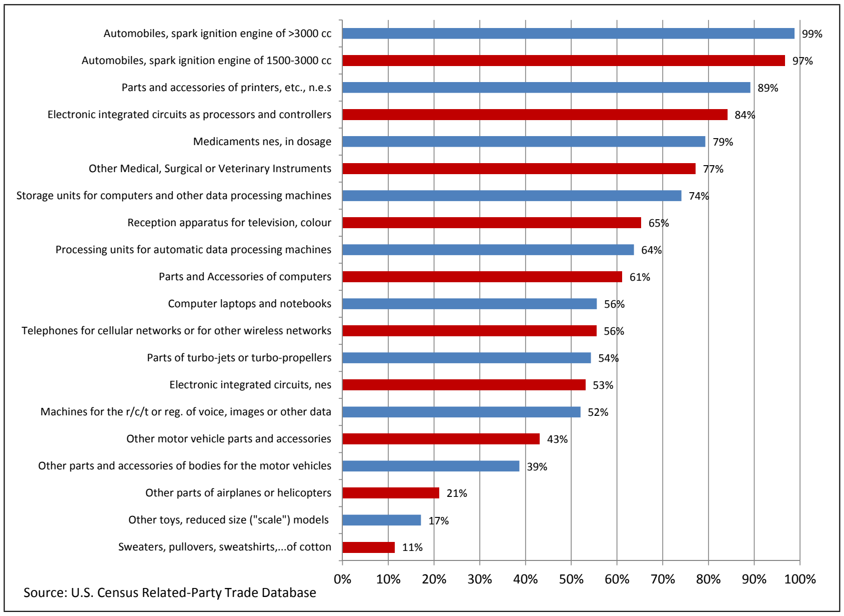

Figure 1.4: Share of U.S. Intrafirm Imports in Largest 20 U.S. Importing Manufacturing Industries in 2011

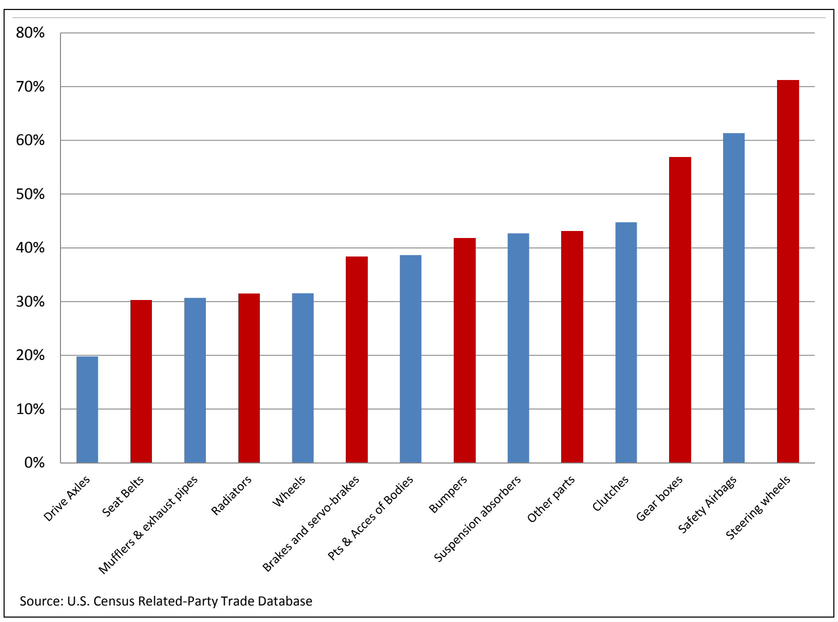

Figure 1.5: Variation in the Share of Intrafirm Trade within HS Sector 8708 (Auto Parts) in 2011

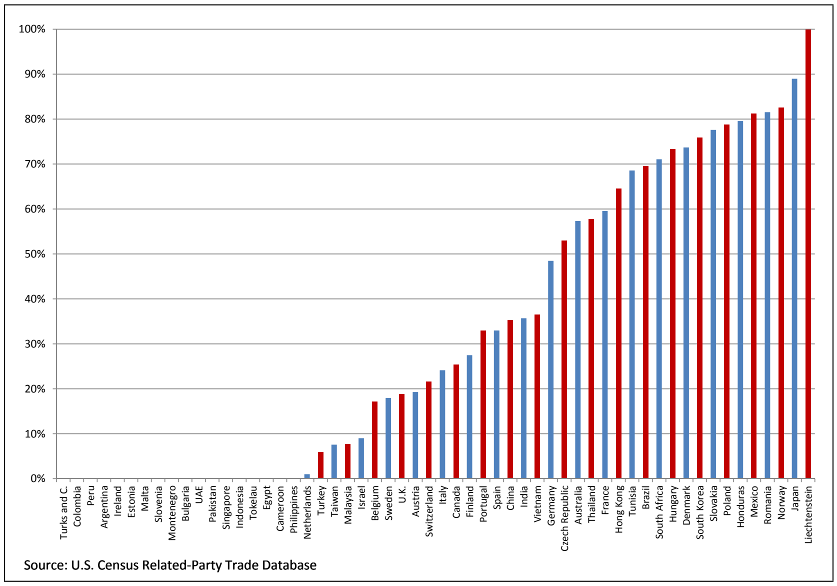

Figure 1.6: Variation in the Share of U.S. Intrafirm Imports within HS 870884 (Steering Wheels) in 2011

largely driven by idiosyncratic factors that cannot possibly be captured by parsimonious models of the organization decisions of firms. If that were the case, however, not only would we observe large variation in the share of intrafirm trade, but we would also expect this variation to be uncorrelated with simple industry or country-level variables. As first demonstrated by Antràs (2003), the evidence suggests otherwise. Chapter 7 will describe in detail several stylized facts regarding the intrafirm component of trade. As a sneak preview, Figures 1.7 and 1.8 illustrate that the share of intrafirm imports in total U.S. imports is significantly higher, the higher the U.S. capital intensity in production of the good being imported and is also significantly higher, the higher the capital-labor ratio of the exporting country. These scatter plots suggest that, as argued above, there may indeed be some common fundamental factors that shape the integration decisions of firms across sectors and countries. The theories of internalization exposited in Chapters 5 and 6 will attempt to shed some light on these factors and will provide a valuable lens through which to study the intrafirm trade data in a more formal and structured manner.

While several features of the U.S. related-party trade database make it

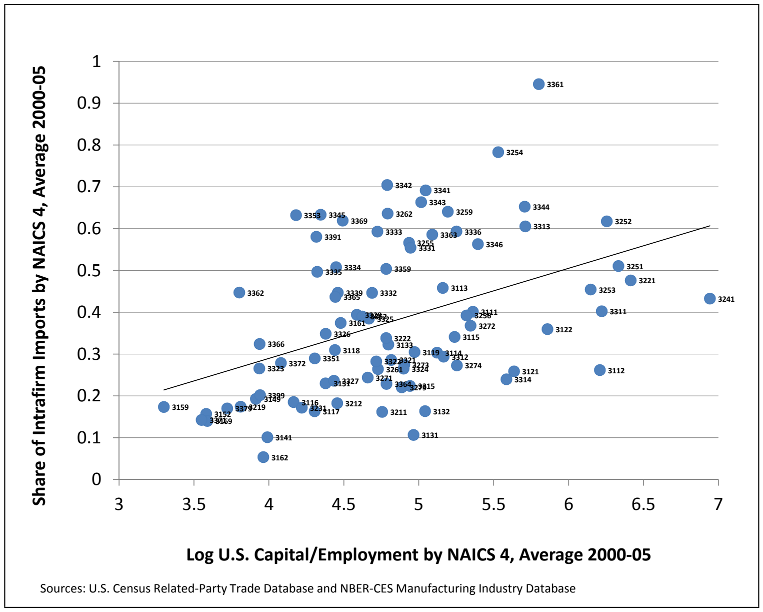

Figure 1.7: The Share of Intrafirm U. S. Imports and Capital Intensity

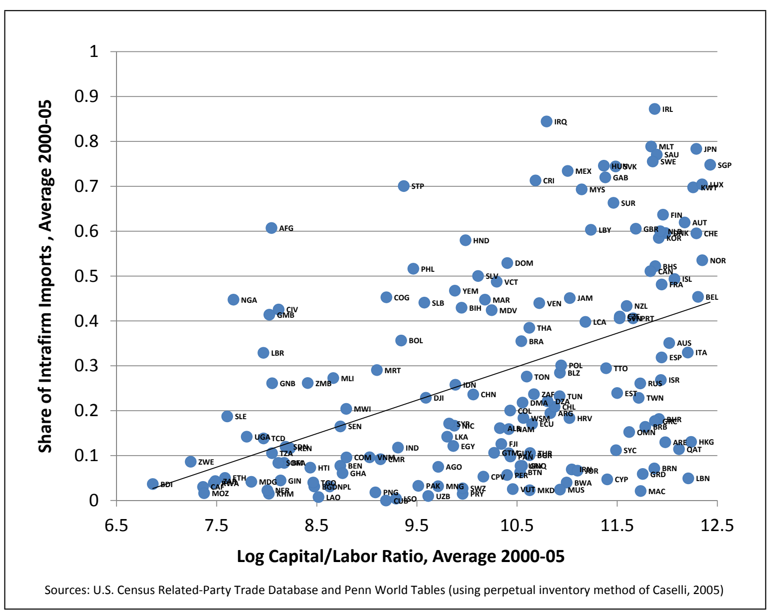

Figure 1.8: The Share of Intrafirm U. S. Imports and Capital Abundance

particularly attractive to empirical researchers, it has some important limitations. Some of the shortcomings of the data relate to the extent to which the characteristics of the data permit a formal test of the theories of internalization developed later in the book, so it is convenient to postpone that discussion until after we have covered those theories in Chapters 5 and 6. Other potential limitations are more fundamental, so it is important to tackle them upfront.

The U.S. database defines 'related-party imports' as import transactions between parties with various types of relationships including "any person directly or indirectly, owning, controlling or holding power to vote, 6 percent of the outstanding voting stock or shares of any organization." A first natural concern is then the 6 percent threshold might be too low for that 'relatedness' to have any significant economic meaning, such as one of the entities having a *controlling* stake in the other entity. In practice, however, extracts from the confidential foreign direct investment dataset collected by the Bureau of Economic Analysis suggest that intrafirm trade is generally associated with one of the entities having a majority-ownership stake in the other entity. More specifically, in 2009, of all U.S. imports associated with U.S. parents purchasing goods from their affiliates, 93.8% of those imports involved majority-owned foreign affiliates, while in the same year, majority-owned U.S. affiliates accounted for 95.5% of U.S. imports by all U.S. affiliates of foreign companies.[20]

A second general concern relates to overall quality of the data. In that respect, the technical documentation that accompanies the dataset stresses that the data are not subject to sampling error, since an indicator of whether the transaction involves or not related parties is required for *all* import or export transactions recorded by the U.S. Bureau of Customs and Border Protection. Despite this requirement, importers and exporters do not always report that information in their shipment documents. Luckily, these transactions are categorized on the data tables as "nonreported," so it is easily verified that these account for a very low share of trade volumes (for instance, just 1.4 percent of total imports in 2011). One might also worry about nonsampling errors related to the imputation of trade values for undocumented shipments and for low-valued transactions (which are sometimes estimated). Nevertheless, quality assurance procedures are performed at every stage of

[20] See Table 9 in http://www.bea.gov/scb/pdf/2011/11%20November/1111_mnc.pdf, and Table I.A.1 in http://www.bea.gov/international/pdf/fdius_2009p/I%20A1%20to%20I%20A9.pdf.

collection, processing, and tabulation, thus there is no reason to believe that these data are any less reliable than U.S. customs data on trade flows.[21]

One way to gain reassurance regarding the usefulness of the data is to see whether it delivers patterns that are consistent with what one would expect based on independent and reliable sources of data. For example, from a quick search of press releases from recent years, one learns that in 2005, Boston-based Gillette Company completed the construction of a 120 million-euro plant in Łódź (Poland), which manufactures disposable razors and other shaving products.[22] Although production was mostly directed to the European market, it seems reasonable to assume that some of the products produced in the plant were shipped back to the U.S., a transaction that would naturally occur within firm boundaries. As shown in Figure 1.9, it is reassuring to observe that the share of intrafirm imports in total U.S. imports from Poland of NAICS code 332211, which is dominated by non-electric razors and razor blades, went up dramatically around the time of the plant opening, jumping from essentially 0 percent in 2004 to close to 100 percent from 2005 onwards.

Back to the Location Decision

We have emphasized above that the internalization decisions of firms in the global economy cannot be understood without appealing to contractual frictions and we have also illustrated the importance of these frictions in the real world. It seems natural, however, to posit that imperfect contracting not only shapes the ownership structure decisions of firms but might also impact their geographical location decisions. As emphasized by neoclassical models of offshoring, profit-maximizing firms will organize production in a cost-minimizing manner, but the effective costs of doing international business are not solely explained by the factors highlighted by neoclassical theory. Certainly, wages will, other things equal, tend to be relatively lower in relatively labor-abundant countries. And, other things equal, costs of production will also tend to be relatively low in countries or regions where the technologies used in production are particularly advanced. Yet, firms might be reluctant to offshore production lines to low-wage countries where suppliers are unreliable and tend not to honor their contracts, and where local courts are unlikely to effectively enforce contracts. Similarly, firms might be

[21] Ruhl (2013) provides a useful overview of alternative U.S. intrafirm trade data sources.

[22] See http://www.paiz.gov.pl/nowosci/?id_news=502.

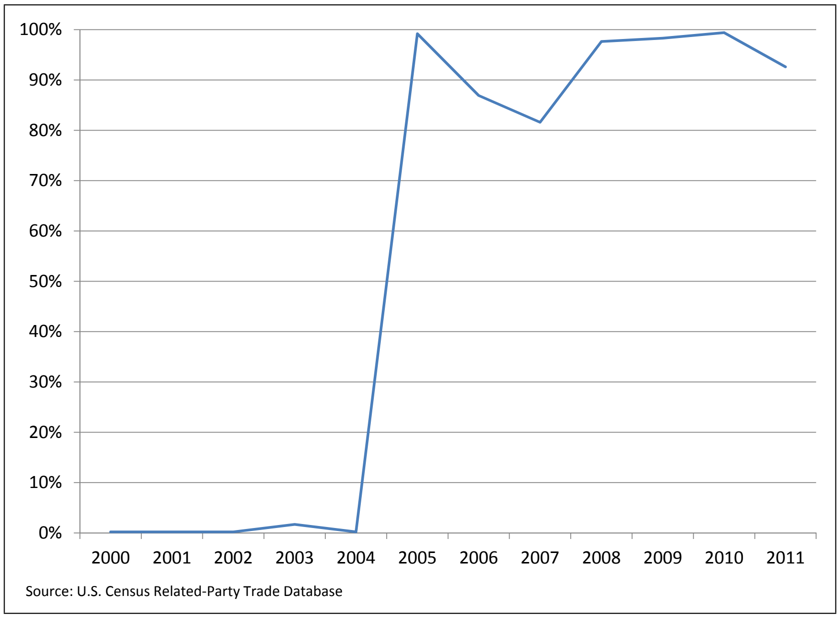

Figure 1.9: Share of Intrafirm Imports of NAICS 332211 (Razors) from Poland

unwilling to operate in countries in which their advanced technologies could be effectively deployed (given the existence of local complementary factors), but in which the contractual environment might not provide enough security to firms, both in terms of quality contracting but also in terms of the risk of intellectual property rights expropriation.

A key factor that makes contractual aspects important for sourcing decisions is the existence of huge variation among countries in judicial quality and contract enforcement. Empirical researchers often resort to easily accessible measures of the quality of the rule of law which are themselves based on weighted averages of various indices of the perceived effectiveness and predictability of courts in different countries. An advantage to these widely used measures, such as the 'Rule of Law' variable produced by the Worldwide Governance Indicators, is that they capture broad features of the contracting environment. A disadvantage is that they are partly based on subjective assessments rather than objective measures of institutional quality. Furthermore, they may provide a useful ordinal measure of legal quality but they are less well equipped to help quantify the existence of cross-country heterogeneity in judicial quality and contract enforcement.

Djankov, La Porta, Lopez-De-Silanes and Shleifer (2003) have proposed an ingenious alternative measure of judicial quality which is narrower in nature but more powerful in illustrating the relevance of differences in the legal system across countries. In particular, Djankov al. (2003) estimate for 109 countries the time it takes a plaintiff using an official court to evict a nonpaying tenant and to collect a bounced check. Figure 1.10 depicts the second of these two variables, which is more likely to be of relevance for firms considering doing business in a particular country. Their estimated total duration of a legal procedure aimed at collecting a bounced check ranges from 7 days in Tunisia to 1003 in Slovenia. Even when focusing on the 43 of the top 50 largest exporters to the U.S. for which they provide data, the estimated duration ranges from 39 days for the Netherlands to 645 days for Italy.

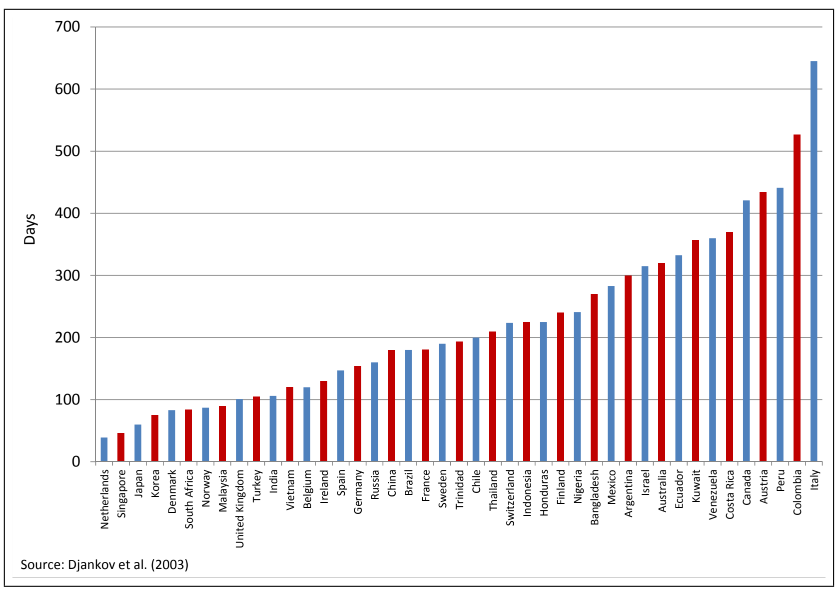

Figure 1.10: Duration of a Legal Procedure Aimed at Collecting a Bounced Check

The extent of contractual insecurity not only varies across countries (or jurisdictions) but it naturally also varies depending on the characteristics of the goods being transacted. For instance, basic goods with low levels of differentiation and which are traded in relatively thick markets can be relatively safely procured even from countries with weak contracting institutions. Conversely, transactions involving highly complex or differentiated goods will tend to be much more 'contract dependent' and one would expect firms to be significantly more sensitive to the institutional environment when choosing the country from which to procure those goods.

A Brief Road Map

This book will study the various ways in which the contracting environment shapes the location and internalization decisions of firms in the global economy. I will focus first on an analysis of the location decision and how it is affected by contracting factors, and only in Part III of the book will I allow firms to optimally decide the extent of control they want to exert over production processes. This does not follow the chronological order in which these topics were developed in the literature, but I will adopt this sequencing for pedagogical reasons.

Before diving into the world of incomplete contracts, it is necessary however to provide an overview of the 'complete-contracting' frameworks that will serve as the basis or skeleton for the models to be developed in future chapters.

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 briefly overview 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 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 sketchy 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 becomes 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 conent 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, 1995, and Davis and Weinstein, 2001).

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 consumer 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 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 tricks 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, het-

[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).

erogeneity in performance was shown to be almost as large within sectors 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 U.S. MNE's foreign affiliates 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

[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.

closely the structure of Krugman (1980). Although, Melitz's framework features 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 $S = 1$. 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 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.[5]

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

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

[5] 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.

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

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

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. Under the mild assumption that any firm with positive production sells some amount of output in their domestic market, we can 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 model in Chaney (2008), with the main difference being that I allow for free entry into each industry, as in Arkolakis, Demidova, Klenow and Rodríguez-Clare (2008) or Helpman, Melitz and Rubinstein (2008).[6] The original model in Melitz (2003) corresponds to the particular case in which $\beta_z = 0$ and $S = 1$, and parameters are fully symmetry 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.[7] As hinted above, the seminal paper of Krugman (1980) 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.[8]

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 result, the potential operating profits for a firm from i with productivity φ

[6] Melitz and Redding (2012) have recently used a model with a very similar structure to navigate the literature on heterogeneous firms and trade.

[7] 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.

[8] 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.

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 under the plausible assumption $f_{ij} > f_{ii}$ for all $j \neq i$, firms will find it harder to profitably sell in foreign markets than in their local market. These intuitive results regarding selection into exporting and productivity differences between exporters and non-exporters are depicted in Figure 2.1.[9]

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.

[9] In some circumstances, marketing costs might be expected to be lower than overhead costs. Note, however, that if a firm from i were to produce only for a particular export market j, then f_{ij} would necessarily have to include overhead costs, and thus we would again have $f_{ij} > f_{ii}$.

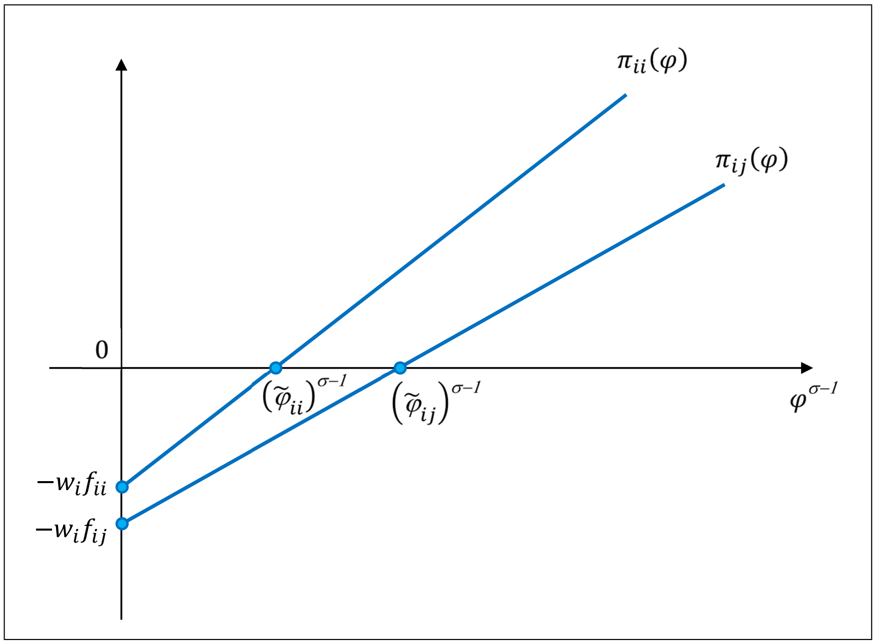

Figure 2.1: Selection into Exporting with Heterogeneous Firms

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 M_i the measure of potential producers from i in that sector (i.e., the set of firms that have paid the fixed cost of entry $w_i f_{ii}$), we have

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

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

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

where M_{ij} 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 Chaney (2008), when productivity is distributed Pareto as in (2.6), integrating (2.12) and using (2.10) to simplify, delivers

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

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 (see below).

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 (2012) (see also Chapter 3 for a related derivation), in the Pareto case, aggregate exports in (2.11) can alternatively be expressed as

$$X_{ij} = \frac{Y_i}{\Theta_i} \left(\frac{\beta E_j}{P_j^{1-\sigma}} \right)^{\frac{\kappa}{(\sigma-1)}} \tau_{ij}^{-\kappa} f_{ij}^{-\frac{\kappa-(\sigma-1)}{\sigma-1}}, \tag{2.13}$$

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.[10] Notice that equation (2.13) 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

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

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 et al. (2012) have also shown that, remarkably, this formula is identical to the one obtained in the Anderson and van Wincoop, Eric (2003), Eaton and Kortum (2002) and Krugman (1980) models.[11]

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. Formally, in terms of the 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 (2012). 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.

---

[11] 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, for more on this).

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 where and how much 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 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

$$C_{ij}(q, \varphi) = f_{ij} w_i + \frac{q}{\varphi} (a_{hi} w_i)^{\eta} (\tau_{ij} a_{mj} w_j)^{1-\eta}. \quad (2.14)$$

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 τ_{ij} with the costs of transporting intermediate inputs across countries, these costs can be interpreted more broadly to reflect other technological barriers associated with international fragmentation, such as communication costs or language barriers. In a similar vein as in Melitz (2003), we will assume that $f_{ij} > f_{ii}$ and $\tau_{ij} > \tau_{ij} = 1$ 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:

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

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 $w^N = 1/a_{zN} > 1/a_{zS} = w^S$.
- 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 Domestic sourcing in the North and Offshoring to South can be written, respectively, as

$$C_D(q,\varphi) = \left(f_D + \frac{q}{\varphi}\right) w_N, \tag{2.15}$$

and

$$C_O(q,\varphi) = f_O w_N + \frac{q}{\varphi}(w_N)^{\eta}(\tau w_S)^{1-\eta}. \tag{2.16}$$

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 (2.15) and (2.16), 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

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

and

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

respectively, where

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

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, $\pi_D(\varphi) > \pi_O(\varphi)$ and offshoring is not a viable option in situations in which domestic sourcing might be. Whenever

$$f_O > \left(\frac{w_N}{\tau w_S}\right)^{(1-\eta)(\sigma-1)} f_D, \tag{2.17}$$

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 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 (2.17) holds. As shown by the Figure, the model features selection into offshoring by which only the most productive firms within an industry, those with productivity $\varphi \geq \tilde{\varphi}_O$, where

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

find it worthwhile to pay the fixed costs of fragmentation to benefit from the lower production costs associated with manufacturing production in the South.[12]

The sorting pattern in Figure 2.2 is consistent with the evidence on selection into importing in Bernard, Jensen, Redding and Schott (2007), who

[12] 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.

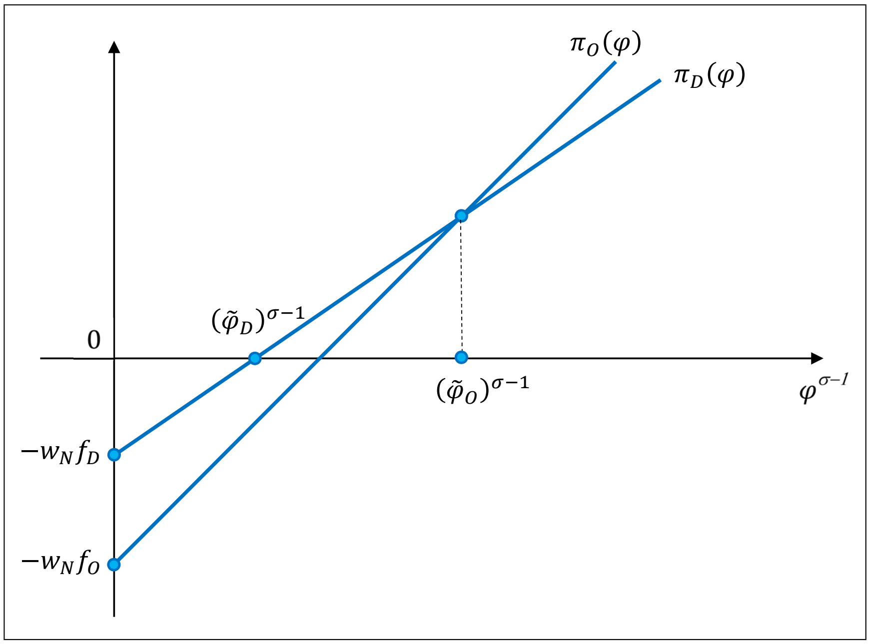

Figure 2.2: Equilibrium Offshoring Sorting with High Wage Differences

show that not only U.S. exporting firms but also U.S. importing firms appear 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. 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.

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.

Figure 2.3 provides further confirmation of this sorting pattern with 2007 data from the Spanish Encuesta sobre Estrategias Empresariales (ESEE). The dataset distinguishes between firms that purchase inputs only from other Spanish producers and firms that purchase inputs from abroad. As is clear from the picture, the distribution of productivity of firms that engage in foreign sourcing is a shift to the right of that of firms that only source locally.[13]

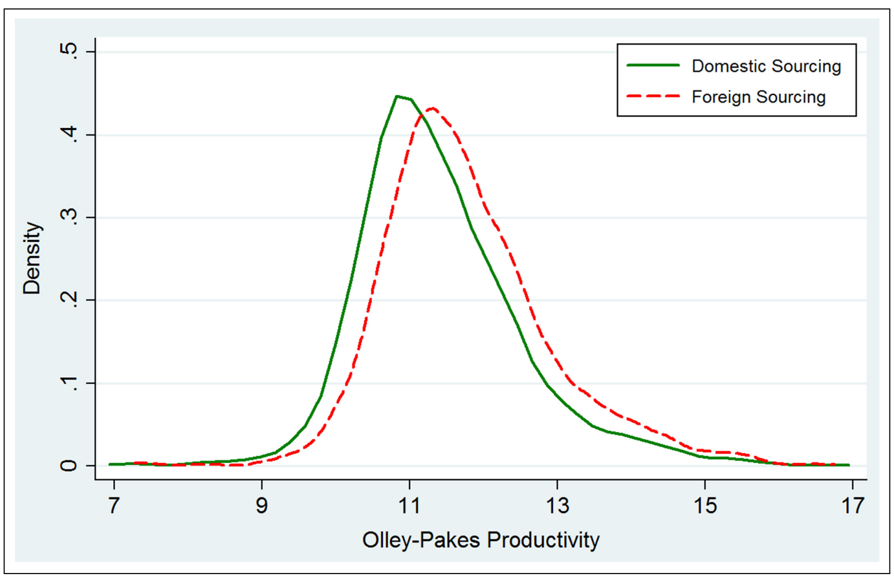

Figure 2.3: Selection into Offshoring in Spain

[13] More details on this dataset and on the measurement of productivity according to the Olley-Pakes method will be provided in Chapter 7.

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 (2.14) and the CES preferences in (2.1), manufacturing input purchases will constitute a share $(\sigma-1)(1-\eta)/\sigma$ of revenue for all firms, while revenue itself will be a multiple σ of firm profits. Using the profit functions (2.15) and (2.16) and cancelling common terms, we can thus express the share of imported manufacturing input purchases in a given industry as

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

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

$$\Upsilon_O = \frac{\left(\frac{w_N}{\tau w_S}\right)^{(1-\eta)(\sigma-1)}}{\left(\frac{\tilde{\varphi}_O}{\tilde{\varphi}_D}\right)^{\kappa-(\sigma-1)} - 1 + \left(\frac{w_N}{\tau w_S}\right)^{(1-\eta)(\sigma-1)}}. \quad (2.18)$$

where

$$\frac{\tilde{\varphi}_O}{\tilde{\varphi}_D} = \left[\frac{f_O/f_D - 1}{\left(\frac{w_N}{\tau w_S}\right)^{(1-\eta)(\sigma-1)} - 1}\right]^{1/(\sigma-1)}. \quad (2.19)$$

As indicated by equation (2.18) and (2.19), the prevalence of offshoring is naturally increasing in the wage gap (w_N/w_S) and decreasing in fragmentation barriers $(f_I - f_D,\ \tau)$. 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.[14]

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 (2.18) and (2.19). 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 (2.14). 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

$$\pi_{ij}(\varphi) = \left((a_{hi}w_i)^{\eta}(\tau_{ij}a_{mj}w_j)^{1-\eta}\right)^{1-\sigma} B\varphi^{\sigma-1} - f_{ij}w_i, \qquad (2.20)$$

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 (2.20) 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.

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

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, 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 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.[15]

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

[15] 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.

(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 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.

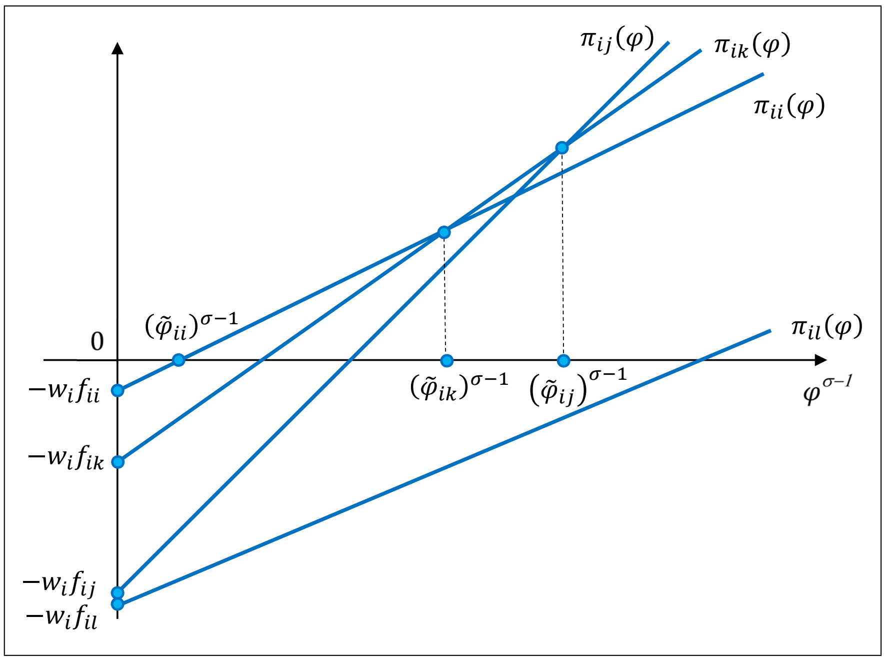

Figure 2.4: Selection into Offshoring with Multiple Countries

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 tricks from a recent paper by Tintelnot (2013), who in turn builds on the seminal work of Eaton and Kortum (2002).[16]

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

$$C_{i\{\ell(v)\}_{v=0}^{1}}(q,\varphi) = w_i \sum_{j\in\mathcal{J}_i(\varphi)} f_{ij} + \frac{q}{\varphi}(a_{hi}w_i)^{\eta}\left(\int_0^1 \left(\tau_{i\ell(v)}a_{m\ell(v)}(v)\,w_{\ell(v)}\right)^{1-\rho}dv\right)^{(1-\eta)/(1-\rho)}, \tag{2.21}$$

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

Following Eaton and Kortum (2002), the key innovation is to treat the (infinite-dimensional) vector of manufacturing productivities $a_{mj}(v)$ as the realization of random variables rather than as being deterministic, as in classical Ricardian models. More formally, by paying the fixed cost f_{ij} of offshoring to country j, a firm headquartered in country i gains the ability of

[16] 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.

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

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

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 the sunk costs of offshoring. Hence, regardless of the different amounts that firms paid to have the ability to source from particular countries, ex-post the choice of location of production of any input v will solve $j^*(v) = \arg\min_{j(v)\in\mathcal{J}_i(\varphi)} \{\tau_{ij} a_{mj}(v) w_j\}$, where remember that $\mathcal{J}_i(\varphi)$ denotes 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

$$\chi_{ij}(\varphi) = \frac{T_j (\tau_{ij} w_j)^{-\theta}}{\sum_{k \in \mathcal{J}_i(\varphi)} T_k (\tau_{ik} w_k)^{-\theta}}. \tag{2.22}$$

With a continuum of inputs, $\chi_{ij}(\varphi)$ also 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 (2.22) 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 appreciates that the extra machinery is starting to pay off. According to expression χ_{ij}, 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. If a location has lower wages w_j, is associated with lower bilateral input trade barriers τ_{ij}, or has better technology T_j it should, other things equal, have a higher market share in the input purchases of firms based in country i for which $j \in \mathcal{J}_i(\varphi)$.

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

$$C_i(q, \varphi, \mathcal{J}_i(\varphi)) = w_i \sum_{j \in \mathcal{J}_i(\varphi)} f_{ij} + \frac{q}{\varphi} (a_{hi} w_i)^{\eta} \left(\gamma \sum_{k \in \mathcal{J}_i(\varphi)} T_k (\tau_{ik} w_k)^{-\theta} \right)^{-(1-\eta)/\theta}, \tag{2.23}$$

where $\gamma = \left[\Gamma\left(\frac{\theta+1-\rho}{\theta}\right)\right]^{\theta/(1-\rho)}$ and Γ is the gamma function.[17] 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 and extra cost draw for all varieties $v \in [0, 1]$. It is thus natural that

[17] 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.

the minimum sourcing cost $\tau_{ij^*}a_{mj^*}(v)\,w_j = \min_{j(v)\in\mathcal{J}_i(\varphi)}\{\tau_{ij}a_{mj}(v)\,w_j\}$ will decrease when a new element is added to the set $\mathcal{J}_i(\varphi)$. In fact, such an addition lowers the expected marginal cost of *all* varieties v, and not just of those that are ultimately sourced from the country being added to $\mathcal{J}_i(\varphi)$.[18]

Following analogous steps as in the previous models to solve for the profit function associated with the cost function in (2.21), 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:

$$\begin{aligned}\pi_{ij}(\varphi) \;=\; \max_{\mathcal{J}_i(\varphi)} \;&\{\,(a_{hi}w_i)^{-\eta(\sigma-1)}\left(\gamma\sum\nolimits_{k\in\mathcal{J}_i(\varphi)}T_k\left(\tau_{ik}w_k\right)^{-\theta}\right)^{(\sigma-1)(1-\eta)/\theta}B\varphi^{\sigma-1}\\ &-w_i\sum\nolimits_{k\in\mathcal{J}_i(\varphi)}f_{ik}\}. \qquad (2.24)\end{aligned}$$

As is clear from equation (2.24), 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 (2.24) is not straightforward to solve since 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, and despite the fact that 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 choice of locations $\mathcal{J}_i(\varphi_0)$ of a firm with productivity φ_0 is 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 low value of $T_l(\tau_{il}w_l)^{-\theta}$, after which the marginal incentive to add further locations might be greatly diminished whenever $(\sigma-1)(1-\eta) < \theta$.[19]

As we show in the Appendix, however, these complications do not arise whenever $(\sigma-1)(1-\eta) \geq \theta$, in which case the addition of a location to the a set of active locations does not decrease the marginal benefit of adding

[18] 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 (2009), Goldberg, Khandelwal, Pavcnik and Topalova (2010), and Gopinath and Neiman (2013).

[19] 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.

further locations.[20] 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.

We can obtain sharper characterizations of the solution to the sourcing strategy problem in (2.24) 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. Fixing the 'Home' country i, we can then rank foreign locations according to their values of $T_j(\tau_{ij} w_j)^{-\theta}$ and denote by r_i the country with the r-th highest value of $T_j(\tau_{ij} w_j)^{-\theta}$. We then have that for any firm with productivity φ from i that offshores to at least one country, $1_i \in \mathcal{J}_i(\varphi)$; for any firm that offshores to at least two countries, we have $2_i \in \mathcal{J}_i(\varphi)$, and so on. In 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

[20] 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.

described in the previous paragraph applies, but when one ranks foreign locations according to the ratio $T_j (\tau_{ij} w_j)^{-\theta} / f_{ij}$.

After having solved the sourcing strategy problem in (2.24), 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

$$M_{ij} = (\sigma - 1)(1 - \eta) \widetilde{B} \int_{\tilde{\varphi}_{ij}}^{\infty} \chi_{ij} (\varphi) \left(\sum\nolimits_{k \in \mathcal{J}_i(\varphi)} T_k (\tau_{ik} w_k)^{-\theta} \right)^{(\sigma-1)(1-\eta)/\theta} \varphi^{\sigma-1} dG (\varphi), \tag{2.25}$$

where $\chi_{ij} (\varphi)$ is given in (2.22), $\widetilde{B} = (a_{hi} w_i)^{-\eta(\sigma-1)} \gamma^{(\sigma-1)(1-\eta)} B$, 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 (2.25) 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 (2.25) 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, (2.25) can be written as

$$M_{ij} = (\sigma - 1)(1 - \eta) \widetilde{B} \Theta_i^{(\sigma-1)(1-\eta)/\theta} \chi_{ij} \int_{\varphi_{\min i}}^{\infty} \varphi^{\sigma-1} dG (\varphi), \tag{2.26}$$

where

$$\Theta_i \equiv \sum_{k=1}^{J} T_k (\tau_{ik} w_k)^{-\theta}$$

and

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

Defining $A_i = \sum_j M_{ij}$ the total absorption of intermediate inputs by firms in i, and by $Q_j = \sum_l M_{lj}$ the total production of intermediates in country j, it is straightforward to verify that (2.26) in fact reduces to

$$M_{ij} = \frac{(\tau_{ij})^{-\theta} \frac{A_i}{\Theta_i}}{\sum_l (\tau_{ij})^{-\theta} \frac{A_l}{\Theta_l}} Q_j,$$

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

Part II

Location

Chapter 3

Contracts and Export Behavior

Noble Group Limited is a global supply chain manager of agricultural and energy products, metals and minerals.[1] In January of 2004, the firm had arranged to export Brazilian soybeans to soybean crushers in China. The contracts signed in January fixed a price for the transaction, even though the delivery was only scheduled to occur in April of that same year. Unfortunately for the buyers in China, prospects for a bumper soybean crop led to a 20 percent decline in soybean prices between the months of January and April. The associated drop in the price of crushed soybeans implied that the Chinese crushers would be operating at substantial losses where they to honor the high prices fixed in their January contracts with Noble. As a result, Chinese buyers began searching for ways to nullify their January contract with Noble. Perhaps not coincidentally, that same month Chinese port authorities discovered a discoloration among a handful of red beans on a 60,000 ton soybean shipment from Brazil, which they claimed indicated the presence of carboxin, a slightly toxic fungicide. Although such discoloration (at least in small quantities) is not unusual in traded soybeans, the Chinese government proceeded to institute a ban on *all* soybean shipments from Brazil, thereby effectively voiding the contract that Noble had signed with the Chinese soybean crushers. As a result, Noble was left with millions of dollars worth of stranded cargo. Noble eventually found other buyers for its shipments, but the incident cost the company around $25 million in demurrage losses.

This unfortunate incident of Noble Group in China exemplifies the con-

[1] The following discussion builds on Foley, Chen, Johnson and Meyer (2009).

tractual insecurity that exporters face in their international transactions, the sources of which were explained earlier in Chapter 1.[2] In the next two chapters, I will begin to discuss simple ways to introduce contractual imperfections in the benchmark models developed in Chapter 2. In this chapter, I will develop simple imperfect-contracting variants of the Melitz (2003) model of exporting. In the next chapter, I will introduce contractual frictions into the model of global sourcing developed in last part of Chapter 2. Along the way, we will also discuss empirical evidence suggestive of the role of these frictions as determinants of the structure of international trade flows and the international organization of production.

Contracting in the Melitz Model

As derived in Chapter 2, in the Melitz (2003) model firms set the volume of output sold and the price charged in each market in a profit-maximizing manner, and as a result, the profits that a firm from country i with productivity φ anticipates obtaining in country j are given by

$$\pi_{ij}(\varphi) = (\tau_{ij} w_i)^{1-\sigma} B_j \varphi^{\sigma-1} - w_i f_{ij}, \tag{3.1}$$

where

$$B_j = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} P_j^{\sigma-1} \beta E_j, \tag{3.2}$$

and E_j is aggregate spending in country j.

It is worth pausing to discuss some key and often overlooked assumptions needed for a firm from i with productivity φ to *actually* realize the profit flow in equation (3.1) when choosing to export in country j. First, it is necessary for the firm to have full information regarding all variables relevant for profits, including its own productivity level φ and the level of (residual) demand implicit in the term B_j. Second, equation (3.1) implicitly assumes that the firm can expand its production in order to meet foreign demand by costlessly hiring additional labor (or the composite factor of production) at a market wage rate w_i which is independent of the firm's operational decisions. Third, the firm is assumed to be be able to costlessly contract with a local

[2] Interestingly, and in line with the internalization response to contractual insecurity highlighted in Chapter 1, in 2005 acquired four soybean processing plants in China.

distributor or importer (an agent, an employee, or a firm) that will collect the sales revenue in country j and will hand them over to the exporter in i.[3]

Some interesting recent work in the field of international trade has been devoted to studying the implications of relaxing the first two assumptions mentioned above. On the one hand, Segura-Cayuela and Vilarrubia (2008), Albornoz, Calvo Pardo, Corcos and Ornelas (2012), and Nguyen (2012) have all fruitfully incorporated foreign demand uncertainty in heterogeneous firm frameworks.[4] On the other hand, a voluminous recent literature, which includes the work among others of Helpman, Itskhoki and Redding (2010) and Amiti and Davis (2012), has studied the implications of imperfect labor markets for the exporting decision, the structure of international trade and the effect of trade liberalization on labor markets, wage inequality and unemployment. As interesting as these contributions are, a treatment of these topics is beyond the scope of the current book. Instead, I will hereafter focus on relaxing the third of the assumptions mentioned above, namely that the contracting between exporters and local distributors or importers is frictionless and allows the exporter to capture the full surplus from the transaction.

Before discussing the implications of contractual imperfections in the Melitz (2003) framework it is necessary to introduce contracting into the framework and this requires us to be a bit more explicit about the agents involved in the model. For simplicity, in this chapter I will restrict attention to situations in which each export transaction involves only two agents, the exporting firm F in country i and the importer M in country j. One can think of the fixed cost of exporting $w_i f_{ij}$ as partly capturing the cost incurred by the exporter in order to be able to contract with importers from j. For the time being, I will also focus on discussing simple contracts taking the following form: at some initial date t_0, the exporting firm F agrees to ship an amount of goods equal to q_{ij}, and in exchange the importer simultaneously agrees to pay the exporter an amount s_{ij} at some later date t_1, perhaps corresponding to the time at which the good is received or perhaps when it has been sold and revenue has been collected. In order to avoid introducing non-essential parameters, I set the discount rate between dates t_0 and t_1 to

[3] Although I will abstract from such a possibility below, one could imagine that the fixed cost of exporting $w_i f_{ij}$ partly reflects the remuneration of the importer for his or her services.

[4] Conversely, models in which firms learn their productivity level φ over time, as in the seminal work of Jovanovic (1982), have not been extensively used in international trade environments.

0. Contracts with alternative timings of payments will be discussed below.

It simplifies the exposition to assume that the opportunity cost of the importer's time is 0, so that the net surplus associated with firm F with productivity φ exporting in country j continues to be given by

$$\pi_{ij}(\varphi) = \left(p_{ij}(\varphi) - \frac{\tau_{ij}}{\varphi} w_i\right) q_{ij}(\varphi) - w_i f_{ij}, \tag{3.3}$$

with $q_{ij}(\varphi) = \beta E_j P_j^{\sigma-1} p_j(\varphi)^{-\sigma}$ as dictated by the demand schedule faced by the exporting firm. In the absence of contractual frictions, the contract will set the quantity of goods $q_{ij}(\varphi)$ shipped to country j and the associated price $p_{ij}(\varphi)$ to maximize the joint surplus in (3.2), thereby leading to the joint profit flow given by

$$\pi_{ij}(\varphi) = (\tau_{ij} w_i)^{1-\sigma} B_j \varphi^{\sigma-1} - w_i f_{ij}, \tag{3.4}$$

which coincides with (3.1). Only when this joint profit flow is expected to be positive, will the exporter decide to invest in being able to export to j.

Even if contracting is frictionless, whether the exporter F is able to realize that entire profit flow in (3.4) will depend on the relative bargaining power of the exporter and the importer. Given the zero reservation value of importers, the equilibrium in the Melitz (2003) framework corresponds to the case in which exporters have all the bargaining power, in the sense that they are assumed to be able to credibly make a take-it-or-leave it offer to importers when contracting with them. To see this more formally, notice that the optimal contract from the point of view of the exporter will solve the exporter's profit subject to the importer's participation constraint, or

$$\begin{aligned} \max_{q_{ij}(\varphi), s_{ij}(\varphi)} \quad & s_{ij}(\varphi) - \frac{\tau_{ij}}{\varphi} w_i q_{ij}(\varphi) - w_i f_{ij} \\ \text{s.t.} \quad & p_{ij}(q_{ij}(\varphi)) q_{ij}(\varphi) - s_{ij}(\varphi) \geq 0, \end{aligned} \tag{3.5}$$

with $p_j(q_{ij}(\varphi)) = \left(\beta E_j P_j^{\sigma-1}\right)^{1/\sigma} q_{ij}(\varphi)^{-1/\sigma}$. Quite naturally, the exporter will find it optimal to make the importer's participation constraint bind, thus implying that $q_{ij}(\varphi)$ will maximize joint profits, and the exporter will end up capturing the profit flow in (3.4), as assumed in the Melitz framework.

The assumption that exporters have all the bargaining power is perhaps a natural one to make given that the model is not explicit about the role of importers in facilitating trade. If these agents have a zero opportunity cost and add no value to exports, why should they be remunerated? In the real world, however, intermediaries serve a central role in linking demand and supply by, among others, alleviating search frictions (see Antràs and Costinot, 2011) and providing quality assurance (see Bardhan, Mookherjee and Tsumagari, 2013 or Tang and Zhang, 2012), and therefore naturally capture a share of the gains from international trade. Although important, a treatment of international trade intermediation is beyond the scope of this book.

Contractual Frictions in the Melitz Model

As simple as the contract discussed above is, our discussion of international contract enforcement in Chapter 1 and the above account of Noble Group's soybean misadventures in China suggests that even those simple contracts are not fully enforceable in the real world. To fix ideas, I will next develop a simple model featuring one such source of contractual insecurity, namely a limited commitment problem on the part of the importer along the lines of the seminal work of Hart and Moore (1994) and Thomas and Worrall (1994).

The lack of commitment on the part of the importer is captured by assuming that at t_1, and before he transfers the collected sale revenue to the exporter, this importer is presented with an opportunity to divert some cash flows away from the exporter. In an extreme case, this might reflect the possibility of the importer absconding with the exporter's goods and attempting to sell them on the side, perhaps at a discount. More generally, the assumption reflects the notion that the initial contract might not compel the parties to honor its terms, thereby tempting the importer to deviate from the contract by underreporting the amount of revenues actually collected, perhaps claiming that those lower revenues were due to the low quality of the goods the exporter shipped. To simplify matters, I will let the share of diverted revenues be a common constant $1 - \mu_{ij} \in [0, 1]$ for all pairs shipping goods from i to j, but I will later briefly discuss the case in which this parameter might vary with productivity.

The parameter μ_{ij} captures the extent to which the importer feels constrained in defaulting on its contractual commitments with exporters from i and thus it is natural to treat this parameter as a measure of the degree of contract enforcement in country j. The fact that the share μ_{ij} also depends on the exporting country i implies that the level of international contract enforcement is allowed to potentially be a function of the nationality of the two agents in the transaction, reflecting perhaps the effects of legal similarity (e.g., common versus civil law countries), a common language, or proximity (cultural or geographical).

I realize that the above modeling of contractual institutions is exceedingly simplistic, with the great complexities and nuances of this type of institutions being reduced to a single parameter μ_{ij} capturing the 'stealing' possibilities of agents residing in the importing country. I will stick to this simple framework for most of this chapter, but let me briefly expand on different mechanisms that might jointly contribute to a country offering a low level of contractual security to firms exporting to it. First, in some institutional environments, agents might face more opportunities to deviate from the initial contract than in other environments. This might be partly due to social norms, but is also explained by the legal environment which might determine how complete formal contracts tend to be. Let us denote by $1 - \rho_{ij}$ the probability with which a 'default' opportunity arises for a $j-$importer transacting with an i-exporter (before we assumed $\rho_{ij} = 0$). When such an opportunity to default does not arise, the importer will necessarily honor the initial contract and deliver all sale revenue to the exporter. When a default opportunity arises, however, the importer will assess the legal ramifications of a contractual breach, and will optimally decide whether to default or not. The legal consequences of a default are in turn shaped by both the probability with which a court of law will rule against a misbehaving importer (denoted by λ_{ij}) and by the amount of damages that it will be required to pay in such an eventuality. It is convenient to model these damages as a multiple d_{ij} of the sale revenues the importer had diverted from the exporter.

Notice that if d_{ij} or λ_{ij} are high enough such that $d_{ij}\lambda_{ij} > 1$, the importer will never default from the exporter and thus the exporter will be able to achieve the same profit flow as in the case with no contractual frictions (see equation (3.4)). Conversely, when $d_{ij}\lambda_{ij} < 1$, if the exporter insisted on demanding the entire sale revenue, the importer would optimally choose to fully default on the exporter because by doing so, it could obtain an expected payoff equal to $\left(1 - \rho_{ij}\right)(1 - \lambda_{ij}d_{ij}) > 0$.

Below, I will focus on the more interesting scenario in which $d_{ij}\lambda_{ij} < 1$. In such a case, the exporter is left, *in expectation*, with a share

$$\mu_{ij} \equiv \rho_{ij} + \left(1 - \rho_{ij}\right)\lambda_{ij}d_{ij} \tag{3.6}$$

of sale revenue. This expression for μ_{ij} summarizes how the prevalence of default opportunities, the competence of courts in ruling against deviating parties, and the size and enforceability of damages jointly shape the perceived contractual security associated with different countries. Equation (3.6) also illustrates how even in situations in which contracts include choice-of-law and forum-of-law clauses (see Chapter 1), thus potentially making ρ_{ij} and λ_{ij} insensitive to j, the importing country institutions may still matter by shaping the extent to which damages set by international courts of law or arbitrators are enforced.

Later in this chapter, I will return to the general formulation of μ_{ij} in (3.6), but for the time being I will focus on the reduced form interpretation of $1-\mu_{ij}$ as capturing the share of sale revenues that importers from j are able to divert from exporters from i.

Implications of Contractual Insecurity

How does the lack of commitment affect contracting between the exporter and the importer? The key new constraint facing the exporter when designing the initial contract, is that any remuneration to the importer lower than $\left(1-\mu_{ij}\right) p_{ij}\left(q_{ij}\left(\varphi\right)\right) q_{ij}\left(\varphi\right)$ would necessarily lead the importer to divert cash flows. As a result, the optimal contracting problem now incorporates a new incentive compatibility (IC) constraint which is necessarily tighter than the participation constraint in the previous optimal contracting program in (3.5).

Formally, and maintaining the assumption that the exporter makes a take-it-or-leave-it offer to the importer, we now have that the date-0 quantity shipped $q_{ij}\left(\varphi\right)$ and the date-1 payment $s_{ij}\left(\varphi\right)$ solve

$$\begin{array}{ll} \max\limits_{q_{ij}(\varphi),s_{ij}(\varphi)} & s_{ij}\left(\varphi\right)-\dfrac{\tau_{ij}}{\varphi}w_i q_{ij}\left(\varphi\right)-w_i f_{ij} \\ \text{s.t.} & p_{ij}\left(q_{ij}\left(\varphi\right)\right) q_{ij}\left(\varphi\right)-s_{ij}\left(\varphi\right)\geq 0 \\ & p_{ij}\left(q_{ij}\left(\varphi\right)\right) q_{ij}\left(\varphi\right)-s_{ij}\left(\varphi\right)\geq \left(1-\mu_{ij}\right) p_{ij}\left(q_{ij}\left(\varphi\right)\right) q_{ij}\left(\varphi\right), \end{array}$$

with $p_j\left(q_{ij}\left(\varphi\right)\right)=\left(\beta E_j P_j^{\sigma-1}\right)^{1/\sigma} q_{ij}\left(\varphi\right)^{-1/\sigma}$. It is straightforward to see that $s_{ij}\left(\varphi\right)$ will now be set to exactly satisfy the (tighter) incentive compatibility constraint, thus implying that the exporter will now only capture a share μ_{ij} of revenues, and will choose $q_{ij}\left(\varphi\right)$ such that

$$\pi_{ij}\left(\varphi\right)=\max_{q_{ij}(\varphi)}\left\{\mu_{ij}\left(\beta E_j P_j^{\sigma-1}\right)^{1/\sigma} q_{ij}\left(\varphi\right)^{(\sigma-1)/\sigma}-\frac{\tau_{ij}}{\varphi}w_i q_{ij}\left(\varphi\right)-w_i f_{ij}\right\}.$$

Solving this problem, the profit function for the exporter can be written as

$$\pi_{ij}(\varphi) = \mu_{ij}^{\sigma}(\tau_{ij}w_i)^{1-\sigma} B_j\varphi^{\sigma-1} - w_i f_{ij}. \tag{3.7}$$

Comparing equations (3.1) and (3.7), it is clear that imperfect contracting reduces the profitability of selling in country j, and the more so the lower is μ_{ij}. The reason for this is twofold: first, the exporter now shares part of the profits obtained in country j with an importer there, and second, the exporter naturally responds to this rent dissipation by reducing the desired amount of goods to ship to country j.

The Margins of Trade, Gravity and Welfare

We now turn to a more formal study of the effects of contract enforcement on the intensive and extensive margins of trade and on aggregate bilateral trade flows across countries. In analogy to the benchmark model with perfect contracting, from equation (3.7) we now have that only firms from i with productivity $\varphi > \tilde{\varphi}_{ij}$, where

$$\tilde{\varphi}_{ij} \equiv \tau_{ij}w_i\left(\frac{w_i f_{ij}}{\mu_{ij}^{\sigma}B_j}\right)^{1/(\sigma-1)}, \tag{3.8}$$

will find it optimal to export to country j. Clearly, for fixed w_i and B_j, the lower is μ_{ij}, the lower will be the measure of firms exporting to country j, and thus the extensive margin of trade is negatively affected by weak contract enforcement. This is illustrated in Figure 3.1 with the shift in the export productivity threshold from $\tilde{\varphi}_{ij}$ to $\tilde{\varphi}_{ij'}$.

Next, aggregating across all firms from i, we find

$$X_{ij} = M_i\int_{\tilde{\varphi}_{ij}}^{\infty}\sigma\mu_{ij}^{\sigma}(\tau_{ij}w_i)^{1-\sigma} B_j\varphi^{\sigma-1}dG_i(\varphi), \tag{3.9}$$

which, as in Chapter 2, we can decompose into an extensive margin M_{ij}, namely the measure of firms from i that export in j, and an average intensive margin $\bar{x}_{ij}$, corresponding to the average export volume across the active exporters:

$$X_{ij} = M_{ij}\cdot\bar{x}_{ij}. \tag{3.10}$$

Whenever firm-level productivity is drawn from a Pareto distribution, we can integrate equation (3.9) and use (3.8) to express $\bar{x}_{ij}$ as

$$\bar{x}_{ij} = \frac{\kappa}{\kappa-\sigma+1}\sigma w_i f_{ij}, \tag{3.11}$$

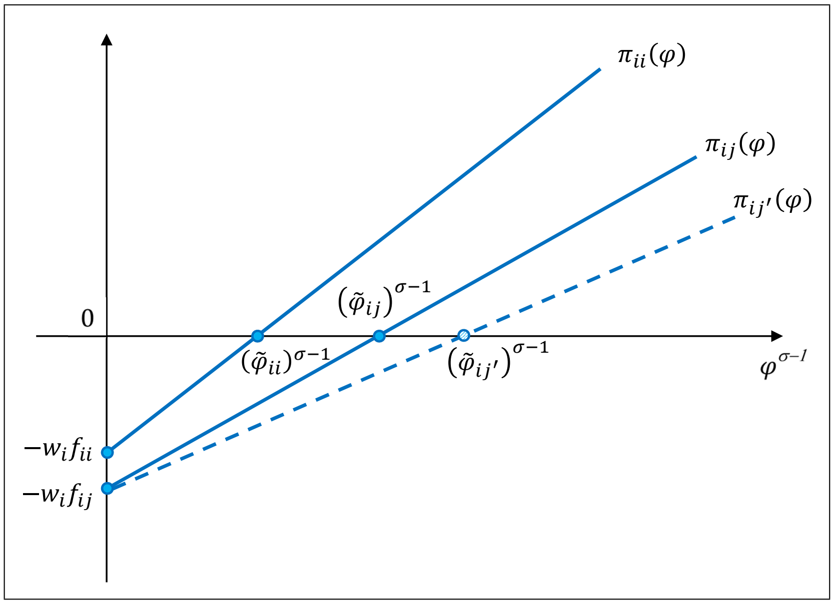

Figure 3.1: Selection into Exporting with Contractual Frictions

just as in Chapter 2. In words, in the special Pareto case, the average intensive margin turns out to be unaffected by the degree of contractual enforcement. It is important to emphasize, however, that this does not imply that the intensive margin of trade *at the firm level* is unaffected by the quality of contracting institutions μ_{ij}. In fact, if a firm were to sell in two markets that differed only in their level μ_{ij}, then the firm would necessarily sell more in the market with better contract enforcement (a higher μ_{ij}), a prediction consistent with the empirical results of Araujo, Mion and Ornelas (2012), who study the effect of the quality of contracting institutions on the cross-section of firm-level export volumes of firms based in Belgium. The insensitivity of $\bar{x}_{ij}$ to μ_{ij} in equation (3.11) is explained by the fact that countries with better contract enforcement attract a disproportionately larger set of relatively small exporters.

We can next follow the steps suggested in Melitz and Redding (2012) to express aggregate sectoral exports from i to j in (3.10) in a slightly more familiar way. In particular, note first that the measure of active exporters in j is given by $M_{ij} = M_i \left(1 - G_i \left(\tilde{\varphi}_{ij}\right)\right)$. Thus plugging (3.11) and the value of the threshold in (3.8) into (3.12), and invoking the Pareto distribution, we have

$$X_{ij} = M_i \left(\frac{\varphi_{\min i}}{\tau_{ij} w_i}\right)^{\kappa} \left(\frac{\mu_{ij}^{\sigma} B_j}{w_i f_{ij}}\right)^{\kappa/(\sigma-1)} \frac{\kappa}{\sigma-\kappa-1} \sigma w_i f_{ij}. \tag{3.12}$$

Next, aggregating over all markets in which firms from i sell (including their domestic market), we can express the aggregate sale revenue obtained by firms in country i as

$$Y_i = \sum_j X_{ij} = M_i \left(\frac{\varphi_{\min i}}{w_i}\right)^{\kappa} \left(\frac{1}{w_i}\right)^{\kappa/(\sigma-1)} \frac{\kappa}{\sigma-\kappa-1} \sigma w_i \Theta_i \tag{3.13}$$

where

$$\Theta_i \equiv \sum_j B_j^{\frac{\kappa}{\sigma-1}} \tau_{ij}^{-\kappa} f_{ij}^{-\frac{\kappa-(\sigma-1)}{(\sigma-1)}} \mu_{ij}^{\sigma\kappa/(\sigma-1)} \tag{3.14}$$

is a measure of country i's market potential (see Redding and Venables, 2004). Plugging equation (3.13) back into (3.12) finally delivers

$$X_{ij} = \frac{Y_i}{\Theta_i} B_j^{\frac{\kappa}{\sigma-1}} \tau_{ij}^{-\kappa} f_{ij}^{-\frac{\kappa-(\sigma-1)}{(\sigma-1)}} \mu_{ij}^{\sigma\kappa/(\sigma-1)}. \tag{3.15}$$

This expression is analogous to equation (2.13) in Chapter 2 except for the last term involving the parameter μ_{ij} (remember that B_j is given by eq. (3.2)).

Equation (3.15) demonstrates that even after introducing contractual frictions, the model continues to deliver a modified sectoral version of the gravity equation for trade flows. This feature can again serve to motivate the widespread use of empirical log-linear specifications of trade flows with exporter and importer fixed effects and measures of bilateral trade frictions. The main new lesson one derives from (3.15) is that such log-linear specifications should include a bilateral measure of the level of contractual security in transactions between producers in country i and country j. We will shortly refer back to equation (3.15) when we review the empirical literature on the effects of institutional quality on trade flows. But before doing so, it is worth addressing one more theoretical matter.

It may appear that the effect of contractual insecurity in the current model is isomorphic to effect of standard variable trade costs τ_{ij} in the benchmark model without contractual frictions. In particular, if one were to define a broad, contract-inclusive measure of trade frictions as

$$\tilde{\tau}_{ij} \equiv \frac{\tau_{ij}}{\mu_{ij}^{\sigma/(\sigma-1)}} > \tau_{ij},$$

it is straightforward to verify that all the equations above exactly correspond to those derived in the benchmark model in Chapter 2 with $\tilde{\tau}_{ij}$ replacing τ_{ij}. One is thus tempted to conclude that, apart from serving to motivate the inclusion of contracting institutions in standard empirical models of export participation, the explicit modeling of contractual frictions has little bearing on the workings of the benchmark model. Such a conclusion is however not warranted because, as mentioned above, contractual frictions not only reduce the profitability of exporting for producers in country i but also transfer exporting surplus to importers from country j. In other words, contractual insecurity not only reduces the overall gains from international trade, but also shapes how those gains are distributed across countries.

This distinction has important bearings for the relationship between contract enforcement and welfare. For instance, in the special case in which there is only one sector in the economy, Demidova and Rodríguez-Clare (2013) have shown that unilateral reductions in variable trade frictions by a small open economy are always welfare enhancing, while the same would not always be true for an increase in μ_{ij} in the framework developed in this chapter. The reason for this is that, although country j would become a more attractive location for foreign exporters if it instituted a higher μ_{ij}, the share of sale proceeds accruing to producers in j would also diminish in that event. For a high enough value of μ_{ij}, the balance of these two effects can be shown to be necessarily negative. This result is analogous to that in Demidova and Rodríguez-Clare (2009) and Felbermayr, Jung and Larch (2013), who show that in a one-sector Melitz (2003) model, each country's unilaterally optimal import tariff is positive. An implication of this result is that countries have a *unilateral* incentive to create some amount of contractual insecurity for producers attempting to sell in their markets. Naturally, however, and as in the case of tariff wars, the unilateral optimality of contractual insecurity is associated with a globally inefficiently low level of contract enforcement.

Preliminary Empirical Evidence

The gravity equation has been one of the most widely-used empirical models of international trade since being introduced by Tinbergen (1962). It is thus not surprising that it has been employed to study the effect of contracting institutions on bilateral international trade flows. The work of Anderson and Marcouiller (2002) is a pioneering study in this literature. Anderson and Marcouiller (2002) start by imposing a model of bilateral trade flows very similar to that in equation (3.15), although not derived from a theoretical model, as we have done above.[5] Following their approach, we next note that if one takes a country, say the United States, as a reference country, one can use equation (3.15), together with the definition of B_j in 3.2, to derive:

$$\frac{X_{ij}}{X_{iUS}} = \left(\frac{P_j^{\sigma-1}E_j}{P_{US}^{\sigma-1}E_{US}}\right)^{\frac{\kappa}{\sigma-1}} \left(\frac{\tau_{ij}}{\tau_{iUS}}\right)^{-\kappa} \left(\frac{f_{ij}}{f_{iUS}}\right)^{-\frac{\kappa-(\sigma-1)}{(\sigma-1)}} \left(\frac{\mu_{ij}}{\mu_{iUS}}\right)^{\sigma\kappa/(\sigma-1)}. \tag{3.16}$$

In words, the ratio of exports of country i to market j relative to the exports of country i to country j is a function of the relative demand or absorption in the two importing countries, as well as different terms capturing the ratio of trade barriers associated with shipping goods from i to j relative to from i to the United States. The main advantage of this approach is that the ratio X_{ij}/X_{iUS} nets out the effect of the exporter country's term Y_i/Θ_i in (3.15) that is common for all destinations j.

Anderson and Marcouiller (2002) estimate a log-linear version of equation (3.16) in which relative bilateral traditional trade barriers (variable and fixed) are proxied by a common border ratio, a common language ratio, a distance ratio and a tariff ratio. The key contract enforcement ratio is proxied by a "composite security" index that corresponds to the average score obtained by each importing country in survey-based measures of transparency and contract enforcement relative to the average score obtained by the United States in those same measures. Note in particular, that Anderson and Marcouiller (2002) assume that the contractual security experienced by exporters in country j relative to that experienced by exporters to the U.S., is common for all exporters, regardless of their country of origin i. This seems to be a restrictive assumption given that one would imagine that differences in legal proximity could make this ratio vary with i, as allowed by (3.15), and so we will revisit this assumption shortly in this chapter. A last important hurdle in estimating equation (3.16) is finding suitable proxies for the first

[5] Anderson and Marcouiller (2005) do study the theoretical links between contractual insecurity and trade flows, but their framework does not predict a gravity equation in trade flows.

term involving the aggregate spending ratio E_j/E_{US} and the price index ratio P_j/P_{US}. The former ratio is proxied with relative measures of GDP, while the latter is approximated with weighted sums of the physical trade cost ratios, in analogy with the "remoteness" variable often present in gravity-style estimations. This is perhaps the least satisfactory element of their empirical design because mismeasurement of these importer-specific terms could lead to important biases in the estimates of the effects of the quality of the importer's contracting institutions on trade flows. We will return to this issue below.

Table 3.1. Importer Contracting Institutions and Relative Exports

	(1)	(2)	(3)
Log GDP ratio	0.855	0.866	0.911
	(0.042)	(0.038)	(0.040)
Relative composite security		0.285	0.279
		(0.073)	(0.081)
Log common border ratio	0.794	0.747	0.665
	(0.155)	(0.163)	(0.186)
Log common language ratio	0.327	0.336	0.358
	(0.080)	(0.082)	(0.109)
Log distance ratio	-1.109	-1.095	-1.133
	(0.058)	(0.056)	(0.056)
Log adjusted tariff ratio	-2.973	-4.814	-4.699
	(1.992)	(2.343)	(2.327)
Number observations	2135	2135	2159
R^2	.69	.70	
Log likelihood			-3865

Robust standard errors with clustering by importer in parentheses. Table reproduced from Table 5 in Anderson and Marcouiller (2002); the regressions also include a log of GDP per capita ratio and remoteness variables for language, border and distance.

With these caveats in mind, the key results of Anderson and Marcouiller (2002), which use 1996 data for 48 importing countries, are reproduced in Table 3.1. Column (1) presents the results of a benchmark gravity equation without institutional variables. As expected, higher relative GDP levels and lower relative traditional trade barriers of any sort are all associated with higher relative export volumes into these countries. When introducing the relative "composite security" index in column (2), this variable has a large and statistically significant effect on relative bilateral trade flows. In column (3), Anderson and Marcouiller confirm the robustness of their results to the use of a Tobit to deal with the large number of zeros in their sample. Although it is not obvious from the nonstandardized point estimates in Table 3.1, the estimates of Anderson and Marcouiller (2002) imply that the effect of weak contracting institutions on trade flows is of a similar order of magnitude as the effect of import tariffs.

As hinted above, two obvious limitations of the Anderson and Marcouiller (2002) study is the insensitivity of their contractual insecurity measure to characteristics of the exporting country and the econometric treatment of the demand terms B_j. These limitations are in fact related to each other. The standard way to address the second concern is to control for these importer-country absorption terms with importer-specific fixed effects. This approach is, however, not feasible when the key explanatory variable of interest only varies across importing countries and thus would be subsumed in the importer fixed effect. This problem clearly applies to Anderson and Marcouiller's (2002) proxy for contractual insecurity. Notice, conversely, that this is not an issue for standard measures of trade barriers, which are defined at the exporter-importer level. A potential way to address both of these limitations is thus to construct a measure of contractual security that is a function of both the exporter and importer country.

What factors might result in particularly high contractual enforcement in transactions between two specific countries i and j? A natural candidate might be a simple measure of whether the exporter and importer country share a common legal origin or not, because that legal relatedness might faciliate the resolution of contractual disputes. A few papers in the literature have explored the role of a common legal origin in affecting bilateral trade flows across countries. Below I focus on the particular contribution of Helpman, Melitz and Rubinstein (2008) because their estimation equation is derived from the Melitz (2003) model in a similar manner than we derived equation (3.15) above, and because their estimation technique allows one to disentangle the effects of particular explanatory variables on both the intensive and extensive margins of trade.

From a theoretical perspective, the only new feature in the framework developed by Helpman, Melitz and Rubinstein (2008) is the introduction of an upper bound in the distribution from which firms draw their productivity level. In specific Pareto case, we now have that

$$G_i(\varphi) = 1 - \left(\frac{\varphi_{\min i}}{\varphi}\right)^{\kappa}, \quad \text{for } \varphi_{\max i} \geq \varphi \geq \varphi_{\min i} > 0. \tag{3.17}$$

The immediate implication of this assumption for our contracting model is that if for some exporting country i, no exporting firm draws a productivity level higher than the threshold $\tilde{\varphi}_{ij}$ defined in (3.8), then bilateral exports from i to j will be zero. This is also easy to see when writing these aggregate bilateral exports as in (3.9), but with the upper limit on the probability distribution:

$$X_{ij} = M_i \int_{\tilde{\varphi}_{ij}}^{\varphi_{\max i}} \sigma \mu_{ij}^{\sigma} (\tau_{ij} w_i)^{1-\sigma} B_j \varphi^{\sigma-1} dG_i(\varphi).$$

Defining aggregate (sectoral) output in i as $Y_i = \sum_j X_{ij}$ and also

$$V_{ij}(\tilde{\varphi}_{ij}) \equiv \int_{\tilde{\varphi}_{ij}}^{\varphi_{\max i}} \varphi^{\sigma-1} dG_i(\varphi),$$

bilateral exports from i to j can be expressed as

$$X_{ij} = \frac{Y_i}{\tilde{\Theta}_i} B_j \tau_{ij}^{1-\sigma} \mu_{ij}^{\sigma} V_{ij}(\tilde{\varphi}_{ij}) \tag{3.18}$$

where

$$\tilde{\Theta}_i = \sum_j \mu_{ij}^{\sigma} (\tau_{ij})^{1-\sigma} B_j V_{ij}(\tilde{\varphi}_{ij}).$$

Equation (3.15) is again a modified version of the gravity equation and it is easily verified that when $\varphi_{\max i} \to \infty$, (3.18) coincides with (3.15) after plugging the value of $\tilde{\varphi}_{ij}$ in (3.8) into $V_{ij}(\tilde{\varphi}_{ij})$.

Equation (3.18) nicely illustrates the existence of an omitted-variable bias in standard gravity-style estimation methods. Even when one partials out the terms $Y_i/\tilde{\Theta}_i$ and B_j with exporter and importer fixed effects, respectively, standard techniques do not take into account the term $V_{ij}(\tilde{\varphi}_{ij})$ capturing the extensive margin of trade from i to j.[6] This omission is likely associated with

[6]The relevance of these biases is clear from the fact that the direct elasticity of trade flows to the index of contractual security is lower in equation (3.18), than in equation (3.18), reflecting the fact that the latter equation does capture the effect of insecurity on the extensive margin of trade.

an upward bias in the elasticity of trade flows to institutional quality because as indicated by the threshold equation (3.8), μ_{ij} has a negative effect on the threshold $\tilde{\varphi}_{ij}$, thus reducing $V_{ij}\left(\tilde{\varphi}_{ij}\right)$.

Helpman et al. (2008) develop a two-step estimation procedure to deal with these biases. In a first stage, a Probit selection equation is derived from the model and the estimates of this equation, together with the model, are used to construct a control variable for the second stage, which is a log-linear model with exporter and importer fixed effects, and various measures of bilateral trade barriers.[7] For the procedure to work, one needs an explanatory variable that enters the first stage (the extensive margin of trade), but not the second stage. Helpman et al. (2008) argue that the cost of creating a business in a particular country satisfies this condition (they also suggest a common religion variable that allows estimation on a larger sample of countries in their sensitivity analysis).

For our purposes, the most relevant feature of Helpman et al.'s (2008) results is that their first and second stages include a variable which is equal to 1 whenever the exporter and importer share a common legal origin as defined by LaPorta, Lopez-de Silanes, Shleifer and Vishny (1999). These authors classify the legal origin of a large cross-section of countries as being either German, Scandinavian, British, French, or Socialist. As argued above, it seems natural that, other things equal, producers located in countries sharing a common legal origin will perceive a higher degree of contractual security when transacting with each other than with producers located in countries with different legal origins. Despite the obvious coarseness of this variable, I will interpret the effect of this variable as reflecting the effect of the contracting institutions term μ_{ij} in specification (3.18).

Table 3.2 reproduces some of the main results in the Helpman et al. (2008) study. The trade data is for 1986 and covers 158 countries. The first column in the table reports the result of the first stage, in which a Probit model is used to predict the probability of positive trade flows from i to j. Not surprisingly, a lower distance and sharing a common language are both positively correlated with the probability that two countries trade with each other.[8] Interestingly, the same is true for the common legal origin variable,

[7] The first stage estimates are also used to include a more standard Heckman-correction term for selection in the second stage.

[8] The original regressions in Helpman et al. (2008) includes six additional controls: whether both countries are islands, whether they are both landlocked, whether they have colonial ties, whether the are members of the same currency union, whether they belong

and the standardized coefficients indicate that the effect of legal institutions is almost half as large as that of a common language. Perhaps surprisingly, zero trade flows appear to be more prevalent for countries that share a land border, perhaps reflecting the incidence of wars. Regulation costs, in turn, appear to have a negative effect on the extensive margin of trade.

Table 3.2: Legal Origin and Bilateral Trade Flows

	(1) Probit	(2) Benchmark	(3) NLS
Distance	-0.213	-1.167	-0.813
	(0.016)	(0.040)	(0.049)
Share a land border	-0.087	0.627	0.871
	(0.072)	(0.165)	(0.170)
Share a common legal origin	0.049	0.535	0.431
	(0.019)	(0.064)	(0.065)
Share a common language	0.101	0.147	-0.030
	(0.021)	(0.075)	(0.087)
Regulation costs ($ amount)	-0.108	-0.146	
	(0.036)	(0.100)	
Regulation costs (days & procedures)	-0.061	-0.216	
	(0.031)	(0.124)	
Firm heterogeneity correction term			0.840
			(0.043)
Sample selection correction			0.240
			(0.099)
Observations	12,198	6,602	6,602
R^2	0.573	0.693	

Robust standard errors clustered by country pair (Bootstrapped for NLS)
Regressions also include exporter and importer fixed effects as well as six other controls (island, landlocked, colonial ties, currency union, FTA, religion).
Marginal effects at sample means and pseudo R^2 reported for Probit.
Table reproduced from Table II in Helpman et al. (2008)

to the same FTA, and a measure of religious proximity. I do not report these coefficients to save space.

Column (2) of Table 3.2 presents the results of a benchmark gravity estimation of the intensive margin of trade that does not correct for the biases identified above. With the exception of the common border variable (which now impacts positively trade flows), the remaining coefficients have the same sign as in the Probit regressions. The effect of a common legal origin on trade flows is positive, large and highly statistically significant. One might worry, however, that the omitted-variable bias discussed above would lead us to overestimate the effects of contractual security on the intensive margin of trade. The results in column (3), which present the second stage in the Helpman et al. (2008) procedure, confirm that such biases exist but the coefficient on legal origins is only reduced by about 20% and remains large and highly significant. We can conclude from these results that, consistently with the simple model we have developed above, contractual insecurity has a significant negative effect on bilateral trade flows and that such effect operates through both an extensive margin as well as an intensive margin.

Responses to Contractual Insecurity

Our theoretical and empirical results so far illustrate that exporters will respond to the perceived contractual insecurity associated with servicing certain foreign markets by reducing their sales or by simply opting out from selling in those markets. As explained in Chapter 1, in practice firms can resort to alternative means to alleviate such contractual insecurity. We will next discuss three of these mechanisms: investing in contract enforcement, repeated interactions with importers, and demanding prepayment from importers.

Consider first the possibility of firms investing in enhancing the contractibility of their transactions. This might involve hiring legal counsel to design the initial contract in a way that makes it more likely to be enforced, or it might be associated with resorting to international arbitration, which would typically also provide the exporter with more contractual security. Without delving into the details of these different legal mechanisms, let us suppose that if a firm from i were to invest a fixed amount $w_i f_c$ of resources in improving contractibility, the share of revenues that an importer from j would be able to divert would be reduced from $1 - \mu_{ij}$ to $1 - \bar{\mu}_{ij}$ with $\bar{\mu}_{ij} > \mu_{ij}$.

Following analogous derivations as those in the previous model, it is straightforward to verify that firms from i will optimally choose to invest in contractibility whenever

$$\bar{\mu}_{ij}^{\sigma}\left(\tau_{ij}w_i\right)^{1-\sigma}B_j\varphi^{\sigma-1}-w_i f_c>\mu_{ij}^{\sigma}\left(\tau_{ij}w_i\right)^{1-\sigma}B_j\varphi^{\sigma-1},$$

which can alternatively be expressed as

$$\varphi^{\sigma-1}>\left(\varphi_{ij^c}\right)^{\sigma-1}\equiv\frac{w_i f_c}{\left(\bar{\mu}_{ij}^{\sigma}-\mu_{ij}^{\sigma}\right)\left(\tau_{ij}w_i\right)^{1-\sigma}B_j}.$$

In words, only the largest, most productive exporters will find it optimal to incur additional legal expenses to reduce their contractual insecurity. This might explain, for instance, why arbitration cases at the International Chamber of Commerce rarely involve disputes over amounts lower than under one million U.S. dollars (see footnote 13 in Chapter 1). The selection of exporters into enhanced contractibility is depicted in Figure 3.2. The figure also shows that the endogenously higher contractibility of large exporters will tend to lead to a more skewed distribution of exports than in the version of the model in which the parameter μ_{ij} is common for all firms within an industry.

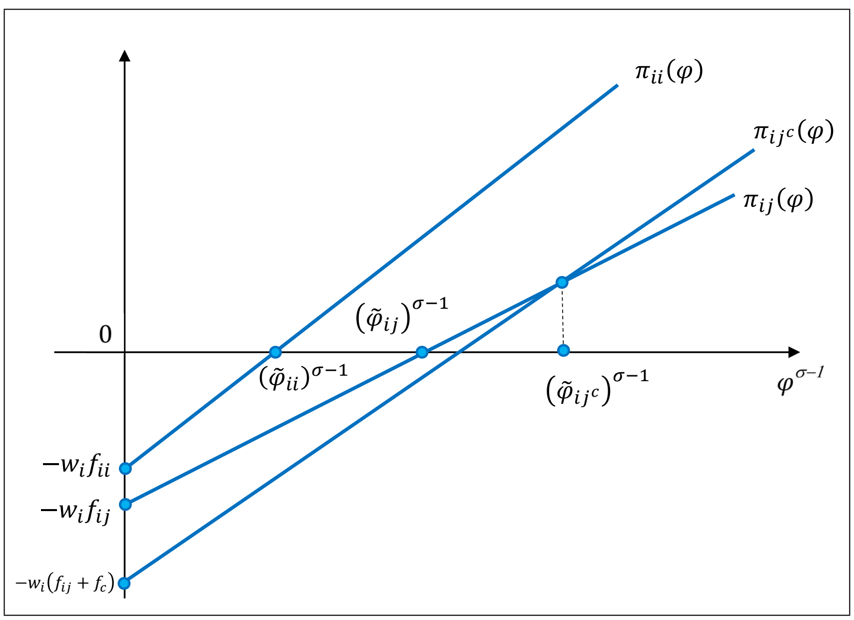

Figure 3.2: Selection into Exporting with Endogenous Contractibility

In the previous setup in which the exporter and the importer transact only once, it is optimal for importers to divert revenue from the exporter if the contracts is not perceived to be enforced. I next briefly explore how the incentives of importers to misbehave might be affected by repeated interactions with a given exporter, and how this affects the dynamics of exporting volumes. In doing so, I build on the work of Araujo, Mion and Ornelas (2012) and Antràs and Foley (2013). To emphasize the differences with the static model, let us assume that the exporter and importer perceive their business relationship to be infinitely repeated. Assume also that importers come in two types: some of them are very patient and discount the future at a very low rate, while the rest are myopic and care only about current payoffs. At the beginning of each period t, the exporter and importer sign an agreement that binds the exporter to ship an amount $q_{ijt}(\varphi)$ of output to the importer in j in exchange for a payment from the importer once the goods have been sold.

With probability $1-\rho_{ij}$, however, the importer is presented with an opportunity to divert all sale revenue and pay nothing to the exporter at the end of the period. We thus adopt here the probabilistic version of contract enforcement discussed earlier in the chapter and that led to expression (3.6) for μ_{ij}, but for simplicity I set $d_{ij}\lambda_{ij}=0$, so that the exporter is left with a zero payoff in case of default. Provided that the discount rate of patient importers is sufficiently low, the folk theorem implies that a trigger-strategy equilibrium exists in which patient importers never choose to default, while myopic importers always do so when an opportunity arises.[9]

To generate nontrivial dynamics, we assume that whether an agent is patient or myopic is private information to that agent. The exporter forms beliefs on the type of the particular importer they are dealing with based on the bilateral transaction history with that importer and on a public signal reflecting the prevalence of patient importers in the population. I will denote this public signal by ξ_0 and I will treat it as an exogenous parameter, although in Araujo, Mion and Ornelas (2012) it is endogenized by specifying a process of matching between exporters and importers.

Notice that a new importer will initially be assigned a probability ξ_0 of being patient, which is associated with a probability $\xi_0+(1-\xi_0)\rho_{ij}$ of the contract being enforced in this initial period. With a history of no defaults, the exporter's belief on the importer's type will improve over time, while an

[9] Of course, this requires that the importer obtains some positive payoff when it chooses to honor the contract. Still, for a discount factor close enough to 1, this required payoff can be made arbitrarily close to 0. This limiting case is considered here for simplicity.

incidence of a default will immediately reveal the importer to be a myopic type. Denoting by ξ_t the particular posterior probability assigned to the importer being patient, repeated application of Bayes' rule delivers

$$\xi_t = \frac{\xi_0}{\xi_0 + (1 - \xi_0)\left(\rho_{ij}\right)^t} \tag{3.19}$$

when there have been no defaults up to length T, and $\xi_t = 0$ otherwise. As a result, the perceived probability of the contract signed at t being enforced is given by $\xi_t + (1 - \xi_t)\,\rho_{ij}$, and naturally rises with a history of no defaults. Having determined this time-varying level of contractual insecurity, the rest of the equilibrium is analogous to that of the static model with the share of revenue accruing to exporters given by

$$\mu_{ij}(t) = \xi_t + (1 - \xi_t)\,\rho_{ij}.$$

Hence, all firm-level equilibrium expressions continue to hold with $\mu_{ij}(t)$ replacing μ_{ij} throughout.

This extension of the model delivers several empirical predictions for the effects of weak contracting institutions on firm-level exports. As in the static model developed above, the extensive margin of trade continues to be negatively impacted by low institutional quality (low ρ_{ij}). This is both because firm are less inclined to begin selling in weak institutions countries, but also because the probability of an export relationship being discontinued is higher the lower is the probability of contracts being enforced. The effects of low formal contract enforcement on the intensive margin of trade are richer. The perceived initial probability of default is given by $(1 - \xi_0)\left(1 - \rho_{ij}\right)$ and thus export relationships in weak contracting environments (countries with low ρ_{ij}) will tend to begin at low volumes. Nevertheless, the negative effect of weak contracting on the intensive margin of trade is predicted to be attenuated over time, resulting in firm-level export volumes that should rise over time.

Araujo, Mion and Ornelas (2012) study the effects of importer-country characteristics on the cross-section of firm-level exports of Belgian firms over the period 1995-2008 and find broad support for these predictions. Other things equal, export entry is higher and export exit is lower in countries with better contracting institutions. Initial firm-level export volumes are increasing in contract enforcement, while firm-level export growth is on average positive. Interestingly, however, this positive growth in exports appears to be lower in countries with weak contracting institutions. The simple model developed above provides a simple rationale for this fact: in countries with high default rates contractual insecurity will be lower, but the exporter will be able to learn the type of the importer at a faster rate than in an environment in which a very low default rate prevents myopic importers from defaulting. Formally, differentiation of (3.19) indicates that for a low enough t, the growth in ξ_t over time t is necessarily decreasing in ρ_{ij}.[10]

Choice of Payment Method: Exporter Institutions Matter

So far we have discussed the role of investments in contractibility and repeated interactions in reducing the extent of contractual insecurity faced by exporters. If the *only* contractual friction in international transactions was the risk of importer default, then a simple solution to this problem would be for the exporter to demand pre-payment from the importer before shipping the goods. Formally, a simple modification of the static contract we have considered so far would suffice to resolve the inefficiencies associated with contractual insecurity: instead of the payment s_{ij} occurring at t_1, the exporter could insist that it was made at t_0. With this simple modification, the exporter would not need to worry about the payment s_{ij} satisfying an incentive compatibility constraint for the importer, and could thus choose s_{ij} to satisfy exactly the importer's participation constraint. The problem would thus reduce to that in (3.5), which we have shown above delivers payoffs identical to those in the Melitz (2003) model without contractual frictions.

Although 'cash-in-advance' transactions are not infrequent in international trade (see Antràs and Foley, 2013, for evidence from a U.S.-based exporter), the available evidence suggests that the majority of international transactions are conducted on open account (or post-shipment payment) terms.[11] These type of transactions roughly correspond to the timing of payments we have assumed so far. A natural question is then: why are cash-in-advance terms not used more often if they effectively eliminate the risk of importer default?

[10] Araujo, Mion and Ornelas (2012) actually show that endogenizing the prior ξ_0, the differentially lower growth of exports in high contract enforcement countries holds true for all values of t.

[11] For instance, using the World Bank Enterprise Survey database, Hoefele, Schmidt-Eisenlohr and Yu (2013) find that the average share of sales on open account terms for the firms in the sample is in excess of 80%.

The key for answering this question is that not only exporters but also importers are exposed to the risk of counterparty misbehavior in international transactions. In particular, a standard concern for importers in cash-in-advance transactions is that, after being paid, the exporter might no longer have the incentive to ship goods in the most advantageous manner for importers, thus intentionally or unintentionally reducing the amount of sale revenues that the importer would obtain when selling the goods in their local market. I next briefly develop a simple model of exporter misbehavior along the lines of the model of limited commitment by importers developed above.[12] The model will serve to illustrate the role of exporter-country institutions in shaping the different margins of international trade.

Suppose that the exporter and importer sign the following simple cash-in-advance contract. At t_0, the exporting firm F agrees to ship an amount of goods equal to q_{ij} in exchange for an amount s_{ij} to be paid upon signing the contract at t_0. After receiving the goods, the importer sells them in her local market and she keeps the collected sale revenue. As argued above, without any type of frictions, the exporter could set an initial payment $s_{ij}(\varphi)$ equal to the sale revenue collected by the importer at time t_1, i.e., $p_{ij}(q_{ij}(\varphi))\,q_{ij}(\varphi)$, thus attaining the frictionless profit flow in (3.4).

Imagine, however, that shortly after signing the contract at t_0, the exporter is presented with an opportunity to deviate from the initial contract in a way that would reduce its costs of production but would also reduce the expected revenues collected by the importer at t_1. Such a deviation might entail shirking in quality-enhancing investments or in the use of shipping methods that best ensure the quality of goods when they reach the importer's market. For the time being, consider the case in which exporter misbehavior takes the extreme form of the exporter incurring no variable production costs and the importer not receiving goods or receiving totally worthless goods. Below, I will consider much less extreme cases. Faced with this opportunity to misbehave, the exporter will consider the legal implications of such a deviation before cheating on the importer. Suppose that when deviating, the importer could sue the exporter in a latter's court of law and win the case with probability λ_{ij}^{exp}, in which case the exporter would be asked to pay the importer an amount in damages equal to a multiple d_{ij}^{exp} of the payment stipulated in

[12] Financial constraints faced by the importer might be another factor limiting the use of cash-in-advance contracts. Although not the focus of her study, Manova (2012) has found that bilateral trade flows are depressed by low quality of financial institutions in importing countries.

the initial contract, i.e., s_{ij}. When $\lambda_{ij}^{exp} d_{ij}^{exp} > 1,$ the exporter would never be tempted to cheat on the importer, so we will focus below on the more interesting case in which $\lambda_{ij}^{exp} d_{ij}^{exp} < 1$. In the latter case, in order for the exporter not be tempted to misbehave, the payment stipulated in the initial contract needs to satisfy the following incentive compatibility (IC) constraint

$$s_{ij}(\varphi) - \frac{\tau_{ij}}{\varphi} w_i q_{ij}(\varphi) \geq \left(1 - \lambda_{ij}^{exp} d_{ij}^{exp}\right) s_{ij}(\varphi). \tag{3.20}$$

Note that it is in the own interest of the exporter to ensure that the initial contract satisfies this IC constraint because otherwise the importer would anticipate misbehavior with probability one, and he or she would not be willing to pay *any* amount of money to the exporter in the initial period, thus leaving both agents with a zero payoff.

Using equation (3.4) and the fact that revenues are a multiple σ of profits, it is straightforward to verify that provided that $\lambda_{ij}^{exp} d_{ij}^{exp} \geq (\sigma - 1)/\sigma$, the constraint in (3.20) will be slack when evaluated at the unconstrained profit-maximizing output level $q_{ij}(\varphi)$, and thus the exporter will still be able to achieve the unconstrained profit flow in (3.4). Conversely, when courts punish deviating agents with a low enough probability or when damages are low enough or unenforceable, so that $\lambda_{ij}^{exp} d_{ij}^{exp} < (\sigma - 1)/\sigma$, the exporter will no longer be able to achieve the unconstrained profit flow in (3.4). Instead, the quantity of output being shipped will need to adjust to ensure that equation (3.20) holds, and the exporter will find it optimal to make that constraint exactly bind. Furthermore, one can show that the exporter will never find it optimal to demand an ex-ante payment lower than the total sale revenues collected by the importer at t_1, so $s_{ij}(\varphi) = p_j(q_{ij}(\varphi)) q_{ij}(\varphi)$, and from equations (3.20) we can infer that

$$p_j(q_{ij}(\varphi)) = \frac{\tau_{ij} w_i}{\lambda_{ij}^{exp} d_{ij}^{exp} \varphi}. \tag{3.21}$$

Using (3.21) together with $q_{ij}(\varphi) = \beta E_j P_j^{\sigma-1} p_j^{-\sigma}$ and the definition of B_j in (3.2), we can then express the profits of the exporter as

$$\pi_{ij}(\varphi) = \mu_{ij}^{exp} (\tau_{ij} w_i)^{1-\sigma} B_j \varphi^{\sigma-1} - w_i f_{ij}, \tag{3.22}$$

where

$$\mu_{ij}^{exp} = \begin{cases} 1 & \text{if } \lambda_{ij}^{exp} d_{ij}^{exp} \geq \frac{(\sigma-1)}{\sigma} \\ \sigma\left(1 - \lambda_{ij}^{exp} d_{ij}^{exp}\right)\left(\frac{\sigma \lambda_{ij}^{exp} d_{ij}^{exp}}{\sigma - 1}\right)^{\sigma-1} < 1 & \text{if } \lambda_{ij}^{exp} d_{ij}^{exp} < \frac{(\sigma-1)}{\sigma}. \end{cases} \tag{3.23}$$

Note that μ_{ij}^{exp} is (weakly) increasing in $\lambda_{ij}^{exp} d_{ij}^{exp}$, with $\mu_{ij}^{exp} = 1$ only when $\lambda_{ij}^{exp} d_{ij}^{exp} \geq (\sigma - 1)/\sigma$. Hence, in the range of parameter values in which the exporter is tempted to misbehave, the profits the exporter will end up obtaining will necessarily be lower than in the unconstrained problem.

Equation (3.22) illustrates that limited commitment problems on the part of the exporter end up affecting the profitability of exporting in a similar manner than limited commitment problems on the importer side. In fact, equation (3.21) is identical to (3.7) except for the term μ_{ij}^{exp} in (3.21) versus μ_{ij}^{σ} in (3.7). An important difference between to equations exists, however. The subscript *exp* in the contracting term in (3.21) serves to emphasize that while in our previous model the quality of the importing country's contracting institutions was key in shaping the profitability of exporting, the extensive and extensive margin of trade, and bilateral trade flows across countries, the same role is now played by the quality of the exporting country's contracting institutions. Of course, one could argue, as we did in the importer limited commitment case, that agents could resort to choice-of-law or choice-of-forum contractual clauses to partly isolate the security of a transaction from weak contracting institutions in the exporting country. Nevertheless, and as explained in Chapter 1, even when disputes are adjudicated by foreign courts, the enforceability of damages is ultimately an issue related to the local legal environment in the exporting country, and particularly whether that country has signed the New York convention.

I will next overview the empirical work linking bilateral trade flows to the quality of the exporter country's contracting institutions, but before doing so I will briefly address two further theoretical points. First, and although it is obvious to see that all equilibrium conditions with exporter limited commitment will be identical to those in (3.7) through (68) with μ_{ij}^{exp} replacing μ_{ij}^{σ}, there is one important, subtle difference in the general equilibrium implications of the two models. Because in this second model, importers always end up with a net payoff of zero, the effects of a low μ_{ij}^{exp} are not isomorphic to an import tariff in the importing country, but instead are analogous to those of an iceberg trade cost. An implication of this difference is that, at least in the one-sector version of the model, improvements in the quality of contracting institutions in a small exporting country will always be beneficial for the importing country (see Demidova and Rodríguez-Clare (2013)). A second point worth making is that our model of exporter misbehavior can easily be extended to the case in which the exporter's temptation to deviate from the contract entails reducing marginal costs by a certain fraction ν_{ij}^{exp} where we now allow $\nu_{ij}^{exp} < 1$. In such a case, the exporter's incentive compatibility constraint becomes

$$\lambda_{ij}^{exp} d_{ij}^{exp} s_{ij}(\varphi) \geq \nu_{ij}^{exp} \frac{\tau_{ij}}{\varphi} w_i q_{ij}(\varphi),$$

and the same expressions (3.21) through (3.23) apply, but with $\lambda_{ij}^{exp} d_{ij}^{exp} / \nu_{ij}^{exp}$ replacing $\lambda_{ij}^{exp} d_{ij}^{exp}$ throughout. Clearly, exporter profits will be higher in that case, but as long as $\lambda_{ij}^{exp} d_{ij}^{exp} / \nu_{ij}^{exp} < (\sigma - 1)/\sigma$, contractual frictions continue to reduce the profitability of exporting.

Exporter-Country Institutions: Empirical Evidence

Earlier in the chapter we discussed the empirical work of Anderson and Marcouiller (2002) establishing a link between bilateral trade flows and the quality of the importer's contracting institutions. It is clear that the empirical strategy in that paper is not applicable to the study of the effects of exporter-country institutions since the relative exports specification in (3.16) effectively partials out exporter-specific variables. Berkowitz, Moenius and Pistor (2006) propose instead a more traditional log-linear gravity specification which can be motivated by a simple variant of equation (3.15) with μ_{ij}^{exp} replacing μ_{ij}^{σ}:

$$X_{ij} = \frac{Y_i}{\Theta_i} B_j^{\frac{\kappa}{\sigma-1}} \tau_{ij}^{-\kappa} f_{ij}^{-\frac{\kappa-(\sigma-1)}{(\sigma-1)}} \left(\mu_{ij}^{exp}\right)^{\kappa/(\sigma-1)}.$$

In order to control for the unobserved multilateral resistance term Θ_i and the price index implicit in B_j, Berkowitz, Moenius and Pistor (2006) introduce exporter and importer fixed effects. The author's measure of the quality of contracting institutions is an average of a country's index of rule of law, expropriation risk, corruption in government, and bureaucratic quality as computed by International Country Risk Guide. This variable is computed for the exporter and the importer in each pair of trading partners and both variables are introduced in the regression, thus allowing for both importer and exporter country institutions to affect bilateral trade flows. In order for the country fixed effects not to absorb these institutional variables, Berkowitz, Moenius and Pistor (2006) use data from 1982 to 1992 and exploit time-series variation in both bilateral trade flows and the perceived quality of contracting institutions. Their specifications also include time fixed effects and controls for GDP, GDP per capita, and various measures of proximity between the exporter and the importer, including a measure of the remoteness related to whether a pair of countries are close to each other but distant from the rest of the world.

Table 3.3: Exporter and Importer Contracting Institutions and Bilateral Exports

	(1)	(2)	(3)	(4)
Type of Goods Included	Overall	Overall	Complex	Simple
GDP importer	0.81	-0.15	0.08	-1.06
	(38.53)	(-0.52)	(0.27)	(-2.52)
GDP exporter	0.76	-0.19	0.32	-1.38
	(39.13)	(-0.65)	(1.08)	(-3.26)
Distance	-1.16	-1.03	-0.98	-1.26
	(-27.97)	(-27.11)	(-24.98)	(-22.72)
Adjacent	0.35	0.40	0.44	0.27
	(2.43)	(2.65)	(2.62)	(1.54)
Links	0.42	0.45	0.54	0.18
	(4.07)	(4.40)	(5.09)	(1.22)
Language similarities	0.09	1.00	1.28	0.11
	(0.51)	(5.74)	(6.77)	(0.40)
Remoteness	0.58	1.79	0.74	6.69
	(6.04)	(2.31)	(0.96)	(5.50)
Quality of importer legal institutions	0.61	0.05	-0.44	0.66
	(5.41)	(0.51)	(-4.24)	(4.42)
Quality of exporter legal institutions	0.91	0.36	0.93	-0.53
	(7.12)	(3.26)	(8.41)	(-3.45)
Country dummies		Yes	Yes	Yes
Time dummies		Yes	Yes	Yes
Number of clusters (country pairs)	2792	2792	2755	2550
R^2	0.70	0.77	0.79	0.38
Number observations	23,564	23,564	22,669	18,948

T-statistics in parentheses (computed w/ robust standard errors, within-group clustering)
Regressions also include exporter and importer GDP per capita and a constant
Table reproduced from Table 2 and 3 in Berkowitz et al. (2006)

The first two columns of Table 3.3 reproduce the results obtained by Berkowitz et al. when running their specification with and without the country and year fixed effects, respectively. As is clear from column (1), when ignoring these fixed effects, all variables affect bilateral trade flows in the expected way and the institutional quality variables related to both the exporter and the importer are highly statistically significant, with exporter institutions appearing to matter more than importer institutions. When introducing the exporter, importer and year fixed effects, a first noteworthy fact is that the effect of GDP on bilateral trade flows vanishes. This is not entirely surprising since the fixed effects were supposed to control for terms in the gravity equation involving GDP. More relevant for the current discussion is the fact that the variable capturing the quality of contracting institutions in the exporting country remains positive and highly statistically, while the importer country institutional quality variable remains positive but loses its significance.

Rather than interpreting their results as suggesting the irrelevance of importer contracting institutions, Berkowitz et al. (2006) acknowledge the potential existence of omitted variable biases in their estimates and suggest an ingenious identification strategy based on the notion that contracting institutions are likely to have a differential effect on different types of goods. More specifically, it seems natural to suppose that the type of contractual difficulties highlighted in this chapter are more likely to apply to complex goods than to simple goods. In fact, it is rather simple to extend the above model of exporter misbehavior to formalize this insight. For that purpose, assume that whether the exporter will be presented with an opportunity to shirk or misbehavior is random and occurs with a probability ρ which is a function of the type of good being traded and is higher for complex, less contractible goods than for simple, homogeneous goods. Provided that both producers know whether such misbehavior is possible or not before they sign the initial contract (but not before the fixed exporting cost is incurred), it is straightforward to show that the equilibrium of the model will be identical to that above, but with

$$\tilde{\mu}_{ij}^{exp} \equiv \rho + (1 - \rho)\,\mu_{ij}^{exp}$$

replacing μ_{ij}^{exp} throughout. It is then clear that the effect of better contract enforcement on profitability, firm-level exports, bilateral exports, and so on is lower, the higher is ρ, i.e., the more complex goods are.

In order to test this prediction, Berkowitz et al. (2006) employ the Rauch (1999) classification of goods into differentiated and homogeneous and run their specification on each set of goods separately. Their results are reproduced in columns (3) and (4) of Table 3.3. A striking feature of their results is that high levels of contract enforcement in the exporting country are shown to increase exports of complex goods but at the same time they *reduce* exports of simple goods. Conversely, and somewhat puzzlingly, good contracting institutions in the importing country enhance imports of simple goods, but reduce those of complex goods!

In order to rationalize their findings, Berkowitz et al. (2006) argue that the quality of contracting institutions will not only affect the security with which international transactions are conducted, but will also shape the efficiency with which the traded goods are produced, thus becoming a source of comparative advantage. Viewed from that perspective, their results in columns (3) and (4) are less surprising. They simply might reflect that countries with strong contracting institutions gain comparative advantage in contract-intensive, complex goods, and as a result they tend to feature disproportionately high levels of exports of complex goods and disproportionately low levels of imports of simple goods.

The idea that the quality of domestic institutions may constitute a source of comparative advantage has featured prominently in the trade literature in recent years. The vast literature on the topic is reviewed in Nunn and Trefler (2013), so I will only sketch a few key contributions here. The earlier papers in that literature were closest in spirit to the work of Berkowitz et al. (2006). Nunn (2007), Levchenko (2007), and Costinot (2009) all explored how domestic contracting institutions shape productivity differentially across sectors depending on characteristics of those sectors. They each proposed a measure of contract intensity different from the dichotomous one used by Berkowitz et al. (2006), and each showed that the effect of contracting institutions on trade flows was disproportionately higher in the industries identified to be relatively contract-intensive. I will use some of these measures in the empirical work in Chapter 4, so I will provide more details on these measures at that point.

Building on the insights of this literature on contracting institutions, other researchers have explored the role of other types of institutions in shaping comparative advantage across sectors. Manova (2008, 2012), for instance, explores the role of the quality of financial institutions in shaping the extensive and intensive margin of trade. Her empirical strategy builds on the seminal work of Rajan and Zingales (1998), who categorized sectors into more or less financially dependent depending on their external finance requirements. Relatedly, Cuñat and Melitz (2012) study how differences in the flexibility of labor market institutions across countries affect comparative advantage by building an industry-level measure of the importance of within-sectoral reallocations of labor as a response to shocks. In a very nicely executed study, Chor (2010) attempts to disentangle the partial effect of each of these institutional determinants of comparative advantage in a unified empirical model.

One cannot conclude from the results in Berkowitz et al. (2006), as well as those of the extensive literature on institution-driven comparative advantage, that international contract enforcement is irrelevant for explaining trade flows across countries. First, the findings of Helpman et al. (2008) regarding the effect of having a common legal origin on aggregate bilateral trade flows are hard to rationalize in models in which international contract enforcement is perfect. Second, Berkowitz et al. (2006) also found that the effects of exporter and importer legal quality appear to be significantly affected by whether countries have ratified the New York convention or not. For instance, their results indicate that for the case of complex goods, the quality of exporter institutions matters disproportionately more when the export partner has not yet signed the New York, and thus international enforcement of damages is more doubtful. For a third illustration of the importance of imperfect contract enforcement, I next briefly return to the choice of payment method decision faced by exporters and importers when negotiating their initial contracts.

Back to Trade Finance

So far we have illustrated how the quality of importer country institutions shapes the profitability and structure of exports whenever contracts are associated with post-shipment payment (or simply, open account) terms, while the quality of exporter country institutions plays a similar role in cash-in-advance transactions. Obviously, this constitutes a simplistic description of the effect of institutions on exporting. It seems natural, for instance, that exporter country institutions will matter even in open account terms to the extent that the consequences of exporter misbehavior might manifest themselves long after the goods have been received by the importer, or even after these goods have been sold to local consumers. Similarly, importer country institutions might affect the profitability of cash-in-advance transactions to the extent that they shape the financing costs faced by exporters. Intuitively, in countries where defaults are not sufficiently punished, not only exporters but also banks will shy away from extending credit to importers.

An active literature in international trade has explored the determinants of the choice of payment mode in international transactions, with a special emphasis on the role of weak contracting institutions. This literature includes among others, the work of Amiti and Weinstein (2011), Antràs and Foley (2013), Ahn (2011), Hoefele, Schmidt-Eisenlohr and Yu (2013), Olsen (2013), and Schmidt-Eisenlohr (2013). Antràs and Foley (2013), in particular, focus on the role of importer country institutions, while allowing these to affect the profitability of both open account transactions (via default risk) as well as cash-in-advance transactions (via financing costs). Their key theoretical finding is that in the plausible case in which local banks in the importing country are better able than exporters to pursue financial claims against importers, one would expect exports to locations characterized by weak contractual enforcement to be more likely to occur on cash in advance as opposed to open account terms.[13]

One of the main challenges in studying the financing arrangements used to support international trade is that detailed data on how different types of transactions are financed are not readily available. Antràs and Foley (2013) overcome this dearth of data by analyzing detailed transaction-level data from a single U.S.-based firm that exports frozen and refrigerated food products, primarily poultry. The data cover roughly $7 billion in sales to more than 140 countries over the 1996-2009 period and contain comprehensive information on the financing terms used in each transaction. A key advantage of the dataset is that by focusing on the sales of a single exporter based in the U.S., any institution-driven variation in the choice of payment mode must be ascribed to importer-country institutions or, following our broader interpretation of the parameter μ_{ij} in (3.6), to legal proximity between the U.S. and the importing country.

Antràs and Foley (2013) find robust evidence that variation in importer country contract enforcement has a strong effect on the method of payment offered to importers. Figure 3.3 reproduces the results in Figure 3 of their paper. For each of the proxies of contractual enforcement in the figure, the share of transactions occurring on cash in advance share is strikingly lower in strong contract enforcement countries than in weak contract enforcement countries. For instance, in common law countries, 4.0% of sales occur on

[13] Antràs and Foley (2013) also consider the possibility of exporters and importers resorting to letters of credit, but these financial instruments mediate a small share of world trade in modern times (see also Olsen, 2013, for more on letter of credit).

cash in advance terms and 79.8% of sales occur on open account terms, while in civil law countries these shares are 63.8% and 20.4%. Similarly stark differences appear when the sample is split using measures of contract viability (from International Country Risk Guide), payment delay (also from ICRG), and the enforceability of contracts (from Knack and Keefer, 1995). Antràs and Foley (2013) show that these patterns persist after controlling for several country-level variables as well as product fixed effects. Consistently with the results of the dynamic model of repeated interactions developed earlier in this chapter, they also find that first-time buyers are disproportionately more likely to be demanded to prepay for their purchases, but that as the exporter establishes a relationship with an importer, the share of cash in advance transactions falls smoothly over time.

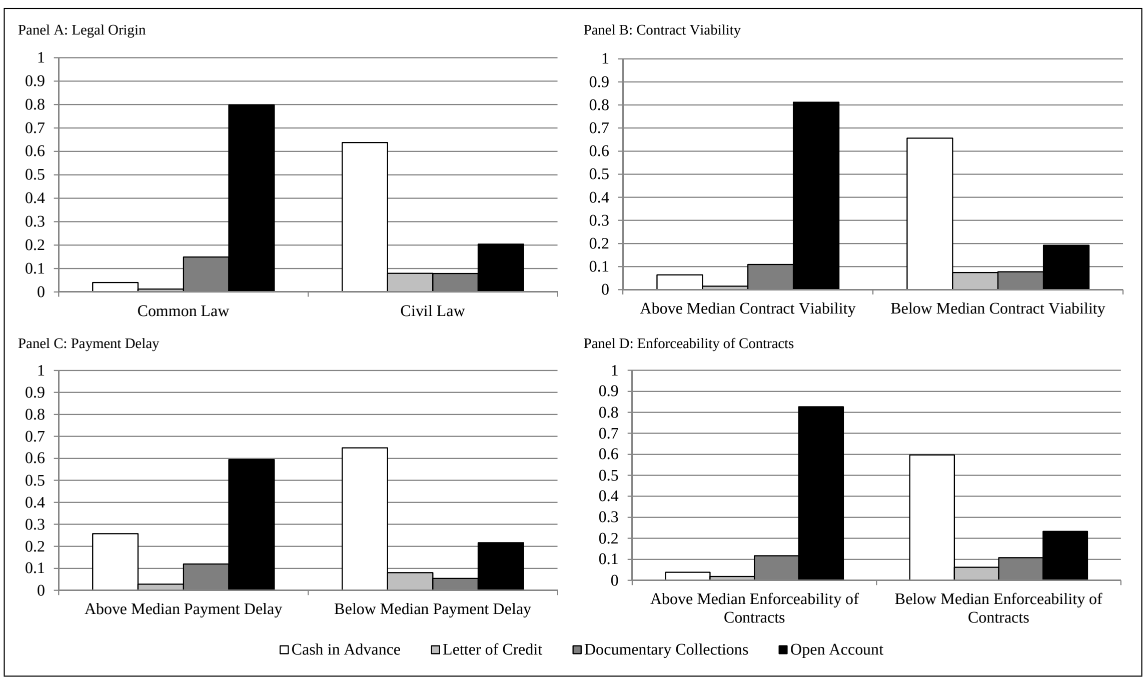

Figure 3.3: Financing Terms and the Enforcement of Contracts

In a recent paper, Hoefele, Schmidt-Eisenlohr and Yu (2013) have employed information from the World Bank Enterprise Survey to study the effects of variation in *exporting* country institutions on the choice of trade finance by firms. Consistently with the models developed in this chapter, Hoefele, Schmidt-Eisenlohr and Yu (2013) find that the use of cash-in-advance terms is more prevalent in exporting countries with strong contracting institutions, in which exporter misbehavior is less of a concern.

Chapter 4

Contracts and Global Sourcing

Part III

Internalization

Chapter 5

Transaction Costs and Internalization

Chapter 6

The Property-Rights Theory

Chapter 7

Internalization: Empirical Evidence

Chapter 8

The Road Ahead

Appendix

Optimal Sourcing Strategy in the Multi-Country Global Sourcing Model in Chapter 2

Remember that the problem is given by

$$\begin{aligned}\max_{\mathcal{J}_i(\varphi)} \pi_{ij}\left(\varphi, \mathcal{J}_i\left(\varphi\right)\right) &= \left(a_{hi}w_i\right)^{-\eta(\sigma-1)}\left(\gamma \sum\nolimits_{k\in\mathcal{J}_i(\varphi)} T_k\left(\tau_{ik}w_k\right)^{-\theta}\right)^{(\sigma-1)(1-\eta)/\theta} B\varphi^{\sigma-1} \\ &\quad -w_i\sum\nolimits_{k\in\mathcal{J}_i(\varphi)} f_{ik}.\end{aligned}$$

With a discrete number of locations, we can rewrite the problem as follows:

$$\begin{aligned}\max_{I_1,I_2,\ldots,I_J\in\{0,1\}^J} \pi_{ij}\left(\varphi, I_1, I_2, ..., I_J\right) &= \left(\sum_{k=1}^{J} I_j T_k\left(\tau_{ik}w_k\right)^{-\theta}\right)^{(\sigma-1)(1-\eta)/\theta} \widetilde{B}\varphi^{\sigma-1} \\ &\quad -w_i\sum_{k=1}^{J} I_j f_{ik},\end{aligned} \tag{8.1}$$

where $\widetilde{B} = \left(a_{hi}w_i\right)^{-\eta(\sigma-1)}\gamma^{(\sigma-1)(1-\eta)}B$. The dummy variable I_j thus takes a value of 1 when $j \in \mathcal{J}_i(\varphi)$ and 0 otherwise.

The key thing to note is that, provided that $(\sigma-1)(1-\eta) > \theta$, the modified objective function in (8.1) features increasing differences in (I_j, I_k) for $j, k \in \{1, ..., J\}$ such that $j \neq k$ and also features increasing differences in (I_j, φ) for any $j \in \{1, ..., J\}$. Invoking standard results in monotone comparative statics, we can then conclude that for $\varphi_1 \geq \varphi_0$, we must have $\left(I_1^*(\varphi_1), I_2^*(\varphi_1), ..., I_J^*(\varphi_1)\right) \geq \left(I_1^*(\varphi_0), I_2^*(\varphi_0), ..., I_J^*(\varphi_0)\right)$. Naturally, this rules out a situation in which $I_j^*(\varphi_1) = 0$ but $I_j^*(\varphi_0) = 1$, and thus we can conclude that $\mathcal{J}_i(\varphi_0) \subseteq \mathcal{J}_i(\varphi_1)$ for $\varphi_1 \geq \varphi_0$.

Bibliography

Ahn, J. (2011), A theory of domestic and international trade finance, IMF Working Paper 11/262, International Monetary Fund.

Albornoz, F., Calvo Pardo, H. F., Corcos, G. and Ornelas, E. (2012), 'Sequential exporting', *Journal of International Economics* **88**(1), 17–31.

Alvarez, F. and Lucas, J. (2007), 'General equilibrium analysis of the Eaton–Kortum model of international trade', *Journal of Monetary Economics* **54**(6), 1726–1768.

Amiti, M. and Davis, D. R. (2012), 'Trade, firms, and wages: Theory and evidence', *The Review of Economic Studies* **79**(1), 1–36.

Amiti, M. and Weinstein, D. E. (2011), 'Exports and financial shocks', *The Quarterly Journal of Economics* **126**(4), 1841–1877.

Anderson, J. E. and Marcouiller, D. (2002), 'Insecurity and the pattern of trade: An empirical investigation', *Review of Economics and Statistics* **84**(2), 342–352.

Anderson, J. E. and Marcouiller, D. (2005), 'Anarchy and autarky: Endogenous predation as a barrier to trade', *International Economic Review* **46**(1), 189–213.

Anderson, J. E. and van Wincoop, Eric (2003), 'Gravity with gravitas: A solution to the border puzzle', *The American Economic Review* **93**(1), 170–192.

Antràs, P. and Chor, D. (2013), 'Organizing the global value chain', *Econometrica* . Forthcoming.

Antràs, P., Chor, D., Fally, T. and Hillberry, R. (2012), 'Measuring the upstreamness of production and trade flows', *American Economic Review* **102**(3), 412–416.

Antràs, P. and Costinot, A. (2011), 'Intermediated trade', *The Quarterly Journal of Economics* **126**(3), 1319–1374.

Antràs, P. and Foley, F. (2013), 'Poultry in motion: A study of international trade finance practices'.

Antràs, P., Garicano, L. and Rossi-Hansberg, E. (2006), 'Offshoring in a knowledge economy', *The Quarterly Journal of Economics* **121**(1), 31–77.

Antràs, P. and Helpman, E. (2004), 'Global sourcing', *Journal of Political Economy* **112**(3), 552–580.

Antràs, P. and Rossi-Hansberg, E. (2009), 'Organizations and trade', *Annual Review of Economics* **1**(1), 43–64.

Antràs, P. and Staiger, R. (2012), 'Offshoring and the role of trade agreements', *American Economic Review* **102**(7), 3140–3183.

Antràs, P. and Yeaple, S. (2013), Multinational firms and international trade structure, *in* 'Handbook of International Economics', Vol. 4, Elsevier.

Araujo, L. F., Mion, G. and Ornelas, E. (2012), Institutions and export dynamics, Technical report, London School of Economics Working Paper.

Arkolakis, C., Costinot, A. and Rodríguez-Clare, A. (2012), 'New trade models, same old gains?', *American Economic Review* **102**(1), 94–130.

Arkolakis, C., Demidova, S., Klenow, P. J. and Rodríguez-Clare, A. (2008), 'Endogenous variety and the gains from trade', *American Economic Review* **98**(2), 444–450.

Atalay, E., Hortacsu, A. and Syverson, C. (2013), 'Vertical integration and input flows'. mimeo University of Chicago.

Baker, G. P. and Hubbard, T. N. (2003), 'Make versus buy in Trucking:Asset ownership, job design, and information', *American Economic Review* **93**(3), 551–572.

Baldwin, R. E. and Forslid, R. (2010), 'Trade liberalization with heterogeneous firms', *Review of Development Economics* **14**(2), 161–176.

Bardhan, P., Mookherjee, D. and Tsumagari, M. (2013), 'Middlemen margins and globalization', *American Economic Journals: Micro* . Forthcoming.

Behar, A. and Freund, C. (2011), 'Factory europe? brainier but not brawnier', *Unpublished Manuscript* .

Berkowitz, D., Moenius, J. and Pistor, K. (2006), 'Trade, law, and product complexity', *Review of Economics and Statistics* **88**(2), 363–373.

Bernard, A. B., Eaton, J., Jensen, J. B. and Kortum, S. (2003), 'Plants and productivity in international trade', *American Economic Review* **93**(4), 1268–1290.

Bernard, A. B., Jensen, J. B., Redding, S. J. and Schott, P. K. (2007), 'Firms in international trade', *Journal of Economic Perspectives* **21**(3), 105–130.

Bernard, A. B., Jensen, J. B. and Schott, P. K. (2009), Importers, exporters, and multinationals: A portrait of firms in the U.S. that trade goods, *in* T. Dunne, J. B. Jensen and M. J. Roberts, eds, 'Producer Dynamics: New Evidence from Micro Data', NBER.

Blaum, J., Lelarge, C. and Peters, M. (2013), 'Non-homothetic import demand: Firm productivity and quality bias', *Unpublished paper* .

Broda, C. and Weinstein, D. E. (2006), 'Globalization and the gains from variety', *The Quarterly Journal of Economics* **121**(2), 541–585.

Campa, J. M. and Goldberg, L. S. (1997), 'The evolving external orientation of manufacturing: A profile of four countries', *Federal Reserve Bank of New York Economic Policy Review* **3**, 53–81.

Chaney, T. (2008), 'Distorted gravity: The intensive and extensive margins of international trade', *American Economic Review* **98**(4), 1707–1721.

Chen, H., Kondratowicz, M. and Yi, K.-M. (2005), 'Vertical specialization and three facts about U.S. international trade', *The North American Journal of Economics and Finance* **16**(1), 35–59.

Chor, D. (2010), 'Unpacking sources of comparative advantage: A quantitative approach', *Journal of International Economics* **82**(2), 152–167.

Clermont, K. M. and Eisenberg, T. (2007), 'Xenophilia or xenophobia in U.S. courts? before and after 9/11', *Journal of Empirical Legal Studies* **4**(2), 441–464.

Coase, R. H. (1937), 'The nature of the firm', *Economica* **4**(16), 386–405.

Costinot, A. (2009), 'On the origins of comparative advantage', *Journal of International Economics* **77**(2), 255–264.

Cuñat, A. and Melitz, M. J. (2012), 'Volatility, labor market flexibility, and the pattern of comparative advantage', *Journal of the European Economic Association* **10**(2), 225–254.

Davis, D. R. and Weinstein, D. E. (2001), 'An account of global factor trade', *American Economic Review* **91**(5), 1423–1453.

Deardorff, A. V. (2001), 'Fragmentation in simple trade models', *The North American Journal of Economics and Finance* **12**(2), 121–137.

Demidova, S. and Rodríguez-Clare, A. (2009), 'Trade policy under firm-level heterogeneity in a small economy', *Journal of International Economics* **78**(1), 100–112.

Demidova, S. and Rodríguez-Clare, A. (2013), 'The simple analytics of the melitz model in a small economy', *Journal of International Economics* **90**(2), 266–272.

Dixit, A. K. and Stiglitz, J. E. (1977), 'Monopolistic competition and optimum product diversity', *The American Economic Review* **67**(3), 297–308.

Djankov, S., La Porta, R., Lopez-De-Silanes, F. and Shleifer, A. (2003), 'Courts', *The Quarterly Journal of Economics* **118**(2), 453–517.

Eaton, J. and Kortum, S. (2002), 'Technology, geography, and trade', *Econometrica* **70**(5), 1741–1779.

Eaton, J., Kortum, S. and Kramarz, F. (2011), 'An anatomy of international trade: Evidence from french firms', *Econometrica* **79**(5), 1453–1498.

Ethier, W. J. (2001), Dixit-stiglitz, trade and growth, *in* S. Brakman and B. J. Heijdra, eds, 'The Monopolistic Competition Revolution in Retrospect', Cambridge University Press.

Fally, T. (2012), *Production Staging: Measurement and Facts.* mimeo University of Colorado.

Feenstra, R. C. (1998), 'Integration of trade and disintegration of production in the global economy', *Journal of Economic Perspectives* **12**(4), 31–50.

Feenstra, R. C. and Hanson, G. H. (1996*a*), Foreign investment, outsourcing and relative wages, *in* R. Feenstra, G. Grossman and D. Irwin, eds, 'The Political Economy of Trade Policy: Papers in Honor of Jagdish Bhagwati', MIT Press.

Feenstra, R. C. and Hanson, G. H. (1996*b*), 'Globalization, outsourcing, and wage inequality', *American Economic Review* **86**, 240–245.

Felbermayr, G., Jung, B. and Larch, M. (2013), 'Optimal tariffs, retaliation, and the welfare loss from tariff wars in the melitz model', *Journal of International Economics* **89**(1), 13–25.

Foley, C. F., Chen, M., Johnson, M. and Meyer, L. (2009), 'Noble group'. 210-021, October 2009.

Gans, J. S. and Shepherd, G. B. (1994), 'How are the mighty fallen: Rejected classic articles by leading economists', *Journal of Economic Perspectives* **8**(1), 165–179.

Goldberg, P. K., Khandelwal, A. K., Pavcnik, N. and Topalova, P. (2010), 'Imported intermediate inputs and domestic product growth: Evidence from india', *The Quarterly Journal of Economics* **125**(4), 1727–1767.

Gopinath, G. and Neiman, B. (2013), 'Trade adjustment and productivity in large crises', *American Economic Review* p. Forthcoming.

Grossman, G. M. and Rossi-Hansberg, E. (2008), 'Trading tasks: A simple theory of offshoring', *American Economic Review* **98**(5), 1978–97.

Grossman, S. J. and Hart, O. D. (1986), 'The costs and benefits of ownership: A theory of vertical and lateral integration', *Journal of Political Economy* **94**(4), 691–719.

Halpern, L., Koren, M. and Szeidl, A. (2009), 'Imported inputs and productivity', *Working Paper, CEU* **8**, 28.

Harrison, A. and Scorse, J. (2010), 'Multinationals and anti-sweatshop activism', *American Economic Review* **100**(1), 247–273.

Hart, O. and Moore, J. (1994), 'A theory of debt based on the inalienability of human capital', *The Quarterly Journal of Economics* **109**(4), 841–879.

Helpman, E. (1984), 'A simple theory of international trade with multinational corporations', *Journal of Political Economy* **92**(3), 451–471.

Helpman, E., Itskhoki, O. and Redding, S. (2010), 'Inequality and unemployment in a global economy', *Econometrica* **78**, 1239–1283.

Helpman, E. and Krugman, P. (1985), *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, The MIT Press.

Helpman, E., Melitz, M. J. and Yeaple, S. R. (2004), 'Export versus FDI with heterogeneous firms', *The American Economic Review* **94**(1), 300–316.

Helpman, E., Melitz, M. and Rubinstein, Y. (2008), 'Estimating trade flows: Trading partners and trading volumes', *The Quarterly Journal of Economics* **123**(2), 441–487.

Hoefele, A., Schmidt-Eisenlohr, T. and Yu, Z. (2013), Payment choice in international trade: Theory and evidence from cross-country firm level data, CESifo Working Paper Series 4350, CESifo Group Munich.

Hummels, D. (2007), 'Transportation costs and international trade in the second era of globalization', *Journal of Economic Perspectives* **21**(3), 131–154.

Hummels, D., Ishii, J. and Yi, K.-M. (2001), 'The nature and growth of vertical specialization in world trade', *Journal of International Economics* **54**, 75–96.

Johnson, R. C. and Noguera, G. (2012*a*), 'Accounting for intermediates: Production sharing and trade in value added', *Journal of International Economics* **86**(2), 224–236.

Johnson, R. C. and Noguera, G. (2012*b*), *Fragmentation and trade in value added over four decades*, Working Paper, Dartmouth College.

Jones, R. W. (2000), *Globalization and the Theory of Input Trade*, MIT Press.

Jones, R. W. and Neary, P. (1984), The positive theory of international trade, *in* R. W. Jones and P. B. Kenen, eds, 'Handbook of International Economics', Vol. Volume 1, Elsevier, p. 1–62.

Jovanovic, B. (1982), 'Selection and the evolution of industry', *Econometrica* **50**(3), 649.

Knack, S. and Keefer, P. (1995), 'Institutions and economic performance: Cross-country tests using alternative institutional measures', *Economics & Politics* **7**(3), 207–227.

Koopman, R., Wang, Z. and Wei, S.-J. (2013), 'Tracing value-added and double counting in gross exports', *American Economic Review* . Forthcoming.

Kremer, M. and Maskin, E. (2006), 'Globalization and inequality', *mimeo Harvard University* .

Krugman, P. (1980), 'Scale economies, product differentiation, and the pattern of trade', *The American Economic Review* **70**(5), 950–959.

Krugman, P. R. (1979), 'Increasing returns, monopolistic competition, and international trade', *Journal of International Economics* **9**(4), 469–479.

Kuhn, T. S. (1996), *The Structure of Scientific Revolutions*, University of Chicago Press.

Lafontaine, F. and Slade, M. (2007), 'Vertical integration and firm boundaries: The evidence', *Journal of Economic Literature* **45**(3), 629–685.

LaPorta, R., Lopez-de Silanes, F., Shleifer, A. and Vishny, R. (1999), 'The quality of government', *Journal of Law, Economics and Organization* **15**(1), 222–279.

Leamer, E. E. (1984), *Sources of International Comparative Advantage: Theory and Evidence*, MIT Press.

Levchenko, A. A. (2007), 'Institutional quality and international trade', *The Review of Economic Studies* **74**(3), 791–819.

Manova, K. (2008), 'Credit constraints, equity market liberalizations and international trade', *Journal of International Economics* **76**(1), 33–47.

Manova, K. (2012), 'Credit constraints, heterogeneous firms, and international trade', *The Review of Economic Studies* p. rds036.

Melitz, M. J. (2003), 'The impact of trade on intra-industry reallocations and aggregate industry productivity', *Econometrica* **71**(6), 1695–1725.

Melitz, M. J. and Ottaviano, G. I. P. (2008), 'Market size, trade, and productivity', *Review of Economic Studies* **75**(1), 295–316.

Melitz, M. J. and Redding, S. (2012), Heterogeneous firms and trade, *in* 'Handbook of International Economics', Vol. Volume 4, Elsevier.

Melitz, M. and Redding, S. (2013), 'Firm heterogeneity and aggregate welfare'. mimeo Harvard University.

Miroudot, S., Lanz, R. and Ragoussis, A. (2009), 'Trade in intermediate goods and services', *OECD Trade Policy Working Papers* (93).

Nguyen, D. X. (2012), 'Demand uncertainty: Exporting delays and exporting failures', *Journal of International Economics* **86**(2), 336–344.

Novy, D. (2013), 'International trade without CES: estimating translog gravity', *Journal of International Economics* **89**(2), 271–282.

Nunn, N. (2007), 'Relationship-specificity, incomplete contracts, and the pattern of trade', *The Quarterly Journal of Economics* **122**(2), 569–600.

Nunn, N. and Trefler, D. (2013), Domestic institutions as a source of comparative advantage, *in* 'Handbook of International Economics', Vol. 4, North Holland.

Olsen, M. (2013), 'How firms overcome weak international contract enforcement: Repeated interaction, collective punishment, and trade finance'.

Rajan, R. G. and Zingales, L. (1998), 'Financial dependence and growth', *American Economic Review* **88**(3), 559–86.

Ramondo, N., Rappoport, V. and Ruhl, K. J. (2013), 'Horizontal versus vertical foreign direct investment: Evidence from U.S. multinationals'. mimeo NYU Stern.

Rauch, J. E. (1999), 'Networks versus markets in international trade', *Journal of International Economics* **48**(1), 7–35.

Redding, S. and Venables, A. J. (2004), 'Economic geography and international inequality', *Journal of International Economics* **62**(1), 53–82.

Rodrik, D. (2000), 'How far will international economic integration go?', *Journal of Economic Perspectives* **14**(1), 177–186.

Ruhl, K. (2013), 'An overview of US intrafirm-trade data sources'. mimeo NYU Stern.

Schmidt-Eisenlohr, T. (2013), 'Towards a theory of trade finance', *Journal of International Economics* **91**(1), 96–112.

Schott, P. K. (2004), 'Across-product versus within-product specialization in international trade', *The Quarterly Journal of Economics* **119**(2), 647–678.

Segura-Cayuela, R. and Vilarrubia, J. M. (2008), Uncertainty and entry into export markets, Banco de España Working Paper 0811, Banco de España.

Simonovska, I. and Waugh, M. E. (2011), The elasticity of trade: Estimates and evidence, Working Paper 16796, National Bureau of Economic Research.

Tang, H. and Zhang, Y. (2012), 'Quality differentiation and trade intermediation'.

Thomas, J. and Worrall, T. (1994), 'Foreign direct investment and the risk of expropriation', *Review of Economic Studies* **61**(1), 81–108.

Tinbergen, J. (1962), *Shaping the World Economy; Suggestions for an International Economic Policy*, New York: Twentieth Century Fund.

Tintelnot, F. (2013), 'Global production with export platforms'. mimeo Penn State University.

Trefler, D. (1993), 'International factor price differences: Leontief was right!', *Journal of Political Economy* **101**(6), 961–987.

Trefler, D. (1995), 'The case of the missing trade and other mysteries', *The American Economic Review* **85**(5), 1029–1046.

Vanek, J. (1968), 'The factor proportions theory: The n-factor case', *Kyklos* **21**(4), 749–756.

Wright, G. C. (2012), 'Revisiting the employment impact of offshoring', *Working Paper, University of Essex* .

Yeats, A. J. (2001), Just how big is global production sharing?, *in* S. W. Arndt and H. Kierzkowski, eds, 'Fragmentation New Production Patterns in the World Economy', Oxford University Press, Oxford.