Trade Policy with Heterogeneous Firms, Variable Markups, and Multinational Production*

Ziran Ding
University of Washington

Version: 1.0

Date: April 10, 2020

Abstract

This paper surveys the main ingredients and results of heterogeneous firms trade policy literature that has been developing since the early 2000s. First, I present in great detail various stylized facts regarding firm heterogeneity, firm-level markups, and the global structure of multinational production. Second, I summarize the results of the recent development of theoretical approaches of modeling the firm-level markups. Third, I discuss the theoretical frameworks that incorporates multinational production into heterogeneous firms framework. Fourth, I review the trade policy literature that features firm heterogeneity, variable markups, and multinational production. Finally, I discuss directions for future research and offer suggestions for further readings.

Keywords: Trade policy, Firm heterogeneity, Variable markups, Multinational production

1 Introduction

Despite its outstanding contribution to the world output growth, ever since the global financial crisis, international trade and investment have provided much less support to economic growth. Trade growth has barely kept up with output growth and has even lagged behind it for a few years. Compared to the persistent growth from 1990 to 2007, with an average rate of 15 percent, global investment has been fluctuating around a 1 percent annual growth rate for the past ten years. Just

*I'm indebted to Theo Eicher, Fabio Ghironi and Mu-Jeung Yang for guidance, encouragement and continuous support. All errors are my own.

as trade was showing some tentative signs of renewed vigor[1], the anti-globalization sentiment has started to pose a significant threat to global economic integration. For example, after more than 50 years of leading efforts to lower international trade barriers, since 2018, the US has enacted several waves of tariff increases on various products, sectors, and countries. Take China, for instance: From 1989 to 2017, the average US-applied tariffs on imports from China gradually decreased from 5 percent to 3 percent. During the same period, China's tariffs on imports from the US declined from 40 percent to 8 percent. However, according to Bown [2019], by the end of 2019, 96.8 percent of US imports from China are subject to tariffs at an average rate of 20 percent. These measures represent the most comprehensive protectionist trade policies implemented by the US since the 1930s. In response, China also imposes retaliatory tariffs on US exports. By mid-December of 2019, 69 percent of China imports from the US are subject to an average tariff rate of 25.9 percent.

Due to the sizes of the economies involved, the magnitudes of the tariff increases, and the breadth of tariffs across sectors, the current trend of returning to protectionism is unprecedented in the postwar era. What is the welfare implication of such a protectionist trade policy? The traditional wisdom believes that producers can benefit from protectionism by avoiding competition with foreign suppliers, but consumers will suffer from having fewer varieties. This perception is based on two premises: (i) Firms in an industry are atomic and they act as price takers. (ii) The primary channel for firms to access foreign markets is through export. In the past two decades, however, the micro-level data have presented quite a different picture. This paper provides an overview of the trade and trade policy literature from three specific angles: firm-level productivity heterogeneity, firm-level variable markups, and multinational production (henceforth, MP).

Since the pioneering work of Melitz [2003], trade economists have been increasingly focusing on the firm as the unit of analysis. This can be partly attributed to the empirical findings using micro-level data on plants and firms. In the first section (Section 2) of this survey, I briefly summarize the empirical evidence that has been established in the heterogeneous firm literature. One set of empirical findings show that firms are heterogeneous in their productivity even within narrowly defined industries. In particular, firms engaging in MP are usually more productive than pure exporters (see, for example, Temouri et al. [2008], and Criscuolo and Martin [2009]). The second set of empirical results highlights the monopoly power across firms within industries. The micro-level

[1]According to UNCTAD [2018].

evidence suggests that the most productive firms charge the highest markups (see De Loecker and Warzynski [2012] for exporters, and see Dobbelaere and Kiyota [2018] for multinational firms). While the previous two strands of empirical studies focus on firm-level characteristics, others emphasize the role of the firm in understanding patterns of trade and the global production structure. Although goods are mobile across borders, their movement is subject to various types of costs: shipping costs, tariffs, and legal barriers to trade. It turns out that these frictions shape a firm's production decision and location choice, which are pivotal to the understanding of global production (see for example Antràs and Yeaple [2014], and Yeaple [2013]).

Although the Melitz-type heterogeneous firm model has been the workhorse in trade literature, it is clearly at odds with two of the most robust empirical findings in the micro-level data: (i) Different firms charge different levels of markups, even within the same industry.(ii) There is a rapid growth of MP throughout the globe. The first discrepancy has spurred an extensive theoretical literature on introducing variable markups into the heterogeneous firm framework. In the next section (Section 3) of this survey, I analyze and contrast the different approaches that trade economists have adopted to model variable markups. In general, there are three ways to introduce variable markups: the demand-side approach, the supply-side approach, and a combination of both. The demand-side approach focuses on modifying the constant elasticity of substitution (henceforth, CES) preference imposed on consumers. The supply-side approach focuses on modifying the monopolistic competition assumption between the producers. Due to its advantage of analytical tractability, the first approach is more prevalent in the literature, but there has been a resurgence in the supply-side approach (see Head and Spencer [2017] for example). The last approach, which is the combination of the previous two, while being undoubtedly realistic, does not seem to generate additional insights compared to the other two approaches (see, for example, Kokovin et al. [2017]).

The second discrepancy is addressed by the burgeoning papers of multinational firms. In Section 4, I briefly review the state of this literature. Given the space constraint, I mainly focus on various theoretical and quantitative approaches that have been adopted to introduce multinational activity into the heterogeneous firm framework. I divide the review into three parts: (i) Horizontal FDI, which refers to the fact that multinational firms duplicate roughly the same activities in different countries, is the dominant format of FDI flows between developed countries. (ii) Vertical FDI, which describes firms' motive to locate different parts of the production process in different countries, has

been increasing since many developing countries are participating in the global supply chain. (iii) Complex FDI, which captures the notion that firms replicate some activities in many countries while concentrating other activities in a few countries, has been brought to light in recent years thanks to the availability of micro-level data of multinational firms.

Given the empirical evidence and the theoretical development in the trade literature, it is necessary and meaningful to resuscitate the trade policy associated with different types of framework. In Section 5, I review the recent developments in the heterogeneous firm trade policy literature. I begin by investigating the trade policy papers in the benchmark Melitz framework. Then I go over the papers that emphasize the role of variable markup within the heterogeneous firm structure. In the end, I discuss a few papers that look at trade policy questions with firm heterogeneity and MP. While most of the papers reviewed in this section are normative papers studying optimal trade policy, due to the ongoing trade war, I also mention a few recent papers that study the consequences of trade policy uncertainty. In contrast to traditional steady-state analysis, this type of trade policy paper is usually carried out in either a dynamic or an open-economy macro framework.

The review here is limited in terms of focus. There is no need to be encyclopedic given the space constraints and the useful surveys of the literature. To the best of my knowledge, this paper is the first to review the theoretical development of variable markups and trade policy in the heterogeneous firm literature. As for the review of firm heterogeneity and multinational production, more comprehensive treatments include the handbook chapters by Melitz and Redding [2014], Antràs and Yeaple [2014], Yeaple [2013]. Although I cannot avoid covering some of the same ground as these earlier surveys, I focus primarily on the theoretical and empirical work in the past 15 years.

The remainder of the paper is structured as follows. Section 2 identifies a number of stylized facts concerning firm heterogeneity, variable markups, and MP. While some of these facts might be already known by readers who are familiar with the trade literature, they nevertheless provide an important benchmark to which I return when addressing the incompatibility between theory and data. Section 3 then surveys the various theoretical approaches that have been used to model variable markups within the heterogeneous firm framework. The papers discussed in this section have the common feature that they all exclude the possibility of MP. Section 4 discusses the papers that introduce MP into the heterogeneous firm framework. The papers in this section all feature

constant markups. Section 5 briefly goes through trade policy papers in the heterogeneous firm literature. Section 6 concludes. All tables and graphs to which this paper refers are included in the appendix.

2 Stylized Facts on Heterogeneous Firms, Variable Markups and FDI

This section discusses robust patterns in the data on heterogeneous firms. The discussion is organized into three areas. First, I discuss the differences between domestic firms, exporters, and multinational firms, focusing on firm-level productivity. Second, I discuss the markups charged by the firms within the same industry, which include entrants, firms that only serve domestic markets, exporters, and multinational firms. Finally, I discuss the structure of global production within multinational firms across countries. The micro-level evidence that supports this part of the discussion comes from firm-level data in both developed and developing economies.

2.1 Firm Heterogeneity

Classical theories of international trade believe that the differences in technologies and factor endowments are the basis for international trade. In these frameworks, countries import goods in one set of industries and export goods in another. Therefore trade only happens across the industries. However, earlier empirical evidence indicates that a significant amount of trade occurs between industries with similar technologies and factor endowments. Krugman [1980] successfully explains this intra-industry trade phenomenon by bringing increasing returns to scale and consumers' love for variety into the traditional framework. Helpman and Krugman [1985] integrate the inter-industry and intra-industry trade into an equilibrium framework and provide a sound explanation for trade patterns across countries and industries. The conventional modeling approach in this literature is the representative firm framework. From the mid-1990s, with the firm-level and plant-level data becoming more accessible than earlier, it has become clear that there is a considerable amount of heterogeneity across firms within an industry. This subsection briefly documents the well-established

empirical evidence regarding firm heterogeneity within an industry.

2.1.1 Export Participation

Countries possess quite different export participation patterns. Bernard et al. [2007] show that only a small portion of plants within an industry export in the US. As shown in **Table 1** column 3, the average industry-level export share is around 18 percent in the US manufacturing sector. There are substantial variations across industries: from only 4 percent in nonmetallic mineral products to almost 40 percent in computer and electronic products. Mayer and Ottaviano [2008] utilize firm-level data from the European Firms in International Markets (EFIM)[2] project and find that European firms have higher export participation rates[3]. As illustrated in **Table 2**, about 55 percent of firms export among countries with nonexhaustive data[4], and about 40 percent of firms export among countries with exhaustive data. They also find that export is granular among these countries: The top 1 percent of exporters account for more than 45 percent of the aggregate exports. Lu [2010] finds that Chinese manufacturing firms' export participation rate is around 30 percent. She also finds that the Chinese export intensity distribution is U-shaped: Fewer than 20 percent of exporters sell less than 10 percent of their output abroad, while about 40 percent of them export more than 90 percent of the total output.

2.1.2 Exporters are Different

Not only are exporters rare, but also are they systematically different from nonexporters. **Table 3** presents the US manufacturing exporters' premia documented in Bernard et al. [2007]. Exporters are larger in employment and sales, and they are more productive in value-added per worker and total factor productivity. They also pay higher wages and are more capital and skill intensive. Similar patterns are confirmed in the European firm-level data by Mayer and Ottaviano [2008], as shown in **Table 4**. In general, compared to nonexporters, exporters are typically more productive, sell much

[2]Now updated to European Firms in Global Economy: `http://bruegel.org/publications/datasets/efige/`.

[3]It should be noted that the France, Germany, Hungary, Italy and the United Kingdom have large firms only in the EFIM dataset; Belgian and Norwegian data are exhaustive.

[4]Using exhaustive data, Eaton et al. [2004] find French manufacturing firms on average have higher export participation rates than US manufacturing firms. Bernard and Wagner [1997] find 44 percent of German manufacturing firms are exporting.

more in the domestic market, and export only a small proportion of their output. A notable exception is the Chinese manufacturing exporters. Using firm-level data from 1998 to 2007, Lu [2010] finds that Chinese manufacturing exporters, compared to nonexporters, are typically less productive (see **Figure 1**), sell less in the domestic market, and export a significant fraction of their output.

2.1.3 MNEs are More Productive

According to the estimates of UNCTAD [2011][5], multinational firms account for 25 percent of the global GDP and one third of international trade. Multinational firms not only are quantitatively important but also enhance our understanding of firm heterogeneity. Benfratello and Sembenelli [2006] exploit Italian manufacturing firm-level data with the GMM-System estimator developed by Blundell and Bond [1998], which allows them to control for firms' unobserved heterogeneity, inputs, and ownership endogeneity as well as measurement errors. The authors find a positive and significant effect for firms under US ownership: They tend to outperform both their domestic counterparts and firms under other ownership. Temouri et al. [2008] take a rich firm-level data set from Amadeus[6] to study the total factor productivity differences across 22 manufacturing and 17 service industries in Germany over the period 1995–2004. They find that multinational firms have significantly higher TFP than their domestic counterparts. Mayer and Ottaviano [2008] find similar multinational firm premia in Belgium, as illustrated in **Figure 2**. Using a sample from the UK Annual Respondents Database, Criscuolo and Martin [2009] find that the US and other foreign-owned plants are on average 42 percent and 30 percent, respectively, more productive than British domestic plants. According to Barefoot and Mataloni Jr [2011], in manufacturing sectors, US parents account for less than a half of 1 percent of enterprises but for over 62 percent of the value added and 58 percent of employment. Bloom et al. [2012] find that in the IT industry, management practice is a robust explanatory variable for US multinationals' exceptional performance compared to non-US multinationals or domestic firms in Europe. All this evidence suggests the superior productivity of multinational firms relative to nonmultinational domestic counterparts.

[5] World Investment Report.

[6] **A**nalyse **M**ajor **D**atabases from **EU**ropean **S**ources. Bureau van Dijk compiles public and private company accounts from so-called regional information providers (IPs) which are either Central Banks, Official statistical offices or a credit rating agency.

2.2 Variable Markups

Since the pathbreaking work of Melitz [2003], the combination of CES preference (of the consumers) and monopolistic competition (among the producers) has been the workhorse model in international economics[7]. Although this combination delivers high tractability, it implies constant markups and a complete pass-through in equilibrium for all the firms. However, thanks to the increased availability of microdata on firms and international trade, one of the most robust findings in the empirical trade literature in the past decade is that heterogeneous firms charge heterogeneous markups. In this subsection, I briefly summarize the literature that presents cross-sectional evidence on the variability of markups at firm or plant level.

2.2.1 Entrant's Markup

One of the most robust empirical observations emerging from the recent industrial organization studies is that new entrants have lower average productivity and higher exit rates than the existing incumbents. However, using a unique data set from US Census of Manufacturing, Foster et al. [2008] find that the observed disadvantage in revenue productivity is mainly due to entrants charging a lower price-cost markup rather than technical inefficiency. More recently, by introducing the demand-side into the structural model of production in the spirit of Hall [1988], Kılınç [2014] is able to estimate markups using firms' nominal sales and input expenditures via a control function approach while controlling for the endogeneity issue of inputs on productivity. Using annual plant-level data from manufacturing industries in Japan from 1985 to 2007, the author finds that entrants on average charge lower markups than incumbents. As illustrated in **Figure 3**, the result is robust for different productivity indices.

2.2.2 Exporter's Markup

There are strong theoretical and empirical supports for the notion that, because exporters are more productive than domestic producers, so they select into export and charge higher markups. Motivated by Hall et al. [1986], De Loecker and Warzynski [2012] notice that under any form of imperfect competition, the relevant markup drives a gap between a firm's input revenue share

[7]For works in international trade, see Bernard et al. [2007]. For works in international macroeconomics, see Ghironi and Melitz [2005].

and its output elasticity. They utilize Slovenian plant-level production data from 1994 to 2000 to estimate the markups. Without specifying the market structure in the product market, they find that exporters charge, on average, higher markups and that markups increase upon export entry. Using French census data from 1998 to 2007, Bellone et al. [2014] apply the methodology developed by De Loecker and Warzynski [2012] and find that markups are significantly higher for exporters across most industries, as represented by **Table 6**. Lately, Hornok and Muraközy [2018] apply the same methodology to estimate firm-level markup in Hungarian manufacturing firms (1995–2003). They find robust and consistent evidence for markup premiums of importers, but not exporters.

The other channel, which claims that exporters become more efficient after export entry, has received mixed empirical evidence[8]. Recently, Voigtlaender and Garcia-Marin [2019] decompose exporters' efficiency gains into changes in revenue productivity, markups, and marginal costs. Using rich plant-level data from Chile, Colombia, and Mexico, they find that markups are stable around the exporters' entry (the efficiency gain is fully passed through to consumers), but are higher for established exporters (limited pass-through of efficiency gain to consumers) after tariff-induced expansions. Their finding suggests that the common use of revenue-based productivity measures ($TFPR$) might be the reason that the literature has struggled to identify export-related efficiency gains within plants.

2.2.3 MNE's Markup

Perhaps due to the fact that data sets on multinationals rarely include detailed information about their activities in multiple countries, only a few studies have focused on the empirical relationship between MNE status and markups. Utilizing Spanish firm-level data over the period 1983–1996, Sembenelli and Siotis [2008] attempt to empirically disentangle the efficiency, spillovers, and competition effects of FDI on firms' markup. They find that after controlling for potential endogeneity biases and economy-wide effects, FDI has a positive long-run effect on the markups of targeted firms, but this result is limited to R&D-intensive sectors. They attribute this increase in markups to cost savings arising from improved efficiency after the merger, which embodies the transfer of superior technology and managerial know-how. The authors also find that the results weakly indicate that the foreign presence dampens the markup in the short-run due to enhanced

[8]See Bernard et al. [2012] for a survey related to this literature.

competition.

Muraközy and Russ [2015] utilize Hungarian firm-level data from 1993 to 2007 and find that the markups of foreign-owned firms are higher in general than those of the domestic firms, especially the greenfield FDI firms. Furthermore, they find that the markups of domestic firms are significantly lower in industries where multinationals have a greater technological edge, suggesting that differences in technology and endogenous markups are indispensable dimensions for a heterogeneous firm model with FDI.

Dobbelaere and Kiyota [2018] use Japanese manufacturing firms from 1994 to 2012 to investigate the heterogeneity in product and labor market imperfections across exporters, nonexporters, multinational enterprises (MNEs), and non-MNEs. They find that when controlling for differences in productivity, FDI firms appear to have lower market power in the product market, but higher market power on the labor market from demand side. The authors claim that it could be the case that offshoring increases the substitution between domestic and foreign workers, flattening the labor demand curve. The opposite picture emerges for exporters.

2.3 Multinational Production

According to UNCTAD[9], while real-world GDP grows at a 2.9 percent annual rate and real-world exports grows by 5.9 percent annually from 1997 through 2017, real-world FDI inflows grows by 10 percent over this same period, as presented in **Figure 4**. Part of this phenomenon is due to the rapid expansion of MP. For instance, Bernard et al. [2009] find that in 2000, the top 1 percent of US exporters account for 81 percent of US exports. These superfirms produce in multiple countries and industries, and their activities go way beyond the mere act of selling domestically produced goods to foreign consumers. According to Antràs and Yeaple [2014], roughly 90 percent of US exports and imports flow through multinational firms, with close to a half of US imports transacted within the boundaries of multinational firms rather than across unaffiliated parties. In this subsection, I briefly present the empirical evidence of different modes of MP.

[9] https://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx.

2.3.1 Horizontal FDI

Horizontal FDI, which involves establishing a foreign affiliate to serve customers in the foreign market, is observed when the cost of doing so is smaller than the cost of producing at home and shipping to the destination market. It is well-known in the FDI empirical literature that the bulk of FDI among the developed economies is horizontal FDI. For example, Egger [2008] applies two-stage generalized least squares (GLS) methods on the US outward FDI data in 7 industries and 69 countries over the period 1989–1999, and finds strong support for horizontal FDI. Using firm-level data from the US Bureau of Economic Analysis, Antràs and Yeaple [2014] combine three different sources of data and find that MP occurs primarily among developed countries, and the less developed countries are more likely to be the destination of MP rather than the origination. Ramondo et al. [2016] find that among the foreign affiliate of US MNEs, the median affiliate ships nothing to the rest of the corporation. This shows that, on the one hand, intra-firm trade is concentrated among a small number of large affiliates within large MNEs. On the other hand, horizontal FDI, compared to vertical FDI, seems to capture the role of most US affiliates abroad better. More recently, Garetto et al. [2019] utilize a data set that includes detailed information on the operations of MNEs in the US and their affiliates abroad from 1987 to 2011, and find that almost all affiliates in the data had some horizontal sales when they were first established in the host country.

2.3.2 Vertical FDI

Vertical FDI, which involves establishing a foreign affiliate that produces inputs for, or provides intermediate services associated with, a final product, is an internalization mode where firms take advantage of differences across countries in production costs or availability of specific factors and inputs. Yeaple [2003] uses the Benchmark Survey of 1994, which covers 39 countries and 50 BEA manufacturing industries, to investigate the structure of the US outward FDI. He finds that in industries with high skilled-labor intensities, US MNEs have a relative preference over skilled-labor-abundant countries, whereas in sectors with low skilled-labor intensities US MNEs have a relative preference over the skill-scarce countries. His results on comparative advantage confirm the existence of vertical FDI and indicate that countries' skilled-labor abundances should be an important determinant of FDI.

While the previous literature thinks that vertical FDI mostly happens between developed and developing countries, using a new firm-level data set provided by Dun & Bradstreet (D&B)[10], Alfaro and Charlton [2009] find the share of vertical FDI is more substantial than commonly thought even between the developed countries. They show that this is due to a significant amount of vertical FDI being misclassified as horizontal FDI. For example, many subsidiaries that supply goods to their parents are located in sectors in which both the input and final goods are in the same two-digit Standard Industrial Classification (SIC) code, and they are counted incorrectly as vertical FDI at four-digit SIC level (See **Figure 5**).

Lately, several empirical studies have shown that there is little, if any, intra-firm trade between the MNEs' parent firms and their vertically related affiliates. Using the US multinational Benchmark Survey of 2004 conducted by BEA, Ramondo et al. [2016] find intra-firm trade clusters among a small number of large affiliates within the MNEs. They also find that the input-output coefficient[11] linking the parent's and affiliate's industries of operation is not related to a corresponding intra-firm flow of goods. Such skewness is also confirmed in French firm-level[12] data, where Berlingieri et al. [2018] report that the median French MNE imports only 9 percent of its transactions from affiliated parties.

2.3.3 Complex FDI, Export-platform FDI and Network FDI

More recently, empirical evidence has suggested the challenge to maintain the two-way division of MP. For example, Hanson et al. [2005] investigate the output composition of US multinationals' foreign affiliates. They find US MNEs focus affiliates on processing imported inputs in countries where wages and trade costs are lower, and markets are smaller (the virtues of vertical FDI). They also appear to focus affiliates on production for local consumers in countries where wages and trade costs are higher and markets are larger (the virtues of horizontal FDI). Therefore, MNE's positioning can be driven by both horizontal and vertical motivations. Here I briefly summarize three additional types of FDI: complex FDI, export-platform FDI, and network FDI.

[10]The dataset includes location, ownership, and detailed sector (at the 4-digit level) for each of more than 650,000 multinational subsidiaries in 400 industries and 90 countries.

[11]A characteristic commonly associated with vertical FDI.

[12]EIIG(Enquête Échanges Internationaux Intragroupe), a single cross section in 1999, which covers information about intrafirm trade of French firms.

Firms whose global organization reflects both types of MP are categorized as complex FDI. Utilizing panel data from US industries and 51 related host countries observed over the 1989–1999 period, Baltagi et al. [2007] utilize the spatial panel data generalized moments (GM) estimator and find that the linkage between host countries is positively related to the goods traded by MNEs, but negatively related to bilateral trade costs. This confirms the importance of third-country effects, providing a justification for the existence of complex FDI.

Export-platform FDI refers to foreign affiliates' production for the purpose of exporting to third countries. Based on a data set with information about US manufacturing affiliates in 39 host countries during the period 1984–2003, Ekholm et al. [2007] explore the compositions of US multinationals' export sales. They find US affiliates in Canada and Mexico concentrate their exports on the US (virtue of vertical FDI), whereas US affiliates in Europe concentrate their exports on third countries (virtue of export-platform FDI).

Network FDI is first introduced by Baldwin and Okubo [2014] to shift the emphasis from the characteristics of parent-affiliate pairs to interactions among the foreign affiliates. Using extensive firm-level information on Japan's foreign affiliates, they find almost all sectors and almost all nations involve some 'vertical-ness' and 'horizontal-ness.' They also find North American affiliates are far more 'horizontal' than those in Asia and Europe. Moreover, using a four-way sales and sourcing split, the authors find a pattern that suggests many affiliates are part of the international production networks, especially in Asia.

3 Heterogeneous Firms with Variable Markups

This section is dedicated to discussing the development of recent theoretical approaches of modeling the firm-level markups. Although there are various ways for economists to introduce variable markups into the heterogeneous firm framework, they can be generally put into three categories. The first approach, and also the most commonly adopted approach, is to deviate from the Dixit-Stiglitz-type preference with the constant elasticity of substitution among all the varieties. The second approach is to deviate from the monopolistic competition assumption. Different types of oligopolistic competition all fit into this category. Lastly, one can also obtain variable markup

through a proper combination of the previous two. The discussion below is organized into three corresponding subsections.

3.1 Demand-side Approach: Deviating from CES Preference

Melitz and Ottaviano [2008] introduce variable markups into Melitz [2003] by using the *quadratic quasi-linear preference* developed by Ottaviano et al. [2002]. Together with the assumption of monopolistic competition, their framework generates a linear demand system and an endogenous markup distribution across firms that respond to the toughness of competition in the market. A tougher environment features more varieties, lower average prices, and tougher selection into heterogeneous producers and exporters. Bigger firms earn higher profits and charge lower markups. In the short-run[13], opening up to trade not only forces the least productive firms to exit and reallocate market shares toward more productive firms, but also produces a reduction of average markup in the economy, highlighting the pro-competitive effects, which are often associated with trade liberalization. However, this pattern might be overturned in the long-run if we allow firm entry over time. For more discussion regarding this issue, please refer to Ding [2020].

Rodriguez-Lopez [2011] incorporates variable markups into Ghironi and Melitz [2005], which is a dynamic open economy macro version of Melitz [2003]. The authors adopt the *translog expenditure function*[14] developed by Bergin and Feenstra [2000]. This preference allows the author to investigate the general equilibrium effect of trade liberalization, which cannot be explored under a nonhomothetic preference (such as quadratic quasi-linear preference). The steady-state version of the model yields a very similar response to trade liberalization as in Melitz and Ottaviano [2008]. The authors then use a dynamic version of the model, together with sticky-wage setting, to demonstrate that firms' decisions on pricing, entry, and exit play an important role in explaining the pass-through and the expenditure-switching effect of exchange rates, thereby successfully resolving the slow and low degree of nominal exchange rate pass-through to consumer import prices among the developed economies[15].

[13]No entry and exit is allowed in this case. The economy features a fixed number of incumbents.

[14]For a dynamic stochastic general equilibrium macro model with application of translog expenditure function, see Bilbiie et al. [2012]. For a more recent application of translog expenditure function, see Feenstra and Weinstein [2017].

[15]For empirical evidence, please see Engel [2002].

Chaney [2008] notices that if productivity is unbounded from above, then the gains from trade in Melitz [2003] come entirely from firms' selection effect. The other two traditional channels, variety effect and pro-competitive effect, are entirely absent. To restore the role of product variety and pro-competitive gains from trade, Feenstra [2018] utilizes the *quadratic mean of order r (QMOR) expenditure function*, which is first introduced by Diewert [1976], to generate variable markups in a heterogeneous firm model with monopolistic competition. The framework encompasses CES, the translog expenditure function, and quadratic quasi-linear preference (without the homogeneous good) as special cases, and generates a tractable and closed-form solution even when the productivity distribution of firms follows a bounded Pareto[16]. A reduction in trade cost leads to a positive impact on the number of varieties, and a pro-competitive effect by reducing the average markup. The author then utilizes US firm-level export data to test the predictions of the model, and shows that product variety and the pro-competitive effect jointly contribute to 75 percent of the welfare gain from trade, whereas firms' selection effect at most contributes to 25 percent of the increase in welfare.

Zhelobodko et al. [2012] develop a more general model of monopolistic competition, with CES preference as a special case. The authors assume that consumers have an *additively separable preference* over the differentiated varieties without specifying the specific functional form of the utility[17]. By focusing on the *relative love for variety (RLV)*(i.e., the elasticity of marginal utility), the authors demonstrate that the market equilibrium responds differently to various degrees of competition: more competing firms, a larger market size, or the combination of both will lead to lower market prices due to the increase in the elasticity of substitution. The authors then embed the preference into Melitz [2003] and show that the cutoff productivity and markup decrease with the size of the market when RLV increases with consumption, even without assuming the productivity distribution of firms[18].

Similarly, Behrens et al. [2014] obtain the variable markup by introducing *additively quasi-separable (AQS) preference*[19] as in Behrens and Murata [2007, 2012]. The framework yields

[16] Helpman et al. [2008] utilize a bounded Pareto distribution to obtain a gravity equation of trade flows that is consistent with the many instances of zero trade volumes between countries. However, similar to the popularity of CES, unbounded Pareto distribution gain its popularity due to its analytical tractability.

[17] For a more recent application of this specification, see Dhingra and Morrow [2019].

[18] In the Melitz-type model, the productivity draw of firms are usually assumed to follow Pareto distribution.

[19] The authors prove that AQS is associated with constant absolute risk aversion (CARA), while multiplicatively quasi-separable (MQS) is associated with constant relative risk aversion (CRRA).

pro-competitive effects, i.e., profit-maximizing prices decreases in the number of competing firms, and a *competitive limit*, i.e., profit-maximizing prices converge to marginal costs when the number of competing firms approaches infinity. Trade integration induces tougher selection for domestic producers, but at the same time makes it easier for exporters to export. The more productive exporters can charge higher markups following liberalization. The exiting varieties (associated with the least productive firms) are compensated by the newly imported varieties, producing more varieties after the integration. For two asymmetric regions, a bilateral trade integration will cause the average markups, varieties, and welfare to converge between the two regions.

Simonovska [2015] employs *nonhomothetic* consumer preference via an utility function that belongs to the *hierarchic-demand* class studied by Jackson [1984]. Variable markups play an important role in this exercise: Due to nonhomothetic preferences, different income levels imply different consumption sets across countries. Since the expenditure shares are nonconstant, the framework generates variable price elasticity of demand for a given variety across countries. More specifically, consumers in countries with higher per-capita income are less responsive to price changes than those with lower per-capita income. For the same variety, although profit-maximizing firms can set higher prices in a richer market, they also face tougher competition in those markets, which serves to push down the markups. Using a unique data set from online retailers, the author then estimates the elasticity of price with respect to per-capita income, and finds per-capita income differences account for a third of the observed cross-country differences in tradable goods' prices.

Similar to Simonovska [2015], Bertoletti et al. [2018] introduce variable markup through *non-homothetic indirectly additive preference* into a heterogeneous firms trade model with monopolistic competition. The novel feature of this framework is that it generates markups that are increasing not only in terms of firm's productivity but also in terms of destination market's per-capita income, which cannot be captured by a homothetic preference, such as in Melitz and Ottaviano [2008]. In contrast to Simonovska [2015], markups in this framework are independent of market size, which is consistent with empirical evidence. The model also generates novel insights into the extensive margin of trade: The extensive margin of trade is increasing in destination per-capita GDP, neutral in destination population, and falling in the trade cost to the destination market. Combined with the fact that cost pass-through is falling in firm productivity, the authors then demonstrate that the welfare gain from trade is much smaller in this framework than the commonly employed frameworks.

Edmond et al. [2018] study the welfare implication of variable markups[20] in a dynamic model with heterogeneous firms engaging in monopolistic competition. The authors adopt the *Klenow-Willis* specification: Input bundles are assembled into final goods via the *Kimball aggregator*, as in Kimball [1995]. In this economy, more productive firms are larger in size and face less elastic demand, which allows them to charge higher markups. In equilibrium, this group of firms grow at the expense of less productive firms, driving up the aggregate markup and reducing the aggregate labor share. The authors then quantitatively demonstrate the distortion at aggregate markup counts for three-quarters of the total welfare cost of variable markups, while the misallocation of production factors counts for one quarter. The welfare cost due to inefficient entry is almost negligible in this framework, which is different from the implications in Bilbiie et al. [2016].

Last but not least, two recent papers provide a general demand structure that encompasses many of the previous approaches. Parenti et al. [2017] impose *Fréchet differentiability* on a general symmetric utility function and derive a parsimonious micro-foundation for the variable elasticity of substitution. The framework is able to predict the market's response to exogenous shocks (such as market size, productivity shocks) through the lens of elasticity of substitution. The predictions are in line with salient features established in the industrial organization literature. When considering firm heterogeneity à la Melitz, the model produces *cost-specific elasticity of substitution*, which means firms that differ in productivity sell varieties that differ in their degree of product differentiation. As a result, the cutoff cost in the market outcome is no longer constant, but varies with the market size. The authors then emphasize that the literature should not take for granted that trade liberalization is always productivity enhancing, as suggested in the basic Melitz model.

Arkolakis et al. [2018] derive a general demand system that encompasses all of the previous demand structures[21], and study the welfare gains from trade when firms engage in monopolistic competition. The framework allows the authors to pin down the welfare gains from trade using three sufficient statistics: share of expenditure on domestic goods, elasticity of imports with respect to variable trade costs (i.e., trade elasticity), and an additional statistic(η) that includes the average

[20]For a systematic study of welfare implication of variable markups in a DSGE model under different specifications, see Bilbiie et al. [2016]. The authors investigate four different kinds of specification: CES, generalized love for variety, translog expenditure function, and exponential love of variety.

[21]More specifically, it should be noted that their demand system covers translog expenditure function, QMOR, additively separable preference, additively quasi-separable, nonhomothetic preference, Klenow-Willis specification. With a slight generalization, the system can also cover quadratic quasi-linear and nonhomothetic indirectly additive preference.

elasticity of markups with respect to firm productivity. When the preference is homothetic, the welfare loss from trade gets passed through to domestic consumers exactly equal to the welfare gain from the reduction in misallocation (due to the trade-induced competition). In this case, $\eta = 0$, and the welfare gain from trade is the same as those in models with CES utility. When preferences are nonhomothetic, the first force dominates the second force, i.e., $\eta > 0$, indicating that the welfare gains from trade are lower than those predicted by models with CES preference. Using highly disaggregated data on bilateral US merchandise imports, the authors structurally estimate the value of η and find it to be slightly above zero (0.06). Hence, they conclude that the pro-competitive effects of trade are elusive.

3.2 Supply-side Approach: Deviating from Monopolistic Competition

3.2.1 Bertrand Competition

Around the same time as Melitz [2003], Bernard et al. [2003] also introduce firm heterogeneity into the trade literature. In contrast to the combination of monopolistic competition and CES preference, the authors adapt a Ricardian model with firm-level comparative advantage into Eaton and Kortum [2002]. Variable markup is generated through firms' Bertrand competition, where goods within an industry are perfect substitutes. Firms' efficiency level follows a Pareto distribution. More efficient producers also tend to have a greater cost advantage over their closest competitor, charge lower prices, set higher markups, sell more, and are more likely to beat rivals in foreign markets. The framework delivers an endogenous distribution of markups that captures the producer-level stylized facts at least qualitatively. Similarly to Melitz [2003], the model predicts aggregate productivity gain due to trade-induced competition. However, due to the specific assumptions in their model, the distribution of markups is invariant to country characteristics and to geographic barriers.

De Blas and Russ [2015] provide a transparent generalization of Bernard et al. [2003] through two modifications: (i) Firms' efficiency level draw follows Fréchet distribution instead of Pareto distribution. (ii) There are finite number of rivals instead of infinite number of rivals. These modifications produce a distribution of markups that, compared to the one in Bernard et al. [2003], preserves the characteristics of the market structure, which are sensitive to the degree of trade openness and differences in technological development across countries. For example, trade

liberalization reduces the markups that domestic firms can charge on domestic sales. Moreover, the model predicts that if the trade barrier meets a specific condition, bilateral trade liberalization can create an anti-competitive effect, increasing market power for exporters on average, while switching from bilateral to multilateral trade agreements can generate a pro-competitive effect, reducing average markups among exporters.

In contrast, Eaton et al. [2012] relax the *continuum* assumption in the standard monopolistic competition model. Building on Melitz [2003] and Eaton et al. [2011], the authors assume that a finite number of firms draw their productivities from a Pareto distribution, meaning that in any realization of the data there may be no firms from country i that have sufficiently high productivity to supply destination market j in industry k. The framework also features Bertrand competition and endogenous entry. The simulated model can reproduce the prevalence of zeros in aggregate trade flows, which cannot be achieved in the basic Melitz model.

3.2.2 Cournot Competition

Instead of competing in prices, Atkeson and Burstein [2008] introduce variable markups through quantity competition á la Cournot[22] into a nested CES demand system as in Helpman and Krugman [1985]. Contrary to Melitz [2003] and Bernard et al. [2003], where there are an infinite number of firms in an industry, Atkeson and Burstein [2008] assume that there are a finite number of firms competing with each other. This setup allows firm-level markup to be positively correlated with the firm's market share, and elasticity of demand to be negatively correlated with the firm's market share. The authors find that the separated markets through trade costs and Cournot competition with variable markups lead to pricing-to-market behavior[23] in those separate markets. These two elements are essential for generating deviations from relative PPP at the aggregate level. This approach has gained increasing popularity in the literature–for example, see Amiti et al. [2014], Edmond et al. [2015], and Gaubert and Itskhoki [2018]. More recently, however, Amiti et al. [2019] find that this type of strategic complementarity and variation of markup at the micro level does not

[22]It should be noted that when the goods are perfect substitutes, and fixed export cost equals to zero, and if the firms engage in price competition, the model resembles Bernard et al. [2003], see Atkeson and Burstein [2007] for more details.

[23]More specifically, in their model, the authors find pricing-to-market at the level of the aggregate price indices only because the pricing practices of the large firms in the model dominate the pricing practices of the small firms. If there is no cost dispersion across firms and all firms export, the authors find no pricing-to-market at the aggregate price level.

necessarily explain the macro-level markup adjustment.

3.3 Supply-side and Demand-side Approach

Kokovin et al. [2017] departs from Melitz [2003] in both dimensions: (i) On the demand side, the authors adopt quadratic quasi-linear preference. (ii) On the supply side, instead of a continuum of monopolistically competitive single-product firms, the authors allow a few large-scale firms to manipulate the market and a continuum of small firms to treat market conditions as given. The authors find that if the demands faced by firms are *single-aggregate*[24],then despite being endowed with the capability to manipulate the market strategically, a large firm may find it rational to disregard this ability and imitate the behavior of small firms. In this case, the market structure is observationally equivalent to monopolistic competition. The results show that when considering the interaction between small firms and large firms, consumers' preferences, rather than producers' costs, play a more significant role in determining the market structure.

More recently, Parenti [2018] studies the impact of trade liberalization with a similar 'mixed' market structure in a partial equilibrium framework. Consumers have a quadratic quasi-linear preference. A small firm produces one variety, while a large firm produces a mass of varieties. The structure of the economy is featured with large firms choosing their product range and outputs, and small firms choosing to enter the market and their outputs simultaneously. In equilibrium, a reduction in trade costs selects the biggest firms into export but also drives out the least productive small firms. The overall number of varieties available to consumers increases, but so does the average price in the industry. This impact results in a decrease in consumer surplus since large firms absorb the decrease in trade costs by charging higher markups. Producer surplus increases, but the total welfare impact is ambiguous.

[24]It means, all the cross-effects in the demand system are captured by a scalar function whose value plays the role of a market aggregate. Many preferences that are mentioned in section 1.3.1 satisfy this condition.

4 Heterogeneous Firms with Multinational Production

This section briefly surveys the state of the literature on multinational firms, with an emphasis on firm heterogeneity. The discussion here maps the empirical evidence presented in Section 2.3. First, I discuss the theoretical frameworks that capture the within-industry motive of cross-border investment of multinational firms. I make a distinction between foreign direct investment and cross-border mergers and acquisitions whenever possible. Second, I investigate the models that study the intra-industry firm-level decision-making across countries through the lens of the organizational mode. Lastly, I discuss the literature that is built on the fact that different forms of integration strategy coexist even within the same industry. However, due to the complexity of this approach, most of the work in this line of research focuses on the quantitative implications.

4.1 Horizontal FDI

Helpman et al. [2004] introduce Melitz-type firm heterogeneity into the traditional horizontal FDI model as in Brainard [1993]. When a home firm considers accessing a foreign market, a *proximity-concentration tradeoff* appears: A firm can export, which entails high variable trade costs and low fixed costs, or it can choose FDI, which entails low variable costs and high fixed costs. In equilibrium, only the most productive firms choose to serve the foreign market through FDI, less productive firms choose to export, and the least productive firms only serve the domestic market. The authors then test the predictions of the model and find that country-specific trade costs have a strong negative effect on export sales relative to FDI. Firm-level heterogeneity provides new insight regarding the relative export and FDI sales: More heterogeneity leads to significantly more FDI sales relative to export sales.

Motivated by the empirical observation that the majority of FDI takes the form of cross-border mergers and acquisitions (M&A), Nocke and Yeaple [2007] introduce cross-border M&A into Helpman et al. [2004]. Firms are heterogeneous in their capabilities, and these capabilities differ in their degree of international mobility. Cross-border M&A is motivated by a firm's desire to exploit complementarities between the local firm's country-specific capability and the acquiring firm's intangible technological advantage. The model suggests that when capabilities become relatively less mobile internationally, cross-border M&A becomes the more popular mode of entry

into foreign markets. Furthermore, firm heterogeneity is a crucial determinant of firms' international organization. For example, in industries where firms differ mainly in their mobile capabilities, the most efficient firms will engage in cross-border M&A, while in industries where firms differ mainly in their country-specific nonmobile capabilities, cross-border M&A will involve the least efficient active firms.

In contrast, de Blas and Russ [2013] study the impact of FDI and cross-border M&A on firms' markup in a generalized version of Bernard et al. [2003][25]. In such a Ricardian model, the authors prove that takeovers by foreign firms increase the technological edge of the acquired firms, allowing the target firms to increase their markup, thereby increasing the average markup in the economy. When trade is costly, the authors utilize the model to demonstrate a firm's motive for taking over a foreign rival to increase its market power in either the foreign market or the home market. Both channels will result in higher prices compared to a world without FDI. Similarly, Muraközy and Russ [2015] use the empirical methods developed by De Loecker and Warzynski [2012] and find that foreign-owned firms charge significantly higher markups than domestic firms (the gap is more evident in industries where MNEs have a technological advantage) in Hungary. The authors then try to use a Ricardian model with variable markups and heterogeneous firms to explain the empirical observation. The framework not only provides an analytical distribution for market shares and markups when goods are imperfect substitutes, but also highlights the role of technology in generating market power. The model explains about half of the observed multinational markup premium, calling for more research on the relation between MNEs and market power.

More lately, Gumpert et al. [2017] use a dynamic[26] version of Helpman et al. [2004] and study the life cycle dynamics of exports and FDI. Assuming that firm productivity evolves according to a Markov process, and MNEs face a sunk entry cost, the model can reproduce salient features of the firm-level data from France and Norway. The authors then use the model to conduct counterfactual analysis and find that the presence of MNEs is crucial for exporters' life cycle dynamics. On the one hand, trade liberalization in a model with MNEs would increase exporters' sales, and decrease their exit rates, but it would barely change their life cycle behavior in the model without MNEs. On the

[25]Essentially, the model is a variant of De Blas and Russ [2015].

[26]There is a growing literature emphasizing on the micro-founded (trade-founded) open economy macro framework, for example, see Ghironi and Melitz [2005] and Alessandria and Choi [2007] for models without MNEs, see Ramondo and Rappoport [2010], Ramondo et al. [2013] and Zlate [2016] for models with MNEs.

other hand, export growth is higher, and the exit rate is lower in the model because lower trade costs induce fast-growing exporters to remain exporters and fast-growing MNEs to switch to exports. The study suggests that ignoring MNEs will result in biased quantitative implications for dynamic trade models.

4.2 Vertical FDI

There are two main streams of literature that focus on the firm-level decision regarding vertical specialization: (i) location decisions, and (ii) organizational mode. Here I briefly survey the literature on the latter[27], since this branch has been rapidly growing following the Melitz revolution.

Motivated by the growing empirical evidence on US firms' outsourcing, Antràs and Helpman [2004] combine the within-sectoral heterogeneity, as in Melitz [2003], with the firms' organizational structure, as in Antràs [2003], into a North-South trade framework to study firms' global sourcing strategies. Firms in the North develop differentiated products, and then they decide whether to integrate the production of intermediates or outsource them. In choosing between a foreign and a domestic supplier of intermediates, a final-product producer faces a trade-off between a lower variable cost in the South and a lower fixed cost in the North. When it comes to choosing between vertical integration and outsourcing, a final-good producer is trading off between an ownership advantage from vertical integration and a better incentive for the independent supplier of intermediates. In equilibrium, low-productivity firms acquire intermediate inputs in the North, whereas high-productivity firms acquire them in the South. Among firms that source their inputs domestically, the low-productivity firms outsource, whereas the high-productivity firms insource. In sectors with a low intensity of headquarters, no firm integrates, high-productivity firms outsource abroad, and low-productivity firms outsource from home. The model suggests that a reduction in the costs of foreign sourcing, on the one hand, raises the fraction of firms that import intermediate inputs, on the other hand, raises the fraction of firms that outsource in each one of the countries. As a consequence, the volume of arm's-length trade increases relative to intra-firm trade.

Antràs and Helpman [2009] incorporate varying degrees of contractual frictions into Antràs and Helpman [2004]. The authors then use the framework to study the effects of variations in

[27]It has been a challenge for the first branch of literature to incorporate firm heterogeneity. Interested readers can refer to Antràs et al. [2017] for a recent breakthrough.

country and industry characteristics on the relative prevalence of firms' organizational forms. In the model, heterogeneous firms decide whether to integrate or outsource intermediate inputs and in which countries to source the inputs. Contractual frictions exist both in an integrated firm and in an arm's-length relationship, i.e., final-good producers and their suppliers make relationship-specific investments that are only partially contractible. The model predicts that better contracting in the South, which tends to raise offshoring, may reduce the relative prevalence of FDI if the institutional improvement affects disproportionately the contractibility of inputs provided by the final-good producer. Moreover, better contractibility in the South may decrease the popularity of foreign outsourcing when the improvements are tilted toward inputs provided by suppliers rather than the final-good producers.

Since then, the global sourcing framework has been extended to incorporate other factors that are also considered as factors for global productions. For instance, Carluccio and Fally [2012] demonstrate that the financial restrictions faced by foreign suppliers could cause firms to integrate these suppliers, especially in a condition when the inputs are noncontractible and complex. Carballo [2018] investigates how the uncertainty over final goods' demand would impact MNEss' organizational behavior. As the micro-level data become increasingly available, this line of research remains to be a promising topic in the future.

4.3 Complex FDI

Grossman et al. [2006] investigate the optimal integration strategy in a three-country (two Northern countries and one Southern country) model where firms possess heterogeneous productivity. In this model, firms that are headquartered in a Northern country supply differentiated final goods to all three countries. Each firm must produce an intermediate input and assemble the inputs so as to generate a final product. The production of intermediate input and assembly can take place in any of the three locations. Foreign activities, which either produce an intermediate input or assemble the inputs, incur a fixed cost, while transport intermediate inputs and final product incur an iceberg-type shipping cost. The optimal integration strategies depend on the fixed costs of foreign subsidiaries, the transportation cost of intermediate and final goods, and the North-South factor price differences. In equilibrium, firms with different productivity levels will choose different organizational forms,

confirming the empirical evidence in Hanson et al. [2005] that different forms of integration strategy coexist within the same industry.

Building on Helpman et al. [2004], Irarrazabal et al. [2013] allow a foreign affiliate to combine local inputs with inputs imported from the headquarters to produce final products. The model therefore generates intra-firm trade in Helpman et al. [2004]. The authors also add firm- and destination-specific taste and entry shocks in the spirit of Eaton et al. [2011]. These two modifications deliver a firm-level gravity equation for export and MP. Utilizing a novel data set from the Norwegian manufacturing sector that provides firm-level observation on both export and MP, the authors are able to infer the level of intra-firm trade. The key parameters of the model are structurally estimated using the maximum likelihood method. Their point estimate of the affiliate's cost share related to purchases from the headquarters is 0.9, indicating strong vertical linkages, as well as other mechanisms that dampen firms' MP as trade costs increase. The counterfactual exercise indicates that impeding MP has substantial effects on trade flows and domestic employment.

Using plant-level data from Indonesia, Rodrigue [2014] finds foreign-owned plants in Indonesia exporting heavily to markets from which most FDI is sourced. The author then builds a dynamic trade model with both export and MP based on Helpman et al. [2004]. In this model, MNEs can set up plants solely to access the foreign market or export back to the home country. A structurally estimated model is then used to assess the influence of policy on firm-level decisions and evaluate the impact of economic policy on aggregate productivity, exports, and MP. According to the counterfactual analysis, if MP is prohibited in Indonesia, aggregate manufacturing productivity will fall by 19.6 percent. Similarly, if Indonesia is cut off from trade, aggregate manufacturing productivity will fall by 8.5 percent. The results suggest that the incentives for trade or MP significantly affect the likelihood of the other activity, and their interaction at the firm level has significant aggregate policy implications.

Ramondo and Rodríguez-Clare [2013] introduce MP into a quantitative Eaton and Kortum [2002]-type Ricardian trade model by allowing a country's technology to be used for production abroad. The model features both tradable intermediate goods and nontradable final-consumption goods. To compete in a foreign final-good's market, MP is the only channel. For intermediate goods, firms can access foreign markets either through export or MP. The framework also allows a firm to use a third country as an export platform. Therefore, the model features horizontal FDI, vertical FDI,

and export-platform FDI. The authors then calibrate the model based on bilateral trade and MP data for OECD countries, as well as data on intra-firm trade flows for US and foreign MNEs operating in the US. Since the model captures the fact that trade can facilitate MP (by allowing MNEs' foreign affiliates to import inputs from their home country), the quantitative exercise suggests that the gains from openness (allowing both trade and MP) are much larger than the gains from trade.

Assuming firms can produce a continuum of products and treating firms' product-location-specific productivities as a random variable, Tintelnot [2017] develops a tractable multi-country general equilibrium model that allows MNEs to engage in export-platform sales while maintaining the fixed-cost assumption of establishing foreign plants. Using German firm-level data, the author shows that both differences in variable production costs across countries and the fixed costs of establishing foreign subsidiaries are important barriers to foreign production for German MNEs. The model is then calibrated using trade and MNE data for 12 European and North American countries. The analysis of counterfactual exercises shows that MNEs play a crucial role in the transmission of technology improvements to foreign countries. The pending Canada-EU trade and investment agreement has a third-country effect, which would be missed if MNEs are excluded or modeled in a more restrictive way.

5 Trade Policy with Heterogeneous Firms

While models of firm heterogeneity have transformed the way economists think about international trade and multinational firms, these models have made surprisingly little contribution to our understanding of trade policy. This section surveys the existing work on the normative side of the international trade literature when heterogeneous firms select into export. The discussion is organized into three areas. The first strand of literature studies the optimal trade policy in the basic Melitz [2003] framework while allowing the number of industries and country size to vary. I also include some recent papers that study the effects of trade policy uncertainty on the aggregate economy, which are typically built on the dynamic version of Melitz [2003]. The second group of papers study the optimal trade policy when heterogeneous firms charge variable markups. Both unilateral and multilateral trade policy are discussed. The last category of papers studies the trade

policy implications when heterogeneous firms can access foreign markets either through export or FDI. To the best of my knowledge, there are only two papers falling into this category: one is built on Helpman et al. [2004], and the other is built on Antràs and Helpman [2004].

5.1 Trade Policy with Firm Heterogeneity

Baldwin and Forslid [2006] study the impact of a reduction in the variable and fixed cost of trade in a two-country, two-sector Melitz [2003] model[28]. When the two countries are symmetric in size, the difference in the fixed cost of domestic production and export could lead to a decrease in the number of varieties following trade liberalization in variable costs. When countries are asymmetric, such an *anti-variety effect* only happens in the large economies when the initial protection level is high. However, when trade liberalization takes the form of a reduction in the fixed cost of export, the authors find a *pro-variety effect* when the technical barrier to trade is sufficiently low. Despite the heterogeneous impacts on varieties, the authors find that trade liberalization always leads to welfare gains in the model.

Demidova and Rodriguez-Clare [2009] analyze the optimal trade policy for a small open economy[29] in a Melitz-type model. Two types of distortion exist in such an economy: There is too little spending on the domestic varieties relative to the social optimum due to markup distortion between home and foreign varieties, and there are too few foreign varieties in the market outcome compared to the social optimum due to consumption externality[30]. Since the first distortion prevails over the second, the first-best allocation can be obtained by a consumption subsidy, an import tariff, or an export tax. This finding challenges the view that export subsidy is optimal in the Melitz-type framework. As far as welfare is concerned, export subsidy causes a negative impact on terms of trade, variety, and curvature[31], which dominates the productivity gain from reallocation.

Ossa [2011] develops a novel theory of GATT/WTO negotiations based on the Krugman [1980] model. In a separate appendix, the author demonstrates that all the results can be derived in a variant of the Melitz [2003] model. The modified Melitz model features two sectors: a manufacturing sector

[28] Alternatively, it's a two-country Helpman et al. [2004] without FDI.

[29] Which means the economy takes as given the price of imports and the demand schedules for its exports.

[30] More specifically, domestic consumers are ignoring the positive impact on aggregate productivity of foreign varieties.

[31] This is a term in the welfare decomposition that captures heterogeneity across varieties.

that produces differentiated varieties, and a numéraire good sector that produces homogeneous goods. In this framework, import tariffs allow the country to attract additional firm entry into the manufacturing sector. If the tariff-induced long-run *delocation effect* dominates the direct effect of the tariff on the price index, then consumers' welfare will increase. The framework generates reasonable noncooperative tariffs and moderate gains from GATT/WTO negotiations.

Felbermayr et al. [2013] extend Demidova and Rodriguez-Clare [2009]'s exercise to a two-country large-economy case. The model features markup and entry distortion, as well as a terms of trade effect. They find that the optimal Nash tariff increases in the relative country size, the relative average productivity, and the degree of firm-level productivity dispersion, but decreases in the nontariff trade barriers. The authors then calibrate and simulate the model to demonstrate the importance of productivity dispersion for the size of the optimal tariff quantitatively. The results indicate that the declining trade cost, higher firm-size dispersion, and country-size convergence have magnified the benefit of the WTO over time.

Haaland and Venables [2016] study both domestic taxes and trade policies for a small open economy containing a flexible monopolistically competitive sector[32] in which firms may be heterogeneous in their productivity level. The welfare gain of policy arises from the interaction of two dimensions: the distortion of quantity due to the monopoly power in the monopolistically competitive sector, and the distortion of price due to the terms of trade effect. The authors find the optimal policy is to subsidize domestic sales, and in some cases, a positive tariff on imports. These policies can generate an expenditure-switching effect toward the monopolistically competitive sector that eventually increases the number of varieties and improves the terms of trade. However, besides the fact that the foreign market's reaction toward policy becoming more price-elastic, the authors find firm heterogeneity does not create any new qualitative implication for policy interventions.

Recently, a small but increasing literature has started to emphasize the uncertain nature of trade policy and its impact on global integration. Handley and Limao [2015] first develop a dynamic heterogeneous firm model to examine the impact of trade policy uncertainty on a firm's investment and export behavior. Building on Dixit [1989]'s insight that price uncertainty creates an option

[32]Flexible here means the size of the monopolistically competitive sector can expand or contract. The flexibility depends on the elasticity of labor supply in the economy. If the labor supply is perfectly inelastic, the economy resembles a one-sector economy as in Demidova and Rodriguez-Clare [2009]. If the labor supply is perfectly elastic, the economy resembles a two-sector economy as in Baldwin and Forslid [2006].

value of waiting before making sunk cost entry investments, the authors demonstrate that trade agreement can reduce trade policy uncertainty and affect export investments. Focusing on the case of Portugal's accession to the European Community, the authors find strong evidence that trade agreements eliminate trade policy uncertainty. The counterfactual analysis shows that the rapid growth of firm entry and export growth is primarily due to the reduction of trade policy uncertainty, not the reduced applied tariff. Similarly, Handley [2014] finds that binding tariff commitments can also reduce trade policy uncertainty and increase firm entry into the export market.

Last but not least, some recent papers have emphasized the macroeconomic consequences of trade policy. Utilizing a primal approach and general Lagrange multiplier methods to characterize optimal wedges, Costinot et al. [2016] investigate the optimal trade policy both at the micro and macro level in a generalized version of the Melitz [2003] model. At the micro level, a welfare-maximizing government should impose firm-level import taxes that discriminate against the most productive foreign exporters. The optimal export taxes can be either discriminative or in favor of the most profitable domestic exporters. At the macro level, terms of trade affect the total level of trade taxes. The more home's terms of trade deteriorates with increases in exports or imports, the larger the trade taxes it should impose. Utilizing the number of anti-dumping investigations as a measure of protectionism, Barattieri et al. [2018] explore the effectiveness of trade policy at the macroeconomic level. For small open economies, both their empirical and model-based[33] simulation analyses show that protectionism is costly even when used temporarily. The protectionist measures have effects similar to an unfavorable supply-side shock that causes inflation to rise and real economic activity to fall.

Two very recent papers have also emphasized the macroeconomic impact due to trade policy uncertainty. Steinberg [2019] uses a three-country, dynamic general equilibrium model with heterogeneous firms, input-output production linkages, and stochastic trade costs to study the quantitative implication of trade policy uncertainty following Brexit. The model suggests that Brexit will have a substantial impact on the British economy, especially in the long-run. Both trade flows with the European Union and consumer welfare will fall significantly over time, but trade policy uncertainty has little impact on the macroeconomy. Focusing on newspaper coverage, firms'

[33]The model combines the essential ingredients from the international business cycle literature (such as endogenous physical capital accumulation and nominal rigidities) with a full endogenous trade structure similar to Ghironi and Melitz [2005].

earnings conference calls, and aggregate data on tariff rates in the US economy, Caldara et al. [2019] find empirical evidence that the increase in trade policy uncertainty reduces investment and real economic activity at both the firm and the aggregate level. Using a two-country general equilibrium model with nominal rigidities and firms' endogenous export participation, the authors find that both news shocks and increased uncertainty reduce investment and output in the model, although the impact of uncertainty is quantitatively limited. Their framework highlights the role of price rigidity and fixed export costs as important channels that magnify the effects of trade policy uncertainty.

5.2 Trade Policy with Firm Heterogeneity and Variable Markups

Demidova [2017] takes the first step to study trade policy with firm heterogeneity and variable markups. Following Arkolakis [2008], the author drops the outside good assumption in the Melitz and Ottaviano [2008] model to investigate the general equilibrium effect of trade policy. On the one hand, the author finds, regardless of country size, that the optimal import tariff is strictly positive from a unilateral perspective. On the other hand, the reduction of iceberg-type trade costs is always welfare-improving, making free trade in this dimension socially optimal. Variable markups, in the presence of firm heterogeneity, create a negative pro-competitive effect on the economy due to the misallocation of resources[34]. Consequently, the welfare gain from trade following liberalization[35] is dampened due to the presence of variable markups.

Traditional frameworks cannot explain WTO's strong restrictions on export subsidy. Introducing ad-valorem tariffs and export subsidies into the original Melitz and Ottaviano [2008] model, Bagwell and Lee [2015] explore the implications that a heterogeneous firm model may provide for the use and treatment of export subsidies in the WTO. The authors find that starting from free trade, a country can gain by imposing either a small import tariff, a small export subsidy, or a combination of the two. However, these policies are all 'beggar-thy-neighbor' interventions. In particular, the home country is incentivized to introduce a positive export subsidy when transportation costs are low and productivity dispersion is great. This finding provides a partial explanation for the WTO's prohibition of export subsidy. The paper finds that free trade is not efficient in general, and that

[34]More productive firms do not pass on their cost advantage to consumers and they end up selling quantities below the socially optimal level.

[35]Both in the form of tariff reduction and iceberg-type trade cost reduction.

the import tariff is higher than the export subsidy in a symmetric Nash equilibrium. Furthermore, starting from a symmetric Nash equilibrium, both countries can gain by reducing trade policy interventions.

More recently, Nocco et al. [2019] study multilateral trade policy in an environment in which countries differ in market access and technology, and firms differ in productivity and market power. Utilizing a multi-country version of the Melitz and Ottaviano [2008] model, the authors find that the free market provides an inefficiently high degree of welfare inequality between advantaged and disadvantaged countries[36] if their differences in terms of market size, technology level, and geography are sufficiently large. Therefore, multilateral trade policies, such as increasing the sales of low-cost firms (especially to disadvantaged countries), decreasing the sales of high-cost firms (especially to disadvantaged countries), and reducing firm entry (especially in disadvantaged countries), are socially welfare-improving.

5.3 Trade Policy with Firm Heterogeneity and Multinational Production

Contrary to previous strategic tariff literature, where in equilibrium all foreign firms are either multinationals or exporters, Cole and Davies [2011] utilize the framework of Helpman et al. [2004] and find an equilibrium condition in which both pure exporters and multinationals coexist. The authors then study optimal tariffs in such an environment and find that the optimal tariff for the social planner is negative, i.e. trade should be subsidized. This is because encouraging trade can foster competition and eliminate the least productive firms in the economy, thereby boosting the average productivity and increasing social welfare. The authors also find that Nash tariffs are higher than the socially optimal tariff, promoting the existence of low-productivity firms, and creating a new source of inefficiency due to tariff competition. When FDI is an option for firms, the Nash tariff is lower. Therefore, FDI is welfare-improving since it can mitigate tariff competition.

Building on the framework introduced by Bernard et al. [2003], de Blas and Russ [2013] investigate the market-power motive of FDI within the heterogeneous firm framework. The authors find that under Bertrand competition, FDI, either in the form of mergers and acquisitions or greenfield investment, can increase markups through the channel of technology transfer. When there is no

[36] In their paper, the authors define advantaged countries as countries with bigger market size, better technology in terms of lower innovation and production cost, and better geography in terms of closer proximity to other countries.

trade in goods, the increased markup is always outweighed by the efficiency gains arising from technology transfer, causing prices to stay the same in the source country but fall in the host country. When trade is costly[37], the authors demonstrate a motive for taking over a foreign rival to increase a firm's market power either in the source country or in the host country. Both of these two cases will cause prices to increase compared to the world with trade but without FDI. However, besides providing an additional incentive for multinational firms to pursue tariff-jumping FDI, the study does not generate any specific trade policy implication.

Díez [2014] presents two novel stylized facts: (i) the US intra-firm import volume depends positively on the US tariff level, and (ii) the US intra-firm import volume depends negatively on the foreign tariff level. He then extends the Antràs and Helpman [2004] model to include incomplete contracts to explain these empirical observations. A tariff imposed by the North on final goods will decrease the market share of offshoring firms and decrease the relative market share of outsourcing firms versus vertically integrated firms in both countries. A tariff imposed by the South on final goods will increase the market shares of offshoring and outsourcing firms in both countries. The model also predicts that if offshoring increases, then Northern imports will increase, and if there is relatively more vertical integration than outsourcing, there will be relatively more intra-firm trade and less arm's-length trade. These predictions are also well supported by robust empirical evidence using data from the Foreign Trade Division of the US. Census Bureau from 2000 to 2009.

6 Discussion and Concluding Remarks

This paper has served two main purposes. First, I present in great detail various stylized facts regarding firm heterogeneity, firm-level markups, and the global structure of MP. In doing so, I present a few observations that are inconsistent with the basic heterogeneous firm trade models. In particular, (i) firms with different productivity levels charge different markups even within the same industry, which clearly contrasts with the constant markups generated in the standard Melitz-type model, and (ii) the growth of MP has been much more phenomenal than the growth of exports. However, most of the trade models do not consider the joint decision made by firms concerning

[37]The authors use the term tariff-jumping in the paper, but they do not model tariff specifically. Rather, they simply treat all the variable trade costs equivalent to tariff.

export and foreign direct investment.

I then survey the heterogeneous firm trade and trade policy literature with a focus on variable markups and MP. I detect two prevailing trends in the literature that are inconsistent with the current micro-level empirical evidence. First, heterogeneous firms trade models with variable markups usually assume that firms only access the foreign market through export, while empirically, they have been widely engaging with foreign direct investment. Secondly, models that allow firms to engage in both export and foreign direct investment are introduced under the combination of CES preference and monopolistic competition, which implies that all the firms will charge identical and constant markups. This is far from what we have observed through the firm-level data.

To the best of my knowledge, Ding [2020] makes the first attempt to develop a framework that overcomes the previous two discrepancies between theory and data. The model features firm heterogeneity, variable markups, and MP. To introduce variable markup, the author adopt a particular preference (quadratic quasi-linear preference) as described in the demand-side approach (see Section 3.1). One should note that other types of preferences can also generate similar results, but they generally suffer from analytical transparency. For example, all the theoretical derivation in Ding [2020] can also be carried out with translog preference or constant absolute risk aversion (CARA) preference, but the closed-form solutions under these preferences always carry a Lambert function[38], and are not as transparent as the results derived under quadratic quasi-linear preference. Regarding the supply side, it is a well-known challenge in the literature to work with models of oligopolistic competition, both analytically and computationally[39].

Despite the choice of approach to introduce variable markups, what makes Ding [2020] different from the work mentioned in Section 3 is the addition of MP. More specifically, the author introduce the horizontal FDI, which is the dominant form of foreign direct investment among the developed economies, into the Melitz and Ottaviano [2008] framework. In the trade policy literature, this is the first study to introduce MP into heterogeneous firms with variable markups. In particular, the presence of MP and its interactions with variable markups have never been studied in the heterogeneous firm framework before.

Ding [2020] builds on the most popular framework in the heterogeneous firm literature with

[38]The byproduct of a mathematical tool utilized to solve transcendental equations.

[39]For more discussions, please refer to Edmond et al. [2018].

MP, Helpman et al. [2004] (henceforth, HMY). Yet there are three notable differences in this study. First, the original HMY model utilizes quasi-linear preference, which, together with monopolistic competition, generates constant markups for all firms. This implication is clearly at odds with the empirical evidence on variable markups presented in Section 2.2. I adapt the quadratic quasi-linear preference into the HMY framework to overcome this discrepancy. Second, the original HMY framework treats trade costs as an iceberg-type cost, and treats trade liberalization as an exogenous decrease of the iceberg cost. This type of modeling approach cannot capture the various types of trade costs faced by exporters. In Ding [2020], the author introduces the *ad valorem* tariff into the modified HMY model. This approach not only is a more realistic way to model trade costs but also enables one to study the policy implications of strategic tariffs. Third, in HMY, besides the iceberg-type trade costs, exporters and multinational firms are differentiated by the difference in the fixed cost. In other words, besides the differences coming from the productivity draw, there is no other difference between an exporter and a multinational firm in terms of their marginal cost of production. This assumption is not in line with the empirical evidence on the costs faced by multinational firms[40]. In Ding [2020], the author introduces an iceberg-type efficiency loss for firms engaging in MP, capturing a different type of friction[41] faced by MNEs.

Last but not least, Ding [2020] is closely related to Cole and Davies [2011]. Despite the apparent similarity between the two frameworks, there are clear differences between the two. First, Cole and Davies [2011] utilize quasi-linear CES preference. Combined with monopolistic competition, their model yields constant markups, and a complete pass-through in equilibrium. In contrast, Ding [2020] utilizes quadratic quasi-linear preference to generate variable markups and incomplete pass-through for different firms, which is more suitable for pricing and welfare analysis. Second, Cole and Davies [2011] completely ignore the possibility that tariffs can impact the level of firm entry in the economy. According to Caliendo et al. [2017], the combination of ad valorem tariff and tariff rebate violates the macro assumption in Arkolakis et al. [2012], and the level of entry will be affected by a tariff change. In Ding [2020], the number of entrants is endogenously affected by the tariff level, generating different welfare implications for protectionist trade policy in the short-run

[40]For example, see Keller and Yeaple [2013].

[41]For example, it could be language difference, technology incompatibility between parent and affiliate firms, or the cost associated with property rights.

and long-run.

Space constraints and my own taste and work clearly influenced the choice of material covered in this survey. For instance, my choice to focus on models that generates closed-form solutions have entirely abstracted from research on quantitative exercises. As another example, most of the paper discussed in this survey is built on Krugman [1980, 1979] international trade theory, extended by Melitz [2003] to introduce heterogeneity. One should also consider the trade policy implications that combine Heckscher-Ohlin theory with firm heterogeneity (such as Bernard et al. [2007]) and other approaches. Much more research is needed on the trade policy implication of firm heterogeneity, including on understanding of markups and how they interact with multinational production.

References

Alessandria, George and Horag Choi, "Do sunk costs of exporting matter for net export dynamics?," *The Quarterly Journal of Economics*, 2007, *122* (1), 289–336.

Alfaro, Laura and Andrew Charlton, "Intra-industry foreign direct investment," *American Economic Review*, 2009, *99* (5), 2096–2119.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings, "Importers, exporters, and exchange rate disconnect," *American Economic Review*, 2014, *104* (7), 1942–78.

__ , __ , and __ , "International Shocks, Variable Markups, and Domestic Prices," *The Review of Economic Studies*, 02 2019.

Antràs, Pol, "Firms, contracts, and trade structure," *The Quarterly Journal of Economics*, 2003, *118* (4), 1375–1418.

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

__ and __ , "Contractual Frictions and Global Sourcing," *The organization of firms in a global economy*, 2009.

Antràs, Pol and Stephen R Yeaple, "Multinational firms and the structure of international trade," *Handbook of International Economics*, 2014, *4*, 55–130.

Antràs, Pol, Teresa C Fort, and Felix Tintelnot, "The margins of global sourcing: Theory and evidence from us firms," *American Economic Review*, 2017, *107* (9), 2514–64.

Arkolakis, Costas, "A Generalized Solution of the Monopolistic Competition Model with Heterogeneous Firms and a Linear Demand (Melitz-Ottaviano)," *Working paper, Department of Economics, Yale University*, 2008.

__ , Arnaud Costinot, and Andres Rodriguez-Clare, "New trade models, same old gains?," *American Economic Review*, 2012, *102* (1), 94–130.

__ , __ , Dave Donaldson, and Andrés Rodríguez-Clare, "The elusive pro-competitive effects of trade," *The Review of Economic Studies*, 2018.

Atkeson, Andrew and Ariel Burstein, "Pricing-to-market in a Ricardian Model of International Trade," *American Economic Review*, 2007, *97* (2), 362–367.

__ and __ , "Pricing-to-market, trade costs, and international relative prices," *American Economic Review*, 2008, *98* (5), 1998–2031.

Bagwell, Kyle and Seung Hoon Lee, "Trade policy under monopolistic competition with firm selection," *Mimeogr., Stanford University*, 2015.

Baldwin, Richard and Toshihiro Okubo, "Networked FDI: Sales and sourcing patterns of Japanese foreign affiliates," *The World Economy*, 2014, *37* (8), 1051–1080.

Baldwin, Richard E and Rikard Forslid, "Trade Liberalization with Heterogenous Firms," Working Paper 12192, National Bureau of Economic Research May 2006.

Baltagi, Badi H, Peter Egger, and Michael Pfaffermayr, "Estimating models of complex FDI: Are there third-country effects?," *Journal of Econometrics*, 2007, *140* (1), 260–281.

Barattieri, Alessandro, Matteo Cacciatore, and Fabio Ghironi, "Protectionism and the Business Cycle," Working

Paper 24353, National Bureau of Economic Research February 2018.

Barefoot, Kevin B and Raymond J Mataloni Jr, "Operations of US multinational companies in the United States and abroad," *Survey of Current Business*, 2011, *91* (11), 29–48.

Behrens, Kristian and Yasusada Murata, "General equilibrium models of monopolistic competition: a new approach," *Journal of Economic Theory*, 2007, *136* (1), 776–787.

__ **and** __ , "Trade, competition, and efficiency," *Journal of International Economics*, 2012, *87* (1), 1–17.

__ **, Giordano Mion, Yasusada Murata, and Jens Südekum**, "Trade, wages, and productivity," *International Economic Review*, 2014, *55* (4), 1305–1348.

Bellone, Flora, Patrick Musso, Lionel Nesta, and Frederic Warzynski, "International trade and firm-level markups when location and quality matter," *Journal of Economic Geography*, 2014, *16* (1), 67–91.

Benfratello, Luigi and Alessandro Sembenelli, "Foreign ownership and productivity: Is the direction of causality so obvious?," *International Journal of Industrial Organization*, 2006, *24* (4), 733–751.

Bergin, Paul R and Robert C Feenstra, "Staggered price setting, translog preferences, and endogenous persistence," *Journal of monetary Economics*, 2000, *45* (3), 657–680.

Berlingieri, Giuseppe, Frank Pisch, and Claudia Steinwender, "Organizing Global Supply Chains: Input Cost Shares and Vertical Integration," Working Paper 25286, National Bureau of Economic Research November 2018.

Bernard, Andrew B and Joachim Wagner, "Exports and success in German manufacturing," *Weltwirtschaftliches Archiv*, 1997, *133* (1), 134–157.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott, *Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods*, University of Chicago Press, January

Bernard, Andrew B, J Bradford Jensen, Stephen J Redding, and Peter K Schott, "Firms in international trade," *Journal of Economic perspectives*, 2007, *21* (3), 105–130.

__ **,** __ **,** __ **, and** __ , "The empirics of firm heterogeneity and international trade," *Annu. Rev. Econ.*, 2012, *4* (1), 283–313.

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

Bertoletti, Paolo, Federico Etro, and Ina Simonovska, "International trade with indirect additivity," *American Economic Journal: Microeconomics*, 2018, *10* (2), 1–57.

Bilbiie, Florin, Fabio Ghironi, and Marc Melitz, "Endogenous Entry, Product Variety, and Business Cycles," *Journal of Political Economy*, 2012, *120* (2), 304–345.

Bilbiie, Florin Ovidiu, Fabio Pietro Ghironi, and Marc J Melitz, "Monopoly Power and Endogenous Product Variety: Distortions and Remedies," *American Economic Journal: Macroeconomics (Forthcoming)*, 2016.

Blas, Beatriz De and Katheryn N Russ, "Understanding markups in the open economy," *American Economic Journal: Macroeconomics*, 2015, *7* (2), 157–80.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen, "Americans do IT better: US multinationals and the productivity miracle," *American Economic Review*, 2012, *102* (1), 167–201.

Blundell, Richard and Stephen Bond, "Initial conditions and moment restrictions in dynamic panel data models," *Journal of econometrics*, 1998, *87* (1), 115–143.

Bown, Chad P, "US-China Trade War: The Guns of August," *Peterson Institute of International Economics, Trade and Investment Policy Watch*, 2019.

Brainard, S Lael, "A simple theory of multinational corporations and trade with a trade-off between proximity and concentration," Technical Report, National Bureau of Economic Research 1993.

Caldara, Dario, Matteo Iacoviello, Patrick Molligo, Andrea Prestipino, and Andrea Raffo, "The Economic Effects of Trade Policy Uncertainty," *MIMEO*, 2019.

Caliendo, Lorenzo, Robert C Feenstra, John Romalis, and Alan M Taylor, "Tariff reductions, entry, and welfare: Theory and evidence for the last two decades," *NBER Working Paper*, 2017, *21768.*

Carballo, Jeronimo, "3 Global Sourcing under Uncertainty," *Developments in Global Sourcing*, 2018, p. 71.

Carluccio, Juan and Thibault Fally, "Global sourcing under imperfect capital markets," *Review of Economics and Statistics*, 2012, *94* (3), 740–763.

Caves, Douglas W, Laurits R Christensen, and W Erwin Diewert, "The economic theory of index numbers and the measurement of input, output, and productivity," *Econometrica: Journal of the Econometric Society*, 1982, pp. 1393–1414.

Chaney, Thomas, "Distorted gravity: the intensive and extensive margins of international trade," *American Economic Review*, 2008, *98* (4), 1707–21.

Cole, Matthew T and Ronald B Davies, "Strategic tariffs, tariff jumping, and heterogeneous firms," *European Economic Review*, 2011, *55* (4), 480–496.

Costinot, Arnaud, Andrés Rodríguez-Clare, and Iván Werning, "Micro to macro: Optimal trade policy with firm heterogeneity," Technical Report, National Bureau of Economic Research 2016.

Criscuolo, Chiara and Ralf Martin, "Multinationals and US productivity leadership: evidence from Great Britain," *The Review of Economics and Statistics*, 2009, *91* (2), 263–281.

de Blas, Beatriz and Katheryn Niles Russ, "Hymer's multinationals," *Journal of Economic Behavior & Organization*, 2013, *94*, 381–392.

Demidova, Svetlana, "Trade policies, firm heterogeneity, and variable markups," *Journal of International Economics*, 2017, *108*, 260–273.

__ **and Andres Rodriguez-Clare**, "Trade policy under firm-level heterogeneity in a small economy," *Journal of International Economics*, 2009, *78* (1), 100–112.

Dhingra, Swati and John Morrow, "Monopolistic competition and optimum product diversity under firm heterogeneity," *Journal of Political Economy*, 2019, *127* (1), 196–232.

Diewert, W Erwin, "Exact and superlative index numbers," *Journal of econometrics*, 1976, *4* (2), 115–145.

Díez, Federico J, "The asymmetric effects of tariffs on intra-firm trade and offshoring decisions," *Journal of International Economics*, 2014, *93* (1), 76–91.

Ding, Ziran, "Firm Heterogeneity, Variable Markups, and FDI," *Doctoral Dissertation*, 2020.

Dixit, Avinash, "Entry and exit decisions under uncertainty," *Journal of political Economy*, 1989, *97* (3), 620–638.

Dobbelaere, Sabien and Kozo Kiyota, "Labor market imperfections, markups and productivity in multinationals and exporters," *Labour Economics*, 2018, *53*, 198–212.

Eaton, Jonathan and Samuel Kortum, "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779.

__ , __ , and Francis Kramarz, "Dissecting trade: Firms, industries, and export destinations," *American Economic Review*, 2004, *94* (2), 150–154.

__ , __ , and __ , "An anatomy of international trade: Evidence from French firms," *Econometrica*, 2011, *79* (5), 1453–1498.

__ , __ , and Sebastian Sotelo, "International Trade: Linking Micro and Macro," in "Advances in Economics and Econometrics, Tenth World Congress," Vol. 2 2012.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "Competition, markups, and the gains from international trade," *American Economic Review*, 2015, *105* (10), 3183–3221.

__ , __ , and __ , "How Costly Are Markups?," Working Paper 24800, National Bureau of Economic Research July 2018.

Egger, Peter, "On the role of distance for outward FDI," *The Annals of Regional Science*, 2008, *42* (2), 375–389.

Ekholm, Karolina, Rikard Forslid, and James R Markusen, "Export-platform foreign direct investment," *Journal of the European Economic Association*, 2007, *5* (4), 776–795.

Engel, Charles, "Expenditure switching and exchange-rate policy," *NBER macroeconomics annual*, 2002, *17*, 231–272.

Feenstra, Robert C, "Restoring the product variety and pro-competitive gains from trade with heterogeneous firms and bounded productivity," *Journal of International Economics*, 2018, *110*, 16–27.

__ and David E Weinstein, "Globalization, markups, and US welfare," *Journal of Political Economy*, 2017, *125* (4), 1040–1074.

Felbermayr, Gabriel, Benjamin Jung, and Mario Larch, "Optimal tariffs, retaliation, and the welfare loss from tariff wars in the Melitz model," *Journal of International Economics*, 2013, *89* (1), 13–25.

Foster, Lucia, John Haltiwanger, and Chad Syverson, "Reallocation, firm turnover, and efficiency: selection on productivity or profitability?," *American Economic Review*, 2008, *98* (1), 394–425.

Garetto, Stefania, Lindsay Oldenski, and Natalia Ramondo, "Multinational Expansion in Time and Space," Working Paper 25804, National Bureau of Economic Research May 2019.

Gaubert, Cecile and Oleg Itskhoki, "Granular Comparative Advantage," Working Paper 24807, National Bureau of Economic Research July 2018.

Ghironi, Fabio and Marc Melitz, "International Trade and Macroeconomic Dynamics with Heterogeneous Firms," *Quarterly Journal of Economics*, 2005, *120*, 865–915.

Grossman, Gene M, Elhanan Helpman, and Adam Szeidl, "Optimal integration strategies for the multinational firm," *Journal of international economics*, 2006, *70* (1), 216–238.

Gumpert, Anna, Andreas Moxnes, Natalia Ramondo, and Felix Tintelnot, "The Life-Cycle Dynamics of Exporters and Multinational Firms," Working Paper 24013, National Bureau of Economic Research November 2017.

Haaland, Jan I and Anthony J Venables, "Optimal trade policy with monopolistic competition and heterogeneous firms," *Journal of International Economics*, 2016, *102*, 85–95.

Hall, Robert E, "The relation between price and marginal cost in US industry," *Journal of political Economy*, 1988, *96* (5), 921–947.

__ , Olivier Jean Blanchard, and R Glenn Hubbard, "Market structure and macroeconomic fluctuations," *Brookings papers on economic activity*, 1986, *1986* (2), 285–338.

Handley, Kyle, "Exporting under trade policy uncertainty: Theory and evidence," *Journal of international Economics*, 2014, *94* (1), 50–66.

__ and Nuno Limao, "Trade and investment under policy uncertainty: theory and firm evidence," *American Economic Journal: Economic Policy*, 2015, *7* (4), 189–222.

Hanson, Gordon H, Raymond J Mataloni Jr, and Matthew J Slaughter, "Vertical production networks in multinational firms," *Review of Economics and statistics*, 2005, *87* (4), 664–678.

Head, Keith and Barbara J Spencer, "Oligopoly in international trade: Rise, fall and resurgence," *Canadian Journal of Economics/Revue canadienne d'économique*, 2017, *50* (5), 1414–1444.

Helpman, Elhanan and Paul R Krugman, *Market structure and foreign trade: Increasing returns, imperfect competition, and the international economy*, MIT press, 1985.

__ , Marc J Melitz, and Stephen R Yeaple, "Export versus FDI with heterogeneous firms," *American Economic Review*, 2004, *94* (1), 300–316.

__ , Marc Melitz, and Yona Rubinstein, "Estimating trade flows: Trading partners and trading volumes," *The quarterly journal of economics*, 2008, *123* (2), 441–487.

Hornok, Cecília and Balázs Muraközy, "Markups of Exporters and Importers: Evidence from Hungary," *The Scandinavian Journal of Economics*, 2018.

Irarrazabal, Alfonso, Andreas Moxnes, and Luca David Opromolla, "The margins of multinational production and the role of intrafirm trade," *Journal of Political Economy*, 2013, *121* (1), 74–126.

Jackson, Laurence Fraser, "Hierarchic demand and the Engel curve for variety," *The Review of Economics and Statistics*, 1984, pp. 8–15.

Keller, Wolfgang and Stephen Ross Yeaple, "The gravity of knowledge," *American Economic Review*, 2013, *103* (4), 1414–44.

Kılınç, Umut, "Estimating entrants' productivity when prices are unobserved," *Economic Modelling*, 2014, *38*, 640–647.

Kimball, Miles S, "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit, and Banking*, 1995, *27* (4).

Kokovin, Sergey, Mathieu Parenti, Jacques-François Thisse, and Philip Ushchev, "On the Dilution of Market Power," *CEPR Discussion Paper No. DP12367*, 2017.

Krugman, Paul, "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 1980, *70*, 950–959.

Krugman, Paul R, "Increasing returns, monopolistic competition, and international trade," *Journal of international*

Economics, 1979, *9* (4), 469–479.

Loecker, Jan De and Frederic Warzynski, "Markups and firm-level export status," *American Economic Review*, 2012, *102* (6), 2437–71.

Lu, Dan, "Exceptional exporter performance? Evidence from Chinese manufacturing firms," *manuscript, University of Chicago*, 2010.

Mayer, Thierry and Gianmarco IP Ottaviano, "The happy few: The internationalisation of european firms," *Intereconomics*, 2008, *43* (3), 135–148.

Melitz, Marc, "The Impact of Trade on Intra-Industry Reallocations and Aggre- gate Industry Productivity," *Econometrica*, 2003, *71*, 1695–1725.

__ **and Giancarlo Ottaviano**, "Market Size, Trade, and Productivity," *Review of Economic Studies*, 2008, *75* (295-316).

Melitz, Marc J and Stephen J Redding, "Heterogeneous firms and trade," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 1–54.

Muraközy, Balázs and Katheryn Niles Russ, "Competition with multinational firms: Theory and evidence," Technical Report, IEHAS Discussion Papers 2015.

Nocco, Antonella, Gianmarco IP Ottaviano, and Matteo Salto, "Geography, competition, and optimal multilateral trade policy," *Journal of International Economics*, 2019.

Nocke, Volker and Stephen Yeaple, "Cross-border mergers and acquisitions vs. greenfield foreign direct investment: The role of firm heterogeneity," *Journal of International Economics*, 2007, *72* (2), 336–365.

Ossa, Ralph, "A "new trade" theory of GATT/WTO negotiations," *Journal of Political Economy*, 2011, *119* (1), 122–152.

Ottaviano, Gianmarco, Takatoshi Tabuchi, and Jacques-François Thisse, "Agglomeration and trade revisited," *International Economic Review*, 2002, pp. 409–435.

Parenti, Mathieu, "Large and small firms in a global market: David vs. Goliath," *Journal of International Economics*, 2018, *110*, 103–118.

__ **, Philip Ushchev, and Jacques-François Thisse**, "Toward a theory of monopolistic competition," *Journal of Economic Theory*, 2017, *167*, 86–115.

Ramondo, Natalia and Andrés Rodríguez-Clare, "Trade, multinational production, and the gains from openness," *Journal of Political Economy*, 2013, *121* (2), 273–322.

__ **and Veronica Rappoport**, "The role of multinational production in a risky environment," *Journal of International Economics*, 2010, *81* (2), 240–252.

__ **,** __ **, and Kim J Ruhl**, "The proximity-concentration tradeoff under uncertainty," *Review of Economic Studies*, 2013, *80* (4), 1582–1621.

__ **,** __ **, and** __ , "Intrafirm trade and vertical fragmentation in US multinational corporations," *Journal of International Economics*, 2016, *98*, 51–59.

Rodrigue, Joel, "Multinational production, exports and aggregate productivity," *Review of Economic Dynamics*, 2014, *17* (2), 243–261.

Rodriguez-Lopez, Jose Antonio, "Prices and exchange rates: A theory of disconnect," *The Review of Economic Studies*, 2011, *78* (3), 1135–1177.

Sembenelli, Alessandro and Georges Siotis, "Foreign Direct Investment and mark-up dynamics: Evidence from Spanish firms," *Journal of International Economics*, 2008, *76* (1), 107–115.

Simonovska, Ina, "Income differences and prices of tradables: Insights from an online retailer," *The Review of Economic Studies*, 2015, *82* (4), 1612–1656.

Steinberg, Joseph B, "Brexit and the macroeconomic impact of trade policy uncertainty," *Journal of International Economics*, 2019, *117*, 175–195.

Temouri, Yama, Nigel L Driffield, and Dolores Anón Higón, "Analysis of productivity differences among foreign and domestic firms: evidence from Germany," *Review of World Economics*, 2008, *144* (1), 32–54.

Tintelnot, Felix, "Global production with export platforms," *The Quarterly Journal of Economics*, 2017, *132* (1), 157–209.

UNCTAD, "World Investment Report 2011," *United Nations Publication*, 2011.

__ , "World Investment Report 2018," *United Nations Publication*, 2018.

Voigtlaender, N and A Garcia-Marin, "Exporting and Plant-Level Efficiency Gains: It's in the Measure," *Journal of Political Economy*, 2019.

Yeaple, Stephen Ross, "The role of skill endowments in the structure of US outward foreign direct investment," *Review of Economics and statistics*, 2003, *85* (3), 726–734.

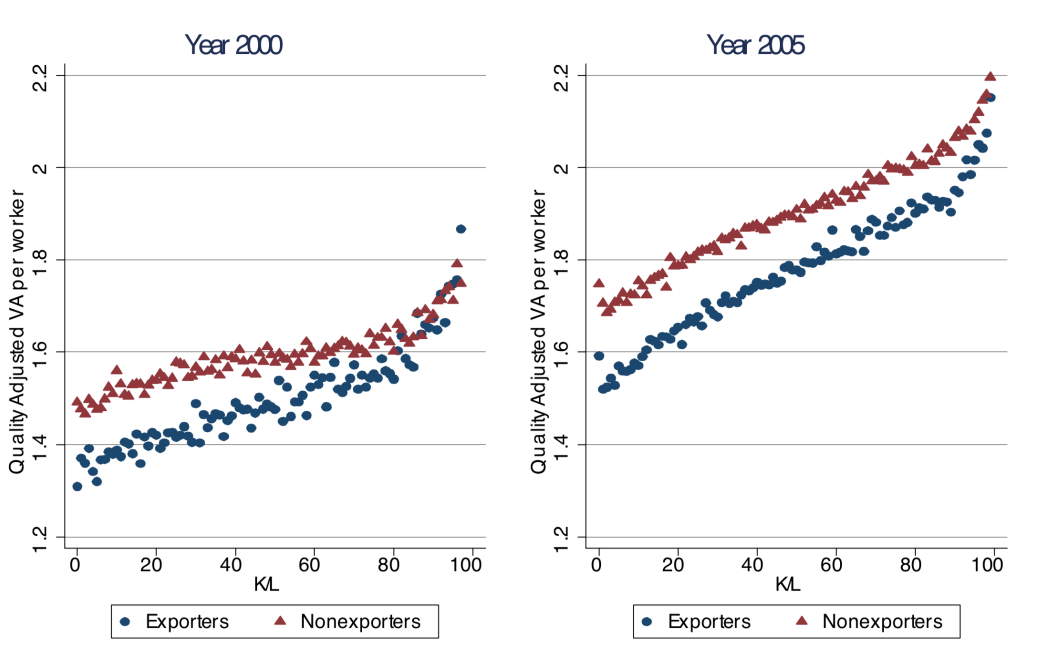
__ , "The multinational firm," *Annu. Rev. Econ.*, 2013, *5* (1), 193–217.

Zhelobodko, Evgeny, Sergey Kokovin, Mathieu Parenti, and Jacques-François Thisse, "Monopolistic competition: Beyond the constant elasticity of substitution," *Econometrica*, 2012, *80* (6), 2765–2784.

Zlate, Andrei, "Offshore production and business cycle dynamics with heterogeneous firms," *Journal of International Economics*, 2016, *100*, 34–49.

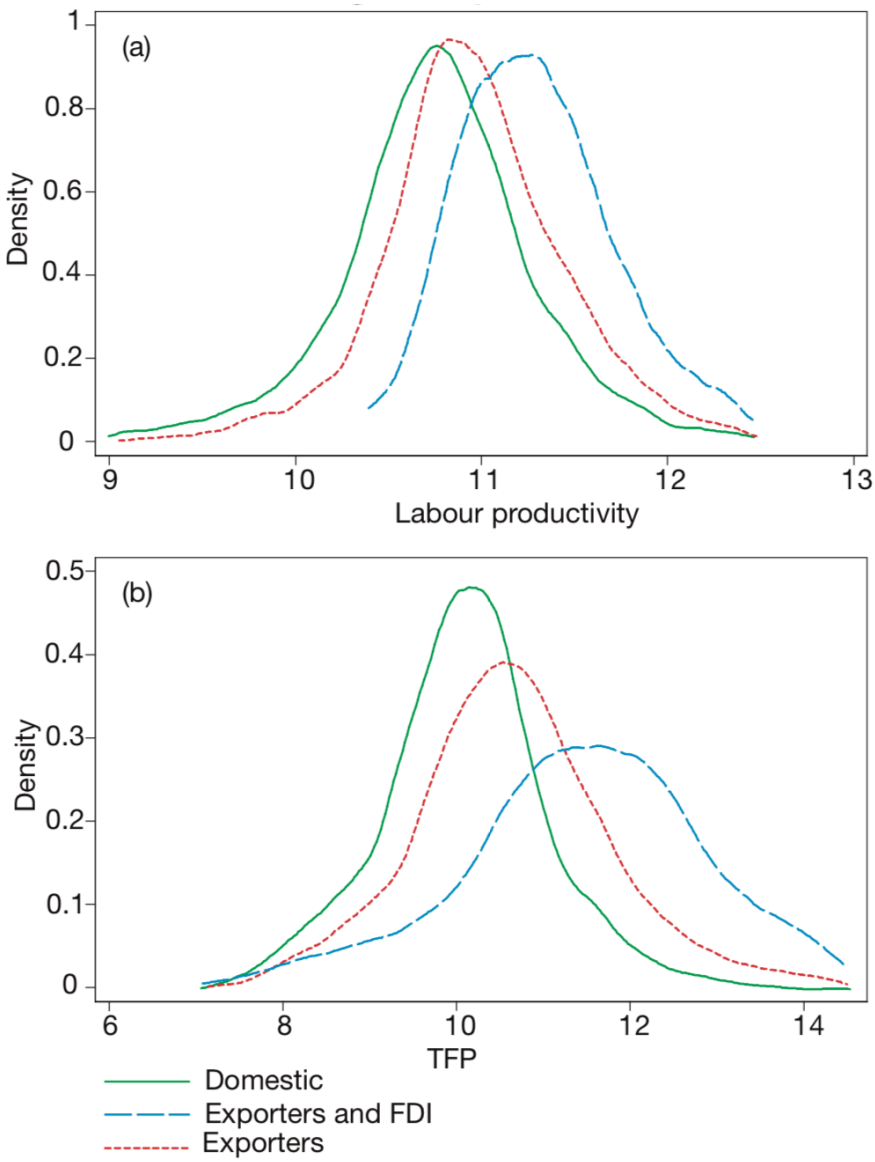
A Figures

Figure 1: Quality Adjusted Value-added per Worker



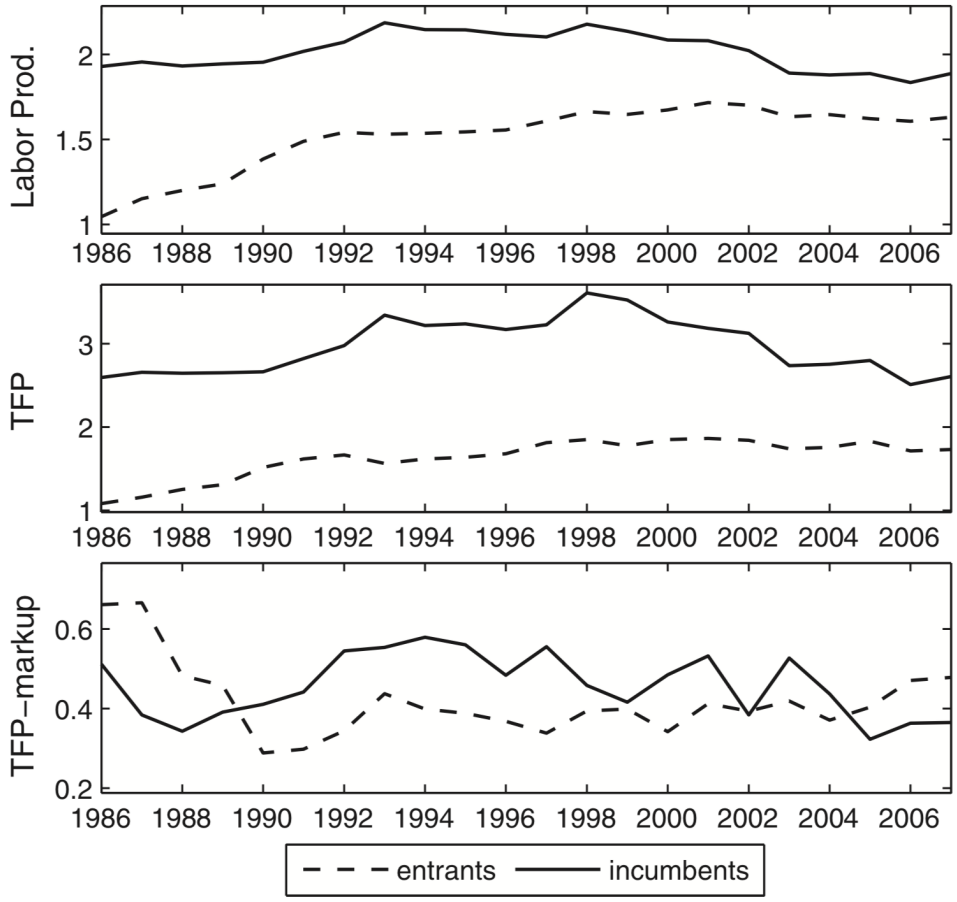
Note: Graph is taken from Lu [2010]. The horizontal axis is the rank of firms by their capital-labor ratio and they are grouped into 100 bins.

Figure 2: Belgian-FDI Makers vs. Belgian Exporters



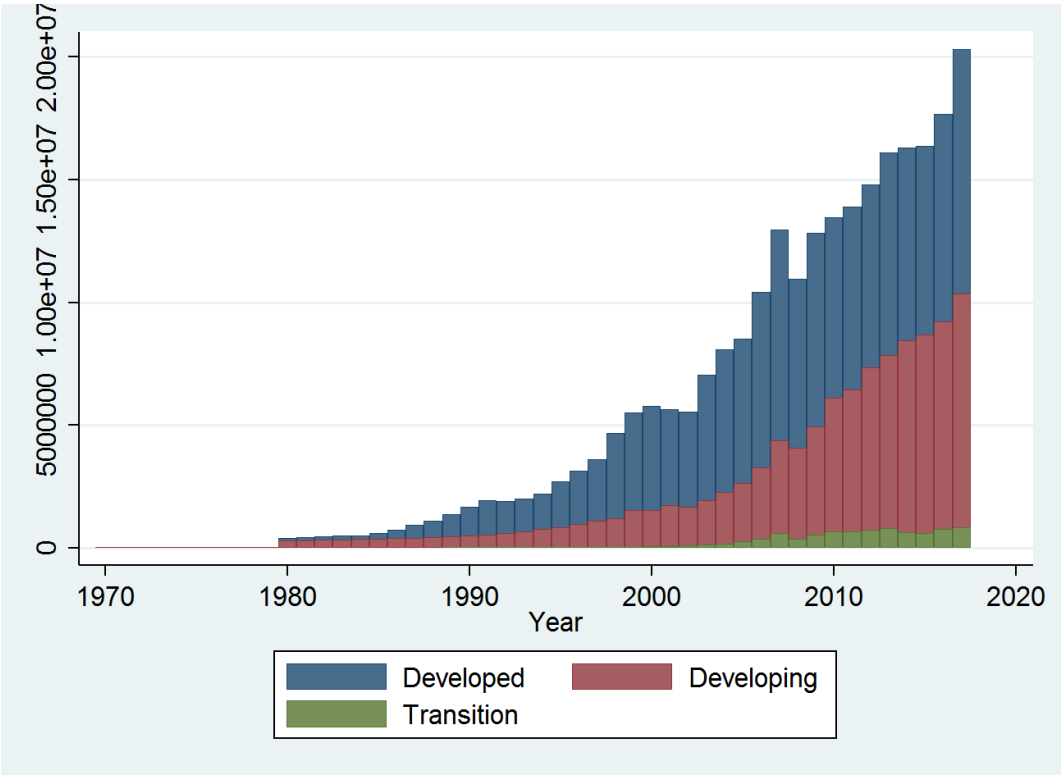
Note: Graph is taken from Mayer and Ottaviano [2008]. The data is from the 2004 Belgian manufacturing firms in the EFIM project. Notice, in this sample, nearly all the Belgian FDI firms are also exporters.

Figure 3: Entrant's vs. Incumbent's Productivity



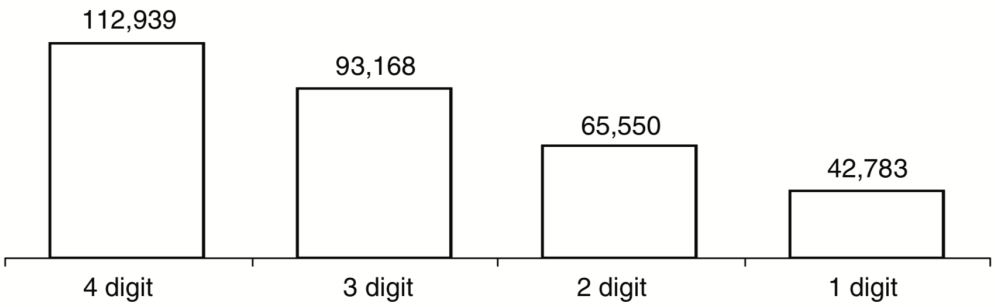
Note: Graph is taken from Kılınç [2014]. The top two panels display a considerable gap in labor productivity and standard TFP between entrants and incumbents. The TFP-markup estimated by the control function approach with entry variation, however, indicates that entrants are on average more productive than incumbents in the first three years of the sample. For the period after 1990, entrants' TFP-markup gradually converge with the incumbents'.

Figure 4: World Inward FDI by Groups from 1970–2017



Note: Graph is constructed by author based on the data from UNCTAD from 1970–2017. According to World Investment Report (2017), foreign direct investment (FDI) is defined as an investment involving a long term relationship and reflecting a lasting interest and control by a resident entity in one economy (foreign direct investor or parent enterprise) in an enterprise resident in an economy other than that of the foreign direct investor (FDI enterprise or affiliate enterprise or foreign affiliate).

Figure 5: Vertical FDI Observed at Different Levels of Aggregation



Note: Graph is taken from Alfaro and Charlton [2009], based on firm-level data from D&B. The numbers on top of each bar indicate the number of manufacturing subsidiaries. The horizontal axis indicates the one-, two-, three-, and four-digit Standard Industrial Classification (SIC) code. At a more disaggregated level, we see the number of vertical FDI subsidiaries increases.

B Tables

Table 1: Export by US Manufacturing Firms (2002)

NAICS industry	Firm distribution	Exporter distribution
311 Food Manufacturing	6.8	12
312 Beverage and Tobacco Product	0.7	23
313 Textile Mills	1.0	25
314 Textile Product Mills	1.9	12
315 Apparel Manufacturing	3.2	8
316 Leather and Allied Product	0.4	24
321 Wood Product Manufacturing	5.5	9
322 Paper Manufacturing	1.4	24
323 Printing and Related Support	11.9	6
324 Petroleum and Coal Products	0.4	18
325 Chemical Manufacturing	3.1	36
326 Plastics and Rubber Products	4.4	28
327 Nonmetallic Mineral Product	4.0	4
331 Primary Metal Manufacturing	1.5	30
332 Fabricated Metal Product	19.9	14
333 Machinery Manufacturing	9.0	33
334 Computer and Electronic Product	4.5	38
335 Electrical Equipment, Appliance	1.7	38
336 Transportation Equipment	3.4	28
337 Furniture and Related Product	6.4	7
339 Miscellaneous Manufacturing	9.1	2
Aggregate manufacturing	**100**	**18**

Note: Table taken from Bernard et al. [2007]. Data are from the 2002 US Census of Manufactures. Column 2 summarizes the distribution of manufacturing firms across the three-digit NAICS manufacturing industries. Column 3 reports the share of firms in each industry that exports.

Table 2: Export by European Manufacturing Firms (2003)

Country of origin	# of firms in the sample	Exporter distribution in the sample
Germany	48325	59.34
France	23691	67.30
United Kingdom	14976	28.33
Italy	4159	74.44
Hungary	6404	49.53
Norway	8125	39.22

Note: Table taken from Mayer and Ottaviano [2008]. Dataset is the EFIM firm-level data. Column 2 summarizes the number of firms in the sample EFIM dataset. Column 3 reports the share of firms within the sample that exports in each country. Germany, Italy, Hungary, the United Kingdom and France have large firms only; Norwegian data are exhaustive.

Table 3: Exporter Premia in US Manufacturing (2002)

	(1)	(2)	(3)
Log employment	1.19	0.97	
Log shipments	1.48	1.08	0.08
Log value-added per worker	0.26	0.11	0.10
Log TFP	0.02	0.03	0.05
Log wage	0.17	0.06	0.06
Log capital per worker	0.32	0.12	0.04
Log skill per worker	0.19	0.11	0.19
Additional covariates	None	Industry fixed effects	Industry fixed effects, log employment

Note: Table taken from Bernard et al. [2007]. Data are for 2002 and are from the US Census of Manufactures. Each row summarizes the average percent difference between exporters and nonexporters for a particular characteristic. All results are from bivariate OLS regressions of firm characteristic in the first column on a dummy variable indicating firm's export status. Columns 2 and 3 include industry fixed effects and industry fixed effects plus log firm employment, respectively, as additional controls. Total factor productivity (TFP) is computed as in Caves et al. [1982]. Capital and skill per worker are capital stock and nonproduction workers per total employment, respectively. All results are significant at the 1 percent level.

Table 4: Exporter Premia in European Manufacturing

Country	Employment	Value-added	Wage	Capital intensity	Skill intensity
Germany	2.99 (4.39)		1.02 (0.06)		
France	2.24 (0.47)	2.68 (0.84)	1.09 (1.12)	1.49 (5.60)	
UK	2.24 (0.47)	1.29 (1.53)	1.15 (1.39)		
Italy	2.42 (2.06)	2.14 (1.78)	1.07 (1.06)	1.01 (0.45)	1.25 (1.04)
Hungary	5.31 (2.95)	13.53 (23.75)	1.44 (1.63)	0.79 (0.35)	
Belgium	9.16 (13.42)	14.80 (21.12)	1.26 (1.15)	1.04 (3.09)	
Norway	6.11 (5.59)	7.95 (7.48)	1.08 (0.68)	1.01 (0.23)	

Note: Table taken from Mayer and Ottaviano [2008]. The table shows premia of the considered variable as the ratio of exporters over non-exporters (standard deviation ratio in brackets). France, Germany, Hungary, Italy and the United Kingdom have large firms only; Belgian and Norwegian data are exhaustive.

Table 5: Exporters vs. Non-Exporters

	Exporters Relative to Non-exporters	
	China	US
Domestic sales	0.96	4.8
Value-added per worker	0.86	1.39
Sales per worker	0.91	1.36

Note: Table taken from Lu [2010]. Each entry is the mean value of exporters divided by mean value of non-exporters. The data used in this table comes from a firm-level data set from the Annual Census of Enterprises by the Chinese National Bureau of Statistics from 1998 to 2007. It includes all the State-Owned Enterprise (henceforth SOE) and non-SOEs with sales over 5 million RMB (about 600,000 US dollars).

Table 6: Exporter Markup Premium in French Manufacturing Industries

	μ	μ_X	μ_{NX}	t_μ
All manufacturing	1.148	1.173	1.136	-47.08
Agro-food	1.097	1.108	1.095	-6.144
Automobile	1.144	1.176	1.112	-11.55
Chemicals	1.304	1.329	1.259	-19.00
Clothing and footwear	0.964	0.978	0.945	-8.125
Electric and electronic components	1.446	1.458	1.433	-4.291
Electric and electronic equipment	1.427	1.397	1.442	10.930
House equipment and furnishings	1.210	1.218	1.206	-4.133
Machinery and mechanical equipment	1.150	1.174	1.134	-19.720
Metallurgy, iron and steel	1.043	1.039	1.048	5.133
Mineral industries	0.993	0.980	0.999	4.919
Pharmaceuticals	1.371	1.388	1.311	-6.920
Printing and publishing	1.182	1.168	1.189	6.958
Textile	1.191	1.207	1.164	-9.451
Transportation machinery	1.088	1.090	1.086	-0.593
Wood and paper	1.261	1.293	1.237	-18.31

Note: Table taken from Bellone et al. [2014]. All values display averages for the period 1998—2007. Greek letter μ stands for markups. Subscripts X and NX denote exporters and non-exporters, respectively. Letter t stands for Student's t, testing the equality of means in the markup of exporters and non-exporters. All t-values indicate significant differences at 1 percent level, with the exception of 'transportation machinery'.