Firm Heterogeneity, Variable Markups, and Multinational Production: A Review from Trade Policy Perspective*

Ziran Ding†

Date: November 11, 2021
First version: November 29, 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 the stylized facts regarding firm heterogeneity, firm-level markups, and multinational production's global structure. I then survey the trade policy papers that build on the workhorse model of firm heterogeneity. Third, I summarize the recent development of theoretical approaches of modeling the firm-level markups and its trade policy implication. Fourth, I discuss the theoretical frameworks that incorporate multinational production into heterogeneous firms' framework and their trade policy implication. Finally, I discuss directions for future research and offer suggestions for further readings.

Keywords: Trade policy, Firm heterogeneity, Variable markups, Multinational production
JEL Codes: F12, F13, F23, F60

---

* I'm indebted to Theo Eicher, Fabio Ghironi and Mu-Jeung Yang for guidance, encouragement and continuous support. Two anonymous referee reports provide lots of constructive suggestions. For helpful comments and discussions, I thank Kristian Behrens, Keith Head, Antonio Rodriguez-Lopez, and seminar participants at the Bank of Lithuania, International and Macro workshop and MTI Brownbag at the University of Washington. The views expressed herein are those of the author and not necessarily those of the Bank of Lithuania or the Eurosystem.

†Bank of Lithuania, and Kaunas University of Technology. Totoriu g. 4, Vilnius, 01121, Lithuania. E-mail: zding@lb.lt

Contents

1 Introduction

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: (i) Even within narrowly defined industries, firms possess monopoly power and charge different markups. (ii) In the era of globalization and global value chains, firms have multiple ways of accessing foreign markets. This paper aims to provide 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 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 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]).

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 3**, I first review the recent developments in the heterogeneous firm trade policy literature, focusing on the benchmark Melitz framework and its extensions. Most of the trade policy discussed here is in the form of tariff or subsidy, with a few exceptions on variable or fixed trade costs. While many papers in this literature focus on steady-state analysis, there are a few exceptions that emphasize the implications of trade policy uncertainty in a dynamic setting. In contrast, another strand of literature incorporating micro-founded trade models into the macro DSGE framework has gained popularity. I also review papers that study trade policy implications along this line.

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 **Section 4**, 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]). Toward the end of this section, I review the trade policy papers that emphasize the role of variable markups in the heterogeneous framework.

The second discrepancy is addressed by the burgeoning papers of multinational firms. In **Section 5**, 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. The section concludes with a summary of the recent papers that study the policy implication of multinational production.

The review here aims to provide a systematic policy assessment in international trade and investment in the light of recent theoretical development in this literature. There is no need to be encyclopedic given the space constraints and the useful surveys of the 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 surveys the trade policy work based on firm heterogeneity. Section 4 surveys the various theoretical approaches that have been used to model variable markups and their trade policy implication. The papers discussed in this section have the common feature that they all exclude the possibility of MP. Section 5 discusses the papers that introduce MP into the heterogeneous firm framework and their trade policy implication. Section 6 concludes.

2 Stylized Facts on Heterogeneous Firms, Variable Markups and Multinational Production

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

Table 1: Export by US Manufacturing Firms (2007)

NAICS industry	Firm distribution	Exporter distribution
311 Food Manufacturing	6.8	23
312 Beverage and Tobacco Product	0.9	30
313 Textile Mills	0.8	57
314 Textile Product Mills	2.7	19
315 Apparel Manufacturing	3.6	22
316 Leather and Allied Product	0.3	56
321 Wood Product Manufacturing	4.8	21
322 Paper Manufacturing	1.5	48
323 Printing and Related Support	11.1	15
324 Petroleum and Coal Products	0.5	34
325 Chemical Manufacturing	3.3	65
326 Plastics and Rubber Products	3.9	59
327 Nonmetallic Mineral Product	4.3	19
331 Primary Metal Manufacturing	1.5	58
332 Fabricated Metal Product	20.6	30
333 Machinery Manufacturing	8.7	61
334 Computer and Electronic Product	3.9	75
335 Electrical Equipment, Appliance	1.7	70
336 Transportation Equipment	3.4	57
337 Furniture and Related Product	6.5	16
339 Miscellaneous Manufacturing	9.3	32
Aggregate manufacturing	**100**	**35**

Note: Table taken from Bernard et al. [2018]. Data are for 2007 and are based on those firms appear in both US Census of manufactures and the Linked-Longitudinal Firm Trade Transactions Database(LFTTD). Column 2 summarizes the distribution of manufacturing firms across the 3-digit NAICS manufacturing industries. Column 3 reports the share of firms in each industry that exports.

Countries possess quite different export participation patterns. Bernard et al. [2018] show that only a fraction of plants within an industry export in the US. As shown in **Table 1** column 3, the average industry-level export share is around 35 percent in the US manufacturing sector. There are substantial variations across industries: from only 15 percent in printing and related support to almost 75 percent in computer and electronic products.

Mayer and Ottaviano [2008] utilize firm-level data from the European Firms in International Markets (EFIM)[1] project and find that European firms have higher export participation rates[2]. As illustrated in **Table 2**, about 55 percent of firms export among countries with nonexhaustive data[3], 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.

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.

2.1.2 Exporters are Different

Table 3: Exporter Premia in US Manufacturing (2007)

	(1)	(2)	(3)
Log employment	1.28	1.11	
Log shipments	1.72	1.35	0.24
Log value-added per worker	0.33	0.19	0.21
Log TFP	0.03	0.04	0.04
Log wage	0.21	0.09	0.10
Log capital per worker	0.28	0.16	0.20
Log skill per worker	0.06	0.01	0.11
Additional covariates	None	Industry fixed effects	Industry fixed effects, log employment

Note: Table taken from Bernard et al. [2018]. Data are for 2007 and are based on those firms appear in both US Census of manufactures and the Linked-Longitudinal Firm Trade Transactions Database(LFTTD). 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.

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

[2] 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.

[3] 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.

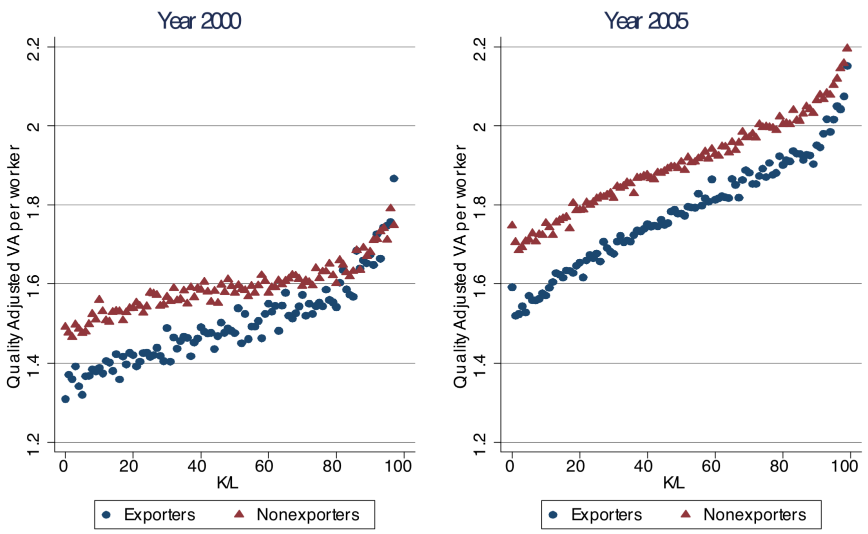
Not only are exporters rare, but also are they systematically different from nonexporters. **Table 3** presents the US manufacturing exporters' premia documented in Bernard et al. [2018]. Exporters are larger in employment and sales, and they are more productive in value-added per worker and total factor 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 more in the domestic market, and export only a small proportion of their output. A notable exception[4] 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.

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.

Figure 1: Quality-adjusted Value-added Per Worker

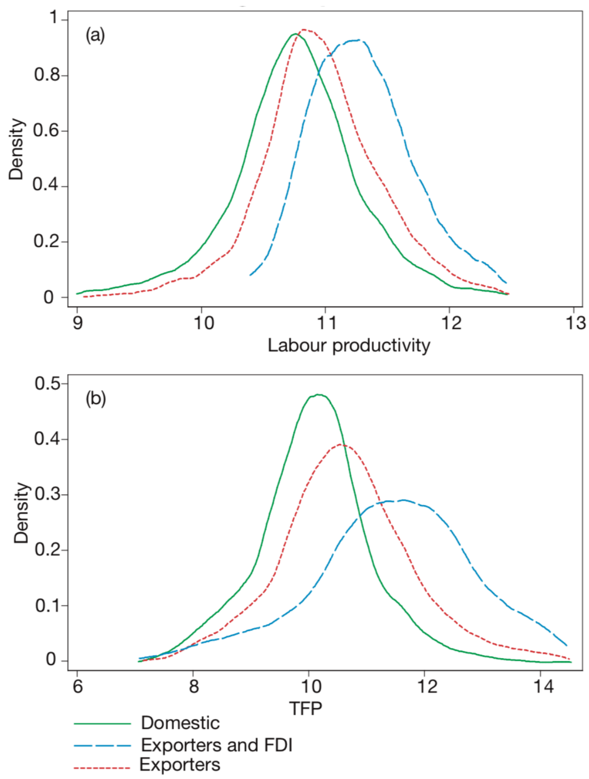


Note: Graph took from Lu [2010]. The horizontal axis is the rank of firms by capital-labor ratio and they are grouped into 100 bins.

[4]More recently, this puzzle has been resolved by Li et al. [2020]. The authors follow Foster et al. [2008] to estimate both revenue based productivity (TFPR) and physical productivity (TFPQ). They find Chinese exporters have lower TFPR, but have higher TFPQ. The difference comes from the fact that exporters charge lower prices than non-exporters.

2.1.3 MNEs are More Productive

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.

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

[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.

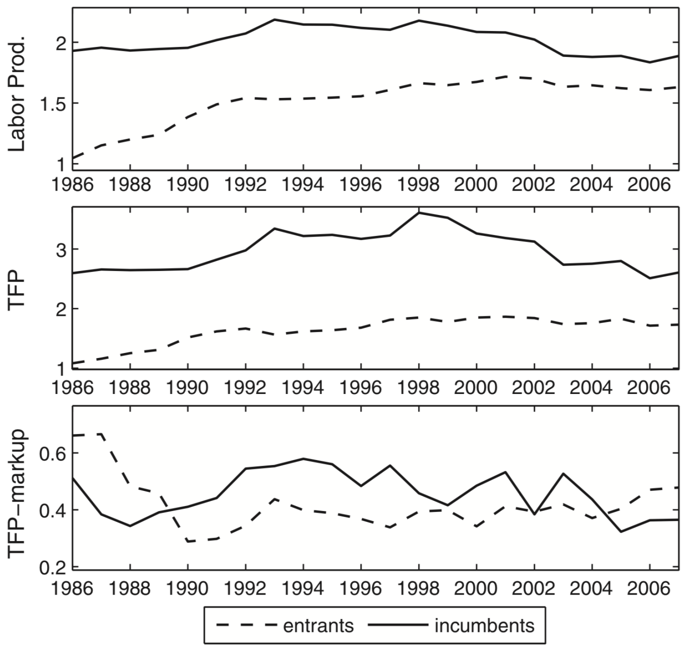
firms relative to nonmultinational domestic counterparts.

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[8] at firm or plant level.

2.2.1 Entrant's Markup

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'.

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, and comparing physical productivity (TFPQ), revenue-based productivity (TFPR) with traditional TFP (that is, the standard revenue-based output measure), 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

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

[8]The survey in this part is the main motivation for many theoretical developments in the trade literature that incorporates variable markups into firm heterogeneity. This is not to say that variable markups is the only reason for the observed price heterogeneity across firms. For example, differences in product quality could also results in price heterogeneity, interesting readers could refer to Gervais [2015] and the references therein.

Table 5: 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'.

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 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 5**. 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[9]. Recently, Voigtlaender and Garcia-Marin [2019] decompose exporters' efficiency gains into changes in revenue productivity, markups[10], 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

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

[10] Again, estimated using the method of De Loecker and Warzynski [2012]

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[11] of targeted firms, but this result is limited to R&D-intensive sectors. They attribute this increase in markups, which is proxied by accounting price-cost margins, 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 competition.

Muraközy and Russ [2015] apply the methodology developed by De Loecker and Warzynski [2012] to the 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.

Different from the studies mentioned above, Dobbelaere and Kiyota [2018] use Japanese manufacturing firms from 1994 to 2012 to investigate the heterogeneity in product and labor market imperfections across exporters, non-exporters, multinational enterprises (MNEs), and non-MNEs. The authors adopt the same methodology in De Loecker and Warzynski [2012] to measure product market markup, and define a new parameter to measure a firm's joint market imperfection in both product and labor market. This new measure varies with worker's bargaining power or the wage elasticity of a firm's labor supply curve, depending on how labor market imperfection is characterized[12]. They find that multinationals are less likely to be characterized by imperfect competition in the product market, even after controlling for productivity differences between MNEs and non-MNEs. On the labor market side, MNEs' wage-setting power is the dominant factor in generating wage dispersion across firms engaging in FDI. The opposite conclusions hold for the exporters.

Two recent papers pay close attention to the impacts of multinational firms on the markups in host countries. Alviarez et al. [2020] study the welfare impacts of multinational acquisitions in the beer and spirits industry in 76 countries. Equipped with a nested CES multi-product oligopoly model as in Atkeson and Burstein [2008], and focusing on the demand-side method developed by Berry [1994], the authors find that the aggregate markups of the largest firms grow by combining big brands under common ownership, rather than by increasing the market shares of the high-markup brands. Using counterfactual analysis, the authors also find that the intervention on cross-border mergers and acquisitions in the beer industry imposed by authorities lowers the beer price quite effectively. Focusing on US multinationals across many industries and using the approach by De Loecker and Warzynski [2012], Keller and Yeaple [2020] find that the foreign affiliate's markup of the multinational firms

[11] In this paper, the authors estimate markups using simple accounting price-cost margins.

[12] The authors investigate three types of labor market settings: (i) perfect competition or right-to-manage bargaining, (ii) efficient bargaining, (iii) monopsony. Case (ii) introduces the worker's bargaining power parameter, whereas case (iii) introduces the wage elasticity of the labor supply curve that a firm faces, which measures the degree of wage-setting power a firm possesses.

is declining in the level of development and competition of the host country. The largest and most productive firms that charge the highest markups tend to locate their affiliates in the wealthiest, most competitive countries, while smaller, lower markup firms tend to operate in less developed economies, consistent with their model's prediction.

2.2.4 Discussion on TFP Measurement and Markup Variation

It is important to note that most of the empirical studies mentioned above, especially in Section 2.1, focus on revenue-based productivity (TFPR), with only a few exceptions[13]. Although this survey is only concerned with the existence of TFP heterogeneity, rather than strictly ranking the firms' by their TFP based on their operation mode, it is important to bear in mind that higher TFPR does not always associate with higher TFPQ. This is because the variation of TFPR could come from factors other than TFPQ.

Consider a firm produces according to a Cobb-Douglas production $q = AK^{\alpha}L^{1-\alpha}$ and faces constraints in either labor or capital. Then the firm's TFPR could be written as follow:

$$TFPR = \frac{1}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}} \frac{|\varepsilon_{pq}|}{|\varepsilon_{pq}|-1} \left(\frac{MRP_K}{r}\right)^{\alpha} \left(\frac{MRP_L}{w}\right)^{1-\alpha}$$

Each component on the right-hand side could results in variation in TFPR: variation in α represents differences in technology, variation in ε_{pq} represents differences in idiosyncratic demand shocks, and variation in $MRP_K/r, MRP_L/w$ differences in capital or labor constraints or heterogeneity in TFP. Let ρ denote the pass-through rate characterized by the demand system, that is $\rho = \partial p/\partial(c/TFP)$, where c/TFP stands for the marginal cost of production. If firms are not facing constraints on input, then it's straightforward to derive the following condition based on the above TFPR equation:

$$\frac{\partial TFPR}{\partial TFP} = \left(\frac{|\varepsilon_{pq}|}{|\varepsilon_{pq}|-1} - \rho\right) \frac{c}{TFP}$$

In Foster et al. [2008], the linear demand systems allow firms to charge variable markups ($\rho < 1 < |\varepsilon|/(|\varepsilon|-1)$), therefore TFPR is increasing in TFP. For the case of constant elasticity of demand, $\rho = |\varepsilon|/(|\varepsilon|-1)$, so TFPR is independent of TFP. This is the case in Hsieh and Klenow [2009], where TFPR measures a firm-specific wedge from the optimum allocation. Finally, if ρ is sufficiently high, then TFP and TFPR could be negatively correlated, which means price fell by more than TFP increased. Diagnosing the causes of TFPR variation from the empirical literature is certainly beyond the scope of this survey, but readers should keep in mind that TFPR variation does not translate into TFP variation in a monotonic way.

2.3 Multinational Production

According to UNCTAD[14], 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[15], measured

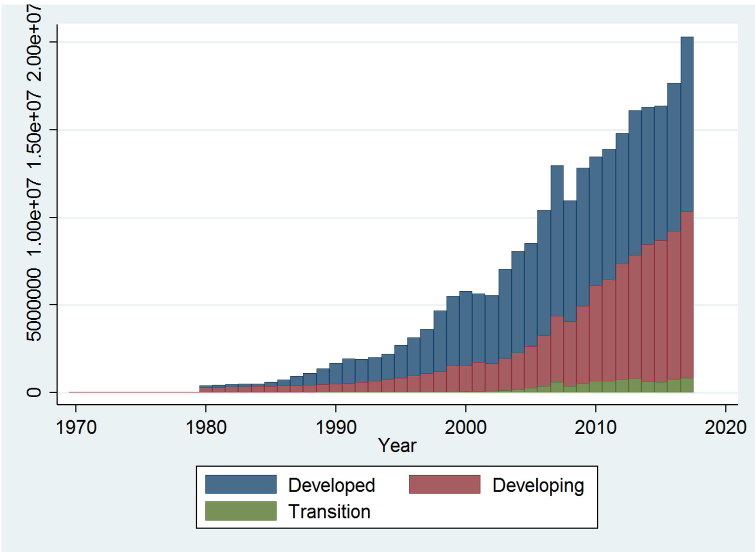
[13] Benfratello and Sembenelli [2006] is more of a causal analysis. Criscuolo and Martin [2009] and Kılınç [2014] try to separate TFPQ and TFPR as illustrated in Foster et al. [2008]. Barefoot and Mataloni Jr [2011] is more of a descriptive analysis, without estimating TFP. Bloom et al. [2012] uses labor productivity in their main analysis, however, they show similar trends hold for TFPR in the appendix of their working paper.

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

[15] It is important to note the difference between FDI data and MNE data. FDI statistics mainly provide information on cross-border capital flows, which may not contribute to the local economy. An example in case is the so-called Special Vehicle Entities (SPE). An increase in SPEs in a country is associated with an increase in inward and outward FDI flows for that country without real economic effects.

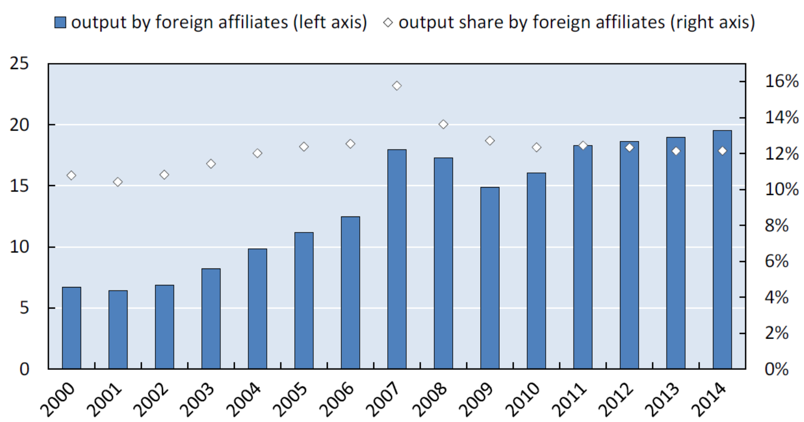
by the gross output by foreign affiliates, as shown in **Figure 5**. For instance, Bernard et al. [2009] find that in 2000, the top 1 percent of US exporters account for 81 percent of US exports. These superfirms produce in multiple countries and industries, and their activities go way beyond the mere 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.

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: Foreign Affiliates' Gross Output and Their Share in Global Output, 2000–2014



Note: Graph taken from Cadestin et al (2018) OECD Science, Technology and Industry Working Papers: *Multinational enterprises and global value chains: New Insights on the trade-investment nexus*. The plot is constructed based on OECD's Analytical AMNE database.

Second, FDI only measures part of what foreign affiliates use to finance their activities, ignoring the sources they draw from the local economy. Lastly, FDI excludes the contribution of labor, hence underestimating MNE activity in countries where labor is relatively more productive

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.

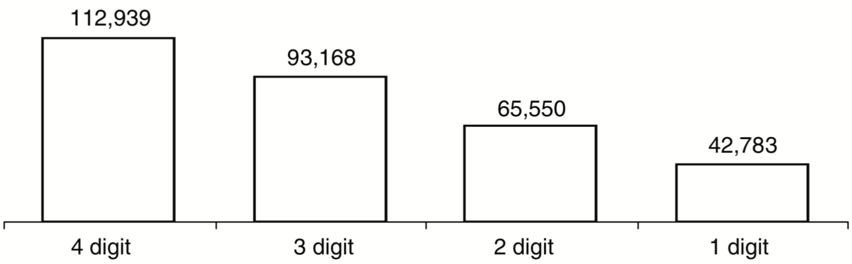
2.3.2 Vertical FDI

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 [2003b] 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 consensus was that horizontal (or market-seeking) FDI was the dominant format of multinational activity among the developed economies, using a new firm-level data set provided by Dun & Bradstreet (D&B)[16], Alfaro and Charlton [2009] find that most FDI flows between advanced economies actually do have a vertical dimension to them. They show that at the two-digit industry level, there is more horizontal FDI than vertical FDI. However, disaggregating to the four-digit level, the authors find that many of the foreign subsidiaries in the same two-digit industry as their parents are actually located in sectors that produce highly specialized inputs for their parents' production. The number of subsidiaries increases as the industry classification gets more disaggregated, as shown in **Figure 6**. A similar observation also appears in the number of employees.

[16]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.

Figure 6: 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.

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[17] 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[18] 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.

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

[17] A characteristic commonly associated with vertical FDI.

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

'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 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 majority of the discussion is dedicated to 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].

3.1 Trade Perspective

Baldwin and Forslid [2010] (henceforth, **BF**) study the impact of a reduction in the variable and fixed cost of trade in a two-country, two-sector Melitz [2003] model[19]. 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] (henceforth, **DRC**) analyze the optimal trade policy for a small open economy[20] in a one-sector Melitz-type model. Policy can affect welfare(utility per capita) in the economy through the following equation:

$$\frac{U}{L} = \text{Productivity Index*TOT Index*Variety Index*Curvature}$$

Two types of *distortions* exist in such an economy: (1) *Markup distortion*–there is too little spending on the domestic varieties relative to the social optimum due to markup distortion between home and foreign varieties. (2) *Entry distortion*–there are too few foreign varieties in the market outcome compared to the social optimum due to consumption externality[21]. Policy instruments affect the four components differently in the equation above. 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. To facilitate comparison across different frameworks, here I present the optimal import tariff formula:

$$t_{opt} = \frac{1}{\rho}\frac{\beta\rho}{\beta-\rho} > 1 \tag{1}$$

where β governs the shape of Pareto distribution(also the degree of firm heterogeneity) and ρ is related to the elasticity of substitution across the varieties. The first term$(1/\rho)$ can be viewed as an import tariff that addresses

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

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

[21]More specifically, domestic consumers are ignoring the positive impact on aggregate productivity through foreign producer entry.

the first distortion, and the second term($\beta\rho/(\beta-\rho)$) is an import subsidy that deals with the second distortion. Since markup distortion dominates the consumer surplus distortion, the optimal import tariff is greater than one. 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[22], which dominates the productivity gain from reallocation.

Felbermayr et al. [2013] (henceforth, **FJL**) extend Demidova and Rodriguez-Clare [2009]'s one-sector small-open-economy 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 non-tariff trade barriers. More specifically, the optimal import tariff for Home is

$$t_{opt} = \underbrace{\frac{\beta}{\beta-\rho}}_{\text{DRC}} + \underbrace{\frac{\rho}{\beta-\rho}\left(w_H^{-\frac{\beta-\rho}{\rho}} LB\eta t_F^{-\frac{\beta}{\rho}}\right)}_{\text{Terms-of-trade externality}} > t_{opt} \text{ in equation (1)} \tag{2}$$

where both β and ρ share the same meaning as in DRC. L stands for relative country size, B represents the relative productivity, and η measures the freeness of trade regarding non-tariff barriers. The first term in equation (2) shares the same expression as in equation (1), representing the product of markup and entry distortion. The second term stands for the terms-of-trade externality. Notice, if Home's relative effective market size goes to zero(Home becomes a small open economy as in DRC), the second term in equation (2) will disappear. The optimal import tariff collapses to the one in equation (1). In other words, optimal import tariff is smaller in a small open economy than in a large economy. This is because the entry distortion is never strong enough to offset the markup distortion in a small open economy. 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](henceforth, **HV**) study both domestic taxes and trade policies for a small open economy containing a flexible monopolistically competitive(MC) sector[23] 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 MC sector that eventually increases the number of varieties and improves the terms-of-trade. To compare with the aforementioned papers, let us focus on the case of import tariff for Home (the last line in their Table 3):

$$t_{opt} = \frac{\lambda_F}{\lambda_F-1}\{1-\frac{1-\mu}{\sigma(1-\mu)+(\sigma-1+\mu)s_D[(1-s_X)\eta+s_X(\lambda_H-1)]}\} \leq t_{opt} \text{ in equation (1)} \tag{3}$$

where $\lambda_i = \beta_i/\rho$, β and ρ share the same definition as in DRC. σ stands for the elasticity of substitution, μ stands for the expenditure share on the MC sector, η stands for the elasticity of labor supply, and s_D, s_X stand for the share of domestic firms' sales and export sales in domestic MC sector respectively. When $\mu = 1$, domestic

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

[23]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 [2010].

consumers only purchase the MC good(as in DRC), the term in the bracket in equation (3) collapses to 1 and (3) gives the same tariff level as in equation (1). 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.

Utilizing a primal approach and general Lagrange multiplier methods to characterize optimal wedges, Costinot et al. [2020] (henceforth, **CRCW**)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.

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] (henceforth, **HL**) 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 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.

Table 6 summarizes the papers surveyed in this subsection in a systematic way, with only one exception: Ossa [2011]. The main results in Ossa [2011] are built on 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. I highlight the results in the summary table but omit the discussion of this paper.

Table 6: Trade Policy with Heterogeneous Firms from Trade Perspective

Authors	Framework	Model Setup	Policy Instrument	Perspective	Policy Implication
BF(2010)	Trade/Static	HMY(2004) without FDI, each country has two sectors (MC, PC)[24] two country can differ in size	Iceberg cost (τ), beachhead costs in domestic and foreign market (F_D, F_X)	Positive	τ: (1) **symmetric country**, $\downarrow \tau$ could lead to anti-variety (pro-) effect if $F_X > (<) F_D$. (2) **asymmetric case**, $\downarrow \tau$ is anti-variety for small country, but still increases welfare; $\downarrow \tau$ is anti-variety for big country only when initial trade barriers are sufficiently high. $\downarrow F_X/F_D$ increases varieties for both small and big country when initial F_X/F_D is sufficiently low. Both nations gain from trade liberalization ($\downarrow \tau$ or $\downarrow F_X/F_D$).
DRC(2009)	Trade/Static	Small open economy one-sector Melitz model	Consumption subsidy (η), import tariff (t), export subsidy(s)	Normative, Positive	**Normative:** socially optimum outcome can be achieved through either one of the three instruments: η,s,t. The size of interventions depends on degree of heterogeneity and the elasticity of substitution across varieties. **Positive:** Assuming government has imposed optimal η, then welfare is decreasing in s, decreasing in t.
Ossa(2011)	Trade/Static	Multi-country version of BF(2010)	Import tariff (t)	Normative	**Two country**: Reciprocal $\downarrow$ ($\uparrow$) t monotonically increases (decreases) welfare in both countries.**Three country**: Multilateral $\downarrow$ ($\uparrow$) t monotonically increases (decreases) welfare in all countries. Any bilateral $\downarrow$ ($\uparrow$) t monotonically increases (decreases) welfare in those two countries at the cost of the third country.
FJL(2013)	Trade/Static	Two-country (H, F) Melitz model with asymmetric labor force and average productivity	Import tariff(t), which is affected by iceberg cost(τ), fixed market access cost(f)	Normative, Positive	**Positive**: $\uparrow t^H$ increases H's wage, domestic productivity cutoffs decline and export productivity cutoffs go up in both H and F. **Normative**: Optimal unilateral import tariff is increasing in relative country size and relative average productivity, decreasing in non-tariff barriers and the tariff level of foreign country. Near the symmetric equilibrium, (1) a higher productivity dispersion or elasticity of substitution increases the tariff level, (2) shutting down the firm selection channel results in higher optimal tariff level.
HV(2016)	Trade/Static	Small open economy version of BF(2010) with flexible labor supply adjustment between two sectors	Domestic subsidy(τ_D), import subsidy(τ_M), export subsidy(τ_X)	Normative	**First-best**: $\tau_{D,FB} < 1$ to subsidize the distortion generated in the domestic MC sector, $\tau_{M,FB} \leq 1$ to subsidize the number of foreign importers that are reduced due to domestic MC sector, $\tau_X = 1$ to control terms-of-trade **Second-best**: With $\tau_M = \tau_X = 1$, optimal $\tau_D < \tau_D^{FB}$ due to loss of imported varieties; With $\tau_D = \tau_X = 1$, optimal τ_M is set to balance the impact from increasing domestic varieties and the changes of terms-of-trade.
CRCW(2020)	Trade/Static	Melitz model with both heterogeneous variable costs and heterogeneous fixed costs[25].	Ad valorem taxes at firm level and at aggregate level	Normative	At the **micro-level**, import subsidies need to be applied to the least profitable foreign firms. Their analysis does not find any support for export subsidies or taxes on the least profitable domestic firms. At the **macro-level**, optimal taxes or subsidies that are uniform across all domestic firms can be useful since domestic consumers do not internalize that buying more from foreign firms may impact the domestic firms.
HL(2015)	Trade/Dynamic	Small open economy version of BF(2010) with an optimal stopping problem of export entry when facing trade policy uncertainty	Foreign import tariff(τ)	Positive	With **deterministic** trade policy, reduction in tariffs increase demand and induce firm entry. With trade policy **uncertainty**(TPU), both the frequency of policy shock arrivals and firm's expectation about future tariffs affect exporter entry. The authors find quantitative evidence that suggest the elimination of TPU generates substantial growth of trade, especially the entry of exporters, following Portugal's accession to European Community.

[24] MC: monopolistically competitive, PC: perfectly competitive

[25] The original paper is also extended to multiple factors of production, multiple countries, strategic tax setting, and with general preferences and technology. The authors show that their results are robust.

3.2 Macro Perspective

Different from the conventional steady-state analysis in the trade literature, Larch and Lechthaler [2013] (henceforth, **LL**) is the first to study optimal trade policy in a full-fledged macro DSGE framework. By introducing non-resource-consuming and revenue-generating tariffs into Ghironi and Melitz [2005], and allowing trade policy makers to focus on a shorter horizon, the authors find that the Nash tariffs in the static model is higher than the Nash tariffs in the dynamic macro model. This is because a decrease in tariff gives consumption a short-run boost through the increase of imports. Consumers enjoy more varieties at lower prices before the inefficient firms are driven out of the market. The shortsighted politicians tend to set lower tariffs than politicians with a longer planning horizon. Hence, in a dynamic setting, the shortsightedness of politicians is actually welfare-improving.

Utilizing the number of anti-dumping investigations as a measure of protectionism, Barattieri et al. [2021] (henceforth, **BCG**) explore trade policy effectiveness at the macroeconomic level. The authors use panel structural VARs estimation and find that temporary trade barriers act as a supply shock, causing the output to reduce and giving inflationary pressure in the short-run, with only a minor positive effect on the trade balance. The authors then build a small open economy version of the Ghironi and Melitz [2005] model and study the aggregate implications for protectionism. Their 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' earnings conference calls, and aggregate data on tariff rates in the US economy, Caldara et al. [2020] (henceforth, **CIMPR**) 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.

The papers mentioned in this subsection are summarized in **Table 7**, with two exceptions: Ghironi and Melitz [2005] (henceforth, **GM**) and Sampson [2016]. The authors in these papers only discuss trade policy implications briefly in their appendices. Hence a detailed description of these papers is omitted here.

Table 7: Trade Policy with Heterogeneous Firms from Macro Perspective

Authors	Framework	Model Setup	Policy Instrument	Perspective	Policy Implication
GM(2005)	Macro/DSGE	A DSGE version of Melitz model	Iceberg costs(τ), fixed export cost(f_X)	Positive	A permanent bilateral reduction in τ or f_X induces an increase in the number of exporting firms, along with a decrease in the export productivity cutoff. Increased export competition induces a small reduction in the number of domestic firms. The total number of varieties increases. Welfare gain from reduction in τ is larger than that from f_X since the latter does not affect any marginal trade costs.
LL(2013)	Macro/DSGE	GM(2005) with revenue-generating import tariffs	Import tariff(t)	Normative	Given the shortsightedness of policymaker, the Nash tariffs that take into account adjustment dynamics is lower than the Nash tariffs based on a static trade model. Shortsighted politicians tend to set lower tariffs than politicians with a longer planning horizon.
Sampson(2016)	Macro/Growth	Dynamic Melitz model with knowledge spillovers from incumbent to entrants	Tax on fixed costs (v) and subsidy on entry (v_e)	Normative, Positive	**Positive**: Either taxing the fixed cost or subsidizing entry leads to higher welfare by raising the firm creation rate since innovators do not internalize the knowledge spillovers that entry generates. **Normative**: with constant return to scale in entry, and if government chooses v and v_e jointly, then the optimal v and v_e are positive. However, if there is congestion in technology diffusion, the optimal v_e might be negative.
BCG(2021)	Macro/DSGE	Small open economy version of GM(2005) with sticky nominal wages	Import tariffs[26]	Positive	For small open economies, a temporary increase in tariffs acts as a supply shock, causing output to fall and inflation to rise in the short-run. It is costly, even when it is only used temporarily, even when economies are facing liquidity traps, and irrespective of the flexibility of exchange rate and trading partners' retaliation.
Steinberg(2019)	Macro/DSGE	Three country, input-output production structure, firm heterogeneity, and trade costs uncertainty	Various post-Brexit trade policy regimes modeled as a combination of import tariffs(τ_d), non-tariff barriers(ε_d)	Positive	In the long run, Brexit will have a large impact on the UK macroeconomy: the announcement of the referendum and the outcome of the vote have little impact on UK in the short run. Once Brexit occurs, export participation, trade flows, and macroeconomic variables all decline to long run levels. Hard Brexit will bring UK the most severe welfare losses. Uncertainty about trade policy only has a very limited impact on the UK economy.
CIMPR(2020)	Macro/DSGE	Two-country general equilibrium model with sticky wages, firm heterogeneity and endogenous export decision	Expectation of future tariffs, uncertainty about future tariffs[27]	Positive	Both higher expected tariffs and increased uncertainty about future tariffs deter investment, with exporters accumulating less capital than non-exporters. Nominal rigidity and fixed costs of export provide important transmission mechanisms of the effects of trade policy uncertainty.

[26] The empirical evidence in this paper is based on antidumping investigations as temporary trade barriers. This barrier is then modeled as ad valorem import tariffs in the model.

[27] The former one is modeled as a first moment shock, and the latter as a second moment shock, both to the tariff process, which is a first-order autoregressive process with stochastic volatility.

4 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, followed by the survey of trade policy when heterogeneous firms charge variable markups.

4.1 Demand-side Approach: Deviating from CES Preference

Melitz and Ottaviano [2008] (henceforth, **MO**) 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[28], 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.

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*[29] 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[30].

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

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

[29] 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].

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

even when the productivity distribution of firms follows a bounded Pareto[31]. 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[32]. 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[33].

Similarly, Behrens et al. [2014] obtain the variable markup by introducing *additively quasi-separable (AQS) preference*[34] as in Behrens and Murata [2007, 2012]. The framework yields 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 *nonhomothetic 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

[31]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.

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

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

[34]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).

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[35] 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, three 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[36], 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 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.

Mrazova et al. [2021] take an important step forward in understanding the assumptions about productivity distribution, demand structure, and their implication for the distribution of markups and other firm outcomes. In the Melitz-type models, the authors find that the CREMR (Constant Revenue Elasticity of Marginal Revenue) demand function is necessary and sufficient for the distributions of firm-level productivity and firm sales to be

[35]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.

[36]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.

members of the same category. Under additive separability, CREMR is also necessary and sufficient for the growth rate of firm sales to be independent of firm size (Gibrat's Law). Utilizing Indian firm-level data, the authors show that this condition is able to generate a better empirical fit to the markup distribution than other better-known demand functions. The authors conclude that the choice between CREMR and other types of demand features is more important than different choices of firm productivity distributions in generating these results.

4.2 Supply-side Approach: Deviating from Monopolistic Competition

4.2.1 Bertrand Competition

Around the same time as Melitz [2003], Bernard et al. [2003a] (henceforth, **BEJK**) 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. [2003a] 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. [2003a], 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.

4.2.2 Cournot Competition

Instead of competing in prices, Atkeson and Burstein [2008] introduce variable markups through quantity competition á la Cournot[37] into a nested CES demand system as in Helpman and Krugman [1985]. Contrary to Melitz [2003] and Bernard et al. [2003a], where there are an infinite number of firms in an industry, Atkeson

[37]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. [2003a], see Atkeson and Burstein [2007] for more details.

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[38] 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 necessarily explain the macro-level markup adjustment.

4.3 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*[39],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.

4.4 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. First, the author finds, regardless of country size, that the optimal import tariff is strictly positive from a unilateral perspective. To compare with DRC and FJL in Section 3.1, the optimal import tariff for Home country is given by the following equation:

$$t_{opt} = \underbrace{\frac{\beta}{\beta+1}}_{>\text{DRC}} + \underbrace{\frac{1}{\beta}\frac{L_H}{L_F}\left(\frac{c_F^M}{c_H^M \tau_{HF}}\right)^{\beta} w^{-\beta}}_{\text{Terms-of-trade externality}} \tag{4}$$

[38]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.

[39]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.

where β stands for the Pareto shape parameter that governs the degree of firm heterogeneity. c_F^M/c_H^M represents the relative productivity, τ_{HF} stands for the openness of the Home country, and L_H/L_F stands for the relative market size. In the case of a small open economy, the terms-of-trade externality will go to zero. However, the remaining term($\beta/(\beta+1)$) is still higher than the tariff level in equation (1). The reason is that protection in the presence of variable markups reallocate resources from high markup firms to low markup ones and reducing misallocation distortion in the economy. So variable markups encourage a higher level of protection. Second, 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[40]. Consequently, the welfare gain from trade following liberalization[41] 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 [2020] 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 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[42] 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.

The papers mentioned in this subsection are summarized in **Table 8**. Whenever possible, the key results in these papers are contrasted with papers in Section 3.1.

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

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

[42]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.

Table 8: Trade Policy with Heterogeneous Firms and Variable Markups

Authors	Model Setup	Policy Instrument	Perspective	Policy Implication	Compare to Section 3.1
Demidova(2017)	Variable markups introduced through demand side: two-country(asymmetric) one-sector MO(2008) (without the homogeneous good sector)	Ad valorem tariff(t), iceberg transportation cost(τ)	Normative, positive	**Normative:** optimal unilateral import tariff is strictly positive, regardless of liberalizing country's size. **Positive:** reduction in τ in Home country improves Home's welfare and can increase Foreign's welfare if both countries are large.	**Compared to DRC(2009)**: if only focus on import tariff, which is explored in both papers, the two papers deliver very similar normative results. In the presence of firm heterogeneity, variable markups reduce welfare gain from trade due to the misallocation distortion they create. **Compared to FJL(2013)**: The optimal tariff level here is higher in Demidova(2017). The survivors due to tariff protection in this model are the ones with the lowest markups. Put differently, protectionism leads to an additional gain by shifting resources from high markup firms to low markup firms and reducing misallocation. So the presence of variable markups encourages a higher level of protection.
BL(2020)	Variable markups introduced through demand side: two-country MO(2008)	Import tariff, export subsidy	Normative/-positive[43]	**Positive:** (1) Starting from free trade, a country can improve its welfare by imposing a small positive import tariff, or a small export subsidy, or a combination of both. All these policies, however, are *beggar-thy-neighbor*.(2) Starting from symmetric Nash equilibrium, a bilateral reduction in import tariff, export tariff[44], or a combination of both can be welfare-improving for both countries. **Normative**: Nash tariffs are inefficient, but politically optimal[45] policies are efficient. Under certain conditions, optimal import tariff and export tariff may both be positive.	**Compared to Ossa(2011) Appendix Section 1**: Both studies find mutual reduction of import tariff is welfare-improving for both countries. However, there are several crucial differences: (1) Ossa's definition of reciprocity eliminates all trade policy externalities. In particular, the number of entrants in the economy is unaffected by reciprocal tariff reduction. In BL, mutual reduction in tariff does affect the number of entrants. (2) BL assumes the model starts from symmetric Nash equilibrium, whereas Ossa's conclusion is not based on the assumption of symmetry.
NOS(2019)	Variable markups introduced through demand side: multi-country MO(2008) with asymmetric sunk entry cost(f), state of technology(c_M), and country size(L)	Variety-specific consumption/production subsidies/taxes, lump-sum transfers for consumers and firms	Normative	**First-best**: To remove product mix distortion, trade by low-cost firms is subsidized whereas trade by high-cost firms is taxed. These transfers are bigger for more disadvantaged firms. To remove product selection and range distortion, firm entry should be reduced in all countries, especially in disadvantaged ones. **Second-best**(without variety/firm-specific taxes):compensate the welfare loss due to product mix distortion with per-unit trade subsidy common to all firms, especially the disadvantaged countries, and a lump-sum profit tax common to all firms. **Third-best**(without lump-sum transfers): compensate the welfare loss due to product mix and product selection via per-unit trade subsidy common to all firms, especially in more disadvantaged countries.	Not available– the papers in Section 3.1 either use an incomparable framework or do not study socially efficient policies.

[43] This study does not provide specific solutions for optimal policies. Due to the complicated feature of quadratic quasi-linear preference, the normative exercise is mainly based on simulation.

[44] This is equivalent to an increase in export subsidy

[45] Politically optimal means a welfare function composed of local and world prices. This welfare function is utilized to illustrate the terms-of-trade rationale for trade agreement in this model.

5 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. The section is concluded with a discussion on the trade policy implications when heterogeneous firms can access foreign markets either through export or FDI.

5.1 Horizontal FDI

Helpman et al. [2004] (henceforth, **HMY**) 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 HMY. 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. [2003a][46]. 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

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

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[47] version of HMY 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 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.

5.2 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[48], 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] (henceforth, **AH**) 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 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

[47] 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.

[48] 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.

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.

5.3 Complex FDI

Building on Yeaple [2003a], 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 HMY, 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 HMY. 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 HMY. 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.

As a wrap-up of Section 5.1 and Section 5.2, Mrázová and Neary [2019] go to the root of different orders of selection effects in this literature. The first-order selection effect, which means whether firms enter or not(such as production or export), is very robust among many trade models. However, the second-order selection effect, which means how firms serve a market(such as horizontal FDI or vertical FDI) conditional on entry, is much less robust. Utilizing the monotone comparative statics advocated by Milgrom and Roberts [1990], the authors find the sufficient condition for the first-order selection effect to hold is that the ex post profit function is monotonically decreasing in marginal production cost. As for the second-order selection effect, the authors find whether a firm's optimal profits are *supermodular* in its marginal production cost and in the marginal cost of serving the market under different access modes is crucial in preserving the conventional second-order selection effect, i.e., more productive firms select into activities with lower marginal costs. By contrast, if profits are *submodular*, then a reverse sorting pattern may occur. When production and market-access costs affect profits multiplicatively, *supermodularity* is only preserved when demand is more convex than the CES demand function with the same elasticity, which is not true for many widely used demand systems.

5.4 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 HMY 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. [2003a], 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 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[49], 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.

Wang [2017] incorporates horizontal FDI into the framework developed by Costinot et al. [2020] to study optimal trade and FDI policies. Focusing on firm-specific taxes, the author finds that optimal trade and FDI policy requires firm-level taxes that discriminate against the most profitable foreign firms, whether they access the Home market through export or FDI. From Home's perspective, it is beneficial to promote the entry of the marginal Foreign firms because the fixed costs that the marginal Foreign firms face are higher than the relevant Home's welfare loss. In contrast, the optimal export tax should be uniform across all exporters. However, different from Costinot et al. [2020], the Lerner symmetry between import tariffs and export taxes breaks down in the presence of FDI. This is mainly due to the cross-border ownership of FDI, which generates international transfers that annihilates trade balance.

The papers mentioned in this subsection are summarized in Table **9**, with only one exception: Chor [2009], where the author's main focus is FDI policy, not trade policy, so I omit this paper in the current survey. Whenever possible, the key results in these papers are contrasted with papers in Section 3.1.

[49]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.

Table 9: Trade Policy with Firm Heterogeneity and Multinational Production

Authors	Model Setup	Policy Instrument	Perspective	Policy Implication	Compare to Section 3.1
Chor(2009)	**Horizontal FDI**: Two-country HMY model with additive separable preference	Import subsidy to Home exporters either on fixed cost(s_f) or variable production cost(s_v)	Normative	Under certain parameter restrictions, the unilateral optimal import subsidy that maximizes welfare in Foreign[50] is strictly positive. The consumption gain from drawing in more firms at the export cutoff margin outweigh the loss of FDI firms at the FDI cutoff margin.	**Compared to DRC(2009)**: Import subsidy will induce some Home firms to start exporting, but also prompt some Home FDI firms to switch from FDI back to export. The markup distortion and consumer surplus distortion appear in both frameworks. In DRC, the former dominates the latter, so the optimal policy is to impose a positive import tariff to raise the consumption of domestic varieties. In Chor, an import tariff will introduce further distortion due to the presence of the homogeneous good sector. To achieve the first-best allocation results, an import subsidy together with a consumption subsidy for domestic varieties are needed.
CD(2011)	**Horizontal FDI**: Two-country HMY model with heterogeneous fixed costs of production for different production modes	Ad-valorem import tariff	Normative	**First-best**: optimal tariff for social planner is a subsidy regardless the presence of FDI firms. **Nash tariff**: the Nash tariff level is higher than the socially optimal tariff regardless the presence of FDI firms. For a specific form of fixed cost, the presence of FDI lowers Nash tariff level.	**Compared to HV(2016) Section 4.1**: In HV, optimal import tariff is strictly positive and decreasing in the elasticity of substitution(σ). In CD, for small σ, optimal import tariff is negative and increasing in σ. Only when σ is big enough, optimal import tariff starts to fall and becomes decreasing in σ. The difference mainly comes from CD's model setup abstracts from the Pareto distribution assumption regarding the firm's productivity as in HV, which guarantees the optimal tariff in HV is always positive.
dBR(2013)	**Horizontal FDI**: two-country, finite number of rivals version of BEJK	Iceberg type trade costs(d)	Positive	For a given trade cost level, high contestability and technology in the exporting country relative to the destination country make it more likely for firms to tariff-jump through cross-border M&A in order to avoid competition from a firm's native market, but with the motivation of increasing the markup charged in the foreign market due to the elimination of trade costs.	Not available since none of the papers in Section 3.1 are based on BEJK's framework
Wang(2017)	**Horizontal FDI**: extend CRCW to include horizontal FDI options for firms	Ad-valorem firm-specific trade taxes/subsidies	Normative	Unilateral optimal taxes for Home should be such that: (1) uniform taxes for all domestic producers; (2) uniform taxes for all exporters, (3) for most profitable Foreign exporters impose uniform import taxes, for least profitable Foreign exporters, impose an import taxes that is increasing in exporter's profitability; (4) taxes scheme similar to (3) should also be applied to Foreign FDI firms	**Compared to CRCW(2020)**: (1) In line with CRCW, Home country finds it optimal to discriminate among the foreign FDI firms by lowering the import tariff or sales taxes for the least profitable ones. (2) Different from CRCW, at the macro-level, Home's problem is no longer a standard terms-of-trade manipulation problem due to the international transfers from foreign FDI firms, which breaks down the Lerner symmetry between import tariffs and export taxes. The presence of the numeraire good sector in this paper should also be noted when it comes to the discussion of the welfare implication of trade policy.
Diez(2014)	**Vertical FDI**: North and South trade model of AH(2004) with offshoring modeled as foreign sourcing of assembly services	Import tariff imposed by North(τ_N) and South(τ_S)	Positive	$\tau_N \uparrow$: decreases the market share of offshoring firms and decreases the relative market share of outsourcing firms vs. vertically integrated firms in both countries. $\tau_S \uparrow$: increases the market shares of offshoring and of outsourcing firms in both countries.	Not available since none of the papers in Section 3.1 is based on AH(2004) framework.

[50] This is also the country that uses FDI policy to attract FDI firms from the other country(Home).

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 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 multinational production. Second, 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 observed through the firm-level data.

Providing policy suggestions but ignoring the previous two discrepancies between theory and data could be misleading. On the one hand, an increase in trade barriers may hurt consumer welfare in the presence of variable markups if protection results in higher market concentration. On the other hand, in a highly integrated global market, foreign firms can avoid trade barriers by locating production within the destination market. Such "tariff-jumping" activities can diminish the market power of domestic producers, thereby substantially mitigate the welfare consequences of the original trade policy. Theoretical frameworks that combine these two prevailing trends are needed in order to have an appropriate policy evaluation.

Careful readers should also notice that most of the trade policy papers surveyed here only have labor as the unique factor for production. Adding other production factors will greatly enrich firm's behavior and can easily overturn the results established in the literature. For example, Spearot [2013] considers a variant of Melitz and Ottaviano [2008] model where firms can expand through capital acquisitions either in the domestic or foreign market. Focusing on the positive analysis and ignoring the selection channel, the author finds that a novel response of the acquisition market may produce surprising trade policy implications. With an acquisition market, relatively unproductive firms may have an incentive to sell their assets to firms that are more efficient in utilizing those resources. In particular, if firms optimally substitute between cross-border acquisitions and no acquisitions, then trade liberalization may reduce aggregate productivity by reducing foreign acquisition demand. This results in a non-monotonic sorting of firms into acquisitions, which is in line with the insights from Mrázová and Neary [2019].

Another important aspect is the Pareto distribution vastly adopted in this literature. As discussed by Feenstra [2018], if the support of cost distribution become bounded, other channels that will affect the pro-competitive effect of trade will begin to work, delivering different welfare implication of trade. On the other hand, the combination of Melitz and Pareto implies that trade cost will only affect export through extensive margin, but this is at odds with the empirical fact that most of the adjustments happen along the intensive margin. Introducing log-normal distribution, as discussed in Fernandes et al. [2017], could be a good starting point. Considering the CREMR demand function developed by Mrazova et al. [2021] would be even better since the results related to the CREMR function are robust across various kinds of firm productivity distributions.

Space constraints clearly limit the choice of material covered in this survey. However, the literature seems also overwhelmingly favor Melitz-type model due to its analytical tractability. Most of the paper discussed in this survey is built on Krugman [1980, 1979] international trade theory, extended by Melitz [2003] to introduce heterogeneity, only a few papers are built on the alternative framework, such as Bernard et al. [2003b]. In addition, many of the trade policy studies surveyed here are based on the assumption of monopolistic competition, and besides entry and exit, trade policy almost has no impact on the strategic behavior of firms. Either Bertrand or Cournot competition could bring more interesting interactions among the firms. As more pieces of firm-level evidence are being discovered, much more research is needed on the trade policy implication of firm heterogeneity, especially on the understanding of markups and how they interact with multinational production in the era of global value chains.

References

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

Alfaro, L. and Charlton, A. (2009). Intra-industry foreign direct investment. *American Economic Review*, 99(5):2096–2119.

Alviarez, V., Head, K., and Mayer, T. (2020). Global giants and local stars: How changes in brand ownership affect competition.

Amiti, M., Itskhoki, O., and Konings, J. (2014). Importers, exporters, and exchange rate disconnect. *American Economic Review*, 104(7):1942–78.

Amiti, M., Itskhoki, O., and Konings, J. (2019). International Shocks, Variable Markups, and Domestic Prices. *The Review of Economic Studies*.

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

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

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

Antràs, P. and Helpman, E. (2009). Contractual frictions and global sourcing. *The organization of firms in a global economy*.

Antràs, P. and Yeaple, S. R. (2014). Multinational firms and the structure of international trade. *Handbook of International Economics*, 4:55–130.

Arkolakis, C. (2008). A generalized solution of the monopolistic competition model with heterogeneous firms and a linear demand (melitz-ottaviano). *Working paper, Department of Economics, Yale University*.

Arkolakis, C., Costinot, A., Donaldson, D., and Rodríguez-Clare, A. (2018). The elusive pro-competitive effects of trade. *The Review of Economic Studies*.

Atkeson, A. and Burstein, A. (2007). Pricing-to-market in a ricardian model of international trade. *American Economic Review*, 97(2):362–367.

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

Bagwell, K. and Lee, S. H. (2020). Trade policy under monopolistic competition with firm selection. *Journal of International Economics*, 127.

Baldwin, R. and Okubo, T. (2014). Networked fdi: Sales and sourcing patterns of japanese foreign affiliates. *The World Economy*, 37(8):1051–1080.

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

Baltagi, B. H., Egger, P., and Pfaffermayr, M. (2007). Estimating models of complex fdi: Are there third-country effects? *Journal of Econometrics*, 140(1):260–281.

Barattieri, A., Cacciatore, M., and Ghironi, F. (2021). Protectionism and the business cycle. *Journal of International Economics*, 129:103417.

Barefoot, K. B. and Mataloni Jr, R. J. (2011). Operations of us multinational companies in the united states and abroad. *Survey of Current Business*, 91(11):29–48.

Behrens, K., Mion, G., Murata, Y., and Südekum, J. (2014). Trade, wages, and productivity. *International Economic Review*, 55(4):1305–1348.

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

Behrens, K. and Murata, Y. (2012). Trade, competition, and efficiency. *Journal of International Economics*, 87(1):1–17.

Bellone, F., Musso, P., Nesta, L., and Warzynski, F. (2014). International trade and firm-level markups when location and quality matter. *Journal of Economic Geography*, 16(1):67–91.

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

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

Berlingieri, G., Pisch, F., and Steinwender, C. (2018). Organizing global supply chains: Input cost shares and vertical integration. Working Paper 25286, National Bureau of Economic Research.

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

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

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

Bernard, A. B., Jensen, J. B., Redding, S. J., and Schott, P. K. (2012). The empirics of firm heterogeneity and international trade. *Annu. Rev. Econ.*, 4(1):283–313.

Bernard, A. B., Jensen, J. B., Redding, S. J., and Schott, P. K. (2018). Global firms. *Journal of Economic Literature*, 56(2):565–619.

Bernard, A. B., Jensen, J. B., and Schott, P. K. (2009). *Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods*, pages 513–552. University of Chicago Press.

Bernard, A. B. and Wagner, J. (1997). Exports and success in german manufacturing. *Weltwirtschaftliches Archiv*, 133(1):134–157.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262.

Bertoletti, P., Etro, F., and Simonovska, I. (2018). International trade with indirect additivity. *American Economic Journal: Microeconomics*, 10(2):1–57.

Bilbiie, F., Ghironi, F., and Melitz, M. (2012). Endogenous entry, product variety, and business cycles. *Journal of Political Economy*, 120(2):304–345.

Bilbiie, F. O., Ghironi, F. P., and Melitz, M. J. (2016). Monopoly power and endogenous product variety: Distortions and remedies. *American Economic Journal: Macroeconomics (Forthcoming)*.

Bloom, N., Sadun, R., and Van Reenen, J. (2012). Americans do it better: Us multinationals and the productivity miracle. *American Economic Review*, 102(1):167–201.

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

Brainard, S. L. (1993). A simple theory of multinational corporations and trade with a trade-off between proximity and concentration. Technical report, National Bureau of Economic Research.

Caldara, D., Iacoviello, M., Molligo, P., Prestipino, A., and Raffo, A. (2020). The economic effects of trade policy uncertainty. *Journal of Monetary Economics*, 109:38–59.

Carballo, J. (2018). 3 global sourcing under uncertainty. *Developments in Global Sourcing*, page 71.

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

Caves, D. W., Christensen, L. R., and Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica: Journal of the Econometric Society*, pages 1393–1414.

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

Chor, D. (2009). Subsidies for fdi: Implications from a model with heterogeneous firms. *Journal of International Economics*, 78(1):113–125.

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

Costinot, A., Rodriguez-Clare, A., and Werning, I. (2020). Micro to macro: Optimal trade policy with firm heterogeneity. *Econometrica*, 88(6):2739–2776.

Criscuolo, C. and Martin, R. (2009). Multinationals and us productivity leadership: evidence from great britain. *The Review of Economics and Statistics*, 91(2):263–281.

de Blas, B. and Russ, K. N. (2013). Hymer's multinationals. *Journal of Economic Behavior & Organization*, 94:381–392.

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

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

Demidova, S. (2017). Trade policies, firm heterogeneity, and variable markups. *Journal of International Economics*, 108:260–273.

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

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

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

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

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

Dobbelaere, S. and Kiyota, K. (2018). Labor market imperfections, markups and productivity in multinationals and exporters. *Labour Economics*, 53:198–212.

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

Eaton, J., Kortum, S., and Kramarz, F. (2004). Dissecting trade: Firms, industries, and export destinations. *American Economic Review*, 94(2):150–154.

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

Eaton, J., Kortum, S., and Sotelo, S. (2012). International trade: Linking micro and macro. In *Advances in Economics and Econometrics, Tenth World Congress*, volume 2.

Edmond, C., Midrigan, V., and Xu, D. Y. (2015). Competition, markups, and the gains from international trade. *American Economic Review*, 105(10):3183–3221.

Edmond, C., Midrigan, V., and Xu, D. Y. (2018). How costly are markups? Working Paper 24800, National Bureau of Economic Research.

Egger, P. (2008). On the role of distance for outward fdi. *The Annals of Regional Science*, 42(2):375–389.

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

Engel, C. (2002). Expenditure switching and exchange-rate policy. *NBER macroeconomics annual*, 17:231–272.

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

Feenstra, R. C. and Weinstein, D. E. (2017). Globalization, markups, and us welfare. *Journal of Political Economy*, 125(4):1040–1074.

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

Fernandes, A., Klenow, P., Pierola, M. D., Rodriguez-Clare, A., and Meleshchuk, S. (2017). The intensive margin in trade.

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

Garetto, S., Oldenski, L., and Ramondo, N. (2019). Multinational expansion in time and space. Working Paper 25804, National Bureau of Economic Research.

Gaubert, C. and Itskhoki, O. (2018). Granular comparative advantage. Working Paper 24807, National Bureau of Economic Research.

Gervais, A. (2015). Product quality and firm heterogeneity in international trade. *Canadian Journal of Economics/Revue canadienne d'économique*, 48(3):1152–1174.

Ghironi, F. and Melitz, M. (2005). International trade and macroeconomic dynamics with heterogeneous firms. *Quarterly Journal of Economics*, 120:865–915.

Grossman, G. M., Helpman, E., and Szeidl, A. (2006). Optimal integration strategies for the multinational firm. *Journal of international economics*, 70(1):216–238.

Gumpert, A., Moxnes, A., Ramondo, N., and Tintelnot, F. (2017). The life-cycle dynamics of exporters and multinational firms. Working Paper 24013, National Bureau of Economic Research.

Haaland, J. I. and Venables, A. J. (2016). Optimal trade policy with monopolistic competition and heterogeneous firms. *Journal of International Economics*, 102:85–95.

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

Hall, R. E., Blanchard, O. J., and Hubbard, R. G. (1986). Market structure and macroeconomic fluctuations. *Brookings papers on economic activity*, 1986(2):285–338.

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

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

Hanson, G. H., Mataloni Jr, R. J., and Slaughter, M. J. (2005). Vertical production networks in multinational firms. *Review of Economics and statistics*, 87(4):664–678.

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

Helpman, E. and Krugman, P. R. (1985). *Market structure and foreign trade: Increasing returns, imperfect competition, and the international economy*. MIT press.

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

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

Hornok, C. and Muraközy, B. (2018). Markups of exporters and importers: Evidence from hungary. *The Scandinavian Journal of Economics*.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Irarrazabal, A., Moxnes, A., and Opromolla, L. D. (2013). The margins of multinational production and the role of intrafirm trade. *Journal of Political Economy*, 121(1):74–126.

Jackson, L. F. (1984). Hierarchic demand and the engel curve for variety. *The Review of Economics and Statistics*, pages 8–15.

Keller, W. and Yeaple, S. R. (2020). Multinationals, markets, and markups. Technical report, Working Paper.

Kılınç, U. (2014). Estimating entrants' productivity when prices are unobserved. *Economic Modelling*, 38:640–647.

Kimball, M. S. (1995). The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit, and Banking*, 27(4).

Kokovin, S., Parenti, M., Thisse, J.-F., and Ushchev, P. (2017). On the dilution of market power. *CEPR Discussion Paper No. DP12367*.

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

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

Larch, M. and Lechthaler, W. (2013). Whom to send to doha? the shortsighted ones! *Review of Economic Dynamics*, 16(4):634 – 649.

Li, Y., Smeets, V., and Warzynski, F. (2020). Physical productivity and exceptional exporter performance: Evidence from a chinese production survey. Technical report, Working Paper, Hong Kong University of Science and Technology.

Lu, D. (2010). Exceptional exporter performance? evidence from chinese manufacturing firms. *manuscript, University of Chicago*.

Mayer, T. and Ottaviano, G. I. (2008). The happy few: The internationalisation of european firms. *Intereconomics*, 43(3):135–148.

Melitz, M. (2003). The impact of trade on intra-industry reallocations and aggre- gate industry productivity. *Econometrica*, 71:1695–1725.

Melitz, M. and Ottaviano, G. (2008). Market size, trade, and productivity. *Review of Economic Studies*, 75(295-316).

Melitz, M. J. and Redding, S. J. (2014). Heterogeneous firms and trade. In *Handbook of international economics*, volume 4, pages 1–54. Elsevier.

Milgrom, P. and Roberts, J. (1990). The economics of modern manufacturing: Technology, strategy, and organization. *The American Economic Review*, pages 511–528.

Mrázová, M. and Neary, J. P. (2019). Selection effects with heterogeneous firms. *Journal of the European Economic Association*, 17(4):1294–1334.

Mrazova, M., Neary, P., and Parenti, M. (2021). Sales and markup dispersion: Theory and empirics. forthcoming. *Econometrica*.

Muraközy, B. and Russ, K. N. (2015). Competition with multinational firms: Theory and evidence. Technical report, IEHAS Discussion Papers.

Nocco, A., Ottaviano, G. I., and Salto, M. (2019). Geography, competition, and optimal multilateral trade policy. *Journal of International Economics*.

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

Ossa, R. (2011). A "new trade" theory of gatt/wto negotiations. *Journal of Political Economy*, 119(1):122–152.

Ottaviano, G., Tabuchi, T., and Thisse, J.-F. (2002). Agglomeration and trade revisited. *International Economic Review*, pages 409–435.

Parenti, M. (2018). Large and small firms in a global market: David vs. goliath. *Journal of International Economics*, 110:103–118.

Parenti, M., Ushchev, P., and Thisse, J.-F. (2017). Toward a theory of monopolistic competition. *Journal of Economic Theory*, 167:86–115.

Ramondo, N. and Rappoport, V. (2010). The role of multinational production in a risky environment. *Journal of International Economics*, 81(2):240–252.

Ramondo, N., Rappoport, V., and Ruhl, K. J. (2013). The proximity-concentration tradeoff under uncertainty. *Review of Economic Studies*, 80(4):1582–1621.

Ramondo, N., Rappoport, V., and Ruhl, K. J. (2016). Intrafirm trade and vertical fragmentation in us multinational corporations. *Journal of International Economics*, 98:51–59.

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

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

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

Sampson, T. (2016). Dynamic selection: an idea flows theory of entry, trade, and growth. *The Quarterly Journal of Economics*, 131(1):315–380.

Sembenelli, A. and Siotis, G. (2008). Foreign direct investment and mark-up dynamics: Evidence from spanish firms. *Journal of International Economics*, 76(1):107–115.

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

Spearot, A. C. (2013). Market access, investment, and heterogeneous firms. *International Economic Review*, 54(2):601–627.

Steinberg, J. B. (2019). Brexit and the macroeconomic impact of trade policy uncertainty. *Journal of International Economics*, 117:175–195.

Temouri, Y., Driffield, N. L., and Higón, D. A. (2008). Analysis of productivity differences among foreign and domestic firms: evidence from germany. *Review of World Economics*, 144(1):32–54.

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

UNCTAD (2011). World investment report 2011. *United Nations Publication*.

Voigtlaender, N. and Garcia-Marin, A. (2019). Exporting and plant-level efficiency gains: It's in the measure. *Journal of Political Economy*.

Wang, S. (2017). Three essays on international trade. *Doctoral Dissertation*.

Yeaple, S. R. (2003a). The complex integration strategies of multinationals and cross country dependencies in the structure of foreign direct investment. *Journal of international economics*, 60(2):293–314.

Yeaple, S. R. (2003b). The role of skill endowments in the structure of us outward foreign direct investment. *Review of Economics and statistics*, 85(3):726–734.

Yeaple, S. R. (2013). The multinational firm. *Annu. Rev. Econ.*, 5(1):193–217.

Zhelobodko, E., Kokovin, S., Parenti, M., and Thisse, J.-F. (2012). Monopolistic competition: Beyond the constant elasticity of substitution. *Econometrica*, 80(6):2765–2784.

Zlate, A. (2016). Offshore production and business cycle dynamics with heterogeneous firms. *Journal of International Economics*, 100:34–49.