Trade Policy with Firm Heterogeneity, Variable Markups, and Foreign Direct Investment

Ziran Ding

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Theo S. Eicher, Chair

Fabio P. Ghironi

Mu-Jeung Yang

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Trade Policy with Firm Heterogeneity, Variable Markups, and Foreign Direct Investment

Ziran Ding

Chair of the Supervisory Committee:
Castor Professor Theo S. Eicher
Economics

The overarching theme of this dissertation is the analysis of trade policy implication in the presence of firm heterogeneity, variable markups, and multinational production.

Chapter 1 surveys the main ingredients and results of heterogeneous firms trade policy literature that has been developing since the early 2000s. First, I present in great detail various stylized facts regarding firm heterogeneity, firm-level markups, and the global structure of multinational production. Second, I summarize the results of the recent development of theoretical approaches of modeling the firm-level markups. Third, I discuss the theoretical frameworks that incorporates multinational production into heterogeneous firms framework. Fourth, I review the trade policy literature that features firm heterogeneity, variable markups, and multinational production. Finally, I highlight the contribution of this dissertation and discuss directions for future research.

Chapter 2 introduces *ad valorem* tariff and horizontal FDI into the Melitz and Ottaviano [2008] framework, producing the first framework in the trade policy literature that incorporates firm heterogeneity, variable markups, and multinational production. The model generates novel equilibrium implications. First, the presence of multinational production generates a competitive effect on the economy, and firms need to be more productive to survive the competition. Second, the *ad valorem* tariff and quadratic quasi-linear preference collectively result in an endogenous level of firm entry. Therefore, the impact of trade/tariff

liberalization depends on the equilibrium number of firms. In the short-run, when the firm entry is prohibited, an increase in import tariff shields the domestic economy from the Foreign competition, making it easier for firms to survive. This result is overturned when firms can enter the market freely in the long-run. In the long-run, an increase in Home's import tariff will make the Home country a more desirable environment to do business, attracting more entrants in the Home market, making the Home market more competitive. Firms need to be more productive to survive. Home's tariff increase also makes it harder for the least productive Foreign exporters to survive, and triggers tariff-jumping FDI among the most productive exporters. Markups also respond to tariff change differently in the short-run vs. long-run, primarily due to the change of competitive environment associated with firm entry.

Chapter 3 studies the welfare implication of tariff and optimal tariffs in an environment features firm heterogeneity, variable markups and FDI. The findings can be broadly summarized in three aspects. First, the quadratic quasi-linear preference generates multiple externalities in this economy, causing market outcome to differ from the socially optimum outcome systematically. Permitting FDI lowers the domestic cutoff levels and reduces the misallocation in the economy. Second, free trade is not always socially optimal. If the domestic cutoff is sufficiently high, an additional firm entry can improve social welfare. In this case, a positive import tariff is welfare-improving because it encourages firm entry. Third, both positively and normatively, the interaction of variable markup and FDI generates novel trade policy insights that are absent if consumers possess CES preference.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

For all the guidance, supporting, and mentoring, I want to express my deepest gratitude to my advisers, professor Theo Eicher, Fabio Ghironi, and Mu-jeung Yang. I'm deeply indebted to Theo for his advice and his dedication to always push me further to seek better answers. I'm very grateful for the weekly workshop Fabio organized throughout the years, in which I obtained insightful and constructive feedback. Last but not least, I'm very thankful for Mu-jeung. I would never forget the days he spends time going through intuitions with me and revising my paper word by word. His patience always supports me on the path of pursuing a Ph.D. Without a doubt, this dissertation would look very different without the invaluable support from these three committee members.

I thank professor Philip Bond, Yu-chin Chen, and Jarrad Harford for their advice and time. I thank my fellow students: An-tsu Chen, Alvin Chen, Guillermo Gallacher, Sunny Huang, Daisoon Kim, Rory Mullen, Pushpak Sakar, and Jeff Tucker. Thanks for answering my questions and being good companions during this course. I also would like to thank Simon Reeve-Parker and Andrew Anderson for all the practical supports.

Lastly, I thank my loving family for their unwavering support. I thank my lovely wife, Luna, for her endless patience and encouragement. I thank fellow believers from the Church in Seattle for all their love and prayers.

...and He made them wander in the wilderness forty years...–Numbers 32:13b

DEDICATION

To my beloved daughter, Esther

Chapter 1

A SYNOPSIS OF THE RELATED LITERATURE

1.1 Introduction

Despite its outstanding contribution to the world output growth, ever since the global financial crisis, international trade and investment have provided much less support to economic growth. Trade growth has barely kept up with output growth and has even lagged behind it for a few years. Compared to the persistent growth from 1990 to 2007, with an average rate of 15 percent, global investment has been fluctuating around a 1 percent annual growth rate for the past ten years. Just as trade was showing some tentative signs of renewed vigor[1], the anti-globalization sentiment has started to pose a significant threat to global economic integration. For example, after more than 50 years of leading efforts to lower international trade barriers, since 2018, the US has enacted several waves of tariff increases on various products, sectors, and countries. Take China, for instance: From 1989 to 2017, the average US-applied tariffs on imports from China gradually decreased from 5 percent to 3 percent. During the same period, China's tariffs on imports from the US declined from 40 percent to 8 percent. However, according to Bown [2019], by the end of 2019, 96.8 percent of US imports from China are subject to tariffs at an average rate of 20 percent. These measures represent the most comprehensive protectionist trade policies implemented by the US since the 1930s. In response, China also imposes retaliatory tariffs on US exports. By mid-December of 2019, 69 percent of China imports from the US are subject to an average tariff rate of 25.9 percent.

Due to the sizes of the economies involved, the magnitudes of the tariff increases, and the breadth of tariffs across sectors, the current trend of returning to protectionism is unprecedented in the postwar era. What is the welfare implication of such a protectionist trade

[1] According to UNCTAD [2018].

policy? The traditional wisdom believes that producers can benefit from protectionism by avoiding competition with foreign suppliers, but consumers will suffer from having fewer varieties. This perception is based on two premises: (i) Firms in an industry are atomic and they act as price takers. (ii) The primary channel for firms to access foreign markets is through export. In the past two decades, however, the micro-level data have presented quite a different picture. This chapter provides an overview of the trade and trade policy literature from three specific angles: firm-level productivity heterogeneity, firm-level variable markups, and multinational production (henceforth, MP).

Since the pioneering work of Melitz [2003], trade economists have been increasingly focusing on the firm as the unit of analysis. This can be partly attributed to the empirical findings using micro-level data on plants and firms. In the first section (Section 1.2) of this chapter, I briefly summarize the empirical evidence that has been established in the heterogeneous firm literature. One set of empirical findings show that firms are heterogeneous in their productivity even within narrowly defined industries. In particular, firms engaging in MP are usually more productive than pure exporters (see, for example, Temouri et al. [2008], and Criscuolo and Martin [2009]). The second set of empirical results highlights the monopoly power across firms within industries. The micro-level evidence suggests that the most productive firms charge the highest markups (see De Loecker and Warzynski [2012] for exporters, and see Dobbelaere and Kiyota [2018] for multinational firms). While the previous two strands of empirical studies focus on firm-level characteristics, others emphasize the role of the firm in understanding patterns of trade and the global production structure. Although goods are mobile across borders, their movement is subject to various types of costs: shipping costs, tariffs, and legal barriers to trade. It turns out that these frictions shape a firm's production decision and location choice, which are pivotal to the understanding of global production (see for example Antràs and Yeaple [2014], and Yeaple [2013]).

Although the Melitz-type heterogeneous firm model has been the workhorse in trade literature, it is clearly at odds with two of the most robust empirical findings in the micro-level data: (i) Different firms charge different levels of markups, even within the same industry.(ii)

There is a rapid growth of MP throughout the globe. The first discrepancy has spurred an extensive theoretical literature on introducing variable markups into the heterogeneous firm framework. In the next section (Section 1.3) of this chapter, I analyze and contrast the different approaches that trade economists have adopted to model variable markups. In general, there are three ways to introduce variable markups: the demand-side approach, the supply-side approach, and a combination of both. The demand-side approach focuses on modifying the constant elasticity of substitution (henceforth, CES) preference imposed on consumers. The supply-side approach focuses on modifying the monopolistic competition assumption between the producers. Due to its advantage of analytical tractability, the first approach is more prevalent in the literature, but there has been a resurgence in the supply-side approach (see Head and Spencer [2017] for example). The last approach, which is the combination of the previous two, while being undoubtedly realistic, does not seem to generate additional insights compared to the other two approaches (see, for example, Kokovin et al. [2017]).

The second discrepancy is addressed by the burgeoning papers of multinational firms. In Section 1.4, I briefly review the state of this literature. Given the space constraint, I mainly focus on various theoretical and quantitative approaches that have been adopted to introduce multinational activity into the heterogeneous firm framework. I divide the review into three parts: (i) Horizontal FDI, which refers to the fact that multinational firms duplicate roughly the same activities in different countries, is the dominant format of FDI flows between developed countries. (ii) Vertical FDI, which describes firms' motive to locate different parts of the production process in different countries, has been increasing since many developing countries are participating in the global supply chain. (iii) Complex FDI, which captures the notion that firms replicate some activities in many countries while concentrating other activities in a few countries, has been brought to light in recent years thanks to the availability of micro-level data of multinational firms.

Given the empirical evidence and the theoretical development in the trade literature, it is necessary and meaningful to resuscitate the trade policy associated with different types of framework. In Section 1.5, I review the recent developments in the heterogeneous firm trade policy literature. I begin by investigating the trade policy papers in the benchmark Melitz framework. Then I go over the papers that emphasize the role of variable markup within the heterogeneous firm structure. In the end, I discuss a few papers that look at trade policy questions with firm heterogeneity and MP, which are closely related to this dissertation. While most of the papers reviewed in this section are normative papers studying optimal trade policy, due to the ongoing trade war, I also mention a few recent papers that study the consequences of trade policy uncertainty. In contrast to traditional steady-state analysis, this type of trade policy paper is usually carried out in either a dynamic or an open-economy macro framework.

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

The remainder of the chapter is structured as follows. Section 1.2 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 1.3 then surveys the various theoretical approaches that have been used to model variable markups within the heterogeneous firm framework. The papers discussed in this section have the common feature that they all exclude the possibility of MP. Section 1.4 discusses the papers that introduce MP into the heterogeneous firm framework. The papers in this section all feature constant markups. Section 1.5 briefly goes through trade policy papers in the heterogeneous firm literature.

Section 1.6 concludes. All tables and graphs to which this chapter refers are included in the appendix.

1.2 Stylized Facts on Heterogeneous Firms, Variable Markups and FDI

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

1.2.1 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.

Export Participation

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

Exporters are Different

Not only are exporters rare, but also are they systematically different from nonexporters. **Table A.3** presents the US manufacturing exporters' premia documented in Bernard et al.

---

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

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

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

[2007]. Exporters are larger in employment and sales, and they are more productive in value-added per worker and total factor productivity. They also pay higher wages and are more capital and skill intensive. Similar patterns are confirmed in the European firm-level data by Mayer and Ottaviano [2008], as shown in **Table A.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 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 A.1**), sell less in the domestic market, and export a significant fraction of their output.

MNEs are More Productive

According to the estimates of UNCTAD [2011][5], multinational firms account for 25 percent of the global GDP and one third of international trade. Multinational firms not only are quantitatively important but also enhance our understanding of firm heterogeneity. Benfratello and Sembenelli [2006] exploit Italian manufacturing firm-level data with the GMM-System estimator developed by Blundell and Bond [1998], which allows them to control for firms' unobserved heterogeneity, inputs, and ownership endogeneity as well as measurement errors. The authors find a positive and significant effect for firms under US ownership: They tend to outperform both their domestic counterparts and firms under other ownership. Temouri et al. [2008] take a rich firm-level data set from Amadeus[6] to study the total factor productivity differences across 22 manufacturing and 17 service industries in Germany over the period 1995–2004. They find that multinational firms have significantly higher TFP than their domestic counterparts. Mayer and Ottaviano [2008] find similar multinational firm premia in Belgium, as illustrated in **Figure A.2**. Using a sample from the UK Annual Re-

[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.

spondents Database, Criscuolo and Martin [2009] find that the US and other foreign-owned plants are on average 42 percent and 30 percent, respectively, more productive than British domestic plants. According to Barefoot and Mataloni Jr [2011], in manufacturing sectors, US parents account for less than a half of 1 percent of enterprises but for over 62 percent of the value added and 58 percent of employment. Bloom et al. [2012] find that in the IT industry, management practice is a robust explanatory variable for US multinationals' exceptional performance compared to non-US multinationals or domestic firms in Europe. All this evidence suggests the superior productivity of multinational firms relative to nonmultinational domestic counterparts.

1.2.2 Variable Markups

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

Entrant's Markup

One of the most robust empirical observations emerging from the recent industrial organization studies is that new entrants have lower average productivity and higher exit rates than the existing incumbents. However, using a unique data set from US Census of Manufacturing, Foster et al. [2008] find that the observed disadvantage in revenue productivity is

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

mainly due to entrants charging a lower price-cost markup rather than technical inefficiency. More recently, by introducing the demand-side into the structural model of production in the spirit of Hall [1988], Kılınç [2014] is able to estimate markups using firms' nominal sales and input expenditures via a control function approach while controlling for the endogeneity issue of inputs on productivity. Using annual plant-level data from manufacturing industries in Japan from 1985 to 2007, the author finds that entrants on average charge lower markups than incumbents. As illustrated in **Figure A.3**, the result is robust for different productivity indices.

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 A.6**. Lately, Hornok and Muraközy [2018] apply the same methodology to estimate firm-level markup in Hungarian manufacturing firms (1995–2003). They find robust and consistent evidence for markup premiums of importers, but not exporters.

The other channel, which claims that exporters become more efficient after export entry, has received mixed empirical evidence[8]. Recently, Voigtlaender and Garcia-Marin [2019] decompose exporters' efficiency gains into changes in revenue productivity, markups, and

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

marginal costs. Using rich plant-level data from Chile, Colombia, and Mexico, they find that markups are stable around the exporters' entry (the efficiency gain is fully passed through to consumers), but are higher for established exporters (limited pass-through of efficiency gain to consumers) after tariff-induced expansions. Their finding suggests that the common use of revenue-based productivity measures ($TFPR$) might be the reason that the literature has struggled to identify export-related efficiency gains within plants.

MNE's Markup

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

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

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

1.2.3 Multinational Production

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

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

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

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.

Vertical FDI

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

While the previous literature thinks that vertical FDI mostly happens between developed and developing countries, using a new firm-level data set provided by Dun & Bradstreet (D&B)[10], Alfaro and Charlton [2009] find the share of vertical FDI is more substantial than

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

commonly thought even between the developed countries. They show that this is due to a significant amount of vertical FDI being misclassified as horizontal FDI. For example, many subsidiaries that supply goods to their parents are located in sectors in which both the input and final goods are in the same two-digit Standard Industrial Classification (SIC) code, and they are counted incorrectly as vertical FDI at four-digit SIC level (See **Figure A.5**).

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

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

[11] A characteristic commonly associated with vertical FDI.

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

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

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

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

1.3 Heterogeneous Firms with Variable Markups

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

1.3.1 Demand-side Approach: Deviating from CES Preference

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

Rodriguez-Lopez [2011] incorporates variable markups into Ghironi and Melitz [2005], which is a dynamic open economy macro version of Melitz [2003]. The authors adopt the *translog expenditure function*[14] developed by Bergin and Feenstra [2000]. This preference allows the author to investigate the general equilibrium effect of trade liberalization, which cannot be explored under a nonhomothetic preference (such as quadratic quasi-linear preference). The steady-state version of the model yields a very similar response to trade liberalization as in Melitz and Ottaviano [2008]. The authors then use a dynamic version of the model,

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

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

together with sticky-wage setting, to demonstrate that firms' decisions on pricing, entry, and exit play an important role in explaining the pass-through and the expenditure-switching effect of exchange rates, thereby successfully resolving the slow and low degree of nominal exchange rate pass-through to consumer import prices among the developed economies[15].

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

Zhelobodko et al. [2012] develop a more general model of monopolistic competition, with CES preference as a special case. The authors assume that consumers have an *additively separable preference* over the differentiated varieties without specifying the specific functional form of the utility[17]. By focusing on the *relative love for variety (RLV)*(i.e., the elasticity of marginal utility), the authors demonstrate that the market equilibrium responds differently to various degrees of competition: more competing firms, a larger market size, or the

---

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

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

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

combination of both will lead to lower market prices due to the increase in the elasticity of substitution. The authors then embed the preference into Melitz [2003] and show that the cutoff productivity and markup decrease with the size of the market when RLV increases with consumption, even without assuming the productivity distribution of firms[18].

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

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

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

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 independent of market size, which is consistent with empirical evidence. The model also generates novel insights into the extensive margin of trade: The extensive margin of trade is increasing in destination per-capita GDP, neutral in destination population, and falling in the trade cost to the destination market. Combined with the fact that cost pass-through is falling in firm productivity, the authors then demonstrate that the welfare gain from trade is much smaller in this framework than the commonly employed frameworks.

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

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

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

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

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

and find it to be slightly above zero (0.06). Hence, they conclude that the pro-competitive effects of trade are elusive.

1.3.2 Supply-side Approach: Deviating from Monopolistic Competition

Bertrand Competition

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

De Blas and Russ [2015] provide a transparent generalization of Bernard et al. [2003] through two modifications: (i) Firms' efficiency level draw follows Fréchet distribution instead of Pareto distribution. (ii) There are finite number of rivals instead of infinite number of rivals. These modifications produce a distribution of markups that, compared to the one in Bernard et al. [2003], preserves the characteristics of the market structure, which are sensitive to the degree of trade openness and differences in technological development across countries. For example, trade liberalization reduces the markups that domestic firms can charge on domestic sales. Moreover, the model predicts that if the trade barrier meets a specific condition, bilateral trade liberalization can create an anti-competitive effect, increasing market power for exporters on average, while switching from bilateral to multilateral trade agreements can generate a pro-competitive effect, reducing average markups among exporters.

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

Cournot Competition

Instead of competing in prices, Atkeson and Burstein [2008] introduce variable markups through quantity competition á la Cournot[22] into a nested CES demand system as in Helpman and Krugman [1985]. Contrary to Melitz [2003] and Bernard et al. [2003], where there are an infinite number of firms in an industry, Atkeson and Burstein [2008] assume that there are a finite number of firms competing with each other. This setup allows firm-level markup to be positively correlated with the firm's market share, and elasticity of demand to be negatively correlated with the firm's market share. The authors find that the separated markets through trade costs and Cournot competition with variable markups lead to pricing-to-market behavior[23] in those separate markets. These two elements are essential for generating deviations from relative PPP at the aggregate level. This approach has gained increasing popularity in the literature–for example, see Amiti et al. [2014], Edmond et al.

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

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

[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.

1.3.3 Supply-side and Demand-side Approach

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

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

---

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

1.3.4 Relation to This Dissertation

This dissertation studies the trade policy implication in an environment that features firm heterogeneity, variable markups, and MP. To introduce variable markup, I adopt a particular preference (quadratic quasi-linear preference) as described in the demand-side approach (see Section 1.3.1). One should note that other types of preferences can also generate similar results, but they generally suffer from analytical transparency. For example, all the theoretical derivation in this dissertation can also be carried out with translog preference or constant absolute risk aversion (CARA) preference, but the closed-form solutions under these preferences always carry a Lambert function[25], and are not as transparent as the results derived under quadratic quasi-linear preference. Regarding the supply side, it is a well-known challenge in the literature to work with models of oligopolistic competition, both analytically and computationally[26].

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

1.4 Heterogeneous Firms with Multinational Production

This section briefly surveys the state of the literature on multinational firms, with an emphasis on firm heterogeneity. The discussion here maps the empirical evidence presented in Section 1.2.3. First, I discuss the theoretical frameworks that capture the within-industry motive of cross-border investment of multinational firms. I make a distinction between for-

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

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

eign 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.

1.4.1 Horizontal FDI

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

Motivated by the empirical observation that the majority of FDI takes the form of cross-border mergers and acquisitions (M&A), Nocke and Yeaple [2007] introduce cross-border M&A into Helpman et al. [2004]. Firms are heterogeneous in their capabilities, and these capabilities differ in their degree of international mobility. Cross-border M&A is motivated by a firm's desire to exploit complementarities between the local firm's country-specific capability and the acquiring firm's intangible technological advantage. The model suggests that when capabilities become relatively less mobile internationally, cross-border M&A becomes the more popular mode of entry into foreign markets. Furthermore, firm heterogeneity is a crucial determinant of firms' international organization. For example, in industries where firms differ mainly in their mobile capabilities, the most efficient firms will engage in cross-border M&A, while in industries where firms differ mainly in their country-specific nonmobile capabilities, cross-border M&A will involve the least efficient active firms.

In contrast, de Blas and Russ [2013] study the impact of FDI and cross-border M&A on firms' markup in a generalized version of Bernard et al. [2003][27]. In such a Ricardian model, the authors prove that takeovers by foreign firms increase the technological edge of the acquired firms, allowing the target firms to increase their markup, thereby increasing the average markup in the economy. When trade is costly, the authors utilize the model to demonstrate a firm's motive for taking over a foreign rival to increase its market power in either the foreign market or the home market. Both channels will result in higher prices compared to a world without FDI. Similarly, Muraközy and Russ [2015] use the empirical methods developed by De Loecker and Warzynski [2012] and find that foreign-owned firms charge significantly higher markups than domestic firms (the gap is more evident in industries where MNEs have a technological advantage) in Hungary. The authors then try to use a Ricardian model with variable markups and heterogeneous firms to explain the empirical observation. The framework not only provides an analytical distribution for market shares and markups when goods are imperfect substitutes, but also highlights the role of technology in generating market power. The model explains about half of the observed multinational markup premium, calling for more research on the relation between MNEs and market power.

More lately, Gumpert et al. [2017] use a dynamic[28] version of Helpman et al. [2004] and study the life cycle dynamics of exports and FDI. Assuming that firm productivity evolves according to a Markov process, and MNEs face a sunk entry cost, the model can reproduce salient features of the firm-level data from France and Norway. The authors then use the model to conduct counterfactual analysis and find that the presence of MNEs is crucial for exporters' life cycle dynamics. On the one hand, trade liberalization in a model with MNEs would increase exporters' sales, and decrease their exit rates, but it would barely change their

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

[28]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.

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.

1.4.2 Vertical FDI

There are two main streams of literature that focus on the firm-level decision regarding vertical specialization: (i) location decisions, and (ii) organizational mode. Here I briefly survey the literature on the latter[29], since this branch has been rapidly growing following the Melitz revolution.

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

[29] 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.

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 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.

1.4.3 Complex FDI

Grossman et al. [2006] investigate the optimal integration strategy in a three-country (two Northern countries and one Southern country) model where firms possess heterogeneous productivity. In this model, firms that are headquartered in a Northern country supply differentiated final goods to all three countries. Each firm must produce an intermediate input and assemble the inputs so as to generate a final product. The production of intermediate input and assembly can take place in any of the three locations. Foreign activities, which either produce an intermediate input or assemble the inputs, incur a fixed cost, while transport intermediate inputs and final product incur an iceberg-type shipping cost. The optimal integration strategies depend on the fixed costs of foreign subsidiaries, the transportation cost of intermediate and final goods, and the North-South factor price differences. In equilibrium, firms with different productivity levels will choose different organizational forms, confirming the empirical evidence in Hanson et al. [2005] that different forms of integration strategy coexist within the same industry.

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

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

Ramondo and Rodríguez-Clare [2013] introduce MP into a quantitative Eaton and Kortum [2002]-type Ricardian trade model by allowing a country's technology to be used for production abroad. The model features both tradable intermediate goods and nontradable final-consumption goods. To compete in a foreign final-good's market, MP is the only channel. For intermediate goods, firms can access foreign markets either through export or MP. The framework also allows a firm to use a third country as an export platform. Therefore, the model features horizontal FDI, vertical FDI, and export-platform FDI. The authors then calibrate the model based on bilateral trade and MP data for OECD countries, as well as data on intra-firm trade flows for US and foreign MNEs operating in the US. Since the model captures the fact that trade can facilitate MP (by allowing MNEs' foreign affiliates to import inputs from their home country), the quantitative exercise suggests that the gains from openness (allowing both trade and MP) are much larger than the gains from trade.

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

1.4.4 Relation to This Dissertation

Given that the goal of this dissertation is to study trade policy, I choose to build on the most popular framework in the heterogeneous firm literature with MP, Helpman et al. [2004] (henceforth, HMY). Yet there are three notable differences in this study. First, the original HMY model utilizes quasi-linear preference, which, together with monopolistic competition, generates constant markups for all firms. This implication is clearly at odds with the empirical evidence on variable markups presented in Section 1.2.2. I adapt the quadratic quasi-linear preference into the HMY framework to overcome this discrepancy. Second, the original HMY framework treats trade costs as an iceberg-type cost, and treats trade liberalization as an exogenous decrease of the iceberg cost. This type of modeling approach cannot capture the various types of trade costs faced by exporters. In this study, I introduce the *ad valorem* tariff into the modified HMY model. This approach not only is a more realistic way to model trade costs but also enables one to study the policy implications of strategic tariffs. Third, in HMY, besides the iceberg-type trade costs, exporters and multinational firms are differentiated by the difference in the fixed cost. In other words, besides the differences coming from the productivity draw, there is no other difference between an exporter and a multinational firm in terms of their marginal cost of production. This assumption is not in line with the empirical evidence on the costs faced by multinational firms[30]. In my model, I introduce an iceberg-type efficiency loss for firms engaging in MP, capturing a different type of friction[31] faced by MNEs.

---

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

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

1.5 Trade Policy with Heterogeneous Firms

While models of firm heterogeneity have transformed the way economists think about international trade and multinational firms, these models have made surprisingly little contribution to our understanding of trade policy. This section surveys the existing work on the normative side of the international trade literature when heterogeneous firms select into export. The discussion is organized into three areas. The first strand of literature studies the optimal trade policy in the basic Melitz [2003] framework while allowing the number of industries and country size to vary. I also include some recent papers that study the effects of trade policy uncertainty on the aggregate economy, which are typically built on the dynamic version of Melitz [2003]. The second group of papers study the optimal trade policy when heterogeneous firms charge variable markups. Both unilateral and multilateral trade policy are discussed. The last category of papers studies the trade policy implications when heterogeneous firms can access foreign markets either through export or FDI. To the best of my knowledge, there are only two papers falling into this category: one is built on Helpman et al. [2004], and the other is built on Antràs and Helpman [2004].

1.5.1 *Trade Policy with Firm Heterogeneity*

Baldwin and Forslid [2006] study the impact of a reduction in the variable and fixed cost of trade in a two-country, two-sector Melitz [2003] model[32]. 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.

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

Demidova and Rodriguez-Clare [2009] analyze the optimal trade policy for a small open economy[33] in a Melitz-type model. Two types of distortion exist in such an economy: There is too little spending on the domestic varieties relative to the social optimum due to markup distortion between home and foreign varieties, and there are too few foreign varieties in the market outcome compared to the social optimum due to consumption externality[34]. Since the first distortion prevails over the second, the first-best allocation can be obtained by a consumption subsidy, an import tariff, or an export tax. This finding challenges the view that export subsidy is optimal in the Melitz-type framework. As far as welfare is concerned, export subsidy causes a negative impact on terms of trade, variety, and curvature[35], which dominates the productivity gain from reallocation.

Ossa [2011] develops a novel theory of GATT/WTO negotiations based on the Krugman [1980] model. In a separate appendix, the author demonstrates that all the results can be derived in a variant of the Melitz [2003] model. The modified Melitz model features two sectors: a manufacturing sector that produces differentiated varieties, and a numéraire good sector that produces homogeneous goods. In this framework, import tariffs allow the country to attract additional firm entry into the manufacturing sector. If the tariff-induced long-run *delocation effect* dominates the direct effect of the tariff on the price index, then consumers' welfare will increase. The framework generates reasonable noncooperative tariffs and moderate gains from GATT/WTO negotiations.

Felbermayr et al. [2013] extend Demidova and Rodriguez-Clare [2009]'s exercise to a two-country large-economy case. The model features markup and entry distortion, as well as a terms of trade effect. They find that the optimal Nash tariff increases in the relative country size, the relative average productivity, and the degree of firm-level productivity dispersion, but decreases in the nontariff trade barriers. The authors then calibrate and simulate the

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

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

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

model to demonstrate the importance of productivity dispersion for the size of the optimal tariff quantitatively. The results indicate that the declining trade cost, higher firm-size dispersion, and country-size convergence have magnified the benefit of the WTO over time.

Haaland and Venables [2016] study both domestic taxes and trade policies for a small open economy containing a flexible monopolistically competitive sector[36] in which firms may be heterogeneous in their productivity level. The welfare gain of policy arises from the interaction of two dimensions: the distortion of quantity due to the monopoly power in the monopolistically competitive sector, and the distortion of price due to the terms of trade effect. The authors find the optimal policy is to subsidize domestic sales, and in some cases, a positive tariff on imports. These policies can generate an expenditure-switching effect toward the monopolistically competitive sector that eventually increases the number of varieties and improves the terms of trade. However, besides the fact that the foreign market's reaction toward policy becoming more price-elastic, the authors find firm heterogeneity does not create any new qualitative implication for policy interventions.

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

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

Similarly, Handley [2014] finds that binding tariff commitments can also reduce trade policy uncertainty and increase firm entry into the export market.

Last but not least, some recent papers have emphasized the macroeconomic consequences of trade policy. Utilizing a primal approach and general Lagrange multiplier methods to characterize optimal wedges, Costinot et al. [2016] investigate the optimal trade policy both at the micro and macro level in a generalized version of the Melitz [2003] model. At the micro level, a welfare-maximizing government should impose firm-level import taxes that discriminate against the most productive foreign exporters. The optimal export taxes can be either discriminative or in favor of the most profitable domestic exporters. At the macro level, terms of trade affect the total level of trade taxes. The more home's terms of trade deteriorates with increases in exports or imports, the larger the trade taxes it should impose. Utilizing the number of anti-dumping investigations as a measure of protectionism, Barattieri et al. [2018] explore the effectiveness of trade policy at the macroeconomic level. For small open economies, both their empirical and model-based[37] 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

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

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

1.5.2 Trade Policy with Firm Heterogeneity and Variable Markups

Demidova [2017] takes the first step to study trade policy with firm heterogeneity and variable markups. Following Arkolakis [2008], the author drops the outside good assumption in the Melitz and Ottaviano [2008] model to investigate the general equilibrium effect of trade policy. On the one hand, the author finds, regardless of country size, that the optimal import tariff is strictly positive from a unilateral perspective. On the other hand, the reduction of iceberg-type trade costs is always welfare-improving, making free trade in this dimension socially optimal. Variable markups, in the presence of firm heterogeneity, create a negative pro-competitive effect on the economy due to the misallocation of resources[38]. Consequently, the welfare gain from trade following liberalization[39] is dampened due to the presence of variable markups.

Traditional frameworks cannot explain WTO's strong restrictions on export subsidy. Introducing ad-valorem tariffs and export subsidies into the original Melitz and Ottaviano [2008] model, Bagwell and Lee [2015] explore the implications that a heterogeneous firm model may provide for the use and treatment of export subsidies in the WTO. The authors find that starting from free trade, a country can gain by imposing either a small import tariff, a small export subsidy, or a combination of the two. However, these policies are all 'beggar-

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

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

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[40] 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.

1.5.3 Trade Policy with Firm Heterogeneity and Multinational Production

Contrary to previous strategic tariff literature, where in equilibrium all foreign firms are either multinationals or exporters, Cole and Davies [2011] utilize the framework of Helpman et al. [2004] and find an equilibrium condition in which both pure exporters and multinationals coexist. The authors then study optimal tariffs in such an environment and find that the optimal tariff for the social planner is negative, i.e. trade should be subsidized. This is because encouraging trade can foster competition and eliminate the least productive firms in the economy, thereby boosting the average productivity and increasing social welfare. The

[40] 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.

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 ineciency due to tariff competition. When FDI is an option for firms, the Nash tariff is lower. Therefore, FDI is welfare-improving since it can mitigate tariff competition.

Building on the framework introduced by Bernard et al. [2003], de Blas and Russ [2013] investigate the market-power motive of FDI within the heterogeneous firm framework. The authors find that under Bertrand competition, FDI, either in the form of mergers and acquisitions or greenfield investment, can increase markups through the channel of technology transfer. When there is no 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[41], 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

[41]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.

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.

1.5.4 Relation to This Dissertation

This dissertation is closely related to Cole and Davies [2011]. Despite the apparent similarity between the two frameworks, there are clear differences between the two. First, Cole and Davies [2011] utilize quasi-linear CES preference. Combined with monopolistic competition, their model yields constant markups, and a complete pass-through in equilibrium. In contrast, my framework utilizes quadratic quasi-linear preference to generate variable markups and incomplete pass-through for different firms, which is more suitable for pricing and welfare analysis. Second, Cole and Davies [2011] completely ignore the possibility that tariffs can impact the level of firm entry in the economy. According to Caliendo et al. [2017], the combination of ad valorem tariff and tariff rebate violates the macro assumption in Arkolakis et al. [2012], and the level of entry will be affected by a tariff change. In the current framework, the number of entrants is endogenously affected by the tariff level, generating different welfare implications for protectionist trade policy in the short-run and long-run.

1.6 Concluding Remarks

This chapter 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 foreign direct investment. Secondly, models that allow firms to engage in both export and foreign direct investment are introduced under the combination of CES preference and monopolistic competition, which implies that all the firms will charge identical and constant markups. This is far from what we have observed through the firm-level data. The next chapter of this dissertation develops a framework that overcomes the previous two discrepancies between theory and data. The model will feature firm heterogeneity, variable markups, and MP.

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

Chapter 2

HETEROGENEOUS FIRMS, VARIABLE MARKUPS, AND FDI

2.1 Introduction

Over the last two decades, firm heterogeneity has been steadily transforming international trade theory. This transformation has been fueled by micro-level empirical studies. The two most robust empirical findings among these studies are: heterogeneous firms charge heterogeneous markups, and many firms are engaging in multinational production. Paradoxically, the workhorse model[1] in the heterogeneous firm trade policy literature assumes monopolistic competition and constant elasticity of substitution (henceforth, CES) preference, which implies a constant markup for all the firms. Furthermore, firms only access foreign markets through export. It goes without saying that such a framework cannot provide an adequate understanding of trade policy analysis.

The goal of this chapter is to provide a theoretical framework that features heterogeneous firms, variable markups, and a particular form of multinational production, horizontal foreign direct investment (henceforth, FDI). To this end, this chapter introduces variable markups through quadratic quasi-linear preference, as in Melitz and Ottaviano [2008], into a two-country (Home and Foreign) model with firm heterogeneity and FDI, as in Helpman et al. [2004]. In the current framework, the economy features two sectors: a perfectly competitive sector that produces homogeneous goods and a monopolistically competitive sector that produces differentiated varieties. Firm entry and exit only happen in the monopolistically competitive sector. A firm needs to pay a fixed cost and draw its marginal production cost (which is inversely related to the firm's productivity) to enter the market. Post-entry, firms

[1] For example, see Demidova and Rodriguez-Clare [2009], Felbermayr et al. [2013], and Ossa [2016].

produce with different marginal cost levels. Exporters encounter two types of costs, iceberg-type trade costs, an *ad valorem* tariff. As for multinationals, they face an iceberg-type of *efficiency loss* as in Keller and Yeaple [2008]. Before the individual firm's productivity is realized, firms formulate entry, export, and FDI decisions based on expected profit. The difference in marginal cost preserves the sorting of firms[2]: the most productive firms access the Foreign market through FDI, the less productive firms export, and the least productive firms only serve their domestic market. An increase in the Home country's import tariff affects the variable profit of Foreign exporters and multinationals, making FDI a more profitable entry mode for the most productive Foreign exporters, inducing tariff-jumping FDI under the heterogeneous firm framework.

The analysis of the findings shows that the economy responds to a tariff change differently in the short-run versus long-run. In the short-run, where the entry has not taken place yet, and fixed costs are sunk, exiting the market is never optimal for firms. In this case, the economy is characterized by a fixed number and distribution of incumbents. These incumbents decide whether they should operate and produce or shut down. If they choose to shut down, they can restart production without incurring the entry cost again. In this short-run case, an increase in the Home import tariff makes it harder for the least productive Foreign exporters to export. These exporters will shut down their export department and only serve their domestic market. At the same time, an increase in Home's import tariff makes export a less desirable entry mode for the most productive Foreign exporters. These firms will switch to FDI simply because the variable profit of FDI is higher than that of export. In the current setup, if the level of tariff is low, the reduction of Foreign exporters will dominate the increase of Foreign multinationals, resulting in a reduction of the total number of Foreign firms in the Home country. In equilibrium, an increase in Home tariff creates an easier environment for Home firms to survive. Therefore, Home's tariff effectively

[2]In Helpman et al. [2004], the sorting of firms is preserved by the combination of fixed cost and variable cost. Here, with bounded marginal utility, high-cost firms will not survive, even without such fixed costs. The difference in marginal cost is sufficient to generate the sorting. Adding fixed cost will substantially degrade the tractability of the model, without generating additional insight.

shields its firms from Foreign competition.

In the long-run, due to the presence of the *ad valorem* tariff and quadratic quasi-linear preference, the number of entrants and the number of firms in the economy are endogenously determined by the tariff level. With the free-entry condition, an increase in Home's import tariff makes the Home country a more desirable place to do business, generating more domestic entry, intensifying competition in the Home market. On the one hand, the increase of Home's tariff makes it harder for the least productive Foreign exporters to export, reducing the number of Foreign exporters. On the other hand, the protection makes it more profitable for the most productive Foreign exporters to do FDI and increases the number of Foreign multinationals. However, the total impact on the number of firms in the Home market is dominated by the domestic entry. Different from the short-run equilibrium, an increase in Home tariff creates more entry, generates more competition in the Home market, and makes it harder for local producers to survive.

In contrast to the traditional combination of monopolistic competition and CES preference, where all the firms charge identical markups, in the current setting, markups also respond to tariff change differently in the short-run versus in the long-run. In the short-run, Home import protectionist tariff reduces the competition in the Home market, causing both Home's domestic firms and Foreign's FDI firms to charge higher markups than before. Foreign exporters' markups vary according to their productivity. For the least productive Foreign exporters, Home's protectionist tariff reduces their markups by giving a competitive edge to the Home firms. However, for the more productive Foreign exporters, Home's protection does not affect them much, and they benefit from the reduced competition by charging a higher markup. These responses are reversed in the long-run. Home's protectionist trade policy makes the Home market a more desirable environment to do business, attracting more firm entry into Home's market. This channel will substantially increase the competition level in the Home market, making it harder for firms to survive and reducing the markups for all the firms that operate in the Home market.

This paper is not the first to address questions regarding trade policy within the heterogeneous firm framework. For example, Demidova and Rodriguez-Clare [2009] use a Melitz-type model to study the trade policy implication in a small open economy. Felbermayr et al. [2013] study the bilateral trade policy implication in a two-country, asymmetric Melitz-type economy. Bagwell and Lee [2015] study trade policy in the Melitz and Ottaviano [2008] model and provide a rationale for the treatment of export subsidies within the World Trade Organization. Costinot et al. [2016] utilize a generalized Melitz model to study the trade policy implication both from a micro and macro perspective. Demidova [2017] studies the optimal tariff in the Melitz and Ottaviano [2008] environment without the numéraire good and finds that protection is always desirable, and reductions in cost-shifting trade barriers are welfare-improving. A common feature of the papers mentioned above is their exclusive focus on domestic producers and foreign exporters. A key message from the current analysis is that ignoring multinational production may provide an incomplete picture of protectionist trade policy.

A recent article by Cole and Davies [2011] is closely related to the current paper. The authors introduce an ad valorem tariff and heterogeneous fixed costs into Helpman et al. [2004], and find equilibria in which both pure exporters and multinationals coexist, resolving a well-known puzzle[3] in the strategic tariff literature in the presence of multinationals. Heterogeneous fixed costs for exporters and multinationals are the critical elements to generate their results. In contrast, the coexistence of exporters and multinationals in the current framework comes from the different iceberg costs they face.

Despite the apparent similarity between the two frameworks, it should be clear that the two exercises are very different. First, Cole and Davies [2011] combine quasi-linear CES preference with monopolistic competition, which produces a constant markup for all the firms. Although being analytically tractable, the combination of CES and monopolistic competition has little merit, even as a first approximation, for welfare analysis. In contrast, the current framework utilizes quadratic quasi-linear preference to generate variable markups

[3] In equilibrium, all foreign firms are either multinationals or exporters.

and incomplete pass-through for firms with different productivity levels. This attribute is more suitable for pricing and welfare analysis. Second, although Cole and Davies [2011]'s model features ad valorem tariff and multiple sectors, the policy implications from their paper are entirely independent[4] of firm's entry level. In the current model, tariff level endogenously affects the number of entrants, generating different implications for protectionist trade policy in both the short-run and the long-run.

The remainder of this chapter is organized as follows. Section 2.2 presents the complete model and the analytical solution of the model. Section 2.3 introduces the short-run version of the model and studies the effect of a unilateral tariff change. Section 2.4 presents the long-run version of the model and investigates its equilibrium features. All the predictions regarding a tariff change are contrasted with those in Cole and Davies [2011]. Section 2.5 discusses the limitations of the current framework and then concludes. All tables and graphs to which this chapter refers are included in the appendix.

2.2 A Model with Heterogeneous Firms, Variable Markups, and FDI

This section introduces quadratic quasi-linear preference into the Helpman et al. [2004] framework. There are two symmetric countries, Home (H) and Foreign (F). The markets are segmented, and international trade entails trade costs that take the form of transportation costs as well as *ad valorem* import tariffs. Tariff revenue is redistributed equally across consumers in the tariff-imposing country. Multinational firms engaging in FDI face an iceberg-type marginal cost (i.e., efficiency loss) in the spirit of Keller and Yeaple [2008]. Different from Cole and Davies [2011], where firms' sorting is induced by different fixed costs, the nonhomothetic preference here induces different productivity cutoffs through different marginal costs. The framework presented here is suitable for this dissertation due to two reasons. First, it enables one to study trade policy in an environment that features firm heterogeneity, variable

[4] To the best of my knowledge, this feature is contradicting with the implications from Balistreri et al. [2011], Arkolakis et al. [2012], and Caliendo et al. [2017].

markups, and multinational production. Second, the model produces tractable analytical solutions due to the specific assumption regarding demand and production structure, yielding quite transparent comparative statics.

2.2.1 Consumers

Each country is endowed with one unit of consumers. In the H economy, each consumer supplies one unit of labor. Consumers in country H maximize their utility by consuming the numéraire good, q_0^H, and the differentiated varieties, q_i^H, subject to their income budget constraint:

$$U^H = q_0^H + \alpha \int_{i\in\Omega^H} q_i^H di - \frac{1}{2}\gamma \int_{i\in\Omega^H} \left(q_i^H\right)^2 di - \frac{1}{2}\eta \left(\int_{i\in\Omega^H} q_i^H di\right)^2$$

$$\text{subject to: } q_0^H + \int_{i\in\Omega^H} p_i^H q_i^H di \leq I^H \equiv w^H + TR^H + \Pi^H$$

where α and η indicate the substitutability between the differentiated varieties and numéraire good, and γ indicates the substitutability among the differentiated varieties. An increase in α and a decrease in η both shift out the demand for the differentiated varieties *relative* to the numéraire good. The degree of product differentiation increases as γ increases since consumers are paying more attention to the distribution of varieties that they consume. All these three demand parameters are positive. Notice that different from Melitz and Ottaviano [2008], the tariff revenue and aggregate profit will enter into consumer's budget constraint through government redistribution.

Assuming consumers have positive demands for the numéraire good $\left(q_0^H > 0\right)$, utility maximization of the previous consumer problem leads to the following inverse demand for each variety i:

$$p_i^H = \alpha - \gamma q_i^H - \eta Q^H \tag{2.1}$$

where $Q^H \equiv \int_{i\in\Omega^H} q_i^H d_i$ is the aggregate consumption of these varieties, and Ω^H is the variety space that is available to the consumers. Invert equation (2.1) to obtain the linear market demand for these varieties:

$$\begin{aligned} q_i \equiv q_i^H &= \frac{\alpha}{\eta N^H + \gamma} - \frac{1}{\gamma} p_i^H + \frac{\eta N^H}{\eta N^H + \gamma} \frac{1}{\gamma} \bar{p}^H \\ &= \frac{1}{\gamma} \left(p_{\max}^H - p_i^H \right) \end{aligned} \tag{2.2}$$

where $p_{\max}^H = (\gamma\alpha + \eta N^H \bar{p}^H)/(\eta N^H + \gamma)$ represents the price at which demand for a variety is driven to 0, $\bar{p}^H \equiv (1/N^H) \int_{i \in \hat{\Omega}^H} p_i^H di$ is the average price of all consumed variety in country H, and $\hat{\Omega}^H$ is the consumed subset of Ω^H. Note that equation (2.1) also implies $p_{\max}^H \leq \alpha$ for positive q_i^H and Q^H. Different from CES preference, where the elasticity of demand across varieties is constant, the price elasticity of demand here is given by:

$$\varepsilon_i^H \equiv \left| \frac{\partial q_i^H}{\partial p_i^H} \times \frac{p_i^H}{q_i^H} \right| = \frac{1}{p_{\max}^H / p_i^H - 1} \tag{2.3}$$

A lower average price $\bar{p}^H$ or a larger number of competing varieties N^H will induce a decrease in the choke price $p_{\max}^H$, and hence an increase in the price elasticity of demand ε_i^H at any given p_i^H. These movements all represent a *tougher* competitive environment, which has received strong empirical support in the industrial organization literature but can not be captured in an environment with constant elasticity of demand.

As in Melitz and Ottaviano [2008], welfare can be evaluated using the following indirect utility function:

$$U^H = I^H + \frac{1}{2} \left(\frac{N^H}{\eta N^H + \gamma} \right) \left(\alpha - \bar{p}^H \right)^2 + \frac{1}{2} \frac{N^H}{\gamma} \sigma_{p^H}^2 \tag{2.4}$$

where $\sigma_{p^H}^2 \equiv \left(1/N^H\right) \int_{i \in \hat{\Omega}^H} \left(p_i^H - \bar{p}^H \right)^2 di$ measures the variance of variety prices. To ensure positive demand levels for the numéraire, I assume that consumer's income is sufficiently high, so that $I^H > \int_{i \in \hat{\Omega}^H} p_i^H q_i^H di = \bar{p}^H Q^H - N^H \sigma_{p^H}^2 / \gamma$. Consumer's utility will be higher when the average price $\bar{p}^H$ is lower, the variance of prices $\sigma_{p^H}^2$ is higher, and the number of varieties N^H is larger.

2.2.2 Firms

Production in the economy only utilizes labor, which is supplied in an inelastic fashion in a competitive market. The economy consists of two sectors, the traditional sector (produces q_0^H) and the modern sector (produces q_i^H). q_0^H is produced under a constant return to scale technology at a unitary cost. Thus the wage[5] in each country equals to one: $w^H = 1$. In the differentiated-good sector, firms are competing with each other in a monopolistically competitive fashion, and each firm produces a single variety.

Entry only happens in the modern sector. Different from Melitz [2003], here exit only happens at the moment when productivity is realized. There is no exogenous per-period death shock in the modern sector due to the one-period nature of the model[6]. To enter the market, a firm needs to pay a sunk entry cost $f_E > 0$ and draws its marginal production cost c, which indicates the unit labor requirement, from a Pareto distribution with cumulative distribution function $G(c) = (c/c_M)^k$, where $k \geq 1$ represents a shape parameter and $c_M > 0$ represents the upper bound of c. When $k = 1$, the marginal cost distribution is uniform on $[0, c_M]$. As k increases, the relative number of low productivity firms increases, and the distribution is more concentrated at these lower productivity levels. As k approaches infinity, the distribution of firm productivity approaches c_M. A lower c_M stands for a higher technology state in the economy. In this dissertation, I assume H and F share the same technology, hence the same upper bound c_M and the same f_E for both countries.

Depending on its productivity draw, a firm enters country H may exit, produce locally, export to country F or engage in the multinational activity. Following Melitz and Ottaviano [2008], since each firm's marginal cost of production does not vary with its production level, the decisions in each market can be made separately. Therefore, all the monopolistically competitive firms make separate decisions about their prices at each market, taking the

[5]One can attempt to drop the numéraire good, then wage will be endogenized and can be pinned down by trade balance condition.

[6]For a dynamic application of Melitz and Ottaviano [2008], see Moon [2015]. Notice the benchmark model in her work is without the numéraire good.

total number of varieties and the average price in a market as given. In what follows, I analyze each type of producer's profit maximization problem.

Domestic Producers

A firm located in country H with cost level c selects its price in the domestic market, p_D^H, to maximize its domestic profit $\pi_D^H(c) = \left[p_D^H(c) - c\right] q_D^H(c)$. Together with equation (2.2), the optimal price, markup, quantity, and profit can be solved as:

$$p_D^H(c) = \frac{1}{2}\left(c_D^H + c\right) \tag{2.5}$$

$$m_D^H(c) = \frac{1}{2c}\left(c_D^H + c\right) \tag{2.6}$$

$$q_D^H(c) = \frac{1}{2\gamma}\left(c_D^H - c\right) \tag{2.7}$$

$$\pi_D^H(c) = \frac{1}{4\gamma}\left(c_D^H - c\right)^2 \tag{2.8}$$

Let $c_D^H \equiv \sup\left\{c : \pi_D^H(c) > 0\right\}$ represent the cost of the firm which is indifferent between exit and remaining in the market. This firm earns zero profit as its price is driven down to marginal cost. Together with equation (2.2), the firm's optimal price is equal to its marginal cost, $p_D^H\left(c_D^H\right) = c_D^H = p_{\max}^H$. Hence, a firm will only serve domestic market if $c \leq c_D^H$. As expected, lower cost firms set lower prices and earn higher profits. However, lower cost firms are also more productive, and have larger market power, therefore they do not pass all of the cost differentials to the consumer and charge higher markups (which is defined as $m(c) = p(c)/MC(c)$[7], decreasing in c).

[7]Different from the Lerner index, where markup is measured in a relative sense, here the markup is measured in absolute sense. These two have no qualitative difference. Here I adopt the latter simply due to analytical convenience.

Exporters

The exporter in country H will face an *ad valorem* import tariff imposed by country F, denoted as $t^F \geq 1$. On top of that, the exporter will also face a per-unit trade cost, denoted by τ^F. More specifically, the delivered cost of a unit cost c to country F is $\tau^F c$ where $\tau^F > 1$. An exporter takes t^F and τ^F as given and maximizes its profit $\pi_X^H(c) = \left[p_X^H(c)/t^F - \tau^F c\right] q_X^H(c)$ by choosing optimal price $p_X^H(c)$. Together with equation (2.2), the optimal price, markup, quantity, and profit are:

$$p_X^H(c) = \frac{t^F \tau^F}{2}\left(c_X^H + c\right) \tag{2.9}$$

$$m_X^H(c) = \frac{t^F}{2c}\left(c_X^H + c\right) \tag{2.10}$$

$$q_X^H(c) = \frac{t^F \tau^F}{2\gamma}\left(c_X^H - c\right) \tag{2.11}$$

$$\pi_X^H(c) = \frac{t^F \left(\tau^F\right)^2}{4\gamma}\left(c_X^H - c\right)^2 \tag{2.12}$$

Let $c_X^H \equiv \sup\left\{c : \pi_X^H(c) > 0\right\}$ denotes the marginal cost of the least productive exporter from H to F, which barely finds export profitable. Combine this threshold with the definition of c_D^F (parallel to c_D^H), this cutoff level then satisfies $c_X^H = c_D^F / t^F \tau^F$. Intuitively, tariffs and transportation cost make it harder for exporters to break even compared to the domestic market.

It should be noted that the presence of iceberg-type transportation cost ensures that when the net tariff is zero, there are still exporters in the economy. Following Melitz and Ottaviano [2008], I abstract from any fixed cost of exporting, which could substantially reduce the tractability of the model without adding additional insights. With the bounded marginal utility, different marginal costs are enough to induce the sorting of firms.

Multinational Firms

To engage in the multinational activity, a firm located in country H with cost level c sets its product price for consumers in country F, denoted as $p^H_{FDI}(c)$. Instead of serving the Foreign market through export, it directly serves locally in country F, but doing so will incur an *efficiency loss*, which effectively increases the marginal cost of production. Here, I assume that the efficiency loss φ^F is greater than τ^F to ensure there are still multinational firms in the economy even when the net tariff is zero. Multinational firm's profit function is as the following:

$$\pi^H_{FDI}(c) = \left[p^H_{FDI}(c) - \varphi^F c\right] q^H_{FDI}(c) \tag{2.13}$$

Together with equation (2.2), the optimal price, markup, quantity, and profit are:

$$p^H_{FDI}(c) = \frac{1}{2}\left(c^F_D + \varphi^F c\right) \tag{2.14}$$

$$m^H_{FDI}(c) = \frac{1}{2\varphi^F c}\left(c^F_D + \varphi^F c\right) \tag{2.15}$$

$$q^H_{FDI}(c) = \frac{1}{2\gamma}\left(c^F_D - \varphi^F c\right) \tag{2.16}$$

$$\pi^H_{FDI}(c) = \frac{1}{4\gamma}\left(c^F_D - \varphi^F c\right)^2 \tag{2.17}$$

Let $c^H_{FDI} = \sup\left\{c : \pi^H_{FDI}(c) > \pi^H_X(c)\right\}$ denote the marginal cost of the least productive multinational firm, which finds it indifferent between export and FDI. Combine this threshold with the definition of c^F_D, the cutoff then satisfies $c^H_{FDI} = \xi^F c^F_D$, where $\xi^F \equiv (1-\sqrt{t^F})/(t^F\tau^F - \sqrt{t^F}\varphi^F)$ is derived by setting $\pi^H_{FDI}(c) = \pi^H_X(c)$. Before moving on to the industry-level entry analysis, I want to discuss a few important modeling features of the multinational firms.

(i) *Efficiency loss* The efficiency loss feature is similar to Keller and Yeaple [2008], who demonstrate that when technologies are complex, it is more difficult for US-owned foreign affiliates to substitute local production with imports from the multinational headquarter. φ^F can also stand for the information costs of working abroad, transaction costs of dealing with

FDI policy barriers[8], the costs of maintaining the affiliate, servicing network costs, and other costs associated with technology costs in offshore production. In a recent quantitative study by Head and Mayer [2019], the authors utilize highly disaggregated automotive industry data and find this type of variable distribution and marketing costs is higher than the conventional trade costs such as tariffs and freight, which is consistent with the assumption that I made here.

(ii) *Productivity sorting* There are two possible answers to equation $\pi^H_{FDI}(c) = \pi^H_X(c)$:

$$c^H_{FDI} = (1 \pm \sqrt{t^F})/(t^F \tau^F \pm \sqrt{t^F}\varphi^F)c^F_D \tag{2.18}$$

However, only one of the answers is interesting and relevant here. From both theoretical and empirical point of view, among those firms that serve foreign markets, multinational firms that engage in FDI are the most productive ones[9]. In the current framework, this implies $c^H_{FDI} < c^H_X < c^H_D$. Compare the expression of c^H_X and c^H_{FDI}, together with the assumption of $\varphi^F > \tau^F$, one can easily verify that both solutions of c^H_{FDI} in equation (2.18) imply that $c^H_{FDI} < c^H_X < c^H_D$. However, for the case of $c^H_{FDI} = (1 + \sqrt{t^F})/(t^F \tau^F + \sqrt{t^F}\varphi^F)c^F_D$, c^H_{FDI} will decrease in response to an increase in t^F, indicating the marginal multinationals will choose to become exporters when tariff increases. This is at odds with the empirical evidence in the literature. For example, Blonigen [2002] finds that tariff-jumping is a realistic option for multinational firms from industrialized countries. Hijzen et al. [2008] find horizontal tariff-jumping M&A evidence for 23 OECD countries for the period 1990–2001. More recently, Alfaro and Chen [2018] also find strong empirical evidence of tariff-jumping FDI through Orbis manufacturing firm-level dataset(60 countries, 2002–2007). Therefore, the other solution $c^H_{FDI} = (1 - \sqrt{t^F})/(t^F \tau^F - \sqrt{t^F}\varphi^F)c^F_D$ is more relevant here since c^H_{FDI} will increase in response to an increase in t^F, which is in line with the empirical evidence of productivity

[8]For example, according to Head and Mayer [2019], foreign car makers complained about the additional costs of daytime running lamps when Canada mandated them for new cars in 1990).

[9]For theoretical work, see Helpman et al. [2004]. For empirical evidence, see Doms and Jensen [1998] for the US and Conyon et al. [2002] for the U.K, for more recent evidence, see Mataloni [2011].

sorting and the tariff-jumping FDI.

(iii) *FDI motivation* Discussion on the multinational firm's FDI motivation here is important. In Helpman et al. [2004], the sorting of firms is induced by the assumption that $f_I > \tau^{\epsilon-1} f_X > f_D$: export incurs a fixed cost (f_X) and a higher marginal cost (τ), but as long as the fixed cost of FDI (f_I) is sufficiently high, the most productive firms are guaranteed to find FDI more desirable than export. This is a classic proximity-concentration trade-off in the spirit of Brainard [1997]. The similar trade-off is also present in Cole and Davies [2011], where the authors embed *ad valorem* tariff and *variable* fixed cost[10] into the Helpman et al. [2004] framework. They find that as the tariff increases, the exporter's variable profit decreases, while the differences in fixed cost remain the same. When the tariff level is sufficiently high, the gain from avoiding the tariff is higher than the fixed cost of becoming multinational, and a firm prefers FDI over export as an entry mode. In the current setup, this is no longer the case. Compare the profit function for an exporter and a multinational firm:

$$\pi_X^H(c) = \left[p_X^H(c)/t^F - \tau^F c\right] q_X^H(c) \tag{2.19}$$

$$\pi_{FDI}^H(c) = \left[p_{FDI}^H(c) - \varphi^F c\right] q_{FDI}^H(c) \tag{2.20}$$

As tariff increases, the revenue of the exporter will drop, making export a less desirable mode of accessing the Foreign market. Eventually, FDI becomes a more desirable entry mode. Although the marginal cost of FDI is higher than export (due to the assumption that $\varphi^F > \tau^F$), the operating profit of FDI exceeds the profit of export. The trade-off between export and FDI is no longer the conventional proximity-concentration trade-off, but a comparison of the profits. I also plot firms' profits as a function of marginal production cost, as in **Figure A.6**. Notice that the presence of a positive net tariff ensures that the

[10] It means firms with different productivity levels will face a different level of fixed cost when accessing the Foreign market.

profit of FDI is strictly higher than the profit of exports, whereas in Helpman et al. [2004] and Cole and Davies [2011], a similar condition is obtained through the combination of fixed cost and variable costs.

Given that the goal of this dissertation is to investigate the trade policy implications in an environment that features both export and multinational production, the selection into FDI (typically introduced through the fixed cost of FDI) margin is not of first-order interest here. A recent paper by Mrázová and Neary [2018] provides a justification for the current framework from a different perspective. They argue that statements like "Only the more productive firms select into the higher fixed-cost activity" are misleading: They are true given super-modularity[11] of the profit function, but otherwise may not hold. They discover that what matters for the direction of second-order selection effects (referring to the choice between export and FDI) is not a trade-off between fixed and variable costs, but whether there is a *complementarity between variable costs of production and trade*. In other words, if we allow FDI to be an equilibrium mode of accessing the Foreign market, then whether firms can afford them or not is independent of the fixed cost of FDI, but depends on the cross-effect of tariffs and production costs on profits. When super-modularity prevails, a more efficient firm has relatively higher operating profits in the FDI case, but when sub-modularity holds, the opposite may hold. The reason that the current setup can preserve the conventional sorting of firm productivity (i.e., second-order selection effect) is precisely due to the super-modularity of profit function[12] since there exists complementarity between the variable costs of production (in this case, the marginal cost of production c) and of trade (in this case, import tariff t).

[11]For example, super-modularity in $\Pi(t, c)$ means a higher tariff (t) reduces, in absolute value, the cost disadvantage of a higher-cost firm (larger c). For more details, please refer to Mrázová and Neary [2018].

[12]For a step-by-step verification, please refer to Mrázová and Neary [2018], their setup is a general preference, so they rely on the fixed cost to generate selection effects. For quadratic quasi-linear preference, they point out the first-order selection effect (according to their description, this is referring to whether serve the foreign market or not) needs the existence of choke price. The second-order effect is taken care of through the assumption we make.

2.2.3 Free Entry Condition

Entry is unrestricted in both countries. Firms choose a production location before entry and pay a sunk cost (f_E) to enter the market. To restrict the analysis on the effects of trade costs differences, I assume that countries share the same technology[13] (i.e., the same entry cost f_E and the same cost distribution $G(c)$). Free entry of domestic firms in country H implies zero expected profits in equilibrium, hence:

$$\int_0^{c_D^H} \pi_D^H(c)\, dG(c) + \int_{c_{FDI}^H}^{c_X^H} \pi_X^H(c)\, dG(c) + \int_0^{c_{FDI}^H} \pi_{FDI}^H(c)\, dG(c) = f_E \tag{2.21}$$

Given the Pareto assumption for cost distribution in both countries, the free entry condition for country H can be rewritten as:

$$\left(c_D^H\right)^{k+2} + \Phi_1^F \left(c_D^F\right)^{k+2} + \Phi_2^F \left(c_D^F\right)^{k+2} = \gamma\phi \tag{2.22}$$

where $\phi \equiv 2(k+1)(k+2)(c_M)^k f_E$ is a technology index that combines the effects of the better distribution of cost draws (lower c_M) and lower entry costs f_E. Moreover,

$$\begin{aligned}
\Phi_1^F \equiv{}& \frac{(k+1)(k+2)\, t^F (\tau^F)^2}{2} \left\{ \left(\frac{1}{t^F\tau^F}\right)^{k+2} - \left(\frac{1}{t^F\tau^F}\right)^2 \left(\xi^F\right)^k \right. \\
& \left. -\frac{2k}{k+1}\left[\left(\frac{1}{t^F\tau^F}\right)^{k+2} - \left(\frac{1}{t^F\tau^F}\right)\left(\xi^F\right)^{k+1}\right] + \frac{k}{k+2}\left[\left(\frac{1}{t^F\tau^F}\right)^{k+2} - \left(\xi^F\right)^{k+2}\right]\right\} \\
\Phi_2^F \equiv{}& \frac{(k+1)(k+2)\left(\xi^F\right)^k}{2}\left[1 - \frac{2k\varphi^F\xi^F}{k+1} + \frac{k\left(\varphi^F\xi^F\right)^2}{k+2}\right]
\end{aligned}$$

are indices that combine the trade-off between tariff and higher marginal cost of FDI. The free entry condition is homogenous to degree $k+2$ regarding the cutoff cost level. This

---

[13]For implications of Ricardian comparative advantage in the heterogeneous firm framework, please refer to the Appendix in Melitz and Ottaviano [2008].

system (for H, F) can then be solved for the cutoffs in both countries:

$$c_D^H = \left[\gamma\phi \frac{1 - \left(\Phi_1^F + \Phi_2^F\right)}{1 - \left(\Phi_1^F + \Phi_2^F\right)\left(\Phi_1^H + \Phi_2^H\right)}\right]^{\frac{1}{k+2}} \tag{2.23}$$

This equilibrium cutoff level differs from the one in Melitz and Ottaviano [2008] in two aspects. First, this cutoff is lower[14] than the closed-economy cutoff in Melitz and Ottaviano [2008]:

$$c_D^H < (\gamma\phi)^{\frac{1}{k+2}} \equiv \text{Closed-economy cutoff in Melitz and Ottaviano [2008]}$$

indicating the opening up of an economy via export and multinational activity will increase the aggregate productivity by forcing the least productive firms to exit. This result is similar to Melitz [2003], but the operating channel is the product market competition, not the factor market competition, as argued in Melitz and Ottaviano [2008]. Second, this cutoff is even lower[15]than the open economy cutoff in Melitz and Ottaviano [2008]:

$$c_D^H < \left(\gamma\phi \frac{1 - \rho^F}{1 - \rho^H \rho^F}\right)^{\frac{1}{k+2}} \equiv \text{Open-economy cutoff in Melitz and Ottaviano [2008]}$$

The intuition is straightforward: the presence of FDI, here the most productive firms in the distribution, intensifies the competitive environment in the economy, forcing the least productive firms to exit and hence further increases aggregate productivity[16]. Notice, the open economy cutoff stated above is slightly different from Melitz and Ottaviano [2008] due to the presence of *ad valorem* tariffs[17].

---

[14]This is based on Proposition 3, see Appendix A.3.3.

[16]This is consistent with the recent empirical evidence discovered in Fons-Rosen et al. [2013], although I abstract from the possibility of any spillover effect.

[17]Please see Appendix A.3.3 for more details.

2.2.4 Prices, Product Variety, Number of Entrants and Welfare

To see more features in the current setup, I first compute $\bar{p}^H$. Notice, the marginal cost of H's operating firms fall into the range $[0, c_D^H]$, which is also the range for delivered cost of exporters ($\tau^F c$), and the effective marginal cost of multinationals production ($\varphi^F c$). They all share identical distributions over the support given by $G^H(c) = (c/c_D^H)^k$. Therefore, the price distributions of H's domestic firms, $p_D^H(c)$, H's exporters producing in F, $p_X^F(c)$, and F's multinationals producing in H, $p_{FDI}^F(c)$, are all identical. The average price in country H is thus given by:

$$\begin{aligned} \bar{p}^H &= \frac{1}{G\left(c_D^H\right)} \int_0^{c_D^H} p_D^H\left(c\right) dG\left(c\right) = \frac{1}{G\left(c_X^F\right)} \int_{c_{FDI}^F}^{c_x^F} p_X^F\left(c\right) dG\left(c\right) \\ &= \frac{1}{G\left(c_{FDI}^F\right)} \int_0^{c_{FDI}^F} p_{FDI}^F\left(c\right) dG\left(c\right) = \frac{2k+1}{2k+2} c_D^H \end{aligned} \tag{2.24}$$

Combining this with the definition of $p_{\max}^H$ and $p_{\max}^F$, the number of firms selling in country H is:

$$N^H = \frac{2\gamma\left(\alpha - c_D^H\right)(k+1)}{\eta c_D^H} \tag{2.25}$$

From this expression, it must be the case that $\alpha > c_D^H$ so that the number of firms selling in country H is positive in equilibrium. The total number of product varieties in country H is composed of domestic producers, exporters, and multinationals from country F. Given a positive mass of entrants N_E in both countries, there are $G(c_D^H)N_E^H$ domestic producers, $[G(c_X^F) - G(c_{FDI}^F)]N_E^F$ Foreign exporters, and $G(c_{FDI}^F)N_E^F$ Foreign multinationals selling in H. Altogether they satisfy the following condition:

$$G\left(c_D^H\right) N_E^H + \left[G\left(c_X^F\right) - G\left(c_{FDI}^F\right)\right] N_E^F + G\left(c_{FDI}^F\right) N_E^F = N^H \tag{2.26}$$

Solving this system (for H and F) will give us the number of entrants in country H:

$$N_E^H = \frac{2\,(c_M)^k\,(k+1)\,\gamma}{\eta\,(1-\delta^H\delta^F)}\left[\frac{\alpha - c_D^H}{(c_D^H)^{k+1}} - \delta^H\frac{\alpha - c_D^F}{(c_D^F)^{k+1}}\right] \tag{2.27}$$

where $\delta^l = (t^l\tau^l)^{-k}$, for $l \in \{H, F\}$. Notice, the condition that ensures positive equilibrium number of varieties (N^l) in the economy, $\alpha > c_D^l$, also guarantees the positive mass of entry (N_E^l) in the equilibrium.

Equation (2.27) marks a crucial difference between the current framework and Cole and Davies [2011]: The number of entrants in the economy is endogenously affected by the tariff level. As first noted by Balistreri et al. [2011], the equilibrium level of firm entry in the Melitz-type model is no longer fixed if: (i) *ad valorem* tariffs are imposed rather than iceberg transport costs, or (ii) there are multiple sectors in the economy. Equilibrium entry-level becomes endogenous in the current framework because of not only the *ad valorem* tariff, but also the two-sector economy. However, it is surprising to note that Cole and Davies [2011] ignore this dimension in their discussion, although their framework also features these two aspects. The authors might have got the impression from the basic Melitz model that trade costs are independent of firm entry, which is true in Melitz [2003]. However, when analyzing the revenue-generating tariff, this perception is no longer valid[18]. This implication is crucial in understanding the equilibrium feature of the model. I will come back to this point in the Section 2.4 of this chapter.

Following Melitz and Ottaviano [2008], combine equation (2.4), (2.24), (2.25) and the definition of $\sigma^2_{p^H}$, it is straightforward to show the consumer welfare in H equals to:

$$U^H = I^H + \frac{\alpha - c_D^H}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_D^H\right) \tag{2.28}$$

Once again, consumer welfare changes monotonically with the domestic cutoff, which cap-

[18]More specifically, according to Arkolakis et al. [2012], one of the 'macro' assumption (R2) is violated, therefore entry become endogenous.

tures the effect of an increase in product variety and a decrease in the average price. Also notice, consumer surplus in country H is given by the second term in equation (2.28).

2.2.5 Tariff Revenue and National Welfare

This part of the model is very important for the analysis of socially optimal tariff and Nash tariff in Chapter 3 of this dissertation. Note that tariff revenue is also a component of consumer income I^H through the redistribution from the government. I define the pre-tax value of country H's import as:

$$\begin{aligned} IM^H &= N_E^F \int_{c_{FDI}^F}^{c_X^F} \frac{p_X^F(c)}{t^H} q_X^F(c)\, dG(c) \\ &= N_E^F \frac{t^H \left(\tau^H\right)^2 \left(c_D^H\right)^{k+2}}{4\gamma (k+2) \left(c_M\right)^k} \left[2\left(\frac{1}{t^H \tau^H}\right)^{k+2} - \frac{k+2}{\left(t^H \tau^H\right)^2} \left(\xi^H\right)^k + k \left(\xi^H\right)^{k+2} \right] \end{aligned} \quad (2.29)$$

Therefore, the total import tariff revenue of country H is defined as

$$\begin{aligned} TR^H &\equiv \left(t^H - 1\right) \times IM^H \\ &= N_E^F \frac{t^H - 1}{t^H} \frac{\left(c_D^H\right)^{k+2}}{4\gamma (k+2) \left(c_M\right)^k} \left[2\left(\frac{1}{t^H \tau^H}\right)^{k} - (k+2) \left(\xi^H\right)^k + k \left(\xi^H\right)^{k+2} \left(t^H \tau^H\right)^2 \right] \end{aligned} \quad (2.30)$$

From the trade-policy perspective, the government will use its policy instrument to maximize consumer welfare:

$$U_n^H = \underbrace{w^H + \left(t^H - 1\right) \times IM^H + \Pi^H}_{\equiv I^H} + \underbrace{\frac{\alpha - c_D^H}{2\eta} \left(\alpha - \frac{k+1}{k+2} c_D^H \right)}_{\equiv CS^H} \quad (2.31)$$

Therefore tariff affects consumer welfare from two channels: (i) consumer surplus, which is directly affected by the change in c_D^H in response to tariff, and (ii) tariff revenue, which is affected by both the tariff level (t^H) and the tariff base (IM^H). Due to the free-entry condition, aggregate profit Π^H will be driven to zero in equilibrium. Notice due to the presence of numéraire good, $w^H = 1$, although consumers will not take t^H into consideration when maximizing their utility, the government does take consumers into consideratoin by choosing the utility maximizing tariff level.

With the model above, I will now discuss the equilibrium features of this economy. All the results will be contrasted with an economy that features heterogeneous firms, FDI, but constant markups, i.e., the one in Cole and Davies [2011].

2.3 Short-run Equilibrium

As discussed in Section 2.2, due to the presence of *ad valorem* tariff and the quadratic quasi-linear preference (which results in multiple sectors in the current setup), the level of firm-entry become endogenous under the free-entry condition. To see how the model directly responds to the trade/tariff liberalization, it is, therefore, necessary to separate the short-run (when the entry is restricted) from the long-run (free entry). In this section, I introduce the short-run version of the model and discuss its equilibrium characteristics.

In the short-run, when entry and exit are prohibited, the economy is characterized by a fixed number of incumbents, and they decide whether to produce or shut-down based on their profits. More specifically, Home country is characterized by a fixed number of incumbents, $\bar{N_I^H}$, with cost distribution $\bar{G^H}$ on $[0, \bar{c}_M]$, where $\bar{c}_M$ is within the long-run technology frontier, c_M. I keep the assumption that the productivity $1/c$ is distributed with Pareto shape k, implying $\bar{G}^H(c) = (c/\bar{c}_M)^k$. The distribution of firm's productivity in the short-run model is briefly displayed in **Figure A.7**.

A Home firm produces if it can earn nonnegative profits from either its domestic market, export market, or FDI market. These decisions based on profits lead to the following short-run cutoff conditions:

$$c_D^H = \sup\{c : \pi_D^H(c) \geq 0 \text{ and } c \leq \bar{c}_M\}$$
$$c_X^H = \sup\{c : \pi_X^H(c) \geq 0 \text{ and } c \leq \bar{c}_M\}$$
$$c_{FDI}^H = \sup\{c : \pi_{FDI}^H(c) \geq \pi_X^H(c) \text{ and } c \leq \bar{c}_M\}$$

Firms with marginal cost $c > c_D^H$ will shut-down. Utilizing the zero-profit conditions, one can establish the following relations between cutoff levels and the number of operating firms in Home and Foreign:

$$N^H = \frac{2\,(k+1)\,\gamma}{\eta} \times \frac{\alpha - c_D^H}{c_D^H}$$

$$N^F = \frac{2\,(k+1)\,\gamma}{\eta} \times \frac{\alpha - t^F \tau^F c_X^H}{t^F \tau^F c_X^H}$$

where N^H and N^F represent the endogenous number of sellers in country H and F in the short-run. Notice that the different cutoffs satisfy the same condition as in the long-run. There are $\bar{N}_I^H \bar{G}\left(c_D^H\right)$ producers in H who sell in their domestic market, $\bar{N}_I^F \left[\bar{G}\left(c_X^F\right) -\bar{G}\left(c_{FDI}^F\right)\right]$ Foreign exporters, and $\bar{N}_I^F \bar{G}\left(c_{FDI}^F\right)$ Foreign FDI firms in H. These numbers must add up to the total number of producers in country H. Similar equation also holds for country F:

$$N^H = \bar{N}_I^H \bar{G}\left(c_D^H\right) + \bar{N}_I^F \left[\bar{G}\left(c_X^F\right) - \bar{G}\left(c_{FDI}^F\right)\right] + \bar{N}_I^F \bar{G}\left(c_{FDI}^F\right)$$

$$N^F = \bar{N}_I^F \bar{G}\left(c_D^F\right) + \bar{N}_I^H \left[\bar{G}\left(c_X^H\right) - \bar{G}\left(c_{FDI}^H\right)\right] + \bar{N}_I^H \bar{G}\left(c_{FDI}^H\right)$$

Combining these two equations with the threshold price conditions yield expressions for the cost cutoffs in both countries:

$$\frac{\alpha - c_D^H}{\left(c_D^H\right)^{k+1}} = \frac{\eta}{2\left(k+1\right)\gamma}\left\{\frac{\bar{N}_I^H}{\bar{c}_M^k} + \left[\left(\frac{1}{t^H\tau^H}\right)^k - \left(\xi^H\right)^k\right]\bar{N}_I^F + \left(\xi^H\right)^k\bar{N}_I^F\right\} \tag{2.32}$$

$$\frac{\alpha - c_D^F}{\left(c_D^F\right)^{k+1}} = \frac{\eta}{2\left(k+1\right)\gamma}\left\{\frac{\bar{N}_I^F}{\bar{c}_M^k} + \left[\left(\frac{1}{t^F\tau^F}\right)^k - \left(\xi^F\right)^k\right]\bar{N}_I^H + \left(\xi^F\right)^k\bar{N}_I^H\right\} \tag{2.33}$$

Note, these two conditions uniquely identify the short-run cutoff levels(c_D^H, c_D^F) with the number of producing firms in each country (N^H, N^F).

Equation (2.32) and (2.33) also clearly highlight the protection role played by import tariff in the short-run. Based on these two equations, we have the following proposition for the short-run equilibrium.

Proposition 1. *In the short-run equilibrium, an increase in Home country's import tariff (t^H) can protect Home producers from Foreign competition, increasing the domestic cost cutoff:*

$$\frac{\partial c_D^H}{\partial t^H}|_{short\text{-}run} > 0$$

Proof. See Appendix A.3.1. □

Intuitively, an increase in H's tariff will make it harder for the Foreign exporters to access the Home market, so the number of exporters from F to H will decrease. At the same time, an increase in H's tariff will induce tariff-jumping FDI among the Foreign exporters, so the number of Foreign firms that access the Home market through FDI will increase. In the current setup, the decrease of exporters surpasses the increase of FDI firms, so the right-hand side of the equation (2.32) is decreasing in t^H, indicating an increase in H's domestic cost cutoff (c_D^H). Therefore, an increase in H's tariff reduces the total number of Foreign firms (exporters and FDI firms) accessing the Home market, making it easier for Home producers to survive.

In other words, import tariff, in the short-run, can effectively shield Home from Foreign competition. This result is similar to the result in Section 3.7 of Melitz and Ottaviano [2008]. However, they obtain the result of an increase in cutoff level through an exogenous variation of trading partner industrial size (N^H or N^F). In the current framework, a change in tariff level alters the relative size of Home and Foreign firms, affecting the cutoff levels. This finding might seem to confirm the findings in Cole and Davies [2011][19], but if we allow firms to enter freely, then the result will be quite different. As we will see in the next section, the classic 'delocation'[20] result will arise.

Based on equation (2.6), (2.10) and (2.15), we can also obtain the following proposition regarding markups in respond to a tariff change:

Proposition 2. *In the short-run equilibrium, an increase in Home country's import tariff (t^H) can increase domestic producer's markup, may decrease or increase Foreign exporter's markup, and increase Foreign FDI firm's markup.*

$$\frac{\partial m_D^H}{\partial t^H}|_{short\text{-}run} > 0, \frac{\partial m_X^F}{\partial t^H}|_{short\text{-}run} ?0, \frac{\partial m_{FDI}^F}{\partial t^H}|_{short\text{-}run} > 0$$

Proof. See Appendix A.3.2. □

Intuitively, for an increase in t^H, tariff affects *Home domestic producer's* markup $m_D^H(c)$ through the equilibrium effect on c_D^H. Protection makes it easier for Home producers to survive and results in a higher c_D^H, meaning a higher markup for all the domestic sellers. For *Foreign exporters*, their markup $m_X^F(c)$ is affected by tariff from two aspects: (i) direct

[19]Specifically, their justification of their equation (11). Similar results are qualitatively identical to those of other similar models in the heterogeneous firm literature, such as Melitz [2003] and Helpman et al. [2004].

[20]The delocation effect has been studied in previous work(see, for example, Venables [1985], Helpman and Krugman [1989], Baldwin et al. [2003]) and here is also confirmed in the heterogeneous firm framework with FDI.

effect—-an increase in t^H directly raises m_X^F, meaning Foreign exporters will pass the tariff burden to the consumers by increasing markup, and (ii) indirect effect—-the tariff indirectly affects m_X^F through the equilibrium effect on c_D^H. With restricted entry, these two effects are in the opposite direction. Which effect is more dominant depends on the individual exporter's productivity. For the least productive Foreign exporters (large c), the direct effect will dominate the indirect effect, resulting in a drop in Foreign exporter's markup, indicating Home's protection will reduce the market power of Foreign exporters. Nevertheless, for the most productive Foreign exporters (small c), the indirect effect will dominate the direct effect, resulting in the bigger market power of Foreign exporters. For *Foreign FDI firms*, tariff affects $m_{FDI}^F(c)$ through the equilibrium effect on c_D^H. Protectionism results in a less competitive Home environment, which benefits the more productive Foreign FDI firms, and allows them to charge higher markups.

In this short-run equilibrium, where additional entry of firms is restricted, the findings in the current framework confirm the previous results in the literature on unilateral trade/tariff liberalization. As we will see in the next section, these results will be reversed with an endogenous level of firm entry in the long-run.

2.4 Long-run Equilibrium

In this section, I analyze the equilibrium features of the model when firms can enter and exit freely. As mentioned in Section 2.2, firm entry, in the long-run, is endogenously affected by the tariff level. This will have crucial implications for trade/tariff liberalization. In what follows, I will illustrate the comparative statics of the model and contrast its benchmark results with Cole and Davies [2011].

First of all, following the discussion of equation (2.23), the presence of FDI in the Melitz and Ottaviano [2008] world will deliver a different equilibrium domestic cutoff in the economy, which can be summarized by the following proposition:

Proposition 3. *The presence of FDI makes the economy more competitive, and the domestic* *cutoff is lower compared to the case when there is no FDI:*

$$c_{D1}^H = \left[\gamma\phi\frac{1-\left(\Phi_1^F+\Phi_2^F\right)}{1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)}\right]^{\frac{1}{k+2}} < c_{D2}^H = \left[\gamma\phi\frac{1-\psi^F}{1-\psi^F\psi^H}\right]^{\frac{1}{k+2}}$$

Proof. See Appendix A.3.3. □

Different from the short-run equilibrium case, here the domestic cost cutoff is pinned down through the long-run free-entry condition (equation (2.22)). In the Appendix A.3.3, I show $\Phi_1^l + \Phi_2^l > \psi^l$ for $l \in \{H, F\}$. The sum of Φs can be viewed as a measure of 'openness'. Intuitively, the presence of FDI makes the country more 'open' compared to the case when FDI is not an option. In Melitz and Ottaviano [2008], ψ^l [21]measures the 'freeness' of trade. The presence of Foreign FDI intensifies the Home country's competitive environment, making it harder for Home producers to survive. The marginally surviving firm needs to be more productive. Openness (either through export or FDI) increases competition[22] in the domestic product market, shifting up residual demand price elasticities for all firms at any given demand level. Therefore, the least productive firms are forced to exit. This effect is very similar to an increase in market size in the closed economy: the increased competition induces a downward shift in the distribution of markups across firms. Although only relatively more productive firms survive (with higher markups than the less productive firms who exit), the average markup is reduced.

With free-entry, firms can freely enter and exit the market in the long-run. Due to the presence of *ad valorem* tariff and the quadratic quasi-linear preference, the number of entrants in the monopolistically competitive sector is endogenously affected by the level of tariff. As we will see soon, this feature has important implications for trade/tariff liberalization. For example, a unilateral change in Home country's import tariff has quite different impacts on the domestic productivity cutoffs of both countries, as can be seen from the following

[21] More precisely, the freeness of trade is measured by τ^{-k} in Melitz and Ottaviano [2008]. Here, due to the presence of tariff, this term is augmented to incorporate tariff, $\tau^{-k}t^{-(k+1)}$.

[22] Compared to the case when export is the only option to access Foreign market.

proposition:

Proposition 4. *An increase in country H's import tariff results in a decrease in the cutoff cost level in country H's domestic market, and an increase in the cutoff cost level in country F's domestic market:*

$$\frac{\partial c_D^H}{\partial t^H} < 0 < \frac{\partial c_D^F}{\partial t^H}$$

Proof. See Appendix A.3.4. □

Different from the short-run outcome, where the H's domestic cost cutoff would increase as a result of tariff protection, in the long-run, H's domestic cost cutoff would decrease in response to tariff protection. Intuitively, although an increase in the import tariff raises the protection level in country H in the short-run, it also fosters a more extensive entry from domestic firms over time. Proposition 7 near the end of this section further demonstrates this point. Protection makes the Home country a more desirable environment for firms to do business in the long-run. With the free-entry condition, the larger entry will generate a higher competition in the domestic market, driving out the least productive firms, and forcing the marginally surviving firms to be more productive.

This result is quite different from Cole and Davies [2011]. They find an increase in the import tariff in country H will raise the protection level in country F, shield country H's firm from the competition, and make the domestic surviving firms less productive, i.e., $\partial c_D^H / \partial t^H > 0$ (their equation (11)). This result is primarily due to the fact that the presence of quadratic quasi-linear preference affects firm-entry, which is entirely absent in Cole and Davies [2011].

A unilateral change in Home's import tariff also affects the exporters in both countries. The impact can be summarized as the following:

Proposition 5. *An increase in country H's import tariff results in an increase in the export cutoff cost level in country H and a decrease in the export cutoff cost level in country F:*

$$\frac{\partial c_X^H}{\partial t^H} > 0 > \frac{\partial c_X^F}{\partial t^H}$$

Proof. See Appendix A.3.5. □

An increase in import tariff in country H will cause the least productive exporters from F to quit exporting, and only serve their domestic market. The reason is that the increase in tariff reduces exporter's revenue (hence profit), making it less desirable for the least productive exporters to serve H's market. With their exit, the marginally surviving exporters are more productive, hence a lower c_X^F. This result is the same as the one obtained in Cole and Davies [2011]. Moreover, although protection generates more entry of Home firms in the long-run, giving a competitive pressure for Home's export market, this impact is dominated by the protective effect of the tariff on exporters, resulting in a higher Home exporter cutoff, and making it easier for Home country to export.

Proposition 6. *An increase in country H's import tariff results in an increase in the FDI cutoff cost level in country F:*

$$\frac{\partial c_{FDI}^F}{\partial t^H} > 0$$

Proof. See Appendix A.3.6. □

The intuition here is straightforward: the most productive exporters from F, when facing an increase in import tariff in H, will find it less desirable to access H's market through export, and hence choose FDI as the entry mode. This result is due to the profit of FDI outweighing the profit of export when t^H increases, as can be seen from equation (2.19) and (2.20). Hence the marginally surviving multinationals from country F are now less productive since previously they are exporters, resulting in a higher c_{FDI}^F. This result is similar to the findings in Cole and Davies [2011].

To sum up these results and contrast them with Cole and Davies [2011], I plot the productivity cutoffs and their responses toward Home country's unilateral change in tariff, which is shown in **Figure A.8**. When t^H increases, from equation (11)–(13) in Cole and Davies [2011], the least productive Foreign exporters exit the domestic market (c_X^F decreases) and the most productive Foreign exporters become multinationals (c_{FDI}^F increases). This change makes the composition of domestic Foreign firms (including F's exporters and multinationals) more productive. Due to the protection, the domestic market is shielded from the Foreign competition. Hence domestic firms find it easier to survive (c_D^H increases).

In the current setup, an increase in t^H will similarly lead the least productive Foreign exporters to exit the domestic market (c_X^F decreases) and the most productive Foreign exporters to become multinationals (c_{FDI}^F increases), making the composition of domestic Foreign firms more productive. However, Home's protection will attract more Home firms to enter the domestic market (N_E^H increases), making the Home country's environment more competitive. Therefore, domestic firms need to be more productive to survive (c_D^H decreases).

In both cases, we have tariff-jumping FDI in response to the increase of t^H. In Cole and Davies [2011], tariff-jumping intensifies the competitive environment in the domestic market of H, but this effect is dominated by the protection effect raised through tariff. So the outcome is an easier-to-survive environment. In the current setup, the tariff-jumping FDI intensifies the competitive environment in the domestic market. The excessive entry generated by protection also makes the domestic environment more competitive. These two effects together result in a tougher environment in Home's domestic market, making it harder for firms to survive.

Corollary 1. *Under the assumption that $\varphi^H > \tau^H$*[23]*, an increase in H's import tariff results in a tougher competitive environment in the domestic market over time, and this effect is exacerbated by the presence of FDI:*

$$\left| \frac{\frac{\partial c_D^H}{\partial t^H}\Big|_{without\ FDI}}{\frac{\partial c_D^H}{\partial t^H}\Big|_{with\ FDI}} \right| < 1$$

Proof. See Appendix A.3.7. □

[23] The domestic cutoff without FDI but with ad valorem tariff is $c_D^H = \left[\gamma\phi\left(1-\rho^F\right)/\left(1-\rho^H\rho^F\right)\right]^{1/(k+2)}$, where $\rho^H = (\tau^H)^{-k}(t^H)^{-(k+1)}$. The domestic cutoff with FDI is defined in equation (2.23).

Finally, the protection also affects the number of entrants and eventually the number of products available in each country. I summarize this result in the following proposition:

Proposition 7. *An increase in H's import tariff results in an increase in the number of entrants in H and a decrease in the number of entrants in F. Over time, this effect contributes to an increase in the number of varieties in H and a decrease in the number of varieties in F:*

$$\frac{\partial N_E^H}{\partial t^H} > 0 > \frac{\partial N_E^F}{\partial t^H}$$
$$\frac{\partial N^H}{\partial t^H} > 0 > \frac{\partial N^F}{\partial t^H}$$

Proof. See Appendix A.3.8. □

The intuition is obvious: Home's tariff protection makes the Home country a more desirable environment to do business for the firms. In the long-run, more firms would choose to enter Home's market, resulting in a larger number of products available in the equilibrium. The opposite condition will hold for the Foreign market. Clearly, this result crucially depends on the fact that tariff can affect the number of entrants in this economy. No similar results are discussed in Cole and Davies [2011].

Proposition 8. *In the long-run equilibrium, an increase in Home country's import tariff (t^H) can decrease domestic producer's markup, decrease Foreign exporter's markup, and decrease Foreign FDI firm's markup.*

$$\frac{\partial m_D^H}{\partial t^H}|_{long\text{-}run} < 0, \frac{\partial m_X^F}{\partial t^H}|_{long\text{-}run} < 0, \frac{\partial m_{FDI}^F}{\partial t^H}|_{long\text{-}run} < 0$$

Proof. See Appendix A.3.9. □

The intuition follows right after the previous proposition. Due to the increase of protectionist tariff, Home becomes a more favorable environment to do business, attracting more firm to enter in the long-run. This effect will increase the competition in the Home market, reducing the markups for all kinds of producers that serve the Home market.

2.5 Concluding Remarks

This chapter introduces *ad valorem* tariff and horizontal FDI into the Melitz and Ottaviano [2008] framework. To the best of my knowledge, this is the first paper in the trade policy literature that incorporates firm heterogeneity, variable markups, and multinational production. The results can be broadly summarized as follows.

First, I find that the presence of multinational production generates a competitive effect on the economy, and firms need to be more productive to survive the competition. Second, I find that the *ad valorem* tariff and quadratic quasi-linear preference collectively result in an endogenous level of firm entry. Therefore, the impact of trade/tariff liberalization will depend on the equilibrium number of firms. In the short-run, when the firm entry is prohibited, an increase in import tariff shields the domestic economy from the Foreign competition, making it easier for firms to survive. This result is overturned when I allow firms to enter the market freely in the long-run. In the long-run, an increase in Home's import tariff will make the Home country a more desirable environment to do business, attracting more entrants in the Home market, making the Home market more competitive. Firms need to be more productive to survive. Home's tariff increase also makes it harder for the least productive Foreign exporters to survive, and triggers tariff-jumping FDI among the most productive exporters. Markups also respond to tariff change differently in the short-run vs. long-run, primarily due to the change of competitive environment brought in through firm entry.

Given these results, there are several potential avenues for future theoretical research. One the one hand, one can deviate from the Pareto distribution assumption. As discussed by Feenstra [2018], if the support of cost distribution becomes bounded, other channels that affect the pro-competitive effect of trade will begin to work, delivering a different welfare implication of trade. On the other hand, the combination of Melitz and Pareto implies that trade costs will only affect export through extensive margin, but this implication is at odds with the empirical fact that most of the adjustments happen along the intensive margin. This phenomenon can be reconciled by introducing log-normal distribution[24].

Another exciting path will be to investigate if the current results on trade/tariff liberalization are robust under alternative demand/supply structures that generate variable markups. Several approaches are readily available. (i) Deviating from the *quadratic quasi-linear preference*. Many preferences that are surveyed in Chapter 1 Section 1.3.1 are great alternatives. I have investigated *translog expenditure function* as in Rodriguez-Lopez [2011], and find that most of the claims in this chapter will still be valid. (ii) Dropping the *monopolistic competition* assumption through the approaches discussed in Chapter 1, Section 1.3.2. I expect this option to be particularly interesting, given the recent increased attention on oligopoly[25]. (iii) Deviating from *quadratic quasi-linear preference* and *monopolistic competition*. However, based on the discussion in Chapter 1 Section 1.3.3, all the existing works adopt quadratic quasi-linear preference, so (iii) is identical to (ii).

Lastly, the implications of trade/tariff liberalization here primarily focus on the monopolistically competitive sector. It would be interesting and relevant to see how a multi-sector framework would affect the results. For example, Spearot [2016] extends Melitz and Ottaviano [2008] by incorporating multiple countries and multiple industries, with heterogeneity in the country-by-industry shape parameters of the Pareto cost distributions, and provides a quantitative implication for unilateral tariff liberalization. The excess entry in the current framework might have different implications in his model since the mass of entrants in a particular sector now depends on the relative expenses on goods produced in that sector. I leave these questions for future work. Now, with a loud fanfare, I turn to the study of optimal trade tariffs in the current framework in the next chapter.

[24] See Fernandes et al. [2017] for a detailed discussion.

[25] See Head and Spencer [2017] for more discussion.

Chapter 3

OPTIMAL TARIFFS WITH FIRM HETEROGENEITY, VARIABLE MARKUPS, AND FDI

3.1 Introduction

What is the welfare implication of protectionist trade policy in an environment that features variable markups and foreign direct investment (henceforth, FDI)? On the one hand, protectionism may hurt consumer welfare in the presence of variable markups if protection results in higher market concentration. This dilemma has been a concern since Adam Smith, and it has received increasing attention in recent years[1]. On the other hand, in a highly-integrated global market, foreign firms can avoid import tariffs by locating production within the destination market. Such 'tariff-jumping' activities can diminish the market power of domestic producers, thereby substantially mitigating the welfare consequences of the original protectionist trade policy.

The goal of this chapter is to utilize the theoretical framework developed in the previous chapter to study the welfare implication of tariff and optimal tariffs. I first follow Nocco et al. [2014] to compare the market allocation with the socially optimum allocation. I find with the free-entry condition, the market outcome in the monopolistically competitive sector[2] is not efficient in several dimensions: (i) The selection is too weak in domestic and export cutoff, but too strong in FDI cutoff. (ii) The market oversupplies high-cost varieties and undersupplies low-cost varieties. (iii) It may feature excessive (insufficient) entry and

---

[1]Outside of the academic literature, increasing market concentration has received significant attention, e.g., *A lapse in concentration* (The Economist, September 2016), CEA [2016]. In the academic literature, see Asker et al. [2017], De Loecker and Eeckhout [2017] for recent evidence.

[2]Since tariff is only imposed in this sector, here I focus exclusively on the policy implication in this sector. There is, indeed, an inter-sector misallocation between the numéraire good sector and the monopolistically competitive sector. Interesting readers should refer to Nocco et al. [2014].

oversupply (undersupply) the total number of varieties. These market failures stem from several externalities: (i) On the supply side, both the markup-pricing and business-stealing effect tend to create too many varieties in the economy. (ii) On the demand side, the 'love of variety' from the quadratic quasi-linear preference tends to create insufficient varieties in the economy. (iii) With variable markup, firm heterogeneity becomes another source of inefficiency: as the demand becomes more inelastic with consumption, low-cost firms charge higher markups, and do not fully transmit their cost advantage to prices. This behavior leaves inefficiently large room for entry and also allows high-cost firms to be inefficiently large. These externalities collectively result in inefficiencies in the market outcome. If the market selection is too weak, then an increase in tariff can improve market selection, reducing the welfare gap between market allocation and the planner's allocation. In contrast, with constant elasticity of substitution (henceforth, CES) preference, the market outcome coincides with the socially optimum outcome.

Several interesting policy implications stand out from the analysis. First, similar to Cole and Davies [2011], I find that free trade is not always socially optimal. The intuition, however, is different. Cole and Davies [2011] find the socially optimal tariff is always to subsidize trade. They reach this conclusion because, in their model, trade liberalization can foster competition and eliminate the least productive firms, increasing aggregate productivity. Here I find whether imposing a tariff is socially optimal depends on the level of market selection. If the domestic cutoff is sufficiently high, which means the selection is too weak, then an additional firm entry can increase social welfare. In this case, a positive tariff is socially optimal because it encourages firm entry[3]. In addition, I find that the degree of firm heterogeneity, which is governed by the Pareto shape parameter k and the upper bound of cost draw c_M, affects the welfare implication of tariff. For example, when the domestic cutoff is sufficiently high, an increase in firm heterogeneity (through an increase in c_M or a decrease in k) is socially inefficient because it reduces the positive externality generated through the firm

[3]Recall in Chapter 2, we established that entry is increasing in tariff, as shown in Proposition 7.

entry, dampening the welfare impact of the tariff.

Second, the Nash tariff is lower than the socially optimal tariff. This result can be analyzed from two perspectives. On the one hand, when Home country sets its uncooperative tariff level, it focuses exclusively on its own tariff revenue and consumer surplus, ignoring the impact on Foreign tariff revenue and consumers. Therefore, Home country will set a higher tariff than the one social planner would choose. On the other hand, different from the intuition in Cole and Davies [2011], a higher Nash tariff level could also arise from Home country's incentive to manipulate the terms of trade. Due to the presence of variable markup, Home's import price varies with its tariff level. Furthermore, the profits of Foreign exporters and multinationals are all affected by Home's tariff level. An increase in Home's import tariff thus generates a terms of trade gain at the cost of Foreign's terms of trade deterioration. This channel is absent in Cole and Davies [2011] due to the constant markup implied by the CES preference.

Third, variable markups, especially their interactions with FDI, yield novel insights on trade policy. While recent studies on the welfare implication of trade liberalization[4] emphasize the importance of variable markup, the insight here is that the role of FDI should not be ignored. A decrease in Home's import tariff makes it easier for the most productive Foreign domestic firms to export, increasing the number of Foreign exporters serving the Home market, and creating downward pressure on the Home average markup. At the same time, the reduction of tariffs also makes it less desirable for the least productive Foreign multinationals to pursue FDI, decreasing the number of Foreign FDI firms and generating upward pressure on the Home average markup. If the initial protection level is sufficiently high, the decrease of multinational firms can dominate the increase of exporters, driving up the average markup in the Home market, and generating a negative pro-competitive effect.

Another insight comes from the interaction of variable markup and FDI related to the Nash tariff level. In Cole and Davies [2011], the presence of FDI effectively reduces the tariff

[4]For example, Edmond et al. [2015] and Arkolakis et al. [2018].

base, mitigating the tariff competition between the two countries. In the current framework, FDI's impact on the tariff level depends on a particular parameter that governs the variable markups: the shape of Pareto distribution, k. A decrease in k means an increase in the degree of firm heterogeneity, which has two impacts on the economy: changing the equilibrium cutoff levels and altering the relative distribution of firms with different marginal costs. When FDI is an option, the easiness of doing FDI (measured by φ) interacts with k, jointly affecting the Nash tariff level: if the degree of firm heterogeneity is big (smaller k), reducing the FDI barrier can lower the Nash tariff level; if the degree of firm heterogeneity is small (larger k), however, promoting FDI can increase the Nash tariff level. Another parameter that also impacts the degree of firm heterogeneity, c_M, only affects the equilibrium cutoffs proportionally, without shifting the relative distribution of firms. Therefore, its interaction with FDI does not change the Nash tariff level.

The remainder of this chapter is organized as follows. Section 3.2 compares the market allocation with the socially optimum allocation. Section 3.3 studies whether free trade is socially optimal in this economy. Section 3.4 contrasts the socially optimal tariff level with the Nash tariff level, and also compares the Nash tariff level with and without FDI. Section 3.5 is dedicated to the role of variable markup in this economy. The policy implication of its interaction with FDI is also discussed in this section. Section 3.6 concludes. All tables and graphs to which this chapter refers are included in the appendix.

3.2 Social Optimum vs. Market Outcome

In this section, I first derive the socially optimum outcome in the framework introduced in Chapter 2. Then I compare the socially optimum outcome with the market outcome derived in the previous chapter. Then I analyze the forces in the economy that result in the departure of market outcome from socially optimum outcome.

The social planner's problem can be described as the following. Since the quadratic quasi-linear utility implies transferable utility, social welfare can be expressed as the sum of all the representative consumers' utilities. Following Nocco et al. [2014], the planner chooses the number of entrants(N_E^H, N_E^F), and production level for homogeneous good and heterogeneous good($q_0^H, q_0^F, q_i^H, q_i^F$) to maximize social welfare subject to aggregate resource budget constraint:

$$\max_{\{N_E^H, q_0^H, q_i^H, N_E^F, q_0^F, q_i^F\}} \mathbb{W} \equiv \mathbb{U}_{\mathbb{H}} + \mathbb{U}_{\mathbb{F}}$$

$$\begin{aligned} \text{s.t. } q_0^H + q_0^F + f\left(N_E^H + N_E^F\right) + N_E^H \int_0^{c_M} \left[c q_D^H(c) + \tau^F c q_X^H(c) + \varphi^F c q_{FDI}^H(c)\right] dG(c) \\ + N_E^F \int_0^{c_M} \left[c q_D^F(c) + \tau^H c q_X^F(c) + \varphi^H c q_{FDI}^F(c)\right] dG(c) = 2 + \bar{q}_0^H + \bar{q}_0^F \end{aligned}$$

where $q_0^H + q_0^F$ stands for the supply of homogeneous good in both countries, $f\left(N_E^H + N_E^F\right)$ the sunk entry cost in the monopolistically competitive sector in H and F, $N_E^H \int_0^{c_M} [c q_D^H(c) + \tau^F c q_X^H(c) + \varphi^F c q_{FDI}^H(c)] dG(c)$ the supply of differentiated varieties in the Home country, and the last term on the left-hand side of the constraint gives the supply for differentiated varieties in the Foreign country. On the right-hand side, we have the endowment of labor and homogeneous good in both countries. The differences between socially optimum outcome and market outcome can be summarized in the following proposition:

Proposition 9. *In the current framework, compared to the socially optimum, the market outcome differs in several dimensions:*

(A) Marginal cost cutoffs

(i) the Home domestic market selection is weaker than the socially optimum selection

$$c_D^{HM} > c_D^{HO}$$

(ii) the Home exporter market selection is weaker than the socially optimum selection

$$c_X^{HM} > c_X^{HO}$$

(ii) the Home FDI market selection is stronger than the socially optimum selection

$$c_{FDI}^{HM} < c_{FDI}^{HO}$$

(B) Intensive margin

(i) Home's domestic producers undersupply varieties with low marginal production cost[5]*,*

$$q_D^{HM} < q_D^{HO} \text{ if } c < \left[2 - (2/\triangle_F)^{1/(k+2)}\right] c_D^{HO}$$

(ii) Home's exporters undersupply varieties with low marginal cost

$$q_X^{HM} < q_X^{HO} \text{ if } c < \left[2 - (2/\triangle_F)^{1/(k+2)}\right] \frac{c_D^{FO}}{\tau^F(2 - t^F)}$$

(iii) Home's FDI firms also undersupply varieties with low marginal production cost

$$q_{FDI}^{HM} < q_{FDI}^{HO} \text{ if } c < \left[2 - (2/\triangle_F)^{1/(k+2)}\right] \frac{c_D^{FO}}{\varphi^F}$$

(C) Extensive margin

Depending on the of domestic cutoff levels (c_D^{HO},c_D^{FO}) in the socially optimum, the market outcome does not always yield the same level of the total number of varieties (N^H, N^F) and the number of entrants (N_E^H, N_E^F) as those in the socially optimum.

Proof. See Appendix A.3.10. □

This proposition says that the market outcome differs from that of the socially optimum in several dimensions. For the Home country, the market selection is weaker than the planner's selection in both the domestic producers' and exporters' market, but is stronger than the planner's selection in the FDI margin, as illustrated in **Figure A.9**. Intuitively, compared to the planner's choice, Home country produces too many varieties in the domestic market,

[5]Please refer to the Appendix for a detailed expression of $\triangle_F$.

imports too many varieties from abroad, and does not have enough varieties from Foreign FDI firms. On the intensive margin, the more productive domestic producers (i.e. firms with low marginal costs) produce less than the socially optimal level. The same results hold true among the exporters and FDI firms, indicating the distribution of products is skewed too much toward high cost varieties. On the extensive margin, the number of entrants and the equilibrium number of varieties are also different from those of the socially optimum level. The market produces too many products from firms with large marginal costs, but not enough from firms with low marginal costs.

Since the trade policy study in this chapter exclusively focuses on the monopolistically competitive sector, the discussion of inter-sector inefficiency is omitted here[6]. The inefficiencies in the market outcome originate from multiple externalities in this economy. Two inefficiencies occur with quadratic quasi-linear preference: (i) The consumers display the 'love of variety' feature, which, however, firms do not consider when making entry decisions. As a result, there are not enough varieties in the economy. (ii) Firms can charge variable markups, and thus firm heterogeneity becomes another source of inefficiency in this economy. Since markup decreases in marginal cost, the low marginal cost firms (more productive) are inefficiently small and high marginal cost firms (less productive) are inefficiently large in the market outcome. The monopoly power in the differentiated-good sector allows a firm to price over its marginal cost. Under the free-entry condition, this externality tends to create too many varieties. The new entrant will take up the market share of existing firms, and this business-stealing effect also tends to create too many varieties. All these externalities work together to generate market failures in the current economy.

According to Dhingra and Morrow [2019], when monopolistic competition is combined with CES preference, the market outcome coincides with the socially optimum outcome. The externalities mentioned above exactly cancel each other out[7]. The current economy, however,

[6]Interesting reader could refer to Nocco et al. [2014] for further discussion.

[7]Note, firm heterogeneity does not create externality since all firms charge identical markup under CES preference.

deviates from this benchmark due to the quadratic quasi-linear preference. The forces that generate externalities do not cancel each other out, and firm heterogeneity becomes a new source of inefficiency in the economy. The market outcome hence differs from the first-best outcome systematically. As discussed in Chapter 2, tariff can affect not only the cutoff levels, entrants, and number of varieties in the equilibrium, but also social welfare. I will come back to this issue again in Section 3.4.

3.3 Free Trade and Its Welfare Implication

In this section, I study the classical question in the trade policy literature: is free trade socially optimal in the current economy? Here the free trade is referring to zero net tariff value. To answer this question, I set $t^H = t^F = 1$ and study the joint welfare in H and F:

$$\mathbb{W} \equiv \mathbb{U}^H|_{t^H=1} + \mathbb{U}^F|_{t^F=1}$$

The results can be summarized in the following proposition:

Proposition 10. *Free trade is, in general, not socially optimal. If H and F start with free trade ($t^H = t^F = 1$), then a small symmetric increment in import tariff raises social welfare if and only if $\tilde{c}_D > \alpha/2$, lowers social welfare if and only if $\tilde{c}_D < \alpha/2$ and has no effect on social welfare if and only if $\tilde{c}_D = \alpha/2$, where $\tilde{c}_D$ is the domestic cutoff under symmetry when net tariff is zero.*

Proof. See Appendix A.3.11. □

Interestingly, in this economy, free trade is not always socially optimal. On the one hand, if $\tilde{c}_D$, the domestic marginal cost cutoff level under symmetry with free trade, is sufficiently high, then a small increase in import tariff will raise the social welfare. This result means that if domestic cutoff is sufficiently high, the market selection too weak, then there are not enough firms competing in the economy. Therefore, the social planner should increase tariff to encourage firm entry. On the other hand, if the domestic cutoff is too low, the market selection too strong, there will be too many firms competing in this economy. In this situation, the social planner should discourage entry by reducing tariff (i.e., subsidize trade). If the domestic marginal cost cutoff level, however, is exactly equal to the threshold value, then free trade is socially optimal. The parameter α here stands for the relative demand of consumers toward differentiated varieties. So whether free trade is socially optimal crucially depends on whether the market outcome could meet consumer's demand on the varieties.

To gain more intuition of this proposition, and to elaborate on the externalities mentioned in Proposition 9, I rewrite the social planner problem as the following:

$$\mathbb{W} = \max_{\{N_E^H, N_E^F\}} I^H + \frac{\alpha - c_D^H}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_D^H\right) + I^F + \frac{\alpha - c_D^F}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_D^F\right)$$

Following Mankiw and Whinston [1986], here I consider a constrained-optimum problem faced by a social planner, who cannot affect the market outcome for any given number of firms. This is particularly relevant under the current heterogeneous firm framework since the first-best outcome cannot be reached due to the presence of externalities in this economy, as discussed in Section 3.2. On top of that, one should also keep in mind that the presence of a numéraire good adds an extra distortion to the model. There is no markup in the numéraire-good sector, but in the differentiated-good sector, producers charge prices above their marginal costs due to their monopoly power. As pointed out by Bhagwati [1969], the presence of distortions can result in the breakdown of Pareto-optimality of laissez-faire.

The planner chooses the optimal level of entry to maximize social welfare. Under free entry condition, firms make entry decisions irrespective of the externalities that they generate on consumers and other firms, so the market entry level might not be socially desirable. Notice, since wage in the economy equals to one, and tariff revenue equals to zero at free trade, maximizing the income is equivalent to maximizing aggregate profits. After imposing symmetry and $t = 1$, the above social welfare function $\mathbb{W}$ can be rewritten as:

$$\max_{\{N_E\}} \mathbb{W} \equiv \Pi + \underbrace{\frac{\alpha - c_D}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_D\right)}_{\text{Consumer Surplus}}$$

where $\Pi = N_E \times (\bar{\pi} - f_E)$. Consumers take the number of entrants as given and maximize their utilities. The above consumer surplus can be rewritten in terms of the optimized choice of variety i as follow:

$$CS = \frac{1}{2}\gamma \int_{i \in \Omega^H} (\hat{q}_i)^2 \, di + \frac{1}{2}\eta \left(\int_{i \in \Omega^H} \hat{q}_i di\right)^2$$

According to Ottaviano et al. [2002], the first term corresponds to the sum of consumer surplus at each variety i, and the second term reflects the variety effect that brings to consumer surplus. To understand the role of the entry in this economy and its impact on the welfare, following Bagwell and Lee [2015], one can similarly rewrite the above equation in the following way:

$$CS = \underbrace{N_E \times \frac{\gamma}{2}\int_0^{\tilde{c}_D} (q_D(c))^2 \, dG(c)}_{N_E \times \text{Average CS for a variety}} + \underbrace{\frac{(\alpha - \tilde{c}_D)}{2\eta}\left[\alpha - \frac{(k+1)\left(1+\tau^{-k}\right)+1}{(k+2)\left(1+\tau^{-k}\right)}\tilde{c}_D\right]}_{\text{Variety Effect (VE)}}$$

where equation (2.31) is utilized to express CS in terms of c_D^H to obtain the exact expression of variety effect. The $\tilde{c}_D$ is the domestic cutoff level under symmetry. Therefore the social planner's problem can be further rewritten as

$$\max_{\{N_E\}} \mathbb{W} \equiv N_E \times \text{Avg.CS} + \text{VE} + \Pi$$

The first order condition related to entry will generate the following condition:

$$\underbrace{\text{Avg. CS} + N_E \frac{\partial \text{Avg. CS}}{\partial N_E} + \frac{\partial \text{VE}}{\partial N_E} + N_E \frac{\partial \bar{\pi}}{\partial N_E}}_{\neq 0 \text{ due quadratic quasi-linear preference}} + \underbrace{\bar{\pi} - f_E}_{\text{Free entry}} = 0 \tag{3.1}$$

The free entry will only take care of the last item, and that is why it is not guaranteed to deliver the socially desirable level of entry. According to the seminal work[8] by Spence [1976] and Mankiw and Whinston [1986], the first term is positive, representing the average consumer surplus gain from a new variety following entry. The second term is negative, representing the average consumer surplus loss for existing varieties when a new variety becomes available (substitution effect). The third item is positive, representing the variety effect/benefit from a new variety. Lastly, the fourth item is negative, which represents the business-stealing effect since it measures how the new entrant affects the average profit of existing firms. These four items added up together give the externality of firms' entry. In the Appendix A.3.12, I show that this externality effect is positive when $\tilde{c}_D > \alpha/2$, which means firm entry increases social welfare. In this case, positive import tariff will increase social welfare by encouraging entry. When $\tilde{c}_D < \alpha/2$, however, the sum of these four terms is negative, indicating that firm entry decreases social welfare. In this case, positive import tariff decreases social welfare by introducing more entry. The optimal thing to do in this case is to subsidize trade and discourage firm entry. Only when $\tilde{c}_D = \alpha/2$, the market entry level coincides with the socially optimum entry level.

The result here is different from that in Cole and Davies [2011], where the authors find that the socially optimal tariff in their setting is always a subsidy. The intuition is that opening up to trade will expose domestic firms to the Foreign competition, driving out the least productive firms and reallocating resources to the more productive firms. When trade barrier is a choice variable, the social planner will have an additional incentive to promote trade since trade-liberalization can boost aggregate productivity. In the current setup, their

[8] For recent related discussions under heterogeneous firms framework, see Dhingra and Morrow [2019], Weinberger [2015], Bagwell and Lee [2015] and Behrens et al. [2018].

conclusion only holds when $\tilde{c}_D < \alpha/2$. The fundamental reason for this difference is the deviation from CES preference. Under CES preference, the first four terms in equation (3.1) always add up to zero. As demonstrated in Dhingra and Morrow [2019], free-entry delivers the first-best outcome. In the current framework, the presence of quadratic quasi-linear preference creates variable elasticity across the varieties, generating multiple externalities in the economy, causing the sum of those four terms to deviate from zero.

More importantly, firm heterogeneity, which does not produce any externality under the CES preference, now becomes a source of inefficiency in this economy. To be more specific, firm heterogeneity is governed by two parameters here: c_M and k. c_M represents the upper bound of the marginal cost distribution. Larger c_M indicates larger region that marginal cost can be drawn from, leading to an increase in firm heterogeneity. k governs the shape of Pareto distribution. When k equals to 1, the marginal cost follows uniform distribution, and different marginal costs can be drawn with equal probability. As k approaches infinity, the marginal cost distribution becomes degenerate at c_M. Therefore, an increase in k means the distribution of firms is skewed toward less productive firms, reducing the degree of firm heterogeneity.

For example, suppose $\tilde{c}_D > \alpha/2$, which means the sum of the first four terms in equation (3.1) is positive, then an additional entry creates positive externality to the society. In this situation, an increase in c_M or a decrease in k (both represent an increase in firm heterogeneity) will reduce the aggregate externality of firm entry. Therefore, an increase in firm heterogeneity is socially inefficient because it holds back the positive externality of firm entry. However, these two dimensions work quite differently. An increase in c_M will reduce the average consumer surplus, the absolute value of substitution effect, the variety effect, and the absolute value of business-stealing effect. The positive terms (average CS and VE) dominate the negative terms (substitution and business-stealing effect), and therefore externality decreases as c_M increases. A decrease in k has the exact opposite impacts on these four terms. The negative terms dominate the positive terms, and therefore externality decreases as k decreases. The fundamental reason behind this observation can be seen from the solutions of cutoffs and the expressions of these four terms[9]. c_M only affects the share of valid varieties on the market $(\tilde{c}_D/c_M)$, but k affects both the share of valid varieties and the average profits. As we will see in the next section, this impact on cutoffs generates important welfare implications of tariff.

3.4 Socially Optimum Tariff and Nash Tariff

In this section, I first follow the discussion in Section 3.2 to study the welfare implication of tariff, with an emphasis on the comparison between market outcome and socially optimum outcome. Then I derive the socially optimal tariff and the Nash tariff in this economy, and investigate the forces that result in the differences between them. Toward the end of this section, I compare the Nash tariff with FDI with the Nash tariff without FDI.

3.4.1 Welfare Implication of Tariff

Imposing symmetry and utilizing the setup in Appendix A.3.10 and Section 2.2, one can write the social welfare in market outcome and socially optimum outcome as functions of the corresponding cutoffs:

$$\mathbb{W}^M = 1 + \bar{q}_0 + \frac{\alpha - c_D^M}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_D^M\right)$$
$$\mathbb{W}^O = 1 + \bar{q}_0 + \frac{1}{2\eta}\left(\alpha - c_D^O\right)^2$$

Therefore, the welfare implication of a tariff change can be summarized in the following proposition:

Proposition 11. *(i) Imposing symmetry, if the domestic cutoff level is sufficiently high, i.e.,* $c_D^M > \alpha(A+B)/2B$[10]*, then a bilateral increase of tariff is welfare-improving; if the domestic*

[9] Please refer to Appendix A.3.12.

[10] Please refer Appendix A.3.13 for the exact expression of A and B.

cutoff level is sufficiently low, i.e., $c_D^M < \alpha(A+B)/2B$, then a bilateral increase of tariff is welfare-deteriorating. (ii) Removing symmetry, then a unilateral increase of Home's import tariff is welfare-improving to the Home country, but welfare-deteriorating to the Foreign country.

Proof. See Appendix A.3.13. □

The first part of this proposition is directly built on Proposition 9 and 10. On the one hand, when the domestic cutoff is sufficiently high, the market selection is too weak. In this case, the market outcome does not have enough varieties. According to Proposition 7, an increase in import tariff can increase the level of entry, foster market selection, reduce the equilibrium domestic cutoff level, and therefore improve the social welfare. On the other hand, if the domestic cutoff is too low, which means the market selection is too strong, then an increase in import tariff will only make the market selection stronger, deteriorating the social welfare. From another perspective, the reduction in domestic cutoff reduces the prices charged by all the firms, but more so for the more productive firms that charge higher markups. This effect reduces the distortions created by all the externalities mentioned in Proposition 9, thereby improving social welfare.

The second part of this proposition is more straightforward. Based on Proposition 4 and 7, an increase in Home's import tariff increases firm entry, resulting in more equilibrium varieties and lower domestic cutoff. These impacts raise the welfare in the Home country. For the Foreign country, Home's import tariff decreases firm entry, reducing the equilibrium number of varieties and producing a higher domestic cutoff. These impacts deteriorate Foreign country's welfare. With these understandings about the welfare implications of tariff in mind, now I turn to the study of optimal tariffs in this economy.

3.4.2 Socially Optimal Tariff vs. Nash Tariff

In this subsection, I derive both the socially optimum tariff and Nash tariff. For the socially optimum tariff, due to the symmetric nature of Home and Foreign, I assume the social planner puts identical weight on the welfare of each country. Thus, the socially optimal tariff maximizes the sum of the two countries' consumer utilities:

$$\max_{\{t^H, t^F\}} \mathbb{W} \equiv \mathbb{U}^H + \mathbb{U}^F = I^H + CS^H + I^F + CS^F$$

Recall that the income $I^l \equiv w^l + (t^l - 1) \times IM^l + \Pi^l$ for $l \in \{H, F\}$, and the equilibrium wages in both countries are equal to one. It is straightforward to verify that the optimal level of t^H satisfies the following condition:

$$\frac{\partial \mathbb{W}}{\partial t^H} = \underbrace{IM^H + (t^H - 1) \times \frac{\partial IM^H}{\partial t^H}}_{\text{Effect on } H\text{'s tariff revenue}} + \underbrace{\left(t^F - 1\right) \times \frac{\partial IM^F}{\partial t^H}}_{\text{Effect on } F\text{'s tariff revenue}} + \underbrace{\frac{\partial CS^H}{\partial t^H} + \frac{\partial CS^F}{\partial t^H}}_{\text{Effect on } CS^H \text{ and } CS^F} = 0 \qquad (3.2)$$

From the socially optimum perspective, the social planner needs to consider Home import tariff's impact on the tariff revenue and consumer surplus in both countries. In contrast, the Nash tariff level for H only focuses on the tariff revenue and consumer surplus of its own country. It is defined as follow:

$$\max_{\{t^H\}} \mathbb{U}^H = I^H + CS^H$$

It is easy to show that the optimal non-cooperative tariff level should satisfy:

$$\frac{\partial \mathbb{U}^H}{\partial t^H} = \underbrace{IM^H + (t^H - 1) \times \frac{\partial IM^H}{\partial t^H}}_{\text{Effect on } H\text{'s tariff revenue}} + \underbrace{\frac{\partial CS^H}{\partial t^H}}_{\text{Effect on } CS^H} = 0 \qquad (3.3)$$

Combining equation (3.2) and (3.3), and if we focus on one of the cases in Proposition 10, $\tilde{c}_D > \alpha/2$, which implies the socially optimal import tariff is greater than one, it then can be shown that the Nash tariff level for country H satisfies the following proposition:

Proposition 12. *When the symmetric domestic cost cutoff is sufficiently high ($\tilde{c}_D > \alpha/2$), the Nash tariff (t_N) is higher than the socially optimal tariff (t_S).*

Proof. See Appendix A.3.14. □

This finding confirms the Proposition 2 in Cole and Davies [2011], but it is established in the Melitz and Ottaviano [2008] framework with the presence of FDI. We can obtain the intuitions from several different angles. To understand the incentive of Home's import tariff, let us investigate F's free-entry condition, which can be expressed in a similar fashion as in equation (2.22):

$$\underbrace{\left(c_D^F\right)^{k+2}}_{\uparrow \text{in } t^H} + \underbrace{\Phi_1^H}_{\downarrow \text{in } t^H} \underbrace{\left(c_D^H\right)^{k+2}}_{\downarrow \text{in } t^H} + \underbrace{\Phi_2^H}_{\uparrow \text{in } t^H} \underbrace{\left(c_D^H\right)^{k+2}}_{\downarrow \text{in } t^H} = \gamma\phi$$

where the first term on the left represents the expected profit of being a domestic producer in F, the second term is the expected profit of being an exporter in F, and the third term is the expected profit of being multinational firm in F. When H country sets its tariff, according to Proposition 4–6, F's exporter cutoff level decreases in t^H. Based on the proof in Appendix A.3.4, Φ_1^H decreases in t^H, and Φ_2^H increases in t^H. So if t^H increases, the expected profit of an exporter in F goes down. In fact, the sum of the expected profit of exporter and multinational in F also goes down. When H sets its unilateral optimal tariff, it ignores the impact of its tariff on F's exporter and multinational. Therefore, the tariff level that solves (3.3) will be negative when evaluated at (3.2), implying H will set a higher tariff than the social planner would choose.

In Cole and Davies [2011], the terms of trade effect is not present for two reasons: (i) Pre-tariff import prices do not change due to the fixed markup over a constant wage. (ii) The quasi-linear utility pushes domestic and overseas income changes onto the numéraire good, leaving the profits from Home exporters or multinationals unaffected by Home's import tariff. In the current setting, neither of these two reasons is valid: (i) Pre-tariff import prices do change due to the variable markups responding to tariff change. (ii) Profits from the Foreign exporters or multinationals do depend on the relevant cutoffs, which are all affected by Home's import tariff level. To understand the terms of trade effect in the current setup, one can easily obtain the following conditions between average prices and their corresponding cutoffs:

$$\bar{p}^H = \frac{2k+1}{2k+2}c_D^H, \quad \bar{p}^F = \frac{2k+1}{2k+2}c_D^F$$

Also notice the IM and CS in equation (3.2) and (3.3) are all functions of c_D^H and c_D^F, so we can rewrite the welfare function and the first-order conditions in terms of $\bar{p}^H$ and $\bar{p}^F$. Based on Proposition 4, it is easy to verify that an increase in Home's import tariff generates a terms of trade gain for itself at the cost of Foreign's terms of trade deterioration. It is evident that the incentive to manipulate the terms of trade also results in the inefficiency of Nash tariff.

3.4.3 Nash Tariff with and without FDI

In this subsection, I compare the symmetric Nash tariff when FDI is an option with the case when it is not. Due to the quadratic quasi-linear preference and the numéraire good, there is no closed-form analytical solution for the socially optimal tariff and the Nash tariff. Therefore, I numerically compute the tariff levels based on equations (3.2) and (3.3) in Mathematica. I follow Behrens et al. [2011] in choosing the parameter values, which are listed in **Table A.7**. To focus on the role of FDI, here I fix the parameters that affect the degree of firm heterogeneity (k and c_M)[11].

For illustration purpose, I plot the computed Nash tariff level as a function of φ and α[12]. When the degree of firm heterogeneity is fixed, the Nash tariff levels are plotted in **Figure A.10**. The yellow plane separates the space: the area above indicates no FDI activity, and the area below indicates where FDI occurs. Since α is chosen such that the optimal tariff when FDI occurs is greater than one, it then makes sense that the blue plane is in-between the yellow plane and the red plane.

[11] I will discuss the role of heterogeneity and its interaction with FDI in Section 3.5. In that case, k and c_M will be varied.

[12] To focus on one case, here α is chosen to be small enough so that the optimal tariff level will be greater than 1.

In **Figure A.11**, I plot a two-dimensional version of **Figure A.10**, and contrast it with the Figure 4 in Cole and Davies [2011]. First of all, given the current parameter choice, the Nash tariff without FDI is always higher than the one with FDI. This confirms the finding in Cole and Davies [2011]. On the one hand, the gain from implementing tariff is smaller due to the tariff-jumping multinationals. On the other hand, based on Corollary 1 in Chapter 2, the presence of FDI causes the domestic cutoff to have a bigger response to tariff change, which affects the consumer surplus component in equation (3.3). These two channels collectively result in the lower Nash tariff level when FDI is present.

Second, φ is in a similar position as the fixed cost parameter (λ) in Cole and Davies [2011], but they have notable differences. φ does not affect the Nash tariff without FDI because it does not have any impact on exporters when FDI is not an option. The fixed cost parameter affects the Nash tariff level, regardless of the presence of FDI. The reason is that both the fixed costs of export and FDI are affected by λ. Hence, the change of λ will have a direct impact on the tariff level. In the current framework, the change in φ will affect the tariff level only when FDI occurs.

Third, as φ increases, the Nash tariff with FDI increases, and gets closer to the Nash Tariff without FDI. One the one hand, if φ approaches infinity, then the FDI cutoff will be zero, indicating Foreign firms only access Home country through exports. Hence, the Nash tariff level returns to the Nash tariff without FDI case. On the other hand, when FDI is an option, Nash tariff level increases in φ. This is similar to the results in Cole and Davies [2011] regarding λ: a higher φ reduces the cutoff of multinational (c_{FDI}), causing the least productive multinationals to switch to export, increasing the tariff base, and hence increasing the incentive of imposing a higher tariff. If φ is sufficiently high, FDI will occur in the equilibrium, confirming the corner solution finding in Cole and Davies [2011]. As we will see in the next subsection, the interaction of φ and the degree of firm heterogeneity generates very interesting policy implications that would have been missing in the CES framework.

3.5 Role of Variable Markup

This subsection is dedicated to the discussion of variable markups. Under quadratic quasi-linear preference, firms with different marginal costs can charge different markups[13]. This feature not only enables the tariff to affect the entire distribution of markups, but also generates misallocation in the economy. As we will see toward the end of this subsection, the interaction between variable markups and FDI unveils important trade policy implications, which would be otherwise absent in the CES framework.

3.5.1 Average Markup

In the current setup, a movement in iceberg trade cost (τ) does not affect the average markup[14] due to the assumption of Pareto cost distribution. However, the *ad valorem* tariff does have the ability to affect the average markup. Note that all the operating firms (N_D^H) serve their domestic market, and on top of that, there are Foreign exporters (N_X^F) and multinationals (N_{FDI}^F). The average markup of all the firms in country H can be expressed as the following:

$$\bar{m}^H = \frac{1}{N_D^H + N_X^F + N_{FDI}^F} \left[N_D^H \int_0^{c_D^H} m_D^H(c) \frac{dG(c)}{G(c_D^H)} \right. \\ \left. + N_X^F \int_{c_{FDI}^F}^{c_X^F} m_X^F(c) \frac{dG(c)}{G(c_X^F)} + N_{FDI}^F \int_0^{c_{FDI}^F} m_{FDI}^F(c) \frac{dG(c)}{G(c_{FDI}^F)} \right] \tag{3.4}$$

To simplify the analysis, here I focus on the symmetric case. This is similar to the bilateral liberalization studied in Section 4.1[15] in Melitz and Ottaviano [2008]. After imposing sym-

[13] There are other ways to generate variable markups, as discussed in Chapter 1, so the implications of variable markups discussed here might not be universal. For example, if variable markups is generated through firms engaging in Bertrand competition, then the implication will be certainly different from what we obtain here.

[14] See Melitz and Ottaviano [2008] Section 3.2.

[15] Note, different from the long-run results established in Chapter 2, bilateral reduction in tariff delivers the same results as in the short-run case: liberalization increases competition and decreases the domestic cutoff level, making it harder for a firm to survive.

metry, the average markup can be rewritten as follow (for detailed derivation, see Appendix A.3.15):

$$\bar{m} = \underbrace{\frac{1}{1+(t\tau)^{-k}} \times \frac{2k-1}{2k-2}}_{\text{weighted expected markup in domestic}} + \underbrace{\underbrace{\frac{(t\tau)^{-k}-\xi^k}{1+(t\tau)^{-k}}}_{\text{share of Foreign exporters}} \times t \underbrace{\left\{\frac{1}{2}\left[1-(t\tau\xi)^k\right] + \frac{k}{2k-2}\left[1-(t\tau\xi)^{k-1}\right]\right\}}_{\text{expected markup of Foreign exporter}}}_{\text{weighted expected markup of Foreign exporters}}$$
$$+ \underbrace{\underbrace{\frac{\xi^k}{1+(t\tau)^{-k}}}_{\text{share of foreign FDI}} \times \underbrace{\left(\frac{k}{2k-2}\frac{1}{\varphi\xi} + \frac{1}{2}\right)}_{\text{expected markup of Foreign FDI}}}_{\text{weighted expected markup of Foreign FDI}}$$

Based on this expression, I obtain the following proposition regarding the impact of a tariff change on the average markup:

Proposition 13. *If the level of protection is high, the increase of tariff-jumping Foreign multinational firms, which creates downward pressure on average markup, can dominate the decrease of Foreign exporter firms, which creates upward pressure on average markup. The average markup in the economy decreases as protection level increases. Therefore, protectionist trade policy can end up reducing Home market's average markup.*

Proof. See Appendix A.3.15. □

As t increases, the weighted expected markup from domestic firms (the first term) increases. This relation is due to the fact that protection reduces the degree of competition and makes it easier for domestic firms to survive. As a result, the expected markup will increase. The weighted expected markup from Foreign exporters (the second term) will decrease as t increases. This effect is due to two channels: the decreasing share of Foreign exporters (extensive margin, based on Proposition 5 and 6), and the expected markup, which increases first and then decreases as t increases (intensive margin, based on Proposition 2). The weighted expected markup of Foreign FDI (the third term) will increase as t increases. This relation also comes from two channels, the increasing share of Foreign FDI (extensive margin, based on Proposition 6) and the increasing expected markup (intensive margin, based on Proposition 2). The first and third term will dominate the second term at the beginning, but as t increases, the second term will eventually dominate the other two terms, dragging down the average markup. For illustration purpose, the average markup without FDI and the average markup with FDI are plotted in **Figure A.12**.

Edmond et al. [2015] suggest that under certain conditions, a reduction in trade barriers (iceberg-type trade costs) can lead to lower domestic markups (as Home producers lose their market share). Combined with higher markups on imported goods (as Foreign producers gain market share), the overall markup dispersion increases and the misallocation in the economy becomes worse. In this case, the pro-competitive gains from trade would be negative. In the current framework, a similar result is found when FDI is an option: the average markup can go up when the tariff level reduces. As the tariff level drops, although the number of imported varieties increases (hence exerting a downward pressure on average markup), the exiting of multinationals (which reduces the competition in the domestic market and exerts an upward pressure on average markup) also contributes to the increase in average markup.

3.5.2 Misallocation

Misallocation is a byproduct of all the inefficiencies discussed in Section 3.2. To gain more insight on the role of variable markups, in this subsection, I follow the approaches developed in Arkolakis et al. [2018] and Hsieh and Klenow [2009] to explore the misallocation in the current economy[16].

According to Arkolakis et al. [2018], variable markups can create a new source of gain or loss from trade liberalization, depending on whether low-cost firms, which charge high markups and under-supply their varieties, end up growing in size. According to their Appendix A.4, the effect of trade liberalization on the welfare of country j depends on two things: (i) the sign of the covariance of the markup, charged by a firm in country j that produces the variety for market i, and (ii) a change in its labor share that is needed to produce

[16] These are the two most prominent papers on misallocation.

the variety for that market. These two forces can be expressed through the following term:

$$\text{cov}\left(m^{i}\left(\omega\right),\frac{dl^{i}\left(\omega\right)}{L^{j}}\right) \tag{3.5}$$

where $l^i(\omega)$ is the total employment associated with a production of variety ω in country j for sales in country i. In other words, if this covariance is positive, then trade liberalization has an additional positive effect on welfare in country j through a reduction in misallocation. In their setup, without considering the choice of FDI, equation (3.5) becomes:

$$\begin{aligned}\text{cov}\left(m^{i}\left(\omega\right),\frac{dl^{i}\left(\omega\right)}{L^{j}}\right) &= N_{D}^{H}\int_{0}^{c_{D}^{H}}\frac{p_{D}^{H}\left(c\right)}{c}\frac{d\left[cq_{D}^{H}\left(c\right)\right]}{L^{H}}\frac{dG\left(c\right)}{G\left(c_{D}^{H}\right)}\\ &+ N_{X}^{H}\int_{0}^{c_{X}^{H}}\frac{p_{X}^{H}\left(c\right)}{\tau^{F}c}\frac{d\left[c\tau^{F}q_{X}^{H}\left(c\right)\right]}{L^{H}}\frac{dG\left(c\right)}{G\left(c_{X}^{H}\right)}\end{aligned}$$

It is important to notice that this covariance is at the firm-level. Therefore, it relates to not only firm's domestic production decision but also its export decision. In their setting, this covariance is negative, so the presence of variable markups reduces the welfare gain from trade. This negative effect is present because a decrease in trade costs makes exporting firms relatively more productive, leading to changes in markups. When demand is log-concave, as in Krugman [1979], higher markups imply incomplete pass-through of changes in marginal costs to prices, lowering the welfare gains from trade.

In the current setup, the covariance term is:

$$\begin{aligned}\text{cov}\left(m^{i}\left(\omega\right),\frac{dl^{i}\left(\omega\right)}{L^{j}}\right) &= N_{D}^{H}\int_{0}^{c_{D}^{H}}\frac{p_{D}^{H}\left(c\right)}{c}\frac{d\left[cq_{D}^{H}\left(c\right)\right]}{L^{H}}\frac{dG\left(c\right)}{G\left(c_{D}^{H}\right)}\\ &+ N_{X}^{H}\int_{c_{FDI}^{H}}^{c_{X}^{H}}\frac{p_{X}^{H}\left(c\right)}{\tau^{F}c}\frac{d\left[c\tau^{F}q_{X}^{H}\left(c\right)\right]}{L^{H}}\frac{dG\left(c\right)}{G\left(c_{X}^{H}\right)}\\ &+ N_{FDI}^{H}\int_{0}^{c_{FDI}^{H}}\frac{p_{FDI}^{H}\left(c\right)}{\varphi^{F}c}\frac{d\left[c\varphi^{F}q_{FDI}^{H}\left(c\right)\right]}{L^{H}}\frac{dG\left(c\right)}{G\left(c_{FDI}^{H}\right)}\end{aligned}$$

There are several differences compared to Arkolakis et al. [2018]. First, trade liberalization takes the form of a tariff reduction in this economy. Second, the covariance has an additional item due to the choice of FDI, which means now the welfare implication of a change in tariff also depends on firm's FDI activity. Third, as discussed in their June 2012 working paper[17], with quadratic quasi-linear preference, the change in welfare depends on the substitutability between the homogeneous good and the differentiated goods. In the current framework, the substitutability is also affected by FDI. In Appendix A.3.16, I analytically derive the covariance term under symmetry:

$$\operatorname{cov}\left(m^{i}\left(\omega\right),\frac{dl^{i}\left(\omega\right)}{L^{j}}\right)=\frac{\left(\alpha-c_{D}^{M}\right)dc_{D}^{M}}{2\eta\left(1+t^{-k}\tau^{-k}\right)}\left\{2k+1+\left[\left(t\tau\right)^{-k}-\xi^{k}\right]\right.$$
$$\left.\times\left[2k+1-k\left(1-t\tau\xi\right)\left(t\tau\xi\right)^{k}-\left(t\tau\xi\right)^{k}\right]+\xi^{k}\left(k+k\varphi+1\right)\right\}$$

And I show that the covariance term is *positive*, indicating a reduction in misallocation through protection. As discussed earlier, when c_D^M is sufficiently high, an increase in tariff is welfare-improving. The presence of variable markup and FDI results in a positive covariance term between the firm-level markup and change in firm-level employment share. This effect means the welfare gain from protection is even larger due to the reduction in misallocation. Intuitively, an increase in tariff will decrease the relative demand for high-cost varieties, and labor will be reallocated toward the low-cost varieties, which include those produced by Foreign FDI firms. Therefore, misallocation is reduced since the market becomes more concentrated, generating a positive correlation between markups and the labor share, and hence increasing the gains from the change in tariff.

Alternatively, we can also follow the way that Hsieh and Klenow [2009] introduce misallocation. For illustration purpose, here I focus on the exporters. A Home exporter with

[17] http://www.econ.uzh.ch/dam/jcr:00000000-0db7-f8ad-0000-00005b2a6145/Arkolakis_Costas_The_Elusive_Pro_Competitive_Effects_of_Trade.pdf.

marginal cost c has the following corresponding $TFPR$:

$$TFPR_X^{HO}(c) \equiv \frac{p_X^{HO}(c)}{c} = \frac{\tau^F c}{c} = \tau^F$$
$$TFPR_X^{HM}(c) \equiv \frac{p_X^{HM}(c)}{c} = \frac{c_D^{FM}/c + \tau^F t^F}{2}$$

In the planner's economy, $TFPR_X^{HO}$ is the same for all the exporters, and there is no misallocation in this case. However, in the market outcome, an exporter with lower marginal cost will have a bigger $TFPR$, implying that the low cost firms are allocated with too little labor. This is consistent with the conclusion in Proposition 9. Hence, misallocation also exists in the market outcome according to the definition in Hsieh and Klenow [2009]. Based on Proposition 4, an increase in t^H will increase $TFPR_X^{HM}$, exacerbating the misallocation among Home's exporters.

Figure A.13 is an attempt to present all the misallocation concepts on the graph. Both panels are demonstrating the responses of cutoffs to an increase in Home's import tariff. The differences between socially optimum outcome and market outcome in Proposition 9 are revealed in the cutoff levels. The positive covariance term in equation (3.5) can be seen from Panel B: following the increase of t^H, c_{FDI}^{FM} and c_D^{HM} both move toward their socially optimum level, indicating that Home's labor are reallocated toward the more productive firms, and hence misallocation is reduced. Finally, the increase of $TFPR_X^{HM}$ can be seen from the widening gap between c_X^{HO} and c_X^{HM}.

3.5.3 Interaction of Variable Markup and FDI

In this subsection, I focus on the interaction of variable markups and FDI through the lens of firm heterogeneity. For the same reason in Section 3.4.3, the analysis in this subsection is based on numerical computation. All the relevant parameter values are chosen from **Table A.7**. As discussed in Section 3.2, the quadratic quasi-linear preference makes firm heterogeneity an additional source of inefficiency in this economy. To see how the interaction of variable markup and FDI affects the Nash tariff level, I focus on the two parameter values that govern the degree of firm heterogeneity: k and c_M. It is also helpful to keep in mind that a country cares for tariff revenue and its consumer surplus, as can be seen from the following objective function:

$$\max_{\{t^H\}} \mathbb{U}^H = I^H + CS^H = 1 + TR^H + CS^H$$

The Nash tariff level is implied by the following first order condition:

$$\frac{\partial TR^H}{\partial t^H} + \frac{\partial CS^H}{\partial t^H} = 0 \tag{3.6}$$

The Role of k

The impact of k can be seen from **Figure A.14** and **Figure A.15**. k governs the shape of Pareto distribution. An increase in k means the distribution of firms is skewed toward less productive firms, reducing the degree of firm heterogeneity.

With the current parameter values, the Nash tariff without FDI decreases in k, which means it increases in the degree of firm heterogeneity. As k increases, two things are happening at the same time: (i) The domestic cutoff level is increasing. (ii) The probability of getting the low marginal cost draws are shrinking. The first channel will affect all cutoff levels proportionally, without changing the relative distribution of firms. The second channel changes the relative distribution of firms, and this channel is particularly important when multinational firms are present.

When FDI is an option, the impact of k on Nash tariff level depends on the size of φ, the parameter that measures the efficiency loss of FDI. First, an increase in k increases domestic cutoff and lowers the consumer surplus, so the second term in equation (3.6) is positive. This implies that the first term in equation (3.6) must be negative. When φ is small, the Nash tariff increases in k, i.e., decreases in firm heterogeneity. Intuitively, when φ is small, many firms access the Foreign market through FDI, so the tariff base is relatively small. In this case, an increase in k lowers the average probability of getting a low marginal cost draw. To maintain the first order condition in equation (3.6), the country needs to charge a higher tariff. This effect can be seen from the bottom panel in **Figure A.15**: when φ is small, the blue plane (bigger k) is above the green plane (smaller k).

When φ is large, however, the Nash tariff decreases in k, i.e., increases in firm heterogeneity. This is because when φ is large, many firms will choose to access the Foreign market through export, so the tariff base is relatively big. In this case, although an increase in k lowers the average probability of getting a low marginal cost draw, the sizable tariff base is big enough to maintain the first order condition in equation (3.6). Therefore, the Nash tariff level is lower. As shown in the bottom panel in **Figure A.15**: when φ is big, the blue plane (bigger k) is below the green plane (smaller k).

The Role of c_M

The impact of k can be seen from **Figure A.16** and **Figure A.17**. c_M represents the lower bound of the marginal cost distribution. An increase in c_M expands the region that marginal cost can be drawn from, raising up the degree of firm heterogeneity.

With the current parameter values, when FDI is an option, the Nash tariff decreases in c_M, which means it decreases in the degree of firm heterogeneity. To understand the intuition, recall that an increase in c_M expands the lower bound of cost draws, which eventually results in higher cutoff levels[18]. The impact is uniform to all the firms, without changing the relative distributions of firms. This is why in **Figure A.17**, the green plane (smaller c_M) is always above the blue plane (bigger c_M), i.e., the relative position of Nash tariff levels under different c_M remains the same. There is no interaction between variable markup (induced by c_M) and FDI (measured by the size of φ).

In summary, both k and c_M affect the degree of firm heterogeneity in the economy. While both of them affect the equilibrium cutoff levels, k also alters the relative distribution of firms with different marginal costs. When FDI is an option, the freeness of doing FDI interacts

[18] This can be easily verified through the closed-form solution of c_D, c_X, and c_{FDI} in Chapter 2.

with k, generating a novel implication for trade policy. If the degree of firm heterogeneity is big (smaller k), reducing the FDI barrier (smaller φ) can effectively lower the Nash tariff level. If the degree of firm heterogeneity is small (bigger k), increasing the FDI barrier (bigger φ) can effectively lower the Nash tariff level.

3.6 Concluding Remarks

This chapter studies the welfare implication of tariff and optimal tariffs in the presence of variable markups and FDI. The conclusions can be broadly summarized as follows. First, I find that quadratic quasi-linear preference generates multiple externalities in this economy, causing market outcome to differ from the socially optimum outcome systematically. Permitting FDI lowers the domestic cutoff levels and reduces the misallocation in the economy. Second, I find that free trade is not always socially optimal. If the domestic cutoff is sufficiently high, an additional firm entry can improve social welfare. In this case, a positive import tariff is welfare-improving because it encourages firm entry. Third, I find that the interaction of variable markup and FDI generates novel trade policy insights that are absent if consumers are under CES preference.

Given these results, there are several interesting questions to ask. First of all, do the trade policy results still hold under an alternative demand or supply structure that generates variable markups? I suspect that the alternative demand structures will produce similar results, but different supply-side structures may generate different outcomes. Second, the presence of numéraire good is a blessing and a curse. It would be interesting to drop the numéraire good and endogenize the wage as in Arkolakis [2008]. It is possible to obtain a closed-form solution of the Nash tariff, as shown by Demidova [2017], and investigate the trade policy implication regarding the labor market in the presence of FDI. Lastly, the trade policy implications here primarily focus on the import tariff. It would be interesting and relevant to study other forms of trade policy, such as export subsidy or corporate taxes.

This chapter provides evidence that the interaction of variable markups and FDI generates interesting trade policy implications. The steady-state analysis, however, might produce very different tariff levels than the actual tariff levels observed in the data. In the light of Larch and Lechthaler [2013], long-run and short-run effects of tariffs may run in opposite directions, implying that an exclusive focus on the steady-state could lead to biased policy conclusions. Carefully disentangling the dynamic effects of tariffs is undoubtedly a fruitful area for future research in the era of globalization .

BIBLIOGRAPHY

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

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

Laura Alfaro and Maggie Chen. Transportation cost and the geography of foreign investment. *Handbook of International Trade and Transportation*, 2018.

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

Mary Amiti, Oleg Itskhoki, and Jozef Konings. International Shocks, Variable Markups, and Domestic Prices. *The Review of Economic Studies*, 02 2019. ISSN 0034-6527. doi: 10.1093/restud/rdz005. URL `https://doi.org/10.1093/restud/rdz005`.

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

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

Pol Antràs and Elhanan Helpman. Contractual frictions and global sourcing. *The organization of firms in a global economy*, 2009.

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

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

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

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

Costas Arkolakis, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. The elusive pro-competitive effects of trade. *The Review of Economic Studies*, 2018.

John Asker, Allan Collard-Wexler, and Jan De Loecker. Market power and production (mis) allocation. a study of the world oil market. *Mimeo, Princeton University*, 2017.

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

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

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

R Baldwin, R Forslid, Ph Martin, and G Ottaviano. F. robert-nicoud (2003). *Economic Geography and Public Policy*, 2003.

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

Richard E Baldwin and Rikard Forslid. Trade liberalization with heterogenous firms. Working Paper 12192, National Bureau of Economic Research, May 2006. URL `http://www.nber.org/papers/w12192`.

Edward J Balistreri, Russell H Hillberry, and Thomas F Rutherford. Structural estimation and solution of international trade models with heterogeneous firms. *Journal of international Economics*, 83(2):95–108, 2011.

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

Alessandro Barattieri, Matteo Cacciatore, and Fabio Ghironi. Protectionism and the business cycle. Working Paper 24353, National Bureau of Economic Research, February 2018. URL `http://www.nber.org/papers/w24353`.

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

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

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

Kristian Behrens, Giordano Mion, and Gianmarco IP Ottaviano. Economic integration and industry reallocations: some theory with numbers. *International Handbook on the Economics of Integration: Competition, Spatial Location of Economic Activity and Financial Issues*, 2:169, 2011.

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

Kristian Behrens, Giordano Mion, Yasusada Murata, and Jens Suedekum. Quantifying the gap between equilibrium and optimum under monopolistic competition. *CEPR DP 11642*, 2018.

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

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

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

Giuseppe Berlingieri, Frank Pisch, and Claudia Steinwender. Organizing global supply chains: Input cost shares and vertical integration. Working Paper 25286, National Bureau of Economic Research, November 2018. URL `http://www.nber.org/papers/w25286`.

Andrew Bernard, Jonathan Eaton, Bradford Jensen, and Samuel Kortum. Plants and productivity in international trade. *American Economic Review*, 93(4):1268–1290, 2003.

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

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

Andrew B. Bernard, J. Bradford Jensen, and Peter K. Schott. *Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods*, pages 513–552. University of Chicago Press, January 2009. URL `http://www.nber.org/chapters/c0500`.

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

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

Jagdish N Bhagwati. *The Generalized Theory of Distortions and Welfare*. Cambridge, MIT, 1969.

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

Florin Ovidiu Bilbiie, Fabio Pietro Ghironi, and Marc J Melitz. Monopoly power and endogenous product variety: Distortions and remedies. *American Economic Journal: Macroeconomics (Forthcoming)*, 2016.

Bruce A Blonigen. Tariff-jumping antidumping duties. *Journal of international Economics*, 57(1):31–49, 2002.

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

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

Chad P Bown. Us-china trade war: The guns of august. *Peterson Institute of International Economics, Trade and Investment Policy Watch*, 2019.

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

S Lael Brainard. An empirical assessment of the proximity-concentration tradeoff between multinational sales and trade. *American Economic Review*, 87(4):520–544, 1997.

Dario Caldara, Matteo Iacoviello, Patrick Molligo, Andrea Prestipino, and Andrea Raffo. The economic effects of trade policy uncertainty. *MIMEO*, 2019.

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

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

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

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

CEA. Benefits of competition and indicators of market power. *Council of Economic Advisers Issue Brief*, page April 2016, 2016.

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

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

Martin J Conyon, Sourafel Girma, Steve Thompson, and Peter W Wright. The productivity and wage effects of foreign acquisition in the united kingdom. *The Journal of Industrial Economics*, 50(1):85–102, 2002.

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

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

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

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

Jan De Loecker and Jan Eeckhout. The rise of market power and the macroeconomic implications. Technical report, National Bureau of Economic Research, 2017.

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

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

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

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

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

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

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

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

Mark E Doms and J Bradford Jensen. Comparing wages, skills, and productivity between domestically and foreign-owned manufacturing establishments in the united states. In *Geography and ownership as bases for economic accounting*, pages 235–258. University of Chicago Press, 1998.

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

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

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

Jonathan Eaton, Samuel Kortum, and Sebastian Sotelo. International trade: Linking micro and macro. In *Advances in Economics and Econometrics, Tenth World Congress*, volume 2, 2012.

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

Chris Edmond, Virgiliu Midrigan, and Daniel Yi Xu. How costly are markups? Working Paper 24800, National Bureau of Economic Research, July 2018. URL `http://www.nber.org/papers/w24800`.

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

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

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

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

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

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

Ana Fernandes, Peter Klenow, Martha Denisse Pierola, Andres Rodriguez-Clare, and Sergii Meleshchuk. The intensive margin in trade, 2017.

Christian Fons-Rosen, Sebnem Kalemli-Ozcan, Bent E Sørensen, Carolina Villegas-Sanchez, and Vadym Volosovych. Quantifying productivity gains from foreign investment. Technical report, National Bureau of Economic Research, 2013.

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

Stefania Garetto, Lindsay Oldenski, and Natalia Ramondo. Multinational expansion in time and space. Working Paper 25804, National Bureau of Economic Research, May 2019. URL `http://www.nber.org/papers/w25804`.

Cecile Gaubert and Oleg Itskhoki. Granular comparative advantage. Working Paper 24807, National Bureau of Economic Research, July 2018. URL `http://www.nber.org/papers/w24807`.

Fabio Ghironi and Marc Melitz. International trade and macroeconomic dynamics with heterogeneous firms. *Quarterly Journal of Economics*, 120:865–915, 2005.

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

Anna Gumpert, Andreas Moxnes, Natalia Ramondo, and Felix Tintelnot. The life-cycle dynamics of exporters and multinational firms. Working Paper 24013, National Bureau of Economic Research, November 2017. URL `http://www.nber.org/papers/w24013`.

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

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

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

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

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

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

Keith Head and Thierry Mayer. Brands in motion: How frictions shape multinational production. *American Economic Review*, 109(9):3073–3124, 2019.

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

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

Elhanan Helpman and Paul R Krugman. *Trade Policy and Market Structure*. MIT press, 1989.

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

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

Alexander Hijzen, Holger Görg, and Miriam Manchin. Cross-border mergers and acquisitions and the role of trade costs. *European Economic Review*, 52(5):849–866, 2008.

Cecília Hornok and Balázs Muraközy. Markups of exporters and importers: Evidence from hungary. *The Scandinavian Journal of Economics*, 2018.

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

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

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

Wolfgang Keller and Stephen R Yeaple. Global production and trade in the knowledge economy. Technical report, National Bureau of Economic Research, 2008.

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

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

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

Sergey Kokovin, Mathieu Parenti, Jacques-François Thisse, and Philip Ushchev. On the dilution of market power. *CEPR Discussion Paper No. DP12367*, 2017.

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

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

Mario Larch and Wolfgang Lechthaler. Whom to send to doha? the shortsighted ones! *Review of Economic Dynamics*, 16(4):634 – 649, 2013. ISSN 1094-2025. doi: https://doi.org/10.1016/j.red.2012.10.006. URL `http://www.sciencedirect.com/science/article/pii/S1094202512000646`.

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

N Gregory Mankiw and Michael D Whinston. Free entry and social inefficiency. *The RAND Journal of Economics*, pages 48–58, 1986.

Raymond J Mataloni. The productivity advantage and global scope of us multinational firms. *BEA Working Paper,WP2011-02*, 2011.

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

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

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

Marc J Melitz and Stephen J Redding. Heterogeneous firms and trade. In *Handbook of international economics*, volume 4, pages 1–54. Elsevier, 2014.

Soojae Moon. The losses from trade restrictions: Policy dynamics with firm selection and endogenous markup. *Review of International Economics*, 23(1):86–110, 2015.

Monika Mrázová and J Peter Neary. Selection Effects with Heterogeneous Firms. *Journal of the European Economic Association*, 17(4):1294–1334, 07 2018. ISSN 1542-4766. doi: 10.1093/jeea/jvy024. URL `https://doi.org/10.1093/jeea/jvy024`.

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

Antonella Nocco, Gianmarco IP Ottaviano, and Matteo Salto. Monopolistic competition and optimum product selection. *American Economic Review*, 104(5):304–09, 2014.

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

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

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

Ralph Ossa. Quantitative models of commercial policy. In *Handbook of commercial policy*, volume 1, pages 207–259. Elsevier, 2016.

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

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

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

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

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

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

Natalia Ramondo, Veronica Rappoport, and Kim J Ruhl. Intrafirm trade and vertical fragmentation in us multinational corporations. *Journal of International Economics*, 98:51–59, 2016.

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

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

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

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

Alan Spearot. Unpacking the long-run effects of tariff shocks: New structural implications from firm heterogeneity models. *American Economic Journal: Microeconomics*, 8(2):128–67, 2016.

Michael Spence. Product selection, fixed costs, and monopolistic competition. *The Review of economic studies*, 43(2):217–235, 1976.

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

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

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

UNCTAD. World investment report 2011. *United Nations Publication*, 2011.

UNCTAD. World investment report 2018. *United Nations Publication*, 2018.

Anthony J Venables. Trade and trade policy with imperfect competition: The case of identical products and free entry. *Journal of International Economics*, 19(1-2):1–19, 1985.

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

Ariel Weinberger. Markups and misallocation with trade and heterogeneous firms. *FRB Dallas Globalization and Monetary Policy Institute Working Paper No. 251*, 2015.

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

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

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

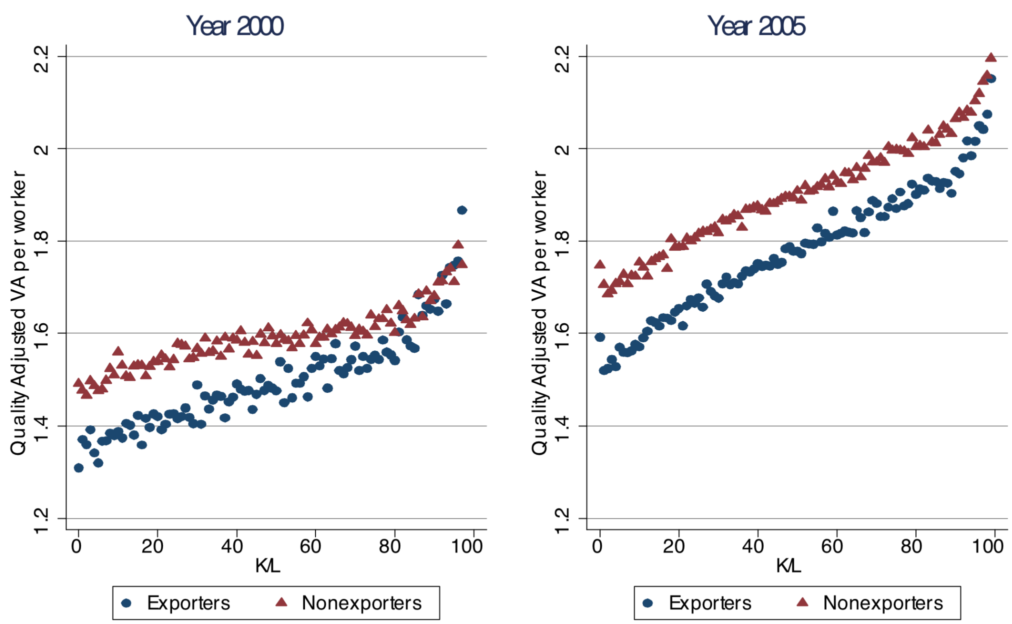
Andrei Zlate. Offshore production and business cycle dynamics with heterogeneous firms. *Journal of International Economics*, 100:34–49, 2016.

Appendix A

APPENDIX

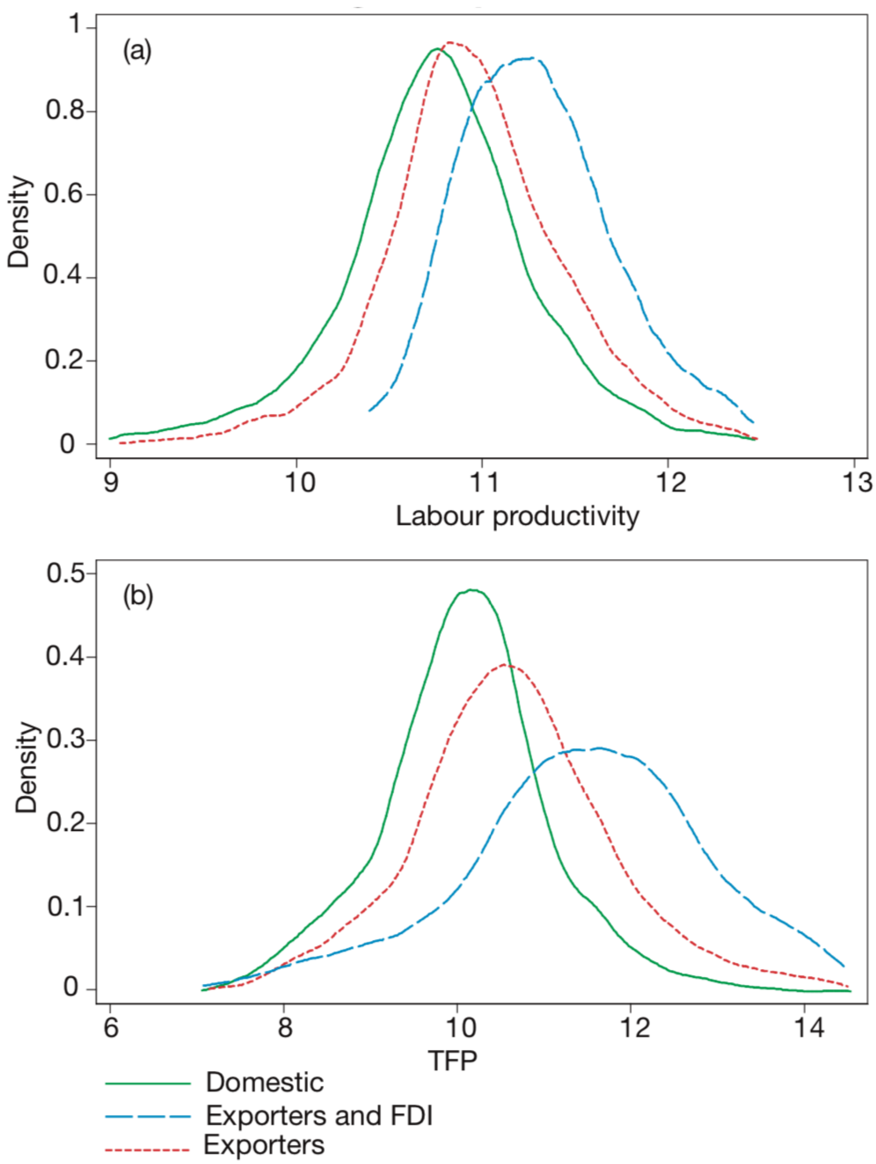
A.1 Figures

Figure A.1: Quality Adjusted Value-added per Worker



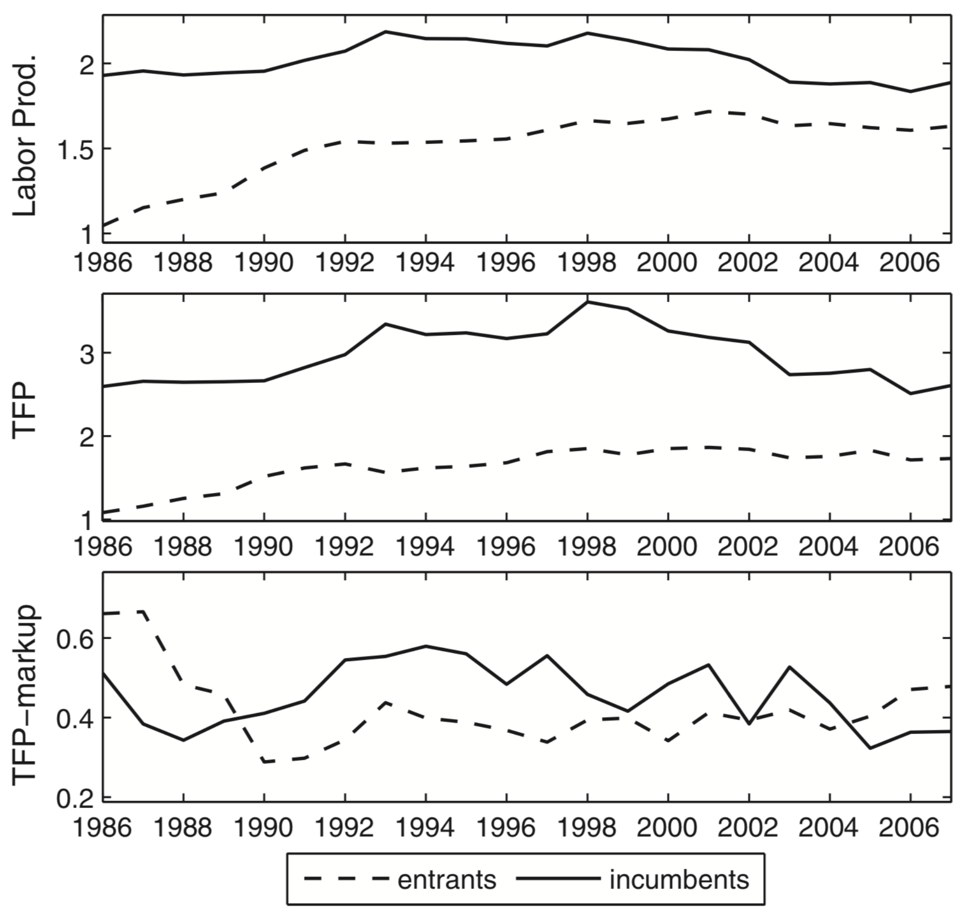
Note: Graph is taken from Lu [2010]. The horizontal axis is the rank of firms by their capital-labor ratio and they are grouped into 100 bins.

Figure A.2: Belgian-FDI Makers vs. Belgian Exporters



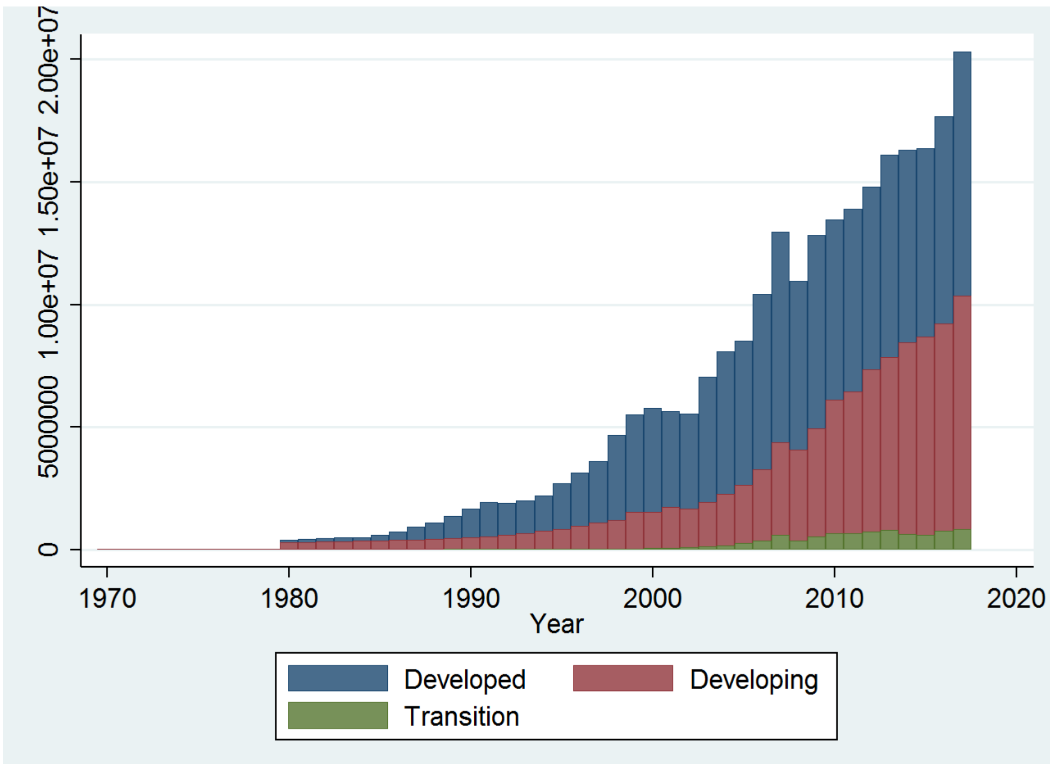
Note: Graph is taken from Mayer and Ottaviano [2008]. The data is from the 2004 Belgian manufacturing firms in the EFIM project. Notice, in this sample, nearly all the Belgian FDI firms are also exporters.

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



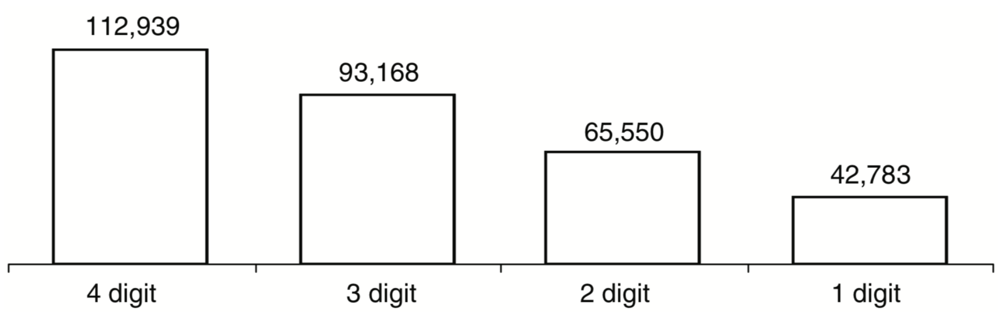
Note: Graph is taken from Kılınç [2014]. The top two panels display a considerable gap in labor productivity and standard TFP between entrants and incumbents. The TFP-markup estimated by the control function approach with entry variation, however, indicates that entrants are on average more productive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

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



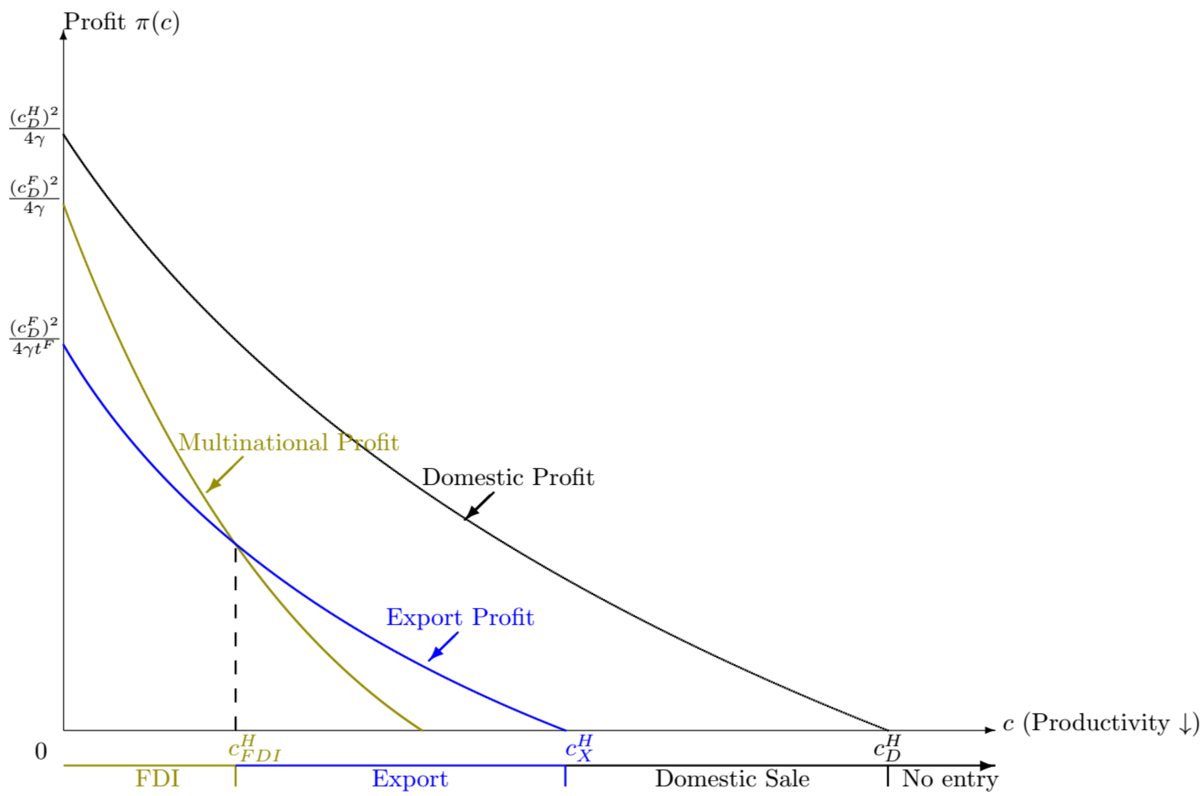
Note: Graph is constructed by author based on the data from UNCTAD from 1970–2017. According to World Investment Report (2017), foreign direct investment (FDI) is defined as an investment involving a long term relationship and reflecting a lasting interest and control by a resident entity in one economy (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 A.5: Vertical FDI Observed at Different Levels of Aggregation



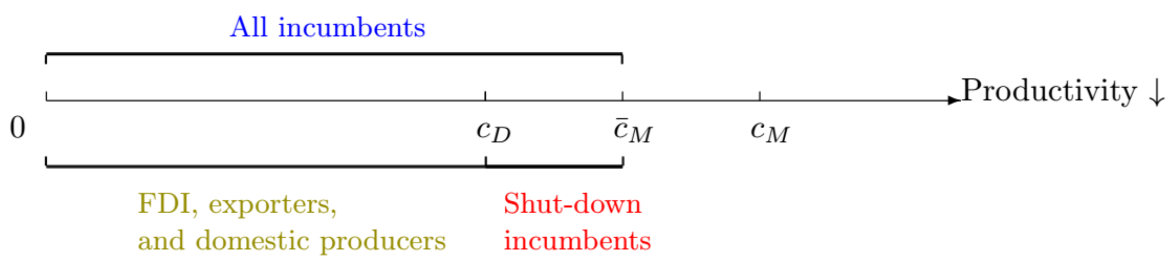
Note: Graph is taken from Alfaro and Charlton [2009], based on firm-level data from D&B. The numbers on top of each bar indicate the number of manufacturing subsidiaries. The horizontal axis indicates the one-, two-, three-, and four-digit Standard Industrial Classification (SIC) code. At a more disaggregated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

Figure A.6: Firms' Profit as a Function of Marginal Cost



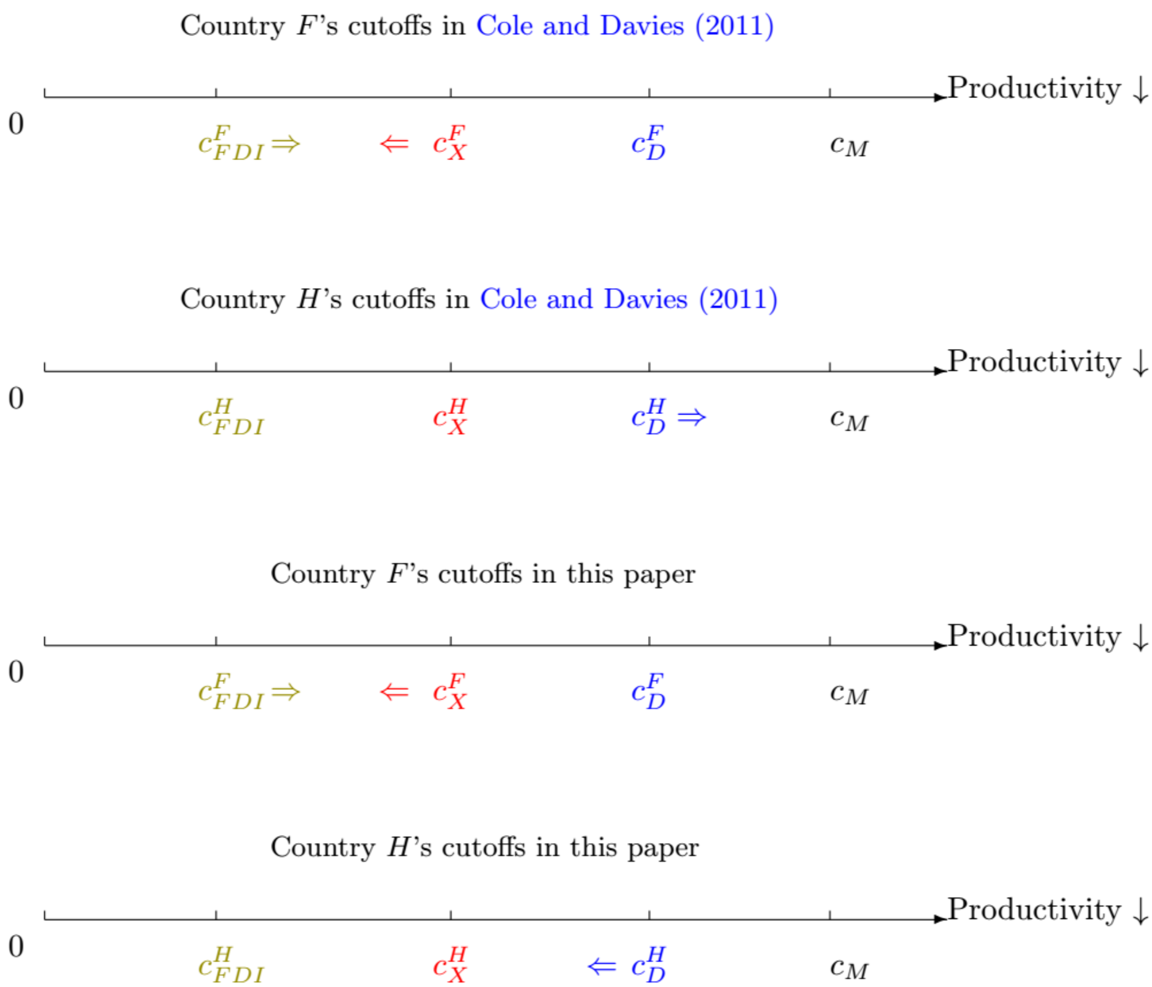
Note: The graph plots the profit function of firms in each market according to the firm's serving mode. The black line represents the firm's profit function in the domestic market (in this case, it is Home). The blue line plots the exporter's profit function in the Foreign market, and the green line plots the profit function of multinational firms.

Figure A.7: Productivity Distribution in the Short-run Equilibrium



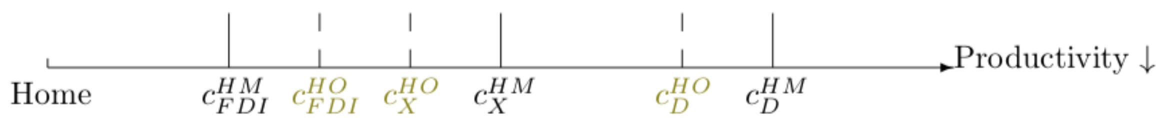
Note: c_M together with f_E stand for the state of the technology in the long-run equilibrium. $\bar{c}_M$ together with f_E stands for the state of the technology in the short-run equilibrium. c_D stands for the domestic cutoff level in the short-run equilibrium. Exporter and FDI cutoffs distributed on the left-hand side of c_D.

Figure A.8: Productivity Cutoffs Responses Comparison in the Long-run Equilibrium



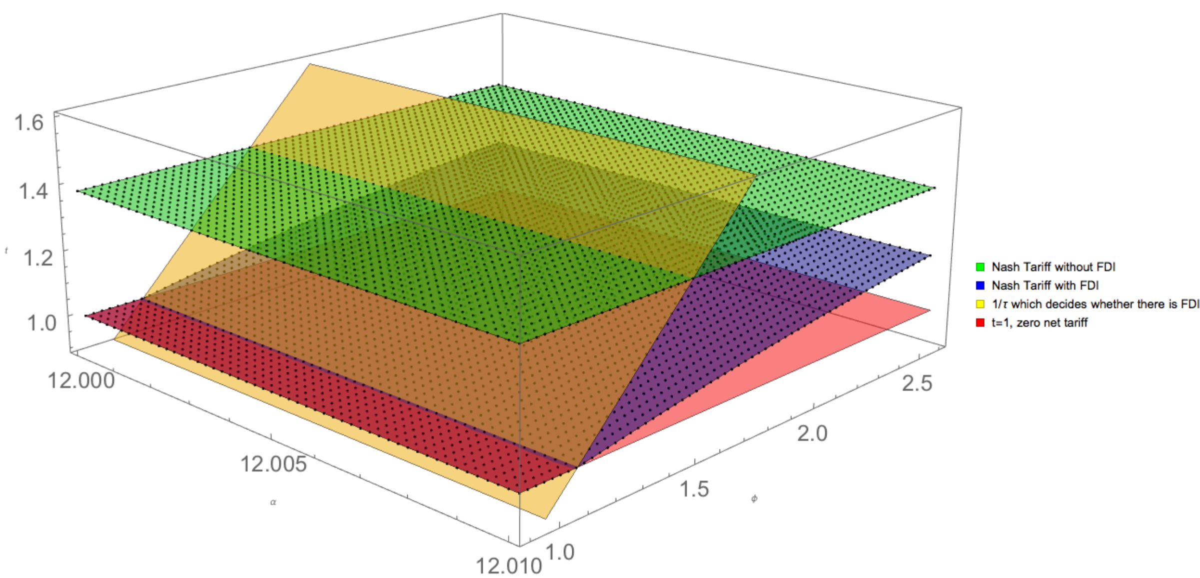
Note: Each axis represents the distribution of productivity within the monopolistically competitive industry. The first two panels are based on equation (11)-(13) from Cole and Davies [2011]. The last two panels are based on Proposition 4 through Proposition 6 in Chapter 2.

Figure A.9: Productivity Cutoffs Comparison between Market Outcome and Socially Optimum Outcome



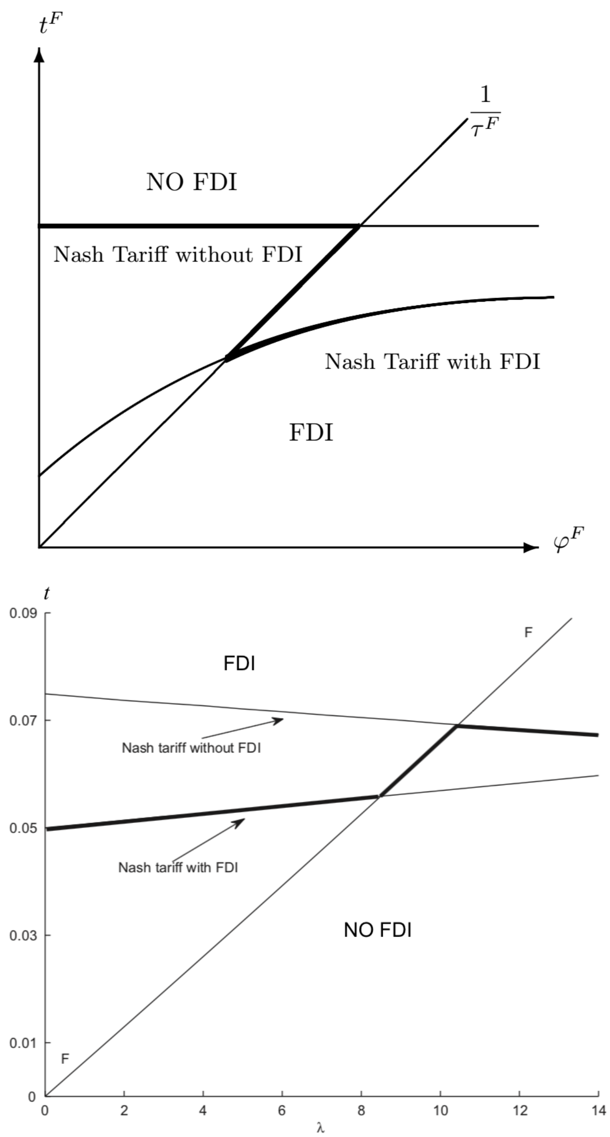
Note: This graph shows the marginal cost distributions of all the domestic firms in Home country. The letter M stands for the market outcome, which are marked with the black color. The letter O stands for the socially optimum outcome, which are marked with the olive color. Note, the indexes here are different from those in **Figure A.13**.

Figure A.10: Three-dimensional Nash Tariff with and without FDI



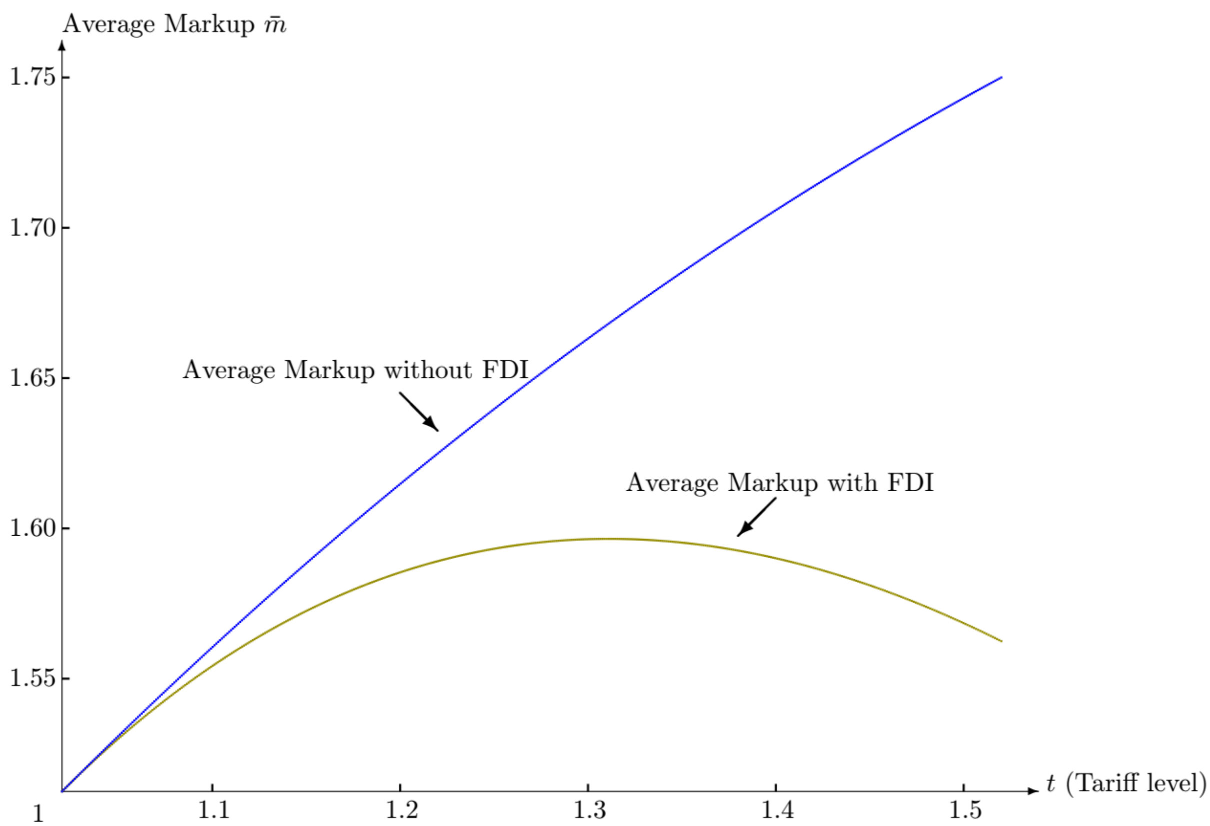
Note: The graph is computed based on the parameter values in **Table A.7**. The green plane plots the Nash tariff without FDI and the blue plane plots the Nash tariff with FDI. The yellow plane separates the space into FDI region and no FDI region. The red plane indicates zero net tariff.

Figure A.11: Two-dimensional Nash Tariff with and without FDI



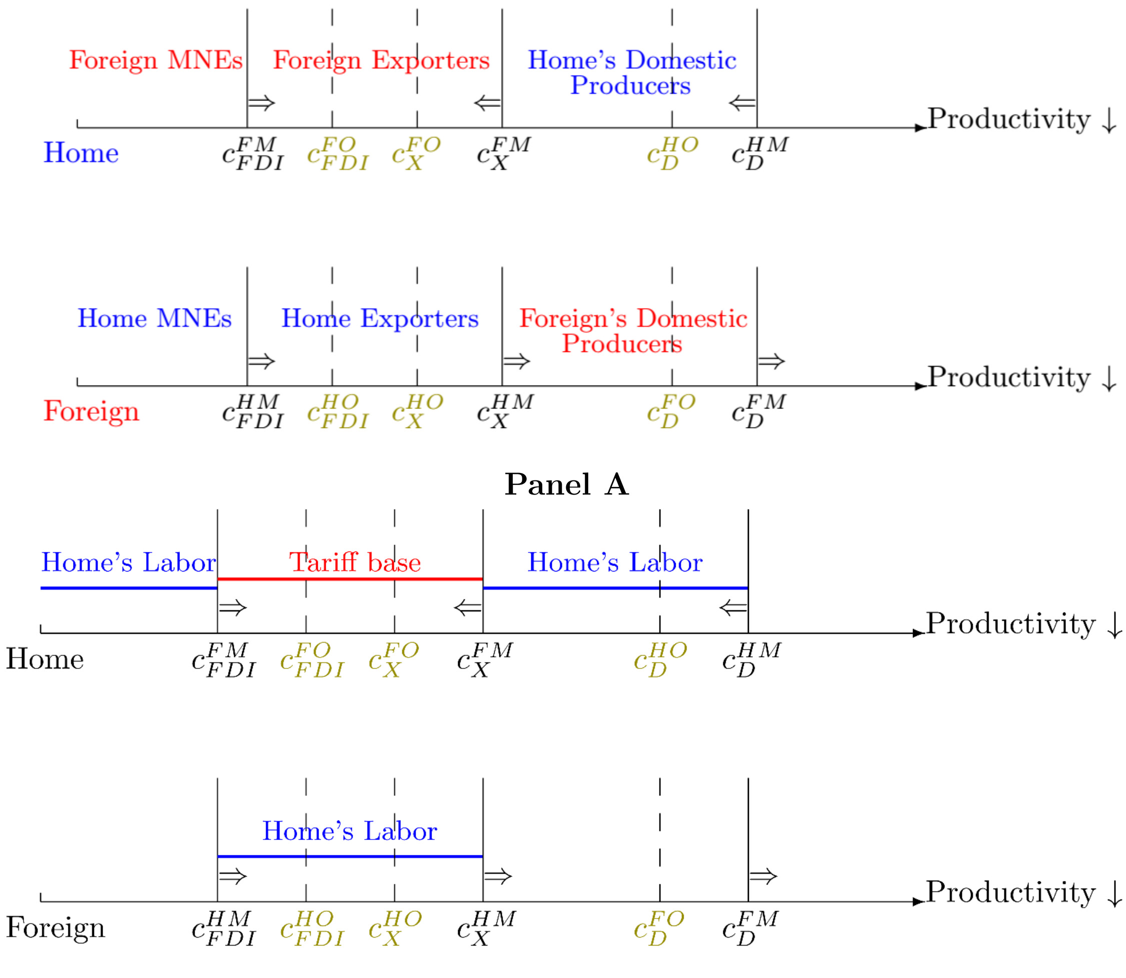
Note: The top panel is a two-dimensional graph of **Figure A.10**. The vertical axis indicates the tariff level, the horizontal axis indicates the easiness of doing FDI, and the $1/\tau^F$ separates the plane into No FDI region and FDI region. The bottom panel is taken from Cole and Davies [2011]. The vertical axis stands for net tariff, the horizontal axis represents the exogeneous component of the fixed cost for all modes of production, and the FF line represents the combination of τ and λ that induces FDI. The bold lines indicate the path of corner solution.

Figure A.12: Average Markup with and without FDI



Note: All the other parameter values are taken from **Table A.7**. The blue curve indicates the weighted average markup without FDI, whereas the green curve indicates the weighted average markup with FDI.

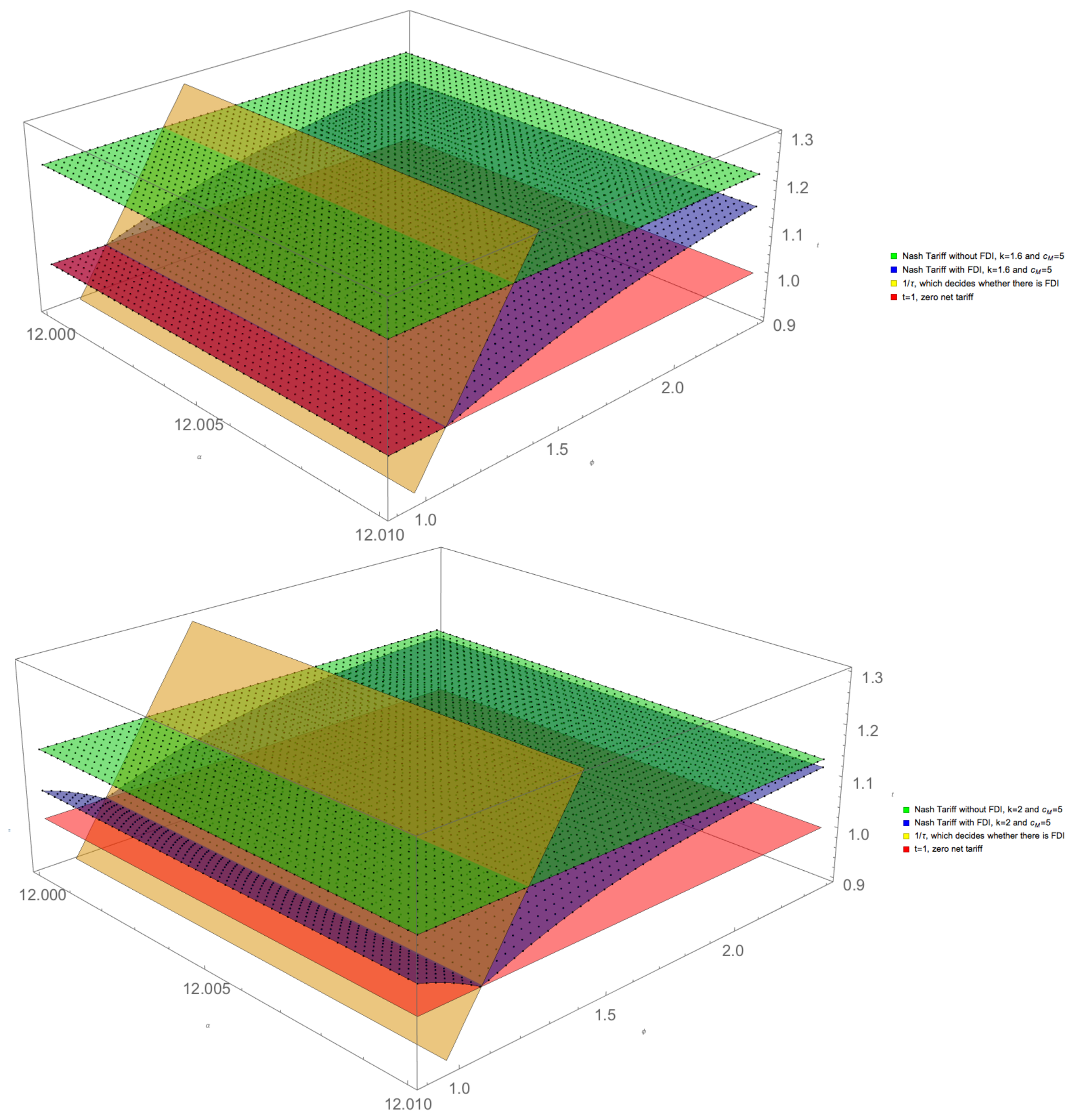
Figure A.13: Productivity Cutoffs Comparison between Market Outcome and Socially Optimum Outcome when Home's Tariff Increases



Panel B

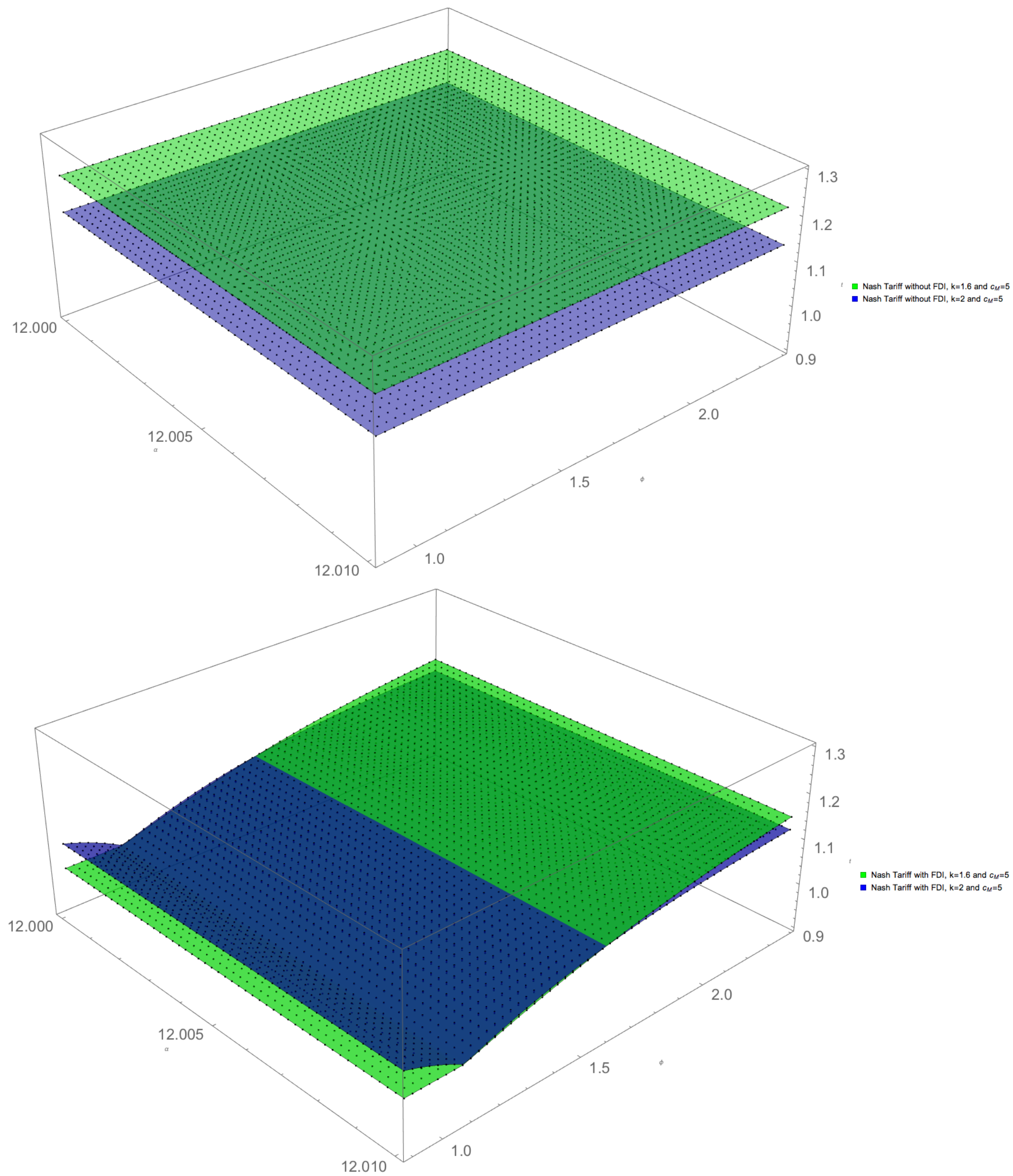
Note: Panel A and Panel B are the same plots with different notations. Each axis indicates the varieties that are available within that country. For example, the first axis indicates Home country has varieties from Foreign MNEs, Foreign exporters, and all the Home firms that serve their domestic market. Panel A focuses on the composition of firms, whereas Panel B focuses on the utilization of labor.

Figure A.14: The Impact of k on the Nash Tariff: Part I



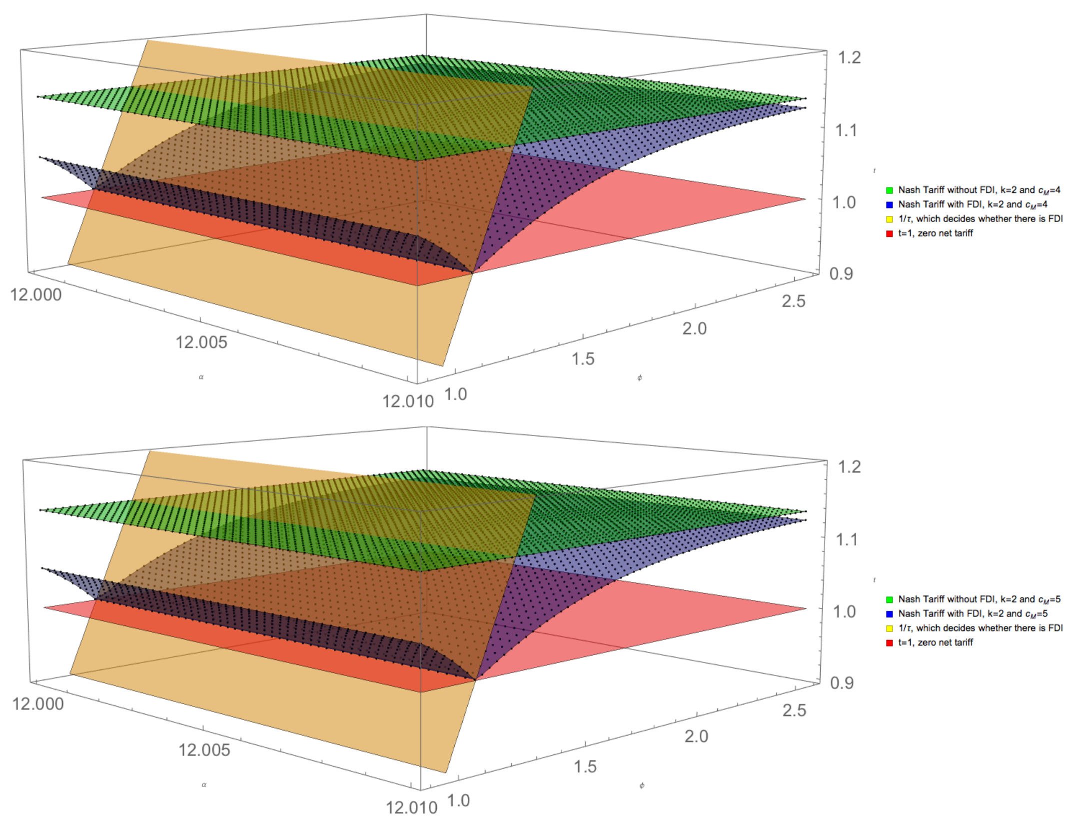
Note: The top panel plots the Nash tariff without FDI and the Nash tariff with FDI when $k = 1.6$. The bottom panel plots the Nash tariff without FDI and the Nash tariff with FDI when $k = 2$. The yellow plane separates the space into FDI region and no FDI region. The red plane indicates zero net tariff.

Figure A.15: The Impact of k on the Nash Tariff: Part II



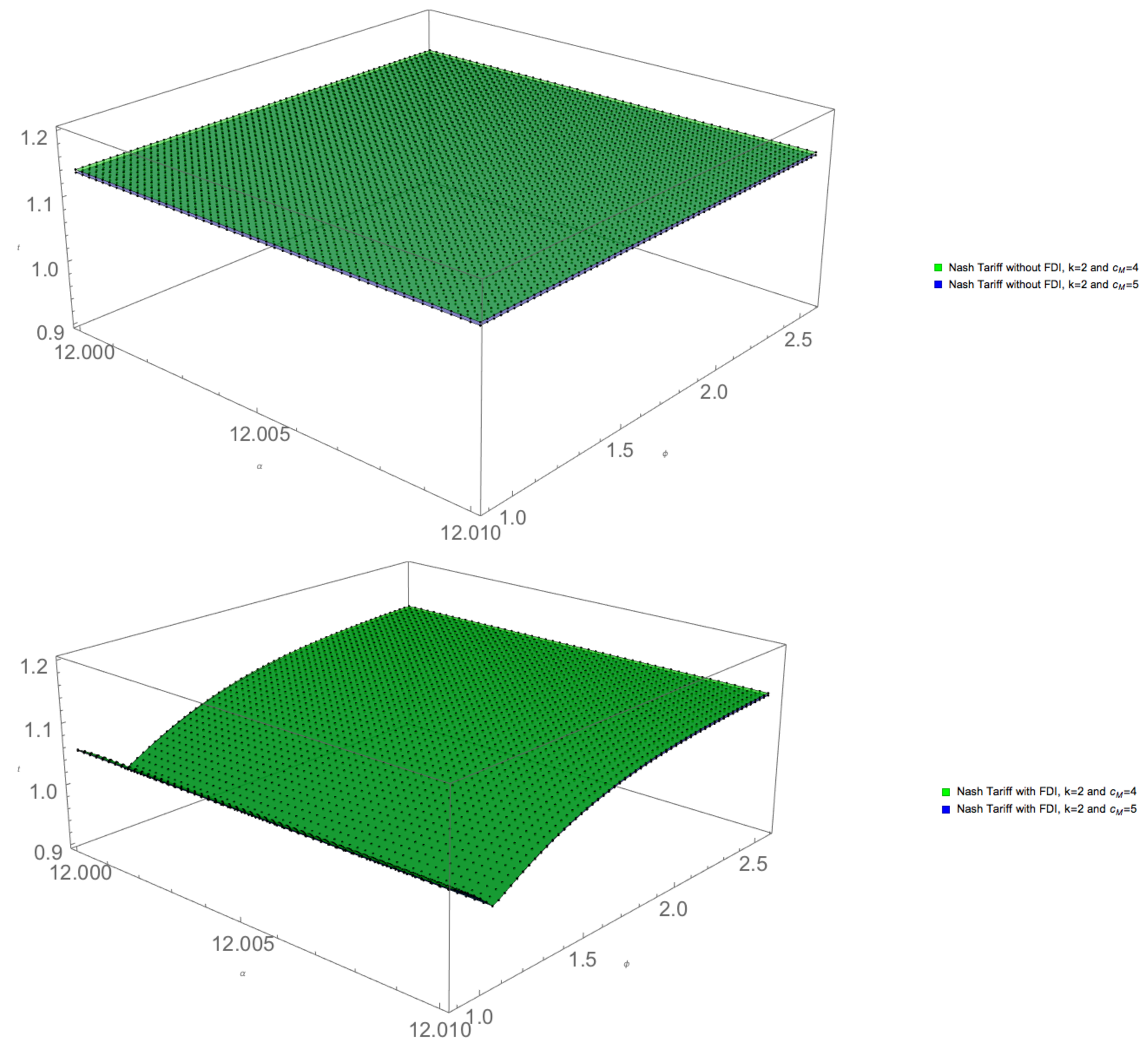
Note: These graphs are extracted from **Figure A.14**. The top panel plots the Nash tariff level without FDI. The green plane, which has a lower k value, stays entirely above the blue plane. The bottom panel plots the Nash tariff level with FDI. In the region where FDI occurs, the green plane, which has a lower k value, stays above the blue plane when φ is big, but stays below the blue plane when φ is small. This graph shows the interaction of FDI and k does affect the Nash tariff level.

Figure A.16: The Impact of c_M on the Nash Tariff: Part I



Note: The top panel plots the Nash tariff without FDI and the Nash tariff with FDI when $c_M = 4$. The bottom panel plots the Nash tariff without FDI and the Nash tariff with FDI when $c_M = 5$. The yellow plane separates the space into FDI region and no FDI region. The red plane indicates zero net tariff.

Figure A.17: The Impact of c_M on the Nash Tariff: Part II



Note: These graphs are extracted from **Figure A.16**. The top panel plots the Nash tariff level without FDI. The green plane, which has a lower c_M value, stays entirely above the blue plane. The bottom panel plots the Nash tariff level with FDI. In the region where FDI occurs, the green plane, which has a lower c_M value, also stays entirely above the blue plane. This graph shows the interaction of FDI and c_M does not change the relative position of Nash tariff.

A.2 Tables

Table A.1: Export by US Manufacturing Firms (2002)

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

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

Table A.2: Export by European Manufacturing Firms (2003)

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

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

Table A.3: Exporter Premia in US Manufacturing (2002)

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

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

Table A.4: Exporter Premia in European Manufacturing

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

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

Table A.5: Exporters vs. Non-Exporters

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

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

Table A.6: Exporter Markup Premium in French Manufacturing Industries

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

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

Table A.7: Paramaterization

α	12	Relative preferences toward the differentiated varieties
η	0.1	Substitutability among the varieties
c_M	5	Upper bound of marginal cost draw in Pareto distribution
γ	0.6	Degree of love for variety
φ	1.9	Iceberg-type efficiency loss of FDI
τ	1.1	Iceberg-type transportation cost
f_E	0.1	Fixed cost of entry

Note: These parameters are the baseline values in Behrens et al. [2011]. All of the computations performed in this dissertation are based on this table. For the comparative statics in Section 3.4 and 3.5, some parameters are varied around the benchmark value listed above.

A.3 Proofs

A.3.1 Proof of Proposition 1

The proof mainly comes from the following equation:

$$\frac{\alpha - c_D^H}{\left(c_D^H\right)^{k+1}} = \frac{\eta}{2\left(k+1\right)\gamma}\left\{\frac{\bar{N}_I^H}{\bar{c}_M^k} + \left[\left(\frac{1}{t^H\tau^H}\right)^k - \left(\xi^H\right)^k\right]\bar{N}_I^F + \left(\xi^H\right)^k\bar{N}_I^F\right\}$$

It is straightforward to show that $\partial\xi^H/\partial t^H > 0$, therefore the whole expression on the right-hand side will decrease as t^H increases. It follows from the equation that it must be true that c_D^H will increase, hence $\partial c_D^H/\partial t^H > 0$ in the short-run. □

A.3.2 Proof of Proposition 2

To prove this proposition, it is helpful to rewrite the markups as follows:

$$\begin{aligned} m_D^H\left(c\right) &= \frac{1}{2c}\left(c_D^H + c\right) \\ m_X^F\left(c\right) &= \frac{t^H}{2c}\left(c_X^F + c\right) = \frac{t^H}{2c}\left(c_D^H/t^H\tau^H + c\right) = \frac{1}{2c}\left(c_D^H/\tau^H + c/t^H\right) \\ m_{FDI}^F\left(c\right) &= \frac{1}{2\varphi^H c}\left(c_D^H + \varphi^H c\right) \end{aligned}$$

It follows from Proposition 1 that $\partial m_D^H/\partial t^H > 0$ and $\partial m_{FDI}^F/\partial t^H > 0$. The responses of m_X^F is less transparent. On the one hand, c_D^H/τ^H increases as t^H increases. On the other hand, c/t^H decreases as t^H increases. The total impact on m_X^F is therefore ambiguous. If c is small, then the first effect will dominate the second effect, $\partial m_X^F/\partial t^H > 0$. If c is big, then the second effect will dominate the first effect, $\partial m_X^F/\partial t^H < 0$. □

A.3.3 Proof of Proposition 3

To prove this proposition, I first prove the following condition:

$$\Phi_1^l + \Phi_2^l > \psi^l \in (0,1) \text{ for } l \in \{H, F\}$$

Given $\psi^l \equiv \left(\tau^l\right)^{-k} \left(t^l\right)^{-(k+1)}$ and

$$\begin{aligned}
\Phi_1^l \equiv & \frac{(k+1)(k+2)\, t^l (\tau^l)^2}{2} \left\{ \left(\frac{1}{t^l \tau^l}\right)^{k+2} - \left(\frac{1}{t^l \tau^l}\right)^2 \left(\xi^l\right)^k \right. \\
& \left. - \frac{2k}{k+1} \left[\left(\frac{1}{t^l \tau^l}\right)^{k+2} - \left(\frac{1}{t^l \tau^l}\right) \left(\xi^l\right)^{k+1} \right] + \frac{k}{k+2} \left[\left(\frac{1}{t^l \tau^l}\right)^{k+2} - \left(\xi^l\right)^{k+2} \right] \right\} \\
\Phi_2^l \equiv & \frac{(k+1)(k+2)\left(\xi^l\right)^k}{2} \left[1 - \frac{2k\varphi^l \xi^l}{k+1} + \frac{k\left(\varphi^l \xi^l\right)^2}{k+2} \right]
\end{aligned}$$

It is then straightforward to show

$$\begin{aligned}
\Phi_1^l + \Phi_2^l = \psi^l + & \frac{(k+1)(k+2)\left(\xi^l\right)^k}{2} \left\{ \left(1 - \frac{1}{t^l}\right) \right. \\
& \left. - \frac{2k}{k+1} \xi^l \left(\varphi^l - \tau^l\right) + \frac{k}{k+2} \left(\xi^l\right)^2 \left(\left(\varphi^l\right)^2 - t^l \left(\tau^l\right)^2 \right) \right\}
\end{aligned}$$

To show that $\Phi_1^l + \Phi_2^l > \psi^l$, it is equivalent to show that

$$\left(1 - \frac{1}{t^l}\right) - \frac{2k}{k+1} \xi^l \left(\varphi^l - \tau^l\right) + \frac{k}{k+2} \left(\xi^l\right)^2 \left(\left(\varphi^l\right)^2 - t^l \left(\tau^l\right)^2 \right) > 0$$

Based on the definition of $\xi^l \equiv \left(\sqrt{t^l}-1\right)/\left(\sqrt{t^l}\varphi^l - t^l\tau^l\right)$, the above equation becomes:

$$\left(1-\frac{1}{t^l}\right) - \frac{2k}{k+1}\frac{\sqrt{t^l}-1}{\sqrt{t^l}\varphi^l - t^l\tau^l}\left(\varphi^l-\tau^l\right) + \frac{k}{k+2}\left(\frac{\sqrt{t^l}-1}{\sqrt{t^l}\varphi^l - t^l\tau^l}\right)^2\left(\left(\varphi^l\right)^2 - t^l\left(\tau^l\right)^2\right) > 0$$

$$\Leftrightarrow \left(1-\frac{1}{t^l}\right) + \frac{k}{k+2}\left(\frac{\sqrt{t^l}-1}{\sqrt{t^l}\varphi^l - t^l\tau^l}\right)^2\left(\left(\varphi^l\right)^2 - t^l\left(\tau^l\right)^2\right) > \frac{2k}{k+1}\frac{\sqrt{t^l}-1}{\sqrt{t^l}\varphi^l - t^l\tau^l}\left(\varphi^l-\tau^l\right)$$

$$\Leftrightarrow \frac{t^l-1}{t^l} + \frac{k}{k+2}\frac{\left(\sqrt{t^l}-1\right)^2}{t^l}\frac{\varphi^l+\sqrt{t^l}\tau^l}{\varphi^l-\sqrt{t^l}\tau^l} > \frac{2k}{k+1}\frac{\sqrt{t^l}-1}{\sqrt{t^l}}\frac{\varphi^l-\tau^l}{\varphi^l-\sqrt{t^l}\tau^l}$$

$$\Leftrightarrow \frac{t^l-1}{t^l} + \frac{k}{k+2}\frac{\left(\sqrt{t^l}-1\right)^2}{t^l}\frac{\varphi^l+\sqrt{t^l}\tau^l}{\varphi^l-\sqrt{t^l}\tau^l} > \frac{2k}{k+1}\frac{\sqrt{t^l}-1}{\sqrt{t^l}}\frac{\varphi^l-\tau^l}{\varphi^l-\sqrt{t^l}\tau^l}$$

Multiply both sides by $(k+1)(k+2)t^l\left(\varphi^l - \sqrt{t^l}\tau^l\right)$, I have

$$\left(k^2+3k+2\right)\left(\sqrt{t^l}+1\right)\left(\varphi^l-\sqrt{t^l}\tau^l\right) +$$
$$\left(k^2+k\right)\left(\sqrt{t^l}-1\right)\left(\varphi^l+\sqrt{t^l}\tau^l\right) > \left(2k^2+4k\right)\sqrt{t^l}\left(\varphi^l-\tau^l\right)$$
$$\Leftrightarrow 2\sqrt{t^l}\varphi^l - 2\sqrt{t^l}\tau^l + 2(k+1)\left(\varphi^l - t^l\tau^l\right) > 0$$
$$\Leftrightarrow 2\sqrt{t^l}\left(\varphi^l-\tau^l\right) + 2(k+1)\left(\varphi^l - t^l\tau^l\right) > 0$$

This is obviously true when $\varphi^l > t^l\tau^l$ (note $t^l > 1$), which is the assumption we made to guarantee the existence of tariff-jumping FDI.

Compare the cutoff expressions, for $l \in \{H, F\}$

$$\text{Open economy, with tariff, export and FDI: } c_{D1}^H = \left[\gamma\phi\frac{1-\left(\Phi_1^F+\Phi_2^F\right)}{1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)}\right]^{\frac{1}{k+2}}$$

$$\text{Open economy, with tariff and export: } c_{D2}^H = \left(\gamma\phi\frac{1-\psi^F}{1-\psi^F\psi^H}\right)^{\frac{1}{k+2}}, \psi^l = \left(\tau^l\right)^{-k}\left(t^l\right)^{-(k+1)}$$

$$\text{Closed economy: } c_{D3}^H = (\gamma\phi)^{\frac{1}{k+2}}, \text{ as in MO(2008) Section 2}$$

With the proved condition, it is straightforward to show that

$$c^H_{D3} > c^H_{D2} > c^H_{D1}$$

□

A.3.4 Proof of Proposition 4

Based on the solution of $c_D^H = \left[\gamma\phi \frac{1-\left(\Phi_1^F+\Phi_2^F\right)}{1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)}\right]^{\frac{1}{k+2}}$, I have

$$\begin{aligned}
\frac{\partial c_D^H}{\partial t^H} &= \frac{\gamma\phi}{k+2}\left[\gamma\phi\frac{1-\left(\Phi_1^F+\Phi_2^F\right)}{1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)}\right]^{-\frac{k+1}{k+2}} \frac{\left(\Phi_1^F+\Phi_2^F\right)\left[1-\left(\Phi_1^F+\Phi_2^F\right)\right]}{\left[1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)\right]^2}\frac{\partial\left(\Phi_1^H+\Phi_2^H\right)}{\partial t^H} \\
&= \underbrace{\frac{1}{k+2}c_D^H\frac{\left(\Phi_1^F+\Phi_2^F\right)}{1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)}}_{>0}\frac{\partial\left(\Phi_1^H+\Phi_2^H\right)}{\partial t^H}
\end{aligned}$$

So the sign crucially depends on $\frac{\partial\left(\Phi_1^H+\Phi_2^H\right)}{\partial t^H}$. It is straightforward to show that

$$\begin{aligned}
\frac{\partial\left(\Phi_1^H+\Phi_2^H\right)}{\partial t^H} &= -\frac{k+1}{2t^H\left(\sqrt{t^H}\varphi^H-t^H\tau^H\right)^2}\left\{\frac{2}{\left(t^H\tau^H\right)^k}\left[t^H\left(\tau^H\right)^2-2\varphi^H\sqrt{t^H}\tau^H+\left(\varphi^H\right)^2\right]\right. \\
&+\left(\xi^H\right)^k\left[-2t^H\left(\tau^H\right)^2-2k\sqrt{t^H}t^H\left(\tau^H\right)^2+k\left(t^H\tau^H\right)^2\right. \\
&\left.\left.+2\left(k+2\right)\sqrt{t^H}\tau^H\varphi^H-\left(k+2\right)\left(\varphi^H\right)^2\right]\right\} \\
&= \underbrace{-\frac{k+1}{2t^H\left(\sqrt{t^H}\varphi^H-t^H\tau^H\right)^2}}_{<0}\left\{2\left(\varphi^H-\sqrt{t^H}\tau^H\right)^2\left[\left(t^H\tau^H\right)^{-k}-\left(\xi^H\right)^k\right]\right. \\
&\left.-\left(\xi^H\right)^k k\left(\varphi^H-t^H\tau^H\right)\left(\varphi^H+t^H\tau^H-2\sqrt{t^H}\tau^H\right)\right\}
\end{aligned}$$

The expression within the big bracket is greater than zero for all $k \in [1, +\infty)$ when

$\varphi^H > t^H\tau^H$, to see this, it is equivalent to show

$$2\left(\varphi^H - \sqrt{t^H}\tau^H\right)^2\left[\left(t^H\tau^H\right)^{-k} - \left(\xi^H\right)^k\right] > \left(\xi^H\right)^k k\left(\varphi^H - t^H\tau^H\right)\left(\varphi^H + t^H\tau^H - 2\sqrt{t^H}\tau^H\right)$$

For $k = 1$, the expression become

$$\begin{aligned}
&2\left(\varphi^H - \sqrt{t^H}\tau^H\right)^2\left[\frac{1}{(t^H\tau^H)} - \xi^H\right] > \xi^H\left(\varphi^H - t^H\tau^H\right)\left(\varphi^H + t^H\tau^H - 2\sqrt{t^H}\tau^H\right)\\
\Leftrightarrow &2\left(\varphi^H - \sqrt{t^H}\tau^H\right)^2\left(\frac{1}{\xi^H t^H\tau^H} - 1\right) > \left(\varphi^H - t^H\tau^H\right)\left(\varphi^H + t^H\tau^H - 2\sqrt{t^H}\tau^H\right)\\
\Leftrightarrow &2\left(\varphi^H - \sqrt{t^H}\tau^H\right)^2\frac{\varphi^H - t^H\tau^H}{\sqrt{t^H}\tau^H\left(\sqrt{t^H} - 1\right)} > \left(\varphi^H - t^H\tau^H\right)\left(\varphi^H + t^H\tau^H - 2\sqrt{t^H}\tau^H\right)\\
\Leftrightarrow &2\left(\varphi^H - \sqrt{t^H}\tau^H\right)^2 > \sqrt{t^H}\tau^H\left(\sqrt{t^H} - 1\right)\left(\varphi^H + t^H\tau^H - 2\sqrt{t^H}\tau^H\right)\\
\Leftrightarrow &2\left(\varphi^H\right)^2 - 3\sqrt{t^H}\tau^H\varphi^H + 3\sqrt{t^H}t^H\left(\tau^H\right)^2 - t^H\tau^H\varphi^H - \left(t^H\tau^H\right)^2 > 0\\
\Leftrightarrow &\left(\varphi^H - t^H\tau^H\right)\left(2\varphi^H + t^H\tau^H - 3\sqrt{t^H}\tau^H\right) > 0
\end{aligned}$$

which is obviously true.

For k approach infinity, to prove the equation, it is equivalent to show

$$\frac{2\left(\varphi^H - \sqrt{t^H}\tau^H\right)^2}{\left(\varphi^H - t^H\tau^H\right)\left(\varphi^H + t^H\tau^H - 2\sqrt{t^H}\tau^H\right)} > \frac{k}{\left(\xi^H t^H\tau^H\right)^{-k} - 1}$$

As $k \to \infty$, the limit of right-hand side is 0. It means as long as the left-hand side is positive, the equation is true for $k \to \infty$. The left hand side is obviously positive given $\varphi^H > t^H\tau^H$. Therefore,

$$\frac{\partial\left(\Phi_1^H + \Phi_2^H\right)}{\partial t^H} < 0 \Rightarrow \frac{\partial c_D^H}{\partial t^H} < 0$$

To show $\dfrac{\partial c_D^F}{\partial t^H}$ is easier, note that

$$\frac{\partial c_D^F}{\partial t^H} = \left[\gamma\phi\frac{1-\left(\Phi_1^H+\Phi_2^H\right)}{1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)}\right]^{-\frac{k+1}{k+2}} \frac{\gamma\phi\left[1-\left(\Phi_1^F+\Phi_2^F\right)\right]/(k+2)}{\left[1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)\right]^2}\frac{-\partial\left(\Phi_1^H+\Phi_2^H\right)}{\partial t^H}$$
$$= \underbrace{\frac{c_D^F}{k+2}\frac{\left[1-\left(\Phi_1^F+\Phi_2^F\right)\right]}{\left[1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)\right]\left[1-\left(\Phi_1^H+\Phi_2^H\right)\right]}}_{>0}\underbrace{\frac{-\partial\left(\Phi_1^H+\Phi_2^H\right)}{\partial t^H}}_{>0} > 0$$

□

A.3.5 Proof of Proposition 5

Based on the cutoff relations, it is straightforward to show

$$\frac{\partial c_X^H}{\partial t^H} = \frac{\partial\left(c_D^F/t^F\tau^F\right)}{\partial t^H} = \frac{1}{t^F\tau^F}\frac{\partial c_D^F}{\partial t^H} > 0$$
$$\frac{\partial c_X^F}{\partial t^H} = \frac{\partial\left(c_D^H/t^H\tau^H\right)}{\partial t^H} = \frac{1}{\left(t^H\right)^2\tau^H}\left(\frac{\partial c_D^H}{\partial t^H}t^H - c_D^H\right)$$

To show $\dfrac{\partial c_D^H}{\partial t^H}t^H - c_D^H < 0$ is equivalent to show

$$\frac{1}{k+2}c_D^H\frac{\left(\Phi_1^F+\Phi_2^F\right)}{1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)}\frac{\partial\left(\Phi_1^H+\Phi_2^H\right)}{\partial t^H}t^H < c_D^H$$
$$\Leftrightarrow \underbrace{\frac{1}{k+2}\frac{\left(\Phi_1^F+\Phi_2^F\right)}{1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)}}_{>0}\underbrace{\frac{\partial\left(\Phi_1^H+\Phi_2^H\right)}{\partial t^H}t^H}_{<0} < 1$$

Which is obviously true. Therefore, Proposition 5 is proved. □

A.3.6 Proof of Proposition 6

This part is relatively easy to prove, notice

$$\begin{aligned}\frac{\partial c_{FDI}^{F}}{\partial t^{H}} &= \frac{\partial\left(c_{D}^{H}\xi^{H}\right)}{\partial t^{H}} = \frac{\partial c_{D}^{H}}{\partial t^{H}}\xi^{H} + \frac{\partial \xi^{H}}{\partial t^{H}}c_{D}^{H} \\ &= \underbrace{\frac{c_{D}^{H}\xi^{H}}{k+2}\frac{\left(\Phi_1^F+\Phi_2^F\right)}{1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)}}_{>0}\underbrace{\frac{\partial\left(\Phi_1^H+\Phi_2^H\right)}{\partial t^{H}}}_{<0} + \underbrace{\frac{\partial \xi^{H}}{\partial t^{H}}c_{D}^{H}}_{>0}\end{aligned}$$

It is straightforward to verify that under the parameter choice in Section 2.3 in Chapter 2, the second term dominates the first term, so Foreign country's FDI cutoff level (c_{FDI}^{F})is strictly increasing as Home country's tariff (t^H) increases. Do notice, the other choice of ξ, which is $\xi = \left(\sqrt{t}+1\right)/\left(\sqrt{t}\varphi + t\tau\right)$ will make the second item negative, thereby making c_{FDI}^{F} decreasing in response to t^H's increase, hence no tariff-jumping FDI. □

A.3.7 Proof of Corollary 1

The proof is straightforward. Utilizing Proposition 3 and 4:

$$\begin{aligned}\left|\frac{\frac{\partial c_D^H}{\partial t^H}\big|_{\text{without FDI}}}{\frac{\partial c_D^H}{\partial t^H}\big|_{\text{with FDI}}}\right| &= \left|\frac{\frac{\gamma\phi(c_{D2}^H)^{-(k+1)}}{k+2}\frac{\psi^F\left[1-\psi^F\right]}{\left[1-\psi^F\psi^H\right]^2}\frac{\partial\psi^F}{\partial t^H}}{\frac{\gamma\phi(c_{D1}^H)^{-(k+1)}}{k+2}\frac{\left(\Phi_1^F+\Phi_2^F\right)\left[1-\left(\Phi_1^F+\Phi_2^F\right)\right]}{\left[1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)\right]^2}\frac{\partial\left(\Phi_1^H+\Phi_2^H\right)}{\partial t^H}}\right| \\ &= \left|\frac{(c_{D1}^H)^{k+1}\frac{\psi^F\left[1-\psi^F\right]}{\left[1-\psi^F\psi^H\right]^2}\frac{\partial\psi^F}{\partial t^H}}{(c_{D2}^H)^{k+1}\frac{\left(\Phi_1^F+\Phi_2^F\right)\left[1-\left(\Phi_1^F+\Phi_2^F\right)\right]}{\left[1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)\right]^2}\frac{\partial\left(\Phi_1^H+\Phi_2^H\right)}{\partial t^H}}\right| \\ &= \left|\left(\frac{c_{D1}^H}{c_{D2}^H}\right)^{k+1}\times\frac{\psi^F}{\left(\Phi_1^F+\Phi_2^F\right)}\times\frac{\left[1-\left(\Phi_1^F+\Phi_2^F\right)\left(\Phi_1^H+\Phi_2^H\right)\right]^2}{\left[1-\psi^F\psi^H\right]^2}\times\frac{\frac{\partial\psi^F}{\partial t^H}}{\frac{\partial\left(\Phi_1^H+\Phi_2^H\right)}{\partial t^H}}\right|\end{aligned}$$

Based on the proof of Proposition 3 and 4, it is obvious that the 1st, 2nd and 3rd term in the above equation are all smaller than 1. The 4th term, however, is greater than 1. It is straightforward to verify that the 4th term will be dominated by the first three terms, therefore, the whole expression is smaller than 1. □

A.3.8 Proof of Proposition 7

Based on equation (2.24), it's straightforward to show

$$\frac{\partial N^H}{\partial t^H} = \frac{2\gamma(k+1)}{\eta} \frac{-\frac{\partial c_D^H}{\partial t^H} c_D^H - \frac{\partial c_D^H}{\partial t^H}\left(\alpha - c_D^H\right)}{\left(c_D^H\right)^2} = -\frac{2\gamma\alpha(k+1)}{\eta\left(c_D^H\right)^2}\frac{\partial c_D^H}{\partial t^H} > 0$$

$$\frac{\partial N^F}{\partial t^H} = \frac{2\gamma(k+1)}{\eta} \frac{-\frac{\partial c_D^F}{\partial t^H} c_D^F - \frac{\partial c_D^F}{\partial t^H}\left(\alpha - c_D^F\right)}{\left(c_D^F\right)^2} = -\frac{2\gamma\alpha(k+1)}{\eta\left(c_D^F\right)^2}\frac{\partial c_D^F}{\partial t^H} < 0$$

Now based on equation (2.26), as t^H increases

$$N_E^F = \frac{2\left(c_M\right)^k(k+1)\gamma}{\eta\left(1-\delta^H\delta^F\right)}\left[\frac{\alpha - c_D^F}{\left(c_D^F\right)^{k+1}} - \delta^F\frac{\alpha - c_D^H}{\left(c_D^H\right)^{k+1}}\right]$$

δ^H decreases (hence the coefficient in front of the bracket decreases), c_D^F increases (the first item in the bracket decreases), c_D^H decreases (the second item in the bracket increases). Hence the whole expression on the right decreases, therefore $\partial N_E^F/\partial t^H < 0$. Now utilizing the free-entry condition, which is equation (2.25)

$$G\left(c_D^H\right)N_E^H + G\left(c_X^F\right)N_E^F = N^H$$

As t^H increases, N^H increases, it means the left-side also needs to increase. Notice N_E^F decreases, c_X^F decreases, c_D^H decreases, it then must be true that N_E^H increases. Hence, $\partial N_E^H/\partial t^H > 0$. □

A.3.9 Proof of Proposition 8

To prove this proposition, it is helpful to rewrite the markups as follows:

$$
\begin{aligned}
m_D^H(c) &= \frac{1}{2c}\left(c_D^H + c\right) \\
m_X^F(c) &= \frac{t^H}{2c}\left(c_X^F + c\right) = \frac{t^H}{2c}\left(c_D^H/t^H\tau^H + c\right) = \frac{1}{2c}\left(c_D^H/\tau^H + c/t^H\right) \\
m_{FDI}^F(c) &= \frac{1}{2\varphi^H c}\left(c_D^H + \varphi^H c\right)
\end{aligned}
$$

It follows from Proposition 4 that $\partial m_D^H/\partial t^H < 0$ and $\partial m_{FDI}^F/\partial t^H < 0$. The responses of m_X^F now is different from what we see in Proposition 2. On the one hand, c_D^H/τ^H decreases as t^H increases. On the other hand, c/t^H decreases as t^H increases. The total impact on m_X^F is therefore negative. Hence, $\partial m_X^F/\partial t^H < 0$. □

A.3.10 Proof of Proposition 9

The social planner solves the following problem

$$
\max_{\{N_E^H, q_0^H, q_i^H, N_E^F, q_0^F, q_i^F\}} \mathbb{W} \equiv \mathbb{U}_H + \mathbb{U}_F
$$

$$
\begin{aligned}
\text{s.t. } q_0^H + q_0^F + f\left(N_E^H + N_E^F\right) + N_E^H \int_0^{c_M} \left[c q_D^H(c) + \tau^F c q_X^H(c) + \varphi^F c q_{FDI}^H(c)\right] dG(c) \\
+ N_E^F \int_0^{c_M} \left[c q_D^F(c) + \tau^H c q_X^F(c) + \varphi^H c q_{FDI}^F(c)\right] dG(c) = 2 + \bar{q}_0^H + \bar{q}_0^F
\end{aligned}
$$

Notice, $\mathbb{W} \equiv \mathbb{U}^H + \mathbb{U}^F$ and since labor has been normalized to 1, $\mathbb{U}^H$ is defined as follow

$$
\begin{aligned}
\mathbb{U}^H \equiv\ & q_0 + \alpha N_E^H \left\{ \int \left[q_D^H(c) + q_X^H(c) + q_{FDI}^H(c)\right] dG(c) \right\} \\
& - \frac{\gamma}{2}\left\{ N_E^H \int \left(q_D^H(c)\right)^2 dG(c) + N_E^F \int \left[\left(q_X^F(c)\right)^2 + \left(q_{FDI}^F(c)\right)^2\right] dG(c) \right\} \\
& - \frac{\eta}{2}\left\{ N_E^H \int q_D^H(c)\, dG(c) + N_E^F \int \left[q_X^F(c) + q_{FDI}^F(c)\right] dG(c) \right\}
\end{aligned}
$$

The first order conditions with respect to q_D, q_X, q_{FDI} deliver the following results for the Home country:

$$q_D^H(c) = \frac{c_D^{HO} - c}{\gamma}, c_D^{HO} = \alpha - \eta Q^{HO}$$
$$q_X^H(c) = \frac{c_X^{HO} - c}{\gamma/\tau^F}, c_X^{HO} = \frac{\alpha - \eta Q^{FO}}{\tau^F}$$
$$q_{FDI}^H(c) = \frac{c_{FDI}^{HO} - c}{\gamma/\varphi^F}, c_{FDI}^{HO} = \frac{\alpha - \eta Q^{FO}}{\varphi^F}$$

The first order condition with respect to N_E delivers the following results

$$Q^{HO} = \frac{N^{HO} + 2N^{FO}}{\gamma + \eta\left(N^{HO} + 2N^{FO}\right)}\left(\alpha - \frac{k}{k+1}\left[\frac{N^{HO} + \left(\tau^H/\varphi^H\right)^{k+1} N^{FO}}{N^{HO} + 2N^{FO}}\right] c_D^{HO}\right)$$

Combine these with the corresponding results for the Foreign country, it's straightforward to obtain Home's domestic cutoff level under the planner's problem

$$c_D^{HO} = \left[\gamma\left(k+1\right)\left(k+2\right) f c_M^k \frac{1 - O_F}{1 - O_F O_H}\right]^{\frac{1}{k+2}}$$

where $O_F \equiv \left(\varphi^F\right)^{-k} + \frac{(k+1)(k+2)}{2}\left[\left(\tau^F\right)^{-k} - \left(\varphi^F\right)^{-k}\right] - k\left(k+2\right)\tau^F\left[\left(\tau^F\right)^{-(k+1)} - \left(\varphi^F\right)^{-(k+1)}\right] + \frac{k(k+1)\left(\tau^F\right)^2}{2}\left[\left(\tau^F\right)^{-(k+2)} - \left(\varphi^F\right)^{-(k+2)}\right]$. All the rest of the equilibrium variables, such as N_E^{HO}, N^{HO}, etc, can be expressed as a function of c_D^{HO} and other parameters. Compare the domestic cutoff from equation (2.23) with the socially optimum cutoff, it is straightforward to show that:

$$\left(\frac{c_D^{HM}}{c_D^{HO}}\right)^{k+2} = \frac{2}{\frac{1 - O_F}{1 - O_F O_H} / \frac{1 - \Phi_1^F - \Phi_2^F}{1 - \left(\Phi_1^F + \Phi_2^F\right)\left(\Phi_1^H + \Phi_2^H\right)}} \equiv \frac{2}{\triangle_F}$$

The term in the denominator of the above expression is defined as $\triangle_F$. For displaying purpose, I then focus on the comparison of all the market outcomes and socially optimum outcomes in the Home market. It follows from the definition of $c_X^{HM}, c_X^{HO}, c_{FDI}^{HM}, c_{FDI}^{HO}$ that the following three equations must hold:

$$c_D^{HM} - c_D^{HO} = \left[\left(\frac{2}{\triangle_F}\right)^{\frac{1}{k+2}} - 1\right] c_D^{HO}$$

$$c_X^{HM} - c_X^{HO} = \left[\left(\frac{2}{\triangle_F}\right)^{\frac{1}{k+2}} - t^F\right] \frac{c_D^{FO}}{\tau^F t^F}$$

$$c_{FDI}^{HM} - c_{FDI}^{HO} = \left[\left(\frac{2}{\triangle_F}\right)^{\frac{1}{k+2}} - \frac{1}{\varphi^F \xi^F}\right] c_D^{FO} \xi^F$$

It then can be verified that if the tariff level is not sufficiently high, $\triangle_F$ will be less than 1, hence both $\left[(2/\triangle_F)^{\frac{1}{k+2}} - 1\right]$ and $\left[(2/\triangle_F)^{\frac{1}{k+2}} - t^F\right]$ are greater than zero. Notice, $\varphi^F \xi^F < 1$, with the assumption that $\varphi^F > t^F \tau^F$, it is straightforward to show that $\left[(2/\triangle_F)^{\frac{1}{k+2}} - 1/\varphi^F \xi^F\right] < 0$. Therefore, part (A) of proposition 9 is proved.

For part (B), we can just follow the definition of production levels. For example, we know

$$q_D^{HM} = \frac{1}{2\gamma}(c_D^{HM} - c), q_D^{HO} = \frac{1}{\gamma}(c_D^{HO} - c)$$

It can be easily verified that $q_D^{HM} < q_D^{HO}$ if $c < \left[2 - (2/\triangle_F)^{1/(k+2)}\right] c_D^{HO}$. The production levels of q_X and q_{FDI} also follow directly from the comparison of market outcome and socially optimum outcome.

For part (C), one can show that

$$N_H^M = \frac{2\gamma(k+1)}{\eta} \times \frac{\alpha - c_D^{HM}}{c_D^{HM}}, N_H^O = \frac{\gamma(k+1)}{\eta} \times \frac{\alpha - c_D^{HO}}{c_D^{HO}}$$

One can then show that $N_H^M > N_H^O$ if

$$c_D^{HO} < \left[\frac{2}{(2/\triangle_F)^{\frac{1}{k+2}}} - 1\right] \alpha$$

For the level of entrants, one can solve them through the following system of equations:

$$N_H^O = N_E^{HO}\left(\frac{c_D^{HO}}{c_M}\right)^k + N_E^{FO}\left(\frac{c_D^{HO}}{c_M}\right)^k (\tau^H)^{-k}$$
$$N_F^O = N_E^{FO}\left(\frac{c_D^{FO}}{c_M}\right)^k + N_E^{HO}\left(\frac{c_D^{FO}}{c_M}\right)^k (\tau^F)^{-k}$$

And obtain the socially optimum entrant level:

$$N_E^{HO} = \frac{\gamma(k+1)c_M^k}{\eta[1-(\tau^H\tau^F)^{-k}]} \times \left[\frac{\alpha - c_D^{HO}}{(c_D^{HO})^{k+1}} - (\tau^H)^{-k}\frac{\alpha - c_D^{FO}}{(c_D^{FO})^{k+1}}\right]$$

Comparing it with the entrant level in the market outcome:

$$N_E^{HM} = \frac{2\,(c_M)^k\,(k+1)\,\gamma}{\eta\,(1-\delta^H\delta^F)} \times \left[\frac{\alpha - c_D^{HM}}{(c_D^{HM})^{k+1}} - \delta^H\frac{\alpha - c_D^{FM}}{(c_D^{FM})^{k+1}}\right]$$

where $\delta^l = (t^l\tau^l)^{-k}$, for $l \in \{H, F\}$. Together with the fact that $c_D^{HO} = c_D^{FO}[(1-O_F)/(1-O_FO_H)]^{\frac{1}{k+2}}$, one can then find a similar threshold of c_D^{HO} where N_E^{HO} differs from N_E^{HM}. That completes the proof of part (C). □

A.3.11 Proof of Proposition 10

Proof: Based on equation (2.28), the social welfare can be rewritten as follow:

$$\mathbb{U}^H + \mathbb{U}^F = I^H + I^F + \frac{\alpha - c_D^H}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_D^H\right) + \frac{\alpha - c_D^F}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_D^F\right)$$

Since consumer receive income from wage (which is equal to 1) and tariff revenue, so the above equation can be rewritten as:

$$\begin{aligned}\mathbb{U}^H + \mathbb{U}^F = {} & 2 + \left(t^H - 1\right) IM^H + \left(t^F - 1\right) IM^F \\ & + \frac{\alpha - c_D^H}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_D^H\right) + \frac{\alpha - c_D^F}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_D^F\right)\end{aligned}$$

To see the welfare implication of free trade, I evaluate the first order condition of the above expression with respect to tariff under symmetry when $t^H = t^F = 1$. Since symmetry implies $IM^F = IM^H, \partial IM^F/\partial t^H = \partial IM^H/\partial t^F, c_D^F = c_D^H, \partial c_D^F/\partial t^H = \partial c_D^H/\partial t^F$, therefore :

$$\frac{\partial\left(\mathbb{U}^H + \mathbb{U}^F\right)}{\partial t}|_{t^H=t^F=1} = \underbrace{(t-1)\left(\frac{\partial IM^H}{\partial t^H} + \frac{\partial IM^F}{\partial t^H}\right)}_{\circledast} + IM^H + \frac{2(k+1)c_D - (2k+3)\alpha}{2\eta(k+2)}\left(\frac{\partial c_D^H}{\partial t^H} + \frac{\partial c_D^F}{\partial t^H}\right)$$

Notice, when $t^H = t^F = 1$, $\circledast = 0$. Based on equation (2.27) and (2.29),

$$\begin{aligned} IM^H|_{t^H=1} &= N_E|_{t^H=1} \times \frac{(c_D)^{k+2}\tau^{-k}}{2\gamma(k+2)(c_M)^k}|_{t=1} \\ &= \frac{2(c_M)^k(k+1)\gamma\left(1-\tau^{-k}\right)}{\eta\left(1-\tau^{-2k}\right)}\frac{(c_D)^{k+2}\tau^{-k}}{2\gamma(k+2)(c_M)^k}|_{t=1} \\ &= \frac{\tau^{-k}(k+1)}{\eta\left(1+\tau^{-k}\right)(k+2)}(\alpha - c_D)\,c_D|_{t=1} \end{aligned}$$

Based on the definition of Φ^H and Φ^F, it is straightforward to show

$$\Phi_1 + \Phi_2 = \tau^{-k}$$

Based on the proof of Proposition 4, it is straightforward to show

$$\begin{aligned} \frac{\partial c_D^H}{\partial t^H} + \frac{\partial c_D^F}{\partial t^H}|_{t=1} &= \frac{-(k+1)c_D^H\tau^{-2k}}{(k+2)\left(1-\tau^{-2k}\right)} + \frac{c_D^F}{k+2}\frac{-(k+1)\tau^{-k}}{\left(1-\tau^{-2k}\right)} \\ &= \frac{\tau^{-k}(k+1)}{\left(1+\tau^{-k}\right)(k+2)}c_D|_{t=1} \end{aligned}$$

Therefore, the original first order condition can be rewritten as

$$\begin{aligned}
\frac{\partial\left(\mathbb{U}^{H}+\mathbb{U}^{F}\right)}{\partial t}|_{t^{H}=t^{F}=1} &= \frac{\tau^{-k}(k+1)}{\eta\left(1+\tau^{-k}\right)(k+2)}\left(\alpha-c_{D}\right)c_{D}|_{t=1} \\
&+\frac{2(k+1)c_{D}-(2k+3)\alpha}{2\eta(k+2)}\frac{\tau^{-k}(k+1)}{\left(1+\tau^{-k}\right)(k+2)}c_{D}|_{t=1} \\
&= \frac{\tau^{-k}(k+1)c_{D}}{2\eta(k+2)^{2}\left(1+\tau^{-k}\right)}\left[2(k+2)\left(\alpha-c_{D}\right)+2(k+1)c_{D}-(2k+3)\alpha\right] \\
&= \frac{\tau^{-k}(k+1)c_{D}}{2\eta(k+2)^{2}\left(1+\tau^{-k}\right)}\left(-2c_{D}+\alpha\right)|_{t=1}
\end{aligned}$$

Define $\tilde{c}_D \equiv c_D|_{t=1}$, then this completes the proof of proposition 10. □

A.3.12 Proof of Second-Best Social Planner Problem

Based on the definition of average consumer surplus, it can be rewritten in terms of $\tilde{c}_D$:

$$\text{Avg. CS} \equiv \frac{\gamma}{2}\int_{0}^{\tilde{c}_{D}}\left(q_{D}(c)\right)^{2}dG(c)=\frac{\left(\tilde{c}_{D}\right)^{k+2}}{4\gamma c_{M}^{k}(k+1)(k+2)}>0$$

Based on equation (2.26), under symmetry and $t = 1$,

$$N_{E}=\frac{2\gamma c_{M}^{k}(k+1)\left(\alpha-\tilde{c}_{D}\right)}{\eta\left(\tilde{c}_{D}\right)^{k+1}\left(1+\tau^{-k}\right)}$$

Then, the variety effect can be defined as the difference between consumer surplus and the sum of average surplus at each variety:

$$\begin{aligned}
\text{VE} &\equiv \text{CS}-N_{E}\times\text{Avg. CS} \\
&= \frac{\alpha-\tilde{c}_{D}}{2\eta}\left(\alpha-\frac{k+1}{k+2}\tilde{c}_{D}\right)-\frac{2\gamma c_{M}^{k}(k+1)\left(\alpha-\tilde{c}_{D}\right)}{\eta\left(\tilde{c}_{D}\right)^{k+1}\left(1+\tau^{-k}\right)}\times\frac{\left(\tilde{c}_{D}\right)^{k+2}}{4\gamma c_{M}^{k}(k+1)(k+2)} \\
&= \frac{\left(\alpha-\tilde{c}_{D}\right)}{2\eta}\left[\alpha-\frac{(k+1)\left(1+\tau^{-k}\right)+1}{(k+2)\left(1+\tau^{-k}\right)}\tilde{c}_{D}\right]
\end{aligned}$$

The expected profit of a firm can be derived from equation (2.20):

$$\begin{aligned}\bar{\pi} &= \int_0^{\tilde{c}_D} \pi_D(c)\, dG(c) + \int_0^{\tilde{c}_X} \pi_X(c)\, dG(c) \\ &= \frac{(\tilde{c}_D)^{k+2}}{2\gamma c_M^k (k+1)(k+2)} + \frac{\tau^2 (\tilde{c}_X)^{k+2}}{2\gamma c_M^k (k+1)(k+2)} = \frac{(\tilde{c}_D)^{k+2}\left(1+\tau^{-k}\right)}{2\gamma c_M^k (k+1)(k+2)}\end{aligned}$$

Notice when $t = 1, c_{FDI} = 0$ and $\tilde{c}_X = \tilde{c}_D/\tau$. With all these components and the fact that $\tilde{c}_D < \alpha$, equation (3.1) can now be properly signed:

$$\begin{aligned}\text{Avg. CS} &= \frac{(\tilde{c}_D)^{k+2}}{4\gamma c_M^k (k+1)(k+2)} > 0 \\ N_E \frac{\partial \text{Avg. CS}}{\partial N_E} &= \frac{(\alpha - \tilde{c}_D)(\tilde{c}_D)^{k+2}}{4\gamma c_M^k (k+1)\left[k\tilde{c}_D - \alpha(k+1)\right]} < 0 \\ \frac{\partial \text{VE}}{\partial N_E} &= \frac{(\tilde{c}_D)^{k+2}\left\{\alpha(k+2)(1+\tau^{-k}) + (\alpha - 2\tilde{c}_D)\left[(k+1)(1+\tau^{-k})+1\right]\right\}}{4\gamma c_M^k (k+1)(k+2)\left[\alpha(k+1) - k\tilde{c}_D\right]} > 0 \\ N_E \frac{\partial \bar{\pi}}{\partial N_E} &= \frac{(\tilde{c}_D)^{k+2}\left(1+\tau^{-k}\right)(\alpha - \tilde{c}_D)}{2\gamma c_M^k (k+1)\left[k\tilde{c}_D - (k+1)\alpha\right]} < 0\end{aligned}$$

Therefore, the externality of entry equals to

$$\begin{aligned}&\text{Avg. CS} + N_E \frac{\partial \text{Avg. CS}}{\partial N_E} + \frac{\partial \text{VE}}{\partial N_E} + N_E \frac{\partial \bar{\pi}}{\partial N_E} \\ =& \frac{\left(1+\tau^{-k}\right)(\tilde{c}_D)^{k+2}}{4\gamma c_M^k (k+1)(k+2)\underbrace{\left[k\tilde{c}_D - (k+1)\alpha\right]}_{<0}} \times (\alpha - 2\tilde{c}_D)\end{aligned}$$

Therefore, the externality will be negative if $\tilde{c}_D < \alpha/2$, will be positive if $\tilde{c}_D > \alpha/2$. □

A.3.13 Proof of Proposition 11

Proof: Under symmetry, it is straightforward to derive the following welfare expression for market outcome and socially optimum outcome:

$$\mathbb{W}^M = 1 + \bar{q}_0 + \frac{\alpha - c_D^M}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_D^M\right)$$
$$\mathbb{W}^O = 1 + \bar{q}_0 + \frac{1}{2\eta}\left(\alpha - c_D^O\right)^2$$

And based on the proof of Proposition 9 that $c_D^M = (2/\triangle)^{\frac{1}{k+2}} c_D^O$, where the $\triangle$ is the symmetric version of $\triangle_F$ and $\triangle_H$. To simplify the expression, here I define:

$$A \equiv \left(\frac{2}{\triangle}\right)^{\dfrac{1}{k+2}}$$
$$B \equiv \frac{k+1}{k+2}$$

Therefore, the welfare gap between market outcome and socially optimum is:

$$\mathbb{W}^M - \mathbb{W}^O = \frac{1}{2\eta}\left[c_D^O(2\alpha - AB\alpha - A\alpha) + (c_D^O)^2(A^2B - 1)\right]$$

It is obvious that the market welfare is smaller than the socially optimum welfare. One can then show the gap is decreasing in t:

$$\frac{\partial(\mathbb{W}^O - \mathbb{W}^M)}{\partial t} = c_D^O \frac{\partial A}{\partial t}\left[\alpha(A+B) - 2ABc_D^O\right]$$

Given that $\partial A/\partial t > 0$, then the whole expression will be negative if $c_D^O > \alpha(A+B)/2AB$, which is equivalent to $c_D^M > \alpha(A+B)/2B$. Therefore part *(i)* of the proposition is proved.

The second part of the proposition is easy. Notice that:

$$\mathbb{W}^H = 1 + \bar{q}_0^H + \frac{\alpha - c_D^H}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_D^H\right)$$
$$\mathbb{W}^F = 1 + \bar{q}_0^F + \frac{\alpha - c_D^F}{2\eta}\left(\alpha - \frac{k+1}{k+2}c_D^F\right)$$

The rest of the proof directly follows from the Proposition 4. □

A.3.14 Proof of Proposition 12

Proof: First, based on the definition of IM in equation (2.29), it is straightforward to show that $\partial IM^H/\partial t^H < 0$. Now compare equation (3.2) and (3.3)

$$t_S^H - 1 = \frac{\frac{\partial CS^H}{\partial t^H} + IM^H + \frac{\partial CS^F}{\partial t^H} + \left(t^F - 1\right) \times \frac{\partial IM^F}{\partial t^H}}{-\frac{\partial IM^H}{\partial t^H}}$$

$$t_N^H - 1 = \frac{\frac{\partial CS^H}{\partial t^H} + IM^H}{-\frac{\partial IM^H}{\partial t^H}}$$

It's straightforward to show that when $\tilde{c}_D > \alpha/2$:

$$\frac{\partial CS^F}{\partial t^H} = \underbrace{\frac{\partial CS^F}{\partial c_D^F}}_{<0} \times \underbrace{\frac{\partial c_D^F}{\partial t^H}}_{>0} < 0$$

$$\left(t^F - 1\right) \times \frac{\partial IM^F}{\partial t^H} = \frac{t^F \left(t^F - 1\right) \left(\tau^F\right)^2}{4\gamma \left(k+2\right) \left(c_M\right)^k} \left[2 \left(\frac{1}{t^F \tau^F}\right)^{k+2} - \frac{k+2}{\left(t^F \tau^F\right)^2} \left(\xi^F\right)^k + k \left(\xi^F\right)^{k+2}\right] \underbrace{\frac{\partial N_E^H \left(c_D^F\right)^{k+2}}{\partial t^H}}_{>0} > 0$$

The first item indicates the distortion on F's consumption generated by t^H, the second item indicates the distortion on F's tariff revenue generated by t^H. It's straightforward to show the distortion on consumption is larger than the distortion on tariff revenue. Therefore the sum of these two items is negative, indicating that when the Nash tariff is evaluated at the socially optimum first-order condition, it creates negative externality. Therefore, it must be the case that $t_S^H < t_N^H$. Hence the Nash tariff is higher than the socially optimal tariff. □

A.3.15 Proof of Average Markup

Once again, to simplify the proof, I assume symmetry, following Appendix A.3.16, $c_D = c_D^H = c_D^F$ and

$$\begin{aligned} N_E &= \frac{2\gamma c_M^k (k+1)(\alpha - c_D)}{\eta (c_D)^{k+1} (1 + t^{-k}\tau^{-k})} \\ N_D &= \frac{2\gamma (k+1)(\alpha - c_D)}{\eta (1 + t^{-k}\tau^{-k}) c_D} \\ N_X &= \frac{2\gamma (k+1)(\alpha - c_D)}{\eta (1 + t^{-k}\tau^{-k}) c_D} \left[(t\tau)^{-k} - \xi^k \right] \\ N_{FDI} &= \frac{2\gamma (k+1)(\alpha - c_D)}{\eta (1 + t^{-k}\tau^{-k}) c_D} \xi^k \end{aligned}$$

Together with equation (2.6), (2.10) and (2.15), the average markup in (3.3) can be written as follow:

$$\begin{aligned} \bar{m} &= \frac{1}{N_D + N_X + N_{FDI}} \left[N_D \int_0^{c_D} m_D(c) \frac{dG(c)}{G(c_D)} \right. \\ &\left. + N_X \int_{c_{FDI}}^{c_X} m_X(c) \frac{dG(c)}{G(c_X)} + N_{FDI} \int_0^{c_{FDI}} m_{FDI}(c) \frac{dG(c)}{G(c_{FDI})} \right] \\ &= \frac{1}{N_D + N_X + N_{FDI}} \left[N_D \int_0^{c_D} \frac{c_D + c}{2c} \frac{kc^{k-1}}{c_D^k} dc \right. \\ &\left. + N_X \int_{c_{FDI}}^{c_X} \frac{t(c_X + c)}{2c} \frac{kc^{k-1}}{c_X^k} dc + N_{FDI} \int_0^{c_{FDI}} \frac{c_D + \varphi^H c}{2\varphi^H c} \frac{kc^{k-1}}{c_{FDI}^k} dc \right] \\ &= \frac{1}{N_D + N_X + N_{FDI}} \left[N_D \times \frac{2k-1}{2k-2} + N_X \times t \right. \\ &\left. \times \left(\frac{2k-1}{2k-2} - \frac{k}{2k-2} (t\tau\xi)^{k-1} - \frac{1}{2} (t\tau\xi)^k \right) + N_{FDI} \times \left(\frac{k}{2k-2} \frac{1}{\varphi\xi} + \frac{1}{2} \right) \right] \end{aligned}$$

$$= \frac{1}{1+(t\tau)^{-k}} \times \frac{2k-1}{2k-2} + \frac{t\left[(t\tau)^{-k} - \xi^{k}\right]}{1+(t\tau)^{-k}} \times \left[\frac{2k-1}{2k-2} - \frac{k}{2k-2}(t\tau\xi)^{k-1} - \frac{1}{2}(t\tau\xi)^{k}\right]$$
$$+ \frac{\xi^{k}}{1+(t\tau)^{-k}} \times \left(\frac{k}{2k-2}\frac{1}{\varphi\xi} + \frac{1}{2}\right)$$
$$= \underbrace{\frac{1}{1+(t\tau)^{-k}} \times \frac{2k-1}{2k-2}}_{\text{weighted expected markup in domestic}}$$
$$+ \underbrace{\frac{1-(t\tau\xi)^{k}}{1+(t\tau)^{-k}} \times \frac{1}{t^{k-1}\tau^{k}}\left\{\frac{1}{2}\left[1-(t\tau\xi)^{k}\right] + \frac{k}{2k-2}\left[1-(t\tau\xi)^{k-1}\right]\right\}}_{\text{weighted expected markup from Foreign exporters}}$$
$$+ \underbrace{\frac{\xi^{k}}{1+(t\tau)^{-k}} \times \left(\frac{k}{2k-2}\frac{1}{\varphi\xi} + \frac{1}{2}\right)}_{\text{weighted expected markup from Foreign FDI}}$$

A.3.16 Proof of Covariance Term

Once again, to simplify the analysis, I imposed symmetry. It's clear from equation (2.26) that

$$N_E = \frac{2\gamma c_M^{k}(k+1)(\alpha - c_D)}{\eta (c_D)^{k+1}(1+t^{-k}\tau^{-k})}$$

where $c_D = c_D^H = c_D^F$. Given equation (2.7), (2.11) and (2.16), and the following expression for the mass of firms:

$$N_D = N_E \times G(c_D) = \frac{2\gamma (k+1)(\alpha - c_D)}{\eta (1+t^{-k}\tau^{-k}) c_D}$$
$$N_X = N_E \times [G(c_X) - G(c_{FDI})] = \frac{2\gamma (k+1)(\alpha - c_D)}{\eta (1+t^{-k}\tau^{-k}) c_D}\left[(t\tau)^{-k} - \xi^{k}\right]$$
$$N_{FDI} = N_E \times G(c_{FDI}) = \frac{2\gamma (k+1)(\alpha - c_D)}{\eta (1+t^{-k}\tau^{-k}) c_D}\xi^{k}$$

the covariance term can be derived as follow:

$$\begin{aligned}
\text{cov}\left(m^i(\omega), \frac{dl^i(\omega)}{L^j}\right) &= N_D \int_0^{c_D} p_D(c)\, d\left[q_D(c)\right] \frac{dG(c)}{G(c_D)} + N_X \int_{c_{FDI}}^{c_X} p_X(c)\, d\left[q_X(c)\right] \frac{dG(c)}{G(c_X)} \\
&+ N_{FDI} \int_0^{c_{FDI}} p_{FDI}(c)\, d\left[q_{FDI}(c)\right] \frac{dG(c)}{G(c_{FDI})} \\
&= \frac{2\gamma(k+1)(\alpha - c_D)}{\eta(1+t^{-k}\tau^{-k})c_D} \int_0^{c_D} \frac{kdc_D(c_D+c)c^{k-1}}{4\gamma c_D^k} dc \\
&+ \frac{2\gamma(k+1)(\alpha - c_D)}{\eta(1+t^{-k}\tau^{-k})c_D} \left[(t\tau)^{-k} - \xi^k\right] \int_{c_{FDI}}^{c_X} \frac{t^2\tau^2 kdc_X(c_X+c)c^{k-1}}{4\gamma c_X^k} dc \\
&+ \frac{2\gamma(k+1)(\alpha - c_D)}{\eta(1+t^{-k}\tau^{-k})c_D} \xi^k \int_0^{c_{FDI}} \frac{kdc_D(c_D+\varphi c)c^{k-1}}{4\gamma c_F^k} dc \\
&= \frac{(\alpha - c_D)\, dc_D}{2\eta(1+t^{-k}\tau^{-k})} \left\{2k+1+\left[(t\tau)^{-k} - \xi^k\right]\right. \\
&\times \left.\left[2k+1-k(1-t\tau\xi)(t\tau\xi)^k - (t\tau\xi)^k\right] + \xi^k(k+k\varphi+1)\right\}
\end{aligned}$$

When $\varphi > t\tau$, it is straightforward to show that $t\tau\xi < 1$. Therefore, the covariance term can be rewritten as

$$\begin{aligned}
\text{cov}\left(m^i(\omega), \frac{dl^i(\omega)}{L^j}\right) &= \frac{(\alpha - c_D)\, dc_D}{2\eta(1+t^{-k}\tau^{-k})} \left\{2k+1+(t\tau)^{-k} \underbrace{\left[1-(t\tau\xi)^k\right]}_{>0}\right. \\
&\times \left.\left[k+k(t\tau\xi)^{k+1} + \underbrace{k - k(t\tau\xi)^k}_{>0} + \underbrace{1-(t\tau\xi)^k}_{>0}\right] + \xi^k(k+k\varphi+1)\right\}
\end{aligned}$$

Notice, under symmetry

$$\begin{aligned}
\frac{dc_D}{dt} &= \frac{d}{dt}\left[\frac{\gamma\phi}{1+\Phi_1+\Phi_2}\right]^{\frac{1}{k+2}} \\
&= \frac{1}{k+2}\left[\frac{\gamma\phi}{1+\Phi_1+\Phi_2}\right]^{-\frac{k+1}{k+2}} \times \frac{-\gamma\phi}{(1+\Phi_1+\Phi_2)^2} \frac{d(\Phi_1+\Phi_2)}{dt} \\
&= -\frac{1}{k+2} \times \frac{c_D}{1+\Phi_1+\Phi_2} \times \frac{d(\Phi_1+\Phi_2)}{dt}
\end{aligned}$$

According to Appendix A.3.4, $d\left(\Phi_1 + \Phi_2\right)/dt < 0$, hence $dc_D/dt > 0$, hence the covariance term is positive for $dt > 0$. $\square$