2009

INTERNATIONAL TRADE AND INDUSTRIAL GEOGRAPHY

Fabien Tondel
University of Kentucky

ABSTRACT OF DISSERTATION

Fabien Tondel

The Graduate School
University of Kentucky
2009

INTERNATIONAL TRADE AND INDUSTRIAL GEOGRAPHY

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Agriculture at the University of Kentucky

By
Fabien Tondel
Lexington, Kentucky

Co-Directors: Dr. David Freshwater, Professor of Agricultural Economics
and: Dr. Michael R. Reed, Professor of Agricultural Economics
Lexington, Kentucky
2009

ABSTRACT OF DISSERTATION

INTERNATIONAL TRADE AND INDUSTRIAL GEOGRAPHY

This dissertation explores the impact of international trade on the geographic location of manufacturing activities and on regional productivity growth patterns within countries. This study develops models of trade with monopolistic competition in the context of a two-region country. It also provides empirical estimates of the effect of tariff policy on the distribution of industrial activities and on productivity growth differentials across Colombia's regions.

The first essay investigates the consequences of trade liberalization for the distribution of manufacturing activities between large and small cities. It presents an extension of the Melitz (2003) model of trade with monopolistic competition and heterogeneous firms where producers' location and export market participation decisions depend on their productivity. As a country's exposure to trade shifts, firms and output are reallocated between large and small urban areas. Data from Colombia's manufacturing sector lend support to theoretical predictions concerning tariff reduction's impact on the repartition of industrial activities between metro- and non-metropolitan areas in this country.

The second essay extends the New Economic Geography, Footloose-Capital model to examine the effect of commercial policy on the distribution of industrial activities between regions within a country. This study aims at distinguishing theoretical cases with regard to the nature of the trade policy change or to the source of asymmetry between regions. It shows that trade liberalization can have adverse consequences for the manufacturing sector of a small or isolated region under bilateral liberalization, but a positive impact under unilateral trade liberalization.

The third essay adapts the Melitz and Ottaviano (2008) model of trade with monopolistic competition, heterogeneous firms, and variable mark-ups to analyze the relationship between trade openness, regional market size, and regional aggregate industry performance. It demonstrates that the impact of trade liberalization on aggregate industry productivity growth varies across regions as a function of regional market size and proximity to foreign markets. A larger region experiences a greater increase in aggregate productivity through intra-industry reallocation of market shares.

Similarly, a region with better access to international markets enjoys a higher productivity growth from tariff reduction. Empirical evidence is obtained from the Colombian manufacturing sector.

KEYWORDS: Colombia, heterogeneous firms, industrialization, spatial distribution of regional economic activity, trade policy

Author's signature: Fabien Tondel

Date: June 29, 2009

INTERNATIONAL TRADE AND INDUSTRIAL GEOGRAPHY

By
Fabien Tondel

Co-Director of Dissertation: David Freshwater

Co-Director of Dissertation: Michael R. Reed

Director of Graduate Studies: Michael R. Reed

Date: June 29, 2009

RULES FOR THE USE OF DISSERTATIONS

DISSERTATION

Fabien Tondel

The Graduate School
University of Kentucky
2009

INTERNATIONAL TRADE AND INDUSTRIAL GEOGRAPHY

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Agriculture at the University of Kentucky

By
Fabien Tondel
Lexington, Kentucky

Co-Directors: Dr. David Freshwater, Professor of Agricultural Economics
and: Dr. Michael R. Reed, Professor of Agricultural Economics
Lexington, Kentucky
2009

To the memory of Raymond Contri

ACKNOWLEDGMENTS

The faculty and staff of the Department of Agricultural Economics at the University of Kentucky, my friends, and my family provided invaluable support to my endeavor.

The Department of Agricultural Economics was tremendously supportive, both financially and materially, during my tenure at the University of Kentucky. Financial support from the Kentucky Opportunity Fellowship is gratefully acknowledged.

I wish to thank the members of my doctoral advisory committee: Dr. David Freshwater, Dr. Michael Reed, Dr. Josh Ederington, and Dr. Timothy Woods. I am greatly indebted to Dr. Reed who gave me the opportunity to pursue a graduate education at the University of Kentucky. His counsel and support as regards both academic and personal matters were helpful in bringing this dissertation to fruition and in my professional development. I am obliged to Dr. Josh Ederington for his guidance at the conception of and through the completion of the first essay, and for his suggestions about the two other essays. I thank Dr. Timothy Woods for his reviewing the manuscript and making worthy remarks. I am grateful to Dr. David Freshwater for his attention and contribution to my dissertation at all stages of this enterprise. The seminal idea leading to this dissertation emerged from our conversations and from studying earlier work of his. I am thankful for his insightful suggestions and the intellectual curiosity he instilled in me.

My appreciation is also addressed to my fellow students, my colleagues, and the staff of the Agricultural Economics department. Emma Bojinova's collaboration on the first essay was very beneficial. Our entertaining conversations with Vijay Subramanian added laughs to our daily routine and generated mutual encouragement. My friendship with Lucia Ona, Ibrahim Bamba, Jean-Marc Gandonou, Tery Lundsford, and others made life at the Department enjoyable and was intellectually and culturally enriching.

I am immensely grateful to my mother, Marielle, my father, Marius, my stepfather, Marc, my sisters, Julia and Mathilde, my grandmother and late grandfather, Josette and Raymond, for the mental and material support they devoted to me throughout my doctoral studies, for their patient encouragement, and for their caring presence through our joyful phone conversations every Sunday. In addition, I am thankful to my other family members and other friends for support of all kinds throughout my studies, especially my uncle Philippe Contri and his wife, and my uncle Philippe Tondel and his wife. Last but not least, I wish to express my gratitude to Chanatip for her encouragement, attention to my concerns, and caring love. Her enduring presence by my sides, as well as that of my faithful companions Coco, Momo, and Matcha, helped me a great deal and I enjoyed all the good moments we shared in this challenging period of our life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Trade, Economic Geography and Development

This dissertation, *International Trade and Industrial Geography*, explores the impact of international trade on the geographic location of manufacturing activities and on regional patterns of productivity growth within countries. This study comprises three independent essays. Each essay develops a theoretical model of trade with monopolistic competition and scale economies in the context of a two-region country to address this issue. Two of the three essays evaluate the theoretical results using data from the manufacturing sector of Colombia.

The aforementioned topic has received much attention in the recent past and numerous contributions have been made to the economic literature. This interest has come at a time when countries worldwide are adjusting to a greater exposure of their economies to international commerce and factor movements, and to the international diffusion of ideas and technologies. In the aftermath of the First World War and the economic crisis of the 1930s, governments around the world imposed protectionist trade policies that led to disruption in the international trading system. Nonetheless, in the 1970s, under favorable economic conditions, national Keynesian policies were giving way—in the US and the UK at least—to a more international and liberal economic agenda, and import substitution policies were being questioned in developing countries. Thus, at that time, countries started to engage in trade liberalization reforms, unilaterally, bilaterally, or through multilateral agreements. Subsequently, trade flows grew at a rapid pace. According to Sachs and Warner [1995], the share of countries described as open increased from 35% in 1980 to 95% in the late 1990s, and the average share of trade in the gross domestic product (GDP) across countries went from 59% up to 74% over this period. Meanwhile, technical progress in the air, land, and maritime transportation, the proliferation of transport infrastructures, and the development of information and communication technologies were lowering the costs of conducting commercial transactions across countries [Hummels, 2007].

These changes happening in the world economy have had noticeable consequences for the geographic location of economic activities within countries. The positive relationship between tariff policy and the geographic concentration of population uncovered by Ades and Glaeser [1995] is often cited as evidence on the link between external trade and internal geography. The shift in the spatial distribution of manufacturing activities in Mexico occurring after this country liberalized its trade policy in the mid 1980s, or the widening gap in economic performance between China's coastal regions and the interior regions have frequently been commented. The 2009 World Development Report of the World Bank [International Bank for Reconstruction and Development/The World Bank, 2009] recognized that rising openness to trade and capital flows has increased sub-national geographic disparities in income and has made them more persistent in developing countries. Not all regions of a country are suited for participating in world markets and coastal and economically dense regions perform better. The analyses in this dissertation generally confirm this claim although they provide a more nuanced assessment.

The spatial concentration of population and productive activities and geographical disparities in income and living standards are ubiquitous phenomena. They arise at different geographical scales, at the local, national, and international levels. At the local level, spatial inequalities are mostly associated with variations in density, or urbanization. Within a region, large cities usually offer higher incomes and living standards than small cities and rural areas. At the national level, population and economic activity tend to concentrate in regions having favorable natural attributes such as proximity to a coast or to other countries. Distance to markets is a major source of disparities in economic outcomes across regions. Disparities among regions are most visible in countries like China, where more than half of the country's GDP in 2005 originated from coastal provinces representing less than a fifth of the country's land area (essentially from the regions of the Bohai Basin, the Pearl River Delta, and the Yangtze River Delta); Brazil, where the south-central states of Minas Gerais, Rio de Janeiro, and São Paulo have been producing more than half of the country's GDP, whereas they comprise less than 15 percent of its land area [International Bank for Reconstruction and Development/The World Bank, 2009]. At the international level, income is concentrated in countries with a large market and wide access to other countries because of lower barriers to trade (tariff barriers, non-tariff barriers, currency transaction costs).

First-nature geography is an important determinant of spatial disparities in population density, production activities, and living standards. However, the evidence indicates that economic development leads to concentration of economic activities and divergence in income across locations. Only at a later stage of development, do spatial disparities in income and living standards diminish. For instance, income inequality across the United States (US) has consistently declined over the twentieth century. During the second half of that century, the South has grown more rapidly than the Northeast. The location pattern of the US manufacturing sector has particularly reflected that trend. The phenomena of divergence and convergence are more rapid at smaller geographical scales. Large cities are established in a shorter period of time than leading regions, and leading regions emerge more rapidly than leading countries. Divergence tends to unfold according to a mechanism of cumulative causation. A geographic shift in production induces population migration, which causes further relocation of productive activities. Convergence happens through neighborhood effects, or positive spillovers, originating from leading locations, where economic activities are concentrated, and benefitting lagging locations (local areas, regions, or countries). These neighborhood effects may be related to the evolution of technology (standardization, improvement in logistics and communications). As transport and communication costs allow firms to spatially separate research and development and finance tasks, in large cities, from production, in small towns, manufacturing activities tend to decentralize.

Societies should be concerned with inequalities in income, labor market outcomes, and living standards across locations within their territory. Such disparities can be put social convergence and cohesion at risk as they contribute to overall inequalities among people, besides inequalities across individuals within geographical areas, and because of the possibility that inequalities among regions within a country coincide with political, ethnic, cultural, or religious divisions. Nevertheless, policy addressing spatial inequalities should allow the efficiency gains from the concentration of economic activities to realize. They should also promote the integration of leading and lagging regions to ensure the convergence of social indicators of well-being across places. Thus, it is desirable to have a better understanding of the effects of trade on the reallocation of manufacturing activities across locations to know better which regions benefit from trade and international economic integration. This would help policy makers to better use the traditional instruments for economic integration within a country such as institutions, infrastructure, and targeted interventions.

The first essay, "Trade Liberalization and the Geographic Location of Industries," investigates the link between trade openness and the distribution of manufacturing activities between large cities and small urban areas. This analysis builds on the Melitz [2003] model of trade with monopolistic competition and heterogeneous firms. The Melitz model extended the Krugman [1980] model of trade with monopolistic competition and scale economies by adding differences in productivity among firms and fixed costs of entering export markets. Productivity differences among manufacturing establishments and firms within industries are well documented [Bartelsman and Doms, 2000]. And the propensity to export has been shown to be strongly correlated with productivity (see references in Melitz [2003]). In the Melitz model, productivity determines whether a firm enters the export market. In this essay, the model assumes that the economy is made of two locations, an urban region and a rural region. These regions exhibit different production cost structures. Producers' location and export market participation decisions depend on their productivity. As a country's exposure to trade shifts, firms and output are reallocated between large and small urban areas. Data from Colombia's manufacturing sector lend support to theoretical predictions concerning tariff reforms' impact on the repartition of industrial activities between metro- and non-metropolitan areas in this country.

The second essay, "Trade Policy, Capital Mobility and Industrial Geography," extends the New Economic Geography (NEG), Footloose-Capital model of Martin and Rogers [1995] to examine the effect of international commerce on the distribution of industrial activities between regions within a country. The NEG framework introduced by Krugman [1991] described the role that scale economies and transport costs play in leading to the concentration of economic activities, magnifying the favorable aspects of a region's natural geography. This framework was used in the Krugman and Livas-Elizondo [1996] to look at the influence of trade openness on the regional distribution of firms. This essay extends that analysis by distinguishing theoretical cases on the basis of the nature of the trade regime change or of the source of regional asymmetry. It shows that trade liberalization can have adverse consequences for the manufacturing sector of a small or isolated region under bilateral liberalization, but a positive impact under unilateral trade liberalization.

The third essay, "The Impact of Trade Liberalization on Productivity Across Regions, adapts the **?** model of trade with monopolistic competition, heterogeneous firms, and variable mark-ups to analyze the relationship between trade openness, regional market size, and regional aggregate industry performance. It demonstrates that the impact of trade liberalization on aggregate industry productivity growth varies across regions as a function of regional market size and proximity to foreign markets. A larger region experiences a greater increase in aggregate productivity through intra-industry reallocation of market shares. Similarly, a region with better access to international markets enjoys a higher productivity growth from tariff reduction. Empirical support is obtained from the Colombian manufacturing sector.

The empirical part of this dissertation investigates the case of Colombia for reasons related to the availability of data, the history of its trade policies, and its geographical characteristics. First, quality data on Colombian manufacturing establishments were readily available. Second, Colombia quickly liberalized its trade policy between the mid 1980s and the early 1990s after more than two decades of protectionism and industrialization through import substitution. Third, this country has a relatively large territory and exhibit disparate natural and economic geographies. Whereas it is bordered by two oceans, the capital city, Bogotá, is situated in the interior of the country. In addition, in the recent past, rural populations migrated in mass to Bogotá.

Chapter 2 Trade Liberalization and the Geographic Location of Industries

2.1 Introduction

In most countries, the uneven spatial distribution of population, infrastructure, natural resources, and institutions has given rise to regional disparities in prosperity and industrial development in particular.[1] As international trade entails reallocations of factors among and within industries, it may also alter the geographic distribution of industrial activities across the regions of a country. The assessment of trade policies should consider their implications for different regions. In this paper, we study one aspect of the link between trade policy and the geographic location of industries within a country. Specifically, we address the question of whether trade liberalization causes a redistribution of manufacturing activities between metropolitan areas and less urbanized regions such as small cities, towns, or rural areas.

Observers have noted that the promotion of industrialization through import substitution after World War II in countries of Latin America bolstered the concentration of population and industrial activities within their metropolises, like in Mexico. The trade policy reforms implemented a few decades later looked as if they spurred the redeployment of manufacturing activities away from these large cities. The era of import substitution in Mexico came to an end as the country swiftly enacted trade policy reforms between 1985 and 1989. Hanson [1998] finds that, throughout the period 1985–93, industries expanded faster in the states closer to the Mexico-U.S. border.[2] In addition, the positive effect of the localization of upstream and downstream industries on industry growth at the state-level weakened in the period 1985–93 compared with 1980–85. While industrial activities relocated to the northern states to cut the costs of transporting goods to the U.S. market, firms, all else equal, also relocated production plants away from the old industrial centers.[3] Argentina carried out a

[1] For instance, it is common to observe absolute and relative factor price differences across regions within a country. Bernard et al. [2005b] report evidence of relative factor endowment and price differences across the states of Mexico. For the United States (U.S.) see Bernard et al. [2005a].

[2] From 1980 through 1993, manufacturing employment in the two states in which lies Mexico City's metropolitan area dropped from 44 to 29% of national manufacturing employment, while the share of the states adjoining the U.S. rose from 21 to 30 percent.

[3] Rodríguez-Pose and Sánchez-Reaza [2005] report that a state's distance to Mexico City had a negative impact on its rate of per-capita GDP growth from 1980 through 1985, but it no longer affected state-level growth after 1985.

comprehensive trade liberalization agenda between 1988 and 1991 and entered the *Mercado Común del Sur* (MERCOSUR) in 1991. Industrial activities had been very clustered within the country's capital and the densely populated province of Buenos Aires until that time. Sanguinetti and Volpe Martincus [2009] document a rise in the shares of manufacturing employment of the peripheral provinces between 1985 and 1994. They also show that industries with lower import tariffs located in provinces that are more distant from the city of Buenos Aires. In an influential article, Krugman and Livas-Elizondo [1996] put forward an explanation for the decentralization of industrial activities in these Latin America countries. They contend that in a relatively closed economy, the tendency of the manufacturing industry to concentrate in one location, which arises from the interplay of scale economies with trade costs, is stronger than the factor causing its dispersion, which stems from competition in the land and labor markets. As the economy opens up to trade, manufacturers export a larger share of their production while consumers substitute foreign products for domestic ones; and thus the incentives for firms and consumers/workers to concentrate in one location wane.[4]

Other studies have focused on the case of developed countries. The exposure of Japan to international trade shifted after a series of episodic currency appreciations in the mid 1980s and early 1990s. Subsequently, the shares of manufactured imports from Asian countries and Japanese affiliates abroad rose substantially. Tomiura [2003] records a negative correlation between regional manufacturing employment growth and a regional index measuring the change in industries' exposure to import competition in the 1990s. Furthermore, his analysis suggests that the growth in imports of intermediate goods from low-wage Asian countries disrupted the vertical linkages within regions and old industrial centers and induced firms to relocate production activities closer to imperfectly mobile factors such as specialized workers and places offering positive externalities such as knowledge spillovers, or away from locations exhibiting congestion.

Hence, some studies provide anecdotal evidence that a greater exposure to trade led to a decline in the concentration of industrial activities in large urban and industrial centers. Moreover, other studies document the decentralization of manufacturing activities, across a variety of industries, from large urban centers to less urbanized, outlying locations such as towns and rural regions from the late 1960s through the early 1980s in the U.S. (Roth 2000) and in the European Community (Keeble et al.

[4] On the basis of variations of this model, Monfort and Nicolini [2000] and Paluzie [2001] refute this claim but Behrens et al. [2006, 2007] confirm it.

1983).

The theoretical part of this paper builds on the Melitz [2003] model of trade under monopolistic competition to investigate the link between external trade openness and the pattern of industry location within a country. In Melitz [2003], firms exhibit different levels of productivity and must incur a fixed cost to export. A self-selection process leads less productive firms to supply only the domestic market and more productive ones to serve both the domestic and foreign markets. As the domestic economy becomes more open to trade, the least productive firms lose market shares and exit the industry, exporters gain market shares, and the most productive non-exporters enter the foreign market. Our model assumes that the domestic economy is made of two locations, the urban and rural regions. The rural location requires a higher fixed cost of production than the urban location whereas the factor price (that is, the marginal cost) is lower in the former. The same process of self-selection at work in Melitz [2003] results in a sorting of firms across locations, according to which more productive firms choose to produce in the rural region.[5] Then, a shift in the country's openness to trade creates incentives for some firms to relocate to another location and thus induces a reallocation of market shares across regions. In a particular case, we find that the most productive urban firms relocate to the rural region as their ability to export improves.

Furthermore, this paper presents corroborating evidence for some aspects of the theory, from the country of Colombia. The commercial and geographical contexts of the Colombian manufacturing sector are appropriate for an evaluation of the theory because quality data are available, Colombia conducted expeditious reforms of its commercial policy from 1985 to 1992, and its economic activities are unevenly distributed over a large territory. Along the lines of the model, we distinguish manufacturing plants located in metropolitan (metro) areas from those sited in nonmetropolitan (nonmetro) regions. We document differences in productivity between metropolitan and nonmetropolitan establishments and find that the reduction in import tariffs entailed a reallocation of manufacturing activities between metro and nonmetro areas.

The remainder of the paper is organized as follows: Section 2.2 describes the setup and equilibrium of the closed economy model. Section 2.3 reports evidence

[5]Our approach differs from that of Krugman and Livas-Elizondo [1996] and related studies applying the New Economic Geography framework of Krugman [1991]. In those studies, the size of the local market and local production costs are endogenously determined by the location of the manufacturing industry. In our model, the spatial difference in production costs is exogenous to the manufacturing sector and instead reflects different levels of economic development, in terms of urbanization, presence of support activities and infrastructure, and so forth.

of productivity differences between metro and nonmetro plants in Colombia prior to the trade policy reforms of the period 1985–91. Section 2.4 characterizes the open economy model and equilibrium. Section 2.5 shows how an increase in trade openness leads to a regional relocation of firms and production. Section 2.6 estimates the impact of a tariff reduction on the distribution of plants and production among metro and nonmetro areas in Colombia. Section 2.7 summarizes our findings.

2.2 Closed Economy Model

The economy consists of two locations, or regions, where production takes place under different conditions and a single market where output is exchanged.

2.2.1 Demand

On the demand side, there is a representative consumer endowed with income R and with preferences expressed as a Dixit-Stiglitz constant elasticity of substitution utility function defined over a continuum of potential varieties of a good:

$$U = \left[\int_{j \in J} x(j)^{\rho} dj \right]^{1/\rho}$$

where j indexes these varieties and J denotes the set of available varieties; $x(j)$ is the quantity of variety j consumed; ρ is a parameter such that $\rho = (\sigma - 1)/\sigma$, where σ is the elasticity of substitution between any pair of varieties. It is assumed that $\sigma > 1$ (varieties are substitutes) or, equivalently, $0 < \rho < 1$. The elasticity of substitution and the elasticity of the residual demand for any particular variety are equal provided that the set of available varieties J is of non-zero measure.

The representative consumer maximizes utility subject to the budget constraint

$$\int_{j \in J} p(j) x(j) dj = R$$

where $p(j)$ is the price of variety j. Demand for that variety is then given by

$$x(j) = \frac{R}{P} \left(\frac{p(j)}{P} \right)^{-\sigma} \tag{2.1}$$

where P is the price index defined as

$$P = \left[\int_{j \in J} p(j)^{1-\sigma} dj \right]^{1/(1-\sigma)} \tag{2.2}$$

2.2.2 Production

The industry is characterized by monopolistic competition. A continuum of firms produces varieties of the good with an increasing returns to scale technology. Each firm produces a single variety and each variety is produced by a single firm. Production employs a composite factor, labeled z, which may combine, in some proportions, various factors and intermediate inputs such as labor, capital, energy, materials, and so forth. The cost function of firm j located in region m, where $m \in \{u, r\}$ denotes one of the two locations, is linear in output:

$$z(j) = f^m + \frac{y(j)}{\theta(j)}$$

where f^m is the fixed cost of production, which is identical across firms in region m; $y(j)$ is the output of firm j; and $\theta(j)$, where $\theta(j) > 0$, is a parameter representing a firm-specific level of factor z productivity. The cost function expressed in nominal terms is $w^m z(j)$, where w^m is the price of factor z in region m. The supply of factor z is infinitely elastic in each location.

From the standpoint of producers of goods, the advantage of highly urbanized locations comes from various sources that together give rise to external returns to scale or agglomeration economies. These factors were identified by Marshall (1920): large cities provide manufacturers with a greater diversity and quality of workers, suppliers of goods and services, and infrastructure than small cities, towns, and rural areas; they also favor knowledge spillovers. Other factors include natural and man-made amenities and the extent of the market. Yet, the costs of urban congestion due to, for instance, higher commuting and housing costs for workers, and greater competition in the land, labor, goods, and services markets may offset the benefits from locating plants in highly urbanized areas. We assume that the factor price and the fixed cost of production differ between the two locations of the model so as to represent the salient differences in the conditions of production between large cities and less urbanized areas. Henceforth we refer to locations u and r as the urban and rural regions, respectively.

First, we make a hypothesis regarding the factor price difference across regions. Studies document that nominal wages increase with city size (for instance, see Glaeser and Maré 2001) Although the variation in workers' education and skills across locations explains a great deal of this pattern, spatial wage differentials are also due to differences in the cost of living and amenities (Roback 1982; Holden and Wertheimer II 1980), employment and training opportunities (Findeis and Jensen 1998; Phimister et al. 2002), and the sway of unions.[6,7] In developing countries, the surplus of agricultural workforce and weak enforcement of labor legislation such as the minimum wage may also lower labor costs in towns and rural areas relative to those in large urban areas. Moreover, there is evidence that manufacturing plants have sought to lower production labor costs by locating outside of U.S. metropolitan areas (Mack and Schaeffer 1993).[8] Thus, the factor price in the urban region is assumed to be higher than that in the rural region: $w^u > w^r$.

Second, we consider the variation in the fixed cost of production across locations. Firms setting production plants in large urban areas may benefit from positive externalities, and thus reduce the transaction costs associated with essential non-production tasks. In particular, proximity to a large, well-educated, and diverse workforce facilitates the recruitment and retention of highly qualified managers, and professional and technical workers. Manufacturers locating away from urban centers forgo such benefits. The 1996 Rural Manufacturing Survey conducted by the U.S. Department of Agriculture (Gale et al. 1999) indicates that plants located in nonmetro U.S. counties were facing difficulties in attracting managers and professionals, and employing local production workers with the adequate skills to implement new technologies and management practices. In consequence, nonmetro plants are likely to spend more resources on training programs. Branch plants of multi-unit firms usually account for a sizable share of manufacturing activities in predominantly rural areas. Distance between headquarters and branch plants may raise transaction costs associated with the coordination of production, the execution of business and financial plans, and the oversight of plant managers. Manufacturing establishments in small communities are also more likely to be distant from potential suppliers of materials, machinery, and equipment, firms providing business services, and major customers, the access of which would be costly. Gale et al. [1999] report that nonmetro plants procured goods and services from outside their local area to a greater extent compared with metro establishments. nonmetro manufacturers may thus incur higher

[6]For instance, Vera-Toscano et al. 2004 find evidence of a significant urban-rural wage gap of Canadian female workers even after controlling for unobserved heterogeneity, self-selection, and individual-specific and job characteristics.

[7]The absence of the United Auto Workers (UAW) union in the southern states of the US has been a key motive behind the decision of Japanese and German automotive manufacturers to set up factories there instead of in the North, which avoided them paying high wages and pension and health benefits and allowed them to implement more efficient production methods without facing the UAW's work rules.

[8]Erickson [1976] finds that branch plants were more likely to locate in nonmetropolitan counties of Wisconsin from 1969 to 1974 than independent firms due to the low level of competition in local labor markets.

search costs to reach suppliers and customers.[9] The lack of infrastructure, suppliers of transportation services, and services supporting emerging technologies (such as a communication systems) and the slow diffusion of knowledge among workers in small cities and rural areas may hinder technology adoption. Forman et al. [2005]'s findings suggest that U.S. business establishments located in small metro and nonmetro areas faced higher costs to adopt complex Internet applications (electronic business and commerce, and so forth) than those situated in large metro areas, within industries. Furthermore, public infrastructure may be insufficient or inadequate in sparsely populated regions, especially in developing countries. Firms choosing to produce in such places may have to cope with outlays to install power generation and water treatment facilities, telecommunication systems, and so forth.[10] Hence, we assume that the fixed cost of a firm producing in the rural region is higher than that of firm producing in the urban location: $w^u f^u < w^r f^r$.[11]

Lastly, producers incur an iceberg trade cost to deliver their output to the market. We assume that this cost is the same for producers in both regions and normalize it to zero.[12]

As every firm faces a residual demand with constant elasticity σ, independently of its productivity level, it maximizes its profit by imposing a mark-up of price over marginal cost identical to that imposed by all other firms. The price charged by firm

[9] Land and property taxes may or may not be an important source of fixed cost relative to other overheads. If production were intensive in land, land requirements would likely vary with plant productivity and output size. More productive and larger plants would need more land and have a greater incentive to locate where land rents are lower, that is, outside of big cities. In this case, the effect of land costs on firms' location decision would be captured in the model by the factor price gap between the two locations.

[10] World Bank surveys of manufacturing plants in Nigeria, rural China, and rural Indonesia report that many owned generators to make up for the unavailability, or unreliability, of public provision of electricity (Lanjouw and Lanjouw 1995). The purchasing and servicing of generators, and the acquisition of water treatment, waste disposal, and telecommunication facilities represented a large share of their capital stock.

[11] This hypothesis is analogous to that of the Cavailhès et al. [2007] model, in which firms locating in a city's secondary business district pay a higher fixed cost than firms locating in its central business district. Our two hypotheses are akin to those of the Antràs and Helpman [2004] model of offshoring, where the South region's wage is lower than the North's, whereas the fixed cost to produce in the North is smaller. This representation of regional differences in production costs also resembles that of technical change in Ederington and McCalman [2008].

[12] This assumption simplifies the notation. The model does not preclude regional differences in the cost of delivering output to the market. A difference between the iceberg trade costs of the two regions can be accounted for by the factor price gap. Nonetheless, the assumption of equal trade costs is justifiable in a context in which the product market is dispersed over the territory of a country. This context prevails if large cities lie distant from each other amidst less urbanized areas and rural hinterlands, and the territory is well served by a transportation infrastructure.

j with productivity $\theta(j)$ and located in region m is given by

$$p^m(j) = \frac{1}{\rho}\frac{w^m}{\theta(j)} \tag{2.3}$$

The operating profit of the firm is

$$\pi^m(\theta(j)) = \left(p^m(j) - \frac{w^m}{\theta(j)}\right) x(j) - w^m f^m \tag{2.4}$$

By substituting the demand function (4.1) and the mark-up pricing formula (2.3) into (2.4), one obtains the maximal value of profit:

$$\pi^m(\theta(j)) = \frac{r^m(\theta(j))}{\sigma} - w^m f^m \tag{2.5}$$

where $r^m(\theta(j))$, the revenue of the firm, is given by

$$r^m(\theta(j)) = \frac{R\,(P\rho)^{\sigma-1}\,\theta(j)^{\sigma-1}}{(w^m)^{\sigma-1}} \tag{2.6}$$

Thus, a firm with a higher productivity level sets a lower price (according to (2.3)), produces more (see (4.1)), and makes a larger profit (see (2.6) and (2.5)) compared with a firm with lower productivity in a given location.

2.2.3 Firm Entry, Location, and Exit

We assume a dynamic economy with an infinite time horizon. Production and consumption take place in every time period, although consumer preferences, technology, and regional factor prices are constant over time. In each period, some firms enter the industry and some exit. Entrants originate from a continuum of identical "entrepreneurs". These entrepreneurs have to incur a fixed sunk cost of entry of nominal value f_e.[13] The determination of the firm-specific productivity parameter follows the Hopenhayn-Melitz modeling of heterogeneous firms (see Melitz 2003). Upon paying the entry fee, an entrepreneur j randomly draws a productivity level $\theta(j)$ from a distribution given by a continuous p.d.f. g, with support $(0, \infty)$, and a c.d.f. G. Once the entrepreneur—now firm j—learned its productivity level, it decides whether to produce the good and where to locate (in the urban region or the rural region) in the current and future periods, or to opt out of the industry without producing if its productivity level is too low.[14] At any time, firms actually entering in the industry

[13]One may interpret the entry cost as an investment in R&D to learn a production process and a business administration start-up cost.

[14] Our model of location choice is similar to the theory behind the technology choice in Bustos [2005].

are, *a priori*, infinitely lived. However, all incumbent firms may be forced to exit with probability δ in every period.[15]

A firm that has just entered the industry with productivity level θ (we now omit the index j for convenience) chooses from the three available strategies the one that maximizes the expected value of its stream of present and future per-period profits. The value function determines the optimal strategy of the firm with respect to θ:

$$v(\theta) = \max\left\{0, \sum_{t=0}^{\infty} (1-\delta)^t \, \pi^u(\theta), \sum_{t=0}^{\infty} (1-\delta)^t \, \pi^r(\theta)\right\} = \max\left\{0, \frac{\pi^u(\theta)}{\delta}, \frac{\pi^r(\theta)}{\delta}\right\} \tag{2.7}$$

where π^u and π^r are given in (2.5). The firm actually selects the strategy yielding the maximal per-period profit level because its productivity is constant over time. The lowest productivity level, or zero cutoff productivity level, of producing firms is given by $\theta^* = \inf\{\theta : v(\theta) \geq 0\}$. Because $\pi^m(0) = -w^m f^m$ is negative for any region $m \in \{u, r\}$ and the profit functions (2.5) are monotonically increasing in θ, then it must be the case that $\max\{\pi^u(\theta^*), \pi^r(\theta^*)\} = 0$. Thus any firm entering the industry with a productivity level strictly smaller than θ^* will exit without producing.

Given that the rural region exhibits a higher fixed cost but a lower marginal cost than the urban region, two cases arise regarding the distribution of firms between the urban and rural regions along the productivity spectrum. In the first case, $\pi^u(\theta^*) = 0$, $\pi^r(\theta^*) < 0$, and there exists a productivity level $\theta_{u,r}$, which we call the urban-rural cutoff productivity level, such that $\theta_{u,r} = \inf\{\theta : \theta > \theta^* \text{ and } \pi^r(\theta) \geq \pi^u(\theta)\}$. Because π^u and π^r are monotonically increasing, this implies $\pi^u(\theta_{u,r}) = \pi^r(\theta_{u,r})$. Thus any firm entering the industry with a productivity level between θ^* and $\theta_{u,r}$ will maximize its profits by locating in the urban region. Any firm receiving a productivity level above $\theta_{u,r}$ will obtain maximal profits by producing in the rural region. This case is illustrated in Figure 2.1, in which the lines labeled π^u and π^r depict the per-period operating profits associated with the urban and rural locations, respectively, as a function of Θ, where $\Theta \equiv \theta^{\sigma-1}$ is a transformed measure of productivity. The π^r line is steeper than the π^u line because, at a given productivity level, the variable cost of producing one unit of output is lower for firms operating in the rural region. The value of the intercept of the π^r line, however, is below that of the π^u line because the fixed cost of rural firms is higher. Figure 2.1 shows that firms with productivity higher than Θ^* but less than $\Theta_{u,r}$ make greater profits by producing in the urban region. At the productivity level $\Theta_{u,r}$ the urban and rural strategies yield equal

[15] Such an exit may result from an adverse shock due to, for instance, unforeseen changes in market conditions depressing the profits of some firms.

profits. Firms with productivity exceeding $\Theta_{u,r}$ obtain greater profits by producing in the rural region. In the second case, $\pi^u(\theta^*) \leq 0$ and $\pi^r(\theta^*) = 0$, so that when firms are productive enough to stay in the industry, they always find it more profitable to carry out production in the rural region no matter what their productivity level is. We will not analyze the second case further as it boils down to the Melitz [2003] model.

Figure 2.1 highlights an implicit relationship between θ^* and $\theta_{u,r}$. This relationship is derived as follows: On one hand, $\pi^u(\theta^*) = 0$ if and only if $R(\rho P)^{\sigma-1}/\sigma = (w^u)^{\sigma-1} w^u f^u/(\theta^*)^{\sigma-1}$ (using the revenue and profit expressions (2.6) and (2.5)). On the other hand, $\pi^u(\theta_{u,r}) = \pi^r(\theta_{u,r})$ if and only if $(\theta_{u,r})^{\sigma-1} = (w^r f^r - w^u f^u) \times \left[(R(\rho P)^{\sigma-1}/\sigma)[(1/w^r)^{\sigma-1} - (1/w^u)^{\sigma-1}]\right]^{-1}$. Substituting the expression of $R(\rho P)^{\sigma-1}/\sigma$ into that of $(\theta_{u,r})^{\sigma-1}$ and solving for $\theta_{u,r}$ yields

$$\theta_{u,r} = \alpha\theta^*, \text{ where } \alpha \equiv \left[\frac{w^r f^r - w^u f^u}{w^u f^u}\left(\frac{(w^u)^{\sigma-1} - (w^r)^{\sigma-1}}{(w^r)^{\sigma-1}}\right)^{-1}\right]^{1/\sigma-1} \tag{2.8}$$

Note that the condition $\alpha > 1$ guarantees the existence of urban firms. This condition is satisfied when the additional fixed cost of operating in the rural region is large relative to the difference between the urban and rural factor prices.

The shape of the equilibrium distribution of urban and rural firms' productivity levels is given by the exogenous distribution, g, and the probability of actual entry into the industry, $1 - G(\theta^*)$. The exit of incumbents from the industry does not shift this equilibrium distribution because the probability that a firm would be forced to exit, δ, is independent of its productivity level and location by assumption. Thus, the equilibrium productivity distribution is given by the p.d.f. g conditional on entry, with support $(0, \infty)$:

$$\mu(\theta) = \begin{cases} \frac{g(\theta)}{1-G(\theta^*)} & \text{if} \quad \theta \geq \theta^*, \\ 0 & \text{otherwise} \end{cases} \tag{2.9}$$

In addition, we can specify the equilibrium productivity distributions of urban and rural firms separately. For urban firms, the p.d.f. is $g(\theta)/[G(\theta_{u,r}) - G(\theta^*)]$ if $\theta^* \leq \theta < \theta_{u,r}$ and 0 otherwise, where $\theta_{u,r}$ is implicitly a function of θ^* as in (2.8). The probability that an actual entrant would choose the urban location is $q^u = [G(\theta_{u,r}) - G(\theta^*)]/[1 - G(\theta^*)]$, which equals the share of the mass of urban firms. Likewise, the p.d.f. of rural firms is given by $g(\theta)/[1 - G(\theta_{u,r})]$ if $\theta \geq \theta_{u,r}$ and 0 otherwise. The probability that a successful entrant would locate in the rural region is $q^r = [1 - G(\theta_{u,r})]/[1 - G(\theta^*)]$, which is equal to the fraction of rural firms from the total mass of firms. The equilibrium distribution of productivity levels depends on the zero cutoff productivity level, θ^*, which is itself endogenous to the decisions of firms about whether to remain in the industry upon entry and where to locate. Firms are not urban or rural by assumption. They select themselves into being urban or rural depending on where they can achieve the highest operating profits given their exogenous productivity level. The sorting of firms across locations determines, in turn, the average productivity levels in the urban and rural locations. Like in Melitz [2003], we analyze a steady-state equilibrium in which the aggregate variables are constant over time and the location decision of every firm is optimal given these aggregate variables.

2.2.4 Equilibrium Conditions

2.2.4.1 Price Index and Average Productivities

In the steady-state equilibrium, the mass of firms is $M = M^u + M^r$, where M^u and M^r are the masses of urban and rural firms, respectively. The distribution of their productivity levels is given by the function μ. In accordance with (3.2), the price index is given by

$$P = \left[\int_0^{\theta_{u,r}} p^u(\theta)^{1-\sigma} M\mu(\theta)d\theta + \int_{\theta_{u,r}}^{\infty} p^r(\theta)^{1-\sigma} M\mu(\theta)d\theta \right]^{1/(1-\sigma)}$$

After substituting in the above expression the mark-up pricing formula (2.3) and the right-hand side of each of these equalities,

$$\int_0^{\theta_{u,r}} \theta^{\sigma-1}\mu(\theta)d\theta = q^u \frac{1}{G(\theta_{u,r}) - G(\theta^*)} \int_{\theta^*}^{\theta_{u,r}} \theta^{\sigma-1} g(\theta)d\theta$$
$$\int_{\theta_{u,r}}^{\infty} \theta^{\sigma-1}\mu(\theta)d\theta = q^r \frac{1}{1 - G(\theta_{u,r})} \int_{\theta_{u,r}}^{\infty} \theta^{\sigma-1} g(\theta)d\theta$$

and recognizing that q^m equals the share of firms in location m, $s^m \equiv M^m/M$, one can rewrite the price index as

$$P = M^{1/(1-\sigma)} \left[s^u p^u(\tilde{\theta^u}(\theta^*, \theta_{u,r}))^{1-\sigma} + s^r p^r(\tilde{\theta^r}(\theta_{u,r}))^{1-\sigma} \right]^{1/(1-\sigma)} \tag{2.10}$$

where $\tilde{\theta^u}$ and $\tilde{\theta^r}$[16] are weighted averages of urban and rural firms' productivity levels, respectively. They are functions of θ^* defined as

$$\tilde{\theta^u}(\theta^*, \theta_{u,r}) = \left[\frac{1}{G(\theta_{u,r}) - G(\theta^*)} \int_{\theta^*}^{\theta_{u,r}} \theta^{\sigma-1} g(\theta)d\theta \right]^{1/(\sigma-1)} \tag{2.11}$$

[16] The condition for $\tilde{\theta^r}(\theta_{u,r})$ to be finite requires that the $(\sigma - 1)$-th uncentered moment of g be finite.

$$\tilde{\theta^r}(\theta_{u,r}) = \left[\frac{1}{1-G(\theta_{u,r})}\int_{\theta_{u,r}}^{\infty}\theta^{\sigma-1}g(\theta)d\theta\right]^{1/(\sigma-1)} \tag{2.12}$$

As more productive firms can sell goods at lower prices, they capture greater shares of the market than less productive firms. Thus, the former exert a greater weight on the price index. The weights in the expressions of the average productivities account for the disproportionate influence of the more productive firms.

The aggregate profit of urban firms can be written as a function of their average productivity:

$$\Pi^u = \int_{\theta^*}^{\theta_{u,r}} \pi^u(\theta)M\mu(\theta)d\theta = M\left[\frac{R(\rho P)^{\sigma-1}}{\sigma(w^u)^{\sigma-1}}\int_{\theta^*}^{\theta_{u,r}}\theta^{\sigma-1}\mu(\theta)d\theta - w^u f^u\int_{\theta^*}^{\theta_{u,r}}\mu(\theta)d\theta\right]$$

As $\int_{\theta^*}^{\theta_{u,r}}\theta^{\sigma-1}\mu(\theta)d\theta = q^u(\tilde{\theta}^u(\theta^*,\theta_{u,r}))^{\sigma-1}$ and $\int_{\theta^*}^{\theta_{u,r}}\mu(\theta)d\theta = q^u$, we obtain $\Pi^u = M^u\pi^u(\tilde{\theta}^u(\theta^*,\theta_{u,r}))$, which entails $\bar{\pi}^u \equiv \Pi^u/M^u = \pi^u(\tilde{\theta}^u(\theta^*,\theta_{u,r}))$. Hence, the profit of the firm with the average urban productivity equals the average profit of urban firms. Similarly, we can express the aggregate profit of rural firms as a function of their average productivity:

$$\Pi^r = \int_{\theta_{u,r}}^{\infty} \pi^r(\theta)M\mu(\theta)d\theta = M\left[\frac{R(\rho P)^{\sigma-1}}{\sigma(w^r)^{\sigma-1}}\int_{\theta_{u,r}}^{\infty}\theta^{\sigma-1}\mu(\theta)d\theta - w^r f^r\int_{\theta_{u,r}}^{\infty}\mu(\theta)d\theta\right]$$

Given that $\int_{\theta_{u,r}}^{\infty}\theta^{\sigma-1}\mu(\theta)d\theta = q^r(\tilde{\theta}^r(\theta_{u,r}))^{\sigma-1}$ and $\int_{\theta_{u,r}}^{\infty}\mu(\theta)d\theta = q^r$, we have $\Pi^r = M^r\pi^r(\tilde{\theta}^r(\theta_{u,r}))$, which implies $\bar{\pi}^r \equiv \Pi^r/M^r = \pi^r(\tilde{\theta}^r(\theta_{u,r}))$. Hence, the profit of the firm with the average rural productivity equals the average profit of rural firms.

2.2.4.2 Cutoff Profit Conditions

According to (2.6), the ratio of revenues of any two urban firms depends only on the ratio of their productivity levels. In particular, for two firms with productivities $\tilde{\theta}^u(\theta^*,\theta_{u,r})$ and θ^*, we have

$$\frac{r^u(\tilde{\theta}^u(\theta^*,\theta_{u,r}))}{r^u(\theta^*)} = \left(\frac{\tilde{\theta}^u(\theta^*,\theta_{u,r})}{\theta^*}\right)^{\sigma-1} \tag{2.13}$$

It follows that $r^u(\tilde{\theta}^u(\theta^*,\theta_{u,r})) = r^u(\theta^*)(\tilde{\theta}^u(\theta^*,\theta_{u,r})/\theta^*)^{\sigma-1}$. Substituting the expression for $r^u(\tilde{\theta}^u(\theta^*,\theta_{u,r}))$ into (2.5) yields an expression for the profit of the firm with the average urban productivity, or, equivalently, the average profit level of urban firms:

$$\bar{\pi}^u(\theta^*,\theta_{u,r}) = \frac{r^u(\theta^*)}{\sigma}\left(\frac{\tilde{\theta}^u(\theta^*,\theta_{u,r})}{\theta^*}\right)^{\sigma-1} - w^u f^u$$

Furthermore, as $\pi^u(\theta^*) = 0$ entails $r^u(\theta^*) = \sigma w^u f^u$ (see (2.5)), the average urban profit level can be expressed as

$$\bar{\pi}^u(\theta^*, \theta_{u,r}) = w^u f^u \left[\left(\frac{\tilde{\theta^u}(\theta^*, \theta_{u,r})}{\theta^*} \right)^{\sigma-1} - 1 \right] \quad (2.14)$$

where, again, $\theta_{u,r}$ is a function of θ^* as in (2.8). Thus, like the average productivity level $\tilde{\theta^u}$, the average urban profit level depends only on the zero cutoff productivity level θ^*.

The average profit of rural firms, derived in a similar way as the average urban profit (see Appendix A.1.1), is given by

$$\bar{\pi}^r(\theta^*, \theta_{u,r}) = w^u f^u \left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta^*} \right)^{\sigma-1} - \left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{u,r}} \right)^{\sigma-1} \right] + w^r f^r \left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{u,r}} \right)^{\sigma-1} - 1 \right] \quad (2.15)$$

and, likewise, it is entirely determined by θ^*. Henceforth equations (2.14) and (2.15) are referred to as the zero cutoff and urban-rural cutoff profit conditions, respectively.

2.2.4.3 Free Entry Condition

The present values of the average profit flows of urban and rural firms are $\sum_{t=0}^{\infty}(1-\delta)^t\bar{\pi}^u = \bar{\pi}^u/\delta$ and $\sum_{t=0}^{\infty}(1-\delta)^t\bar{\pi}^r = \bar{\pi}^r/\delta$, respectively. Thus, the expected value of future profits net of entry cost, or value of entry, v_e, is expressed as

$$v_e(\theta^*, \theta_{u,r}) = [G(\theta_{u,r}) - G(\theta^*)] \frac{\bar{\pi}^u}{\delta} + [1 - G(\theta_{u,r})] \frac{\bar{\pi}^r}{\delta} - f_e$$

In equilibrium, under the assumption of free entry, the value of entry will be driven to zero because there is an infinite mass of potential firms. Hence, we have the following relationship between the average profits and the cutoff productivity levels:

$$\frac{G(\theta_{u,r}) - G(\theta^*)}{1 - G(\theta^*)} \bar{\pi}^u + \frac{1 - G(\theta_{u,r})}{1 - G(\theta^*)} \bar{\pi}^r = \frac{\delta f_e}{1 - G(\theta^*)} \quad (2.16)$$

The left-hand side of (3.7) is the *ex post* (that is, conditional upon entry) per-period average profit in the industry.

2.2.5 Closed Economy Equilibrium

The two cutoff profit conditions and the free entry condition imply three relationships: first, between θ^* and the average urban profit (see (2.14)); second, between θ^* and the average rural profit (see (2.15)); and third, between θ^*, the average urban profit, and the average rural profit (see (3.7)). As shown in Appendix A.1.2, there exists one and only one combination of productivity and profit values (θ^*,$\bar{\pi}^u$,$\bar{\pi}^r$) that satisfies these three conditions. In addition, the equilibrium zero cutoff productivity level determines, according to (2.8), a unique urban-rural cutoff productivity level, $\theta_{u,r}$.

The masses of urban and rural firms must be constant over time in the steady-state equilibrium. Thus, in each period, there must be a mass M_e of firms incurring the sunk cost of entry and drawing a productivity level, such that the mass of actual entrants in region $m \in \{u, r\}$ matches the mass δM^m of incumbent firms forced to exit from region m:

$$q^u M_e = \delta M^u \quad \text{and} \quad q^r M_e = \delta M^r$$

where q^u and q^r were defined in subsection 2.2.3 as the probabilities of locating in the urban and rural regions, respectively, conditional upon actual entry in the industry. The movement of firms in and out of the industry does not shift the equilibrium distribution of productivity levels because actual entrants' and exiting incumbents' productivities are identically distributed.

The mass of producing firms in region m can be expressed as a function of aggregate consumer spending and the average profits in the two regions.

$$M^m = q^m M = q^m \frac{R}{\bar{r}}$$

where $\bar{r} = \sigma[q^u(\bar{\pi}^u + w^u f^u) + q^r(\bar{\pi}^r + w^r f^r)]$ is the average revenue over all firms. Finally, the equilibrium price index is determined by the masses of urban and rural firms as in (2.10).

To summarize the workings of the closed-economy model, firms choose to settle in either the urban or the rural region on the basis of their productivity level; given the cost structures of these two locations, the more productive firms locate in the rural region. As a consequence, rural firms are more productive, produce more output, and earn a greater profit than urban firms. At first, this result seems to be at odd with the empirical evidence documented in the literature. It is usually found that firms or establishments are more productive and bigger in terms of sales in larger markets, in particular large cities. However, there are differences across sectors and time periods. For instance, Holmes and Stevens [2004] report that the sales of the wholesale trade, finance and insurance, and professional, scientific, and technical services industries are concentrated in bigger cities, in the US.[17] Unlike services, manufacturing output is concentrated in smaller cities. This pattern in the location of manufacturing

[17]Using data from the 1997 Economic Census, they sort Primary Metropolitan Statistical Areas (that is, subdivisions of Metropolitan Statistical Areas), which they called "cities", into three classes: large (population over 2 million), medium (between half a million and 2 million), and small (under

production emerged following the shift of manufacturing activities to smaller cities during the second half of the twentieth century. Furthermore, Holmes and Stevens 2004 report that manufacturing establishments are larger (in terms of sales) in smaller cities whereas services establishments are larger in bigger cities. This pattern is also documented in Holmes and Stevens [2002], along with the strong positive correlation between the size of manufacturing plants and the degree of specialization of a location. Thus, the prediction of our model is consistent with the evidence from the location of manufacturing activities in the US.

2.3 Plant-Level Evidence on Geographical Productivity Differences in Colombia

We now look at the manufacturing sector of Colombia to see whether the implication of the model (more productive firms locate in the rural region) is held up by the facts. Specifically, we compare the productivity level of manufacturing plants located in the large urban areas of Colombia with that of plants found in the rest of the country. We obtain evidence that differences in productivity between metro and nonmetro manufacturers are commonplace, and that plants located in small urban areas are more productive than those established in major cities in some instances.

2.3.1 Trade Policy in Colombia

Colombia's trade policy became much less protective towards imports of manufactured goods between the late 1970s and the early 1990s [Goldberg and Pavcnik, 2005, Fernandes, 2007]. In the post-War period, Colombia pursued a strategy of industrialization through import substitution and, to this end, restrained trade by imposing high import tariffs, prohibiting the importation of some products, and granting import licenses for others. Between 1977 and 1981, while the country was planning to join the General Agreement on Tariffs and Trade and seeking to ensure macroeconomic stability, the Colombian government progressively lowered tariffs and removed items from the lists of prohibited imports and of goods requiring a license. Nevertheless, from 1982 to 1984, in a recessionary context, the government attempted to stem the rising fiscal and current account deficits and support industrial production by tightening commercial policies. Thus, according to the data from the *Departamento Nacional de Planeación* [DNP], the average tariff for industrial goods rose from a

half a million). They use the location quotient for a city class (that is, the share of sales in the cities of a class divided by the population share of that class) to measure the concentration of an industry.

through of 33% in 1981 to a peak of 55% in 1984. In 1985, the Colombian government resumed trade policy reforms, which led to a substantial decrease in tariffs, the elimination of licenses, and the relaxation of administrative barriers to trade. The comparison of manufacturers across geographic areas uses data covering 1981–84 period, when the Colombian economy was still relatively closed to international trade.

2.3.2 Data and Empirical Approach

We use an unbalanced longitudinal database of Colombian manufacturing establishments for the period 1981–84. The data were collected by the *Departamento Administrativo Nacional de Estadística* (DANE) through the annual Colombian census of the manufacturing sector.[18] A four-digit industrial classification (ISIC), similar to that used in the U.S., identifies the industries to which plants belong. This data set contains information about plant characteristics such as employment and labor costs, the capital stock, energy consumption, raw materials and intermediate goods utilization, the value of production, and so forth.[19]

The empirical investigation relies on the distinction between metropolitan (metro) and nonmetropolitan (nonmetro) areas of Colombia. The Colombian territory is comprised of cities of different population sizes, of sparsely populated, rural areas, and of unsettled land. Metro areas are the largest cities and conurbations.[20, 21] Nonmetro areas make up the rest of the territory, including medium-size and small cities (henceforth, we shall refer to nonmetro areas as small cities or small urban areas). Metro and nonmetro areas are the counterparts of the model's urban and rural region, respectively.

The bulk of industrial activities takes place in metro areas. Thus, 84.6% of the 6258 establishments recorded in the 1984 census were located in metro areas. That year, manufacturing real production tallied 108.5 billion 1980 pesos, 77.5% of which originated from metro plants; metro manufacturers generated 50.4% of the 6.4 billion 1980 pesos of real export sales; and metro plants employed 81.2% of the 463,581 manufacturing workers. These aggregate numbers conceal substantial variation across

[18] Mark Roberts provided the data. See Roberts [1996] for a comprehensive description of the data.

[19] Note that the 1983 and 1984 censuses excluded the plants with less than ten employees.

[20] Metropolitan areas are defined as administrative entities consisting of a cluster of two or more municipalities integrated around an urban center and linked by dense transport systems and strong economic relationships, and engaging in a coordinated administration of public services and the joint planning and implementation of economic and social policies [DANE].

[21] Throughout the period 1981-84, the metro areas of Colombia were: Bogota D.E., Soacha; Cali, Yumbo; Medellin, Valle de Aburra; Manizales, Villamaria; Barranquilla, Soledad; Bucaramanga, Giron, Floridablanca; Pereira, Santa Rosa de Cabal, Dosquebradas; and Cartagena.

industries (see Table 2.1 reporting summary statistics for variables defined at the four-digit-level of the ISIC). Hence, on average, over all industries, 83.5% of plants in an industry are located in metro areas, with a standard deviation of 18.6%. The smallest and the largest percentages of metro plants in an industry are 18.9% and 100%, respectively. The location quotient of metro areas on the basis of production (that is, the within-industry share of metro production) has a mean of 84.1% and a standard deviation of 21.6%, and varies between 0.3% and 100%. The location quotient of metro areas on the basis of export sales averages out to 79.2%. The size quotient of metro plants with respect to production (or, size of the average metro plant relative to that of the typical plant in the industry) is 1.012, on average, with a standard deviation of 0.272, and is comprised between 0.009 and 1.667. With regard to export sales, the mean size quotient of metro plants is 0.948. This suggests that nonmetro plants export slightly more than metro plants.

For each three-digit-level industry, we estimate a production function and compare productivity estimates between metro and nonmetro plants. As in the productivity studies by Olley and Pakes [1996], Pavcnik [2002], Levinsohn and Petrin [2003], and [Fernandes, 2007], we assume that every producer j in a perfectly competitive industry i employ a Cobb-Douglas technology; and we obtain the following econometric model of production function:

$$y^i_{j,t} = \beta_0 + \beta_l l^i_{j,t} + \beta_k k^i_{j,t} + \beta_e e^i_{j,t} + \beta_m m^i_{j,t} + \omega^i_{j,t} + \eta^i_{j,t} \tag{2.17}$$

where $y^i_{j,t}$ is the logarithm of plant j's output in year t; $l^i_{j,t}$ is the log of the labor input; $k^i_{j,t}$, the log of its capital stock; $e^i_{j,t}$ and $m^i_{j,t}$, the logs of energy consumption and of intermediate inputs, respectively. The output level also depends on a plant-year-specific stochastic error term, $\omega^i_{j,t} + \eta^i_{j,t}$. The term $\omega^i_{j,t}$, which varies over time, is known (or predictable) to the plant's decision maker—but unobservable by anyone else—when he or she decides on factor input levels; as a consequence, it is contemporaneously correlated with the demand for inputs. An estimator of the production function that would not take this correlation into account (the ordinary least squares (OLS) estimator for instance) would produce biased and inconsistent estimates. This issue is commonly referred to as the simultaneity bias. The other part of the error term, $\eta^i_{j,t}$, embodies a random productivity shock materializing after the choice of inputs, as well as measurement errors. The sum of $\omega^i_{j,t}$ and $\eta^i_{j,t}$ stands for total factor productivity (TFP).

In the theoretical model, firms select one region or the other depending on their productivity level because of the trade-off involved between fixed cost and factor cost.

As a result, high-productivity firms locate in the high-fixed-cost and low-factor-cost, rural region. Hence, that model suggests that the production function of metro plants differ from that of nonmetro plants. To parameterize the difference between the metro and the nonmetro production functions, we include an indicator variable of location in the regression model, as follows:

$$\begin{aligned} y^i_{j,t} = \beta_0 + \beta_l l^i_{j,t} + \beta_k k^i_{j,t} + \beta_e e^i_{j,t} + \beta_m m^i_{j,t} + \tilde{\beta}_l metro_j \times l^i_{j,t} \\ + \tilde{\beta}_k metro_j \times k^i_{j,t} + \tilde{\beta}_e metro_j \times e^i_{j,t} + \tilde{\beta}_m metro_j \times m^i_{j,t} + \omega^i_{j,t} + \eta^i_{j,t} \end{aligned} \quad (2.18)$$

where the time-invariant variable $metro_j$ equals one if plant j is located in a metro area and zero otherwise.[22] The coefficients of the terms of interaction between $metro_j$ and inputs measure factor productivity differences between metro and nonmetro plants.

Various methods have been developed to deal with the simultaneity bias and with other problems that plague the estimation of production functions (selection bias, imperfect competition, and so forth). For instance, the approaches of Olley and Pakes [1996] and Levinsohn and Petrin [2003] have proved effective in estimating production functions like (2.17). However, applying their methods to our modified version of the production function is beyond the scope of this section. So, instead, we estimate (2.18) in a simpler way. Because the period we consider, 1981–84, is relatively short, plants are unlikely to experience large shifts in productivity. For this reason, we assume that their unobserved productivity, $\omega^i_{j,t}$, is time invariant, that is, $\omega^i_{j,t} \equiv \omega^i_j$. With this assumption, the fixed-effect estimator consistently estimates (2.18) even if the regressors are correlated with ω^i_j. Specifically, we use the within estimator, or OLS estimator applied to the mean-differenced data. This estimator also addresses the selection bias caused by the exit of plants during the period covered by the data set, as long as exit decisions are determined by the level of ω^i_j [Ackerberg et al., 2007].

2.3.3 Regression Results

Production function (2.18) is separately estimated for twenty-nine three-digit-level industries. Most industry panels are unbalanced because of plants entering, exiting, and switching industries during the period 1981–84. We add year fixed effects to the specification and use standard errors robust to heteroskedasticity and serial

[22] See Appendix A.5.1 for details on this variable and the construction of the output and factor inputs variables.

correlation within plants (cluster-robust standard errors) to make inferences about the parameters.[23] Table 2.2 reports, for the twenty-nine industries, the number of occurrences where the estimated coefficients of the interaction terms are statistically insignificant, significantly positive, and significantly negative at the 5% level. A significantly positive coefficient estimate entails that metro plants have a higher estimated factor productivity than nonmetro plants. In addition, for each industry, we recover estimates of ω_j^i and then regress it on the metro variable. From the coefficient estimates for $metro_j$ and their robust standard errors obtained from this regression, for all industries, Table 2.2 shows the frequency according to which the metro variable's estimated coefficient is statistically insignificant, significantly positive, and significantly negative at the 5% level. This allows us to compare metro and nonmetro plants with respect to TFP.

In some industries, metro and nonmetro plants exhibit different levels of labor, capital, energy, or intermediate inputs productivity, or even TFP. The metro-labor interaction term's estimated coefficient is insignificant in 24 out of the 29 industries. However, this coefficient is positive and significant at the 5% level in four industries, meaning that metro manufacturers in these industries exhibit a higher labor productivity than their nonmetro competitors'. And in one industry, nonmetro plants have a higher labor productivity compared with metro producers. Results for the metro-capital interaction term are similar to those obtained for the previous variable. In five (one) industries, metro establishments employ capital more (less) productively than their nonmetro peers. Thus, metro plants have superior labor and capital productivities compared with nonmetro plants more often that the opposite. We observe a reverse pattern with respect to the energy and intermediate inputs factors. In the case of energy, nonmetro plants use this factor more (less) productively than metro plants in four (three) industries. Nonmetro plants are also more (less) productive in the use of intermediate inputs in six (four) industries. Concerning TFP, the bottom line in table shows that the null hypothesis that metro and nonmetro plants are equally productive cannot be rejected in six industries; metro plants have an estimated TFP higher than that of nonmetro plants in 14 industries; and nonmetro plants are estimated to be more productive in nine industries.

These results only partially support the theoretical model and its prediction that more productive firms locate in the rural region. Nonmetro plants are more productive than their metro rivals only in some industries. In other industries, metro plants are more productive. And in many industries, there are not any productivity

[23] We use the command `xtreg` *depvar indepvars*, `fe cluster`(*varname*) of the Stata software.

differences between these two types of plants. These different outcomes may be due to the fact that the factors influencing the location choice vary among industries. For instance, one would expect unskilled-labor intensive industries or industries intensive in the use of agricultural raw commodities to weigh location attributes differently compared with skilled-labor intensive or high-technology industries. Thus, the model may realistically represent plants' location decision problem for only a subset of industries. We acknowledge this limitation of the model before proceeding to the open economy analysis.

2.4 Open Economy Model

In this section, our goal is to characterize the equilibrium when the country described in the closed economy section has the opportunity to trade with another country. Henceforth the former is referred to as the domestic country (d) and the latter is called the foreign country (f). These two countries are symmetric. In particular, the urban and rural prices of the factor and the fixed costs in the foreign country are identical to those in the domestic country. In each country, firms set the price of output sold in their home market according to the same rule as in the closed economy model. For instance, a domestic firm with productivity θ and located in region $m \in \{u, r\}$ sets the price of output sold in the domestic market at $p_d^m(\theta) = w^m/\rho\theta$ and the revenue it garners from domestic sales is

$$r_d^m(\theta) = \frac{R\,(\rho P)^{\sigma-1}\,\theta^{\sigma-1}}{(w^m)^{\sigma-1}} \tag{2.19}$$

Firms incur both a per-unit trade cost and a fixed cost[24] to sell output abroad. This per-unit trade cost takes the form of an iceberg trade cost, so that a domestic firm must ship $\tau > 1$ units of output to deliver one unit to the foreign market. Thus, the mark-up pricing rule applied by a domestic firm for output sold in the foreign market is given by $p_f^m(\theta) = \tau w^m/\rho\theta = \tau p_d^m(\theta)$. The firm's revenue from export sales is

$$r_f^m(\theta) = \frac{R\,(\rho P)^{\sigma-1}\,\theta^{\sigma-1}}{(\tau w^m)^{\sigma-1}} = \tau^{1-\sigma} r_d^m(\theta) \tag{2.20}$$

[24] To enter a foreign market, a firm must incur search and information costs associated with seeking foreign partners and customers, marketing costs such as the cost of establishing a distribution network, costs of meeting local regulatory constraints, and other possible costs associated with doing business abroad. Previous studies (see for example Roberts and Tybout [1997] and Clerides et al. [1998]) have shown that manufacturing firms entering export markets have to make significant outlays that do not depend on the volume of their exports.

where R and P, the aggregate expenditure and price index in the foreign country, respectively, are equal to R and P in the domestic country as both countries are identical and trade must be balanced. Hence, the total revenue of an exporting firm located in region m is $r^m(\theta) = r_d^m(\theta) + r_f^m(\theta) = (1+\tau^{1-\sigma})r_d^m(\theta)$. The total revenue of a firm serving only the domestic market is just the revenue from its domestic sales, $r^m(\theta) = r_d^m(\theta)$.

2.4.1 Firm Entry, Location, and Export Market Participation

The conditions of entry and exit are the same as in the closed economy model. In particular, firms entering the industry draw their productivity level at random from the distribution g. Upon learning their productivity level, firms decide whether they will export goods to the foreign country (firms can accurately foresee their future foreign sales), and simultaneously select their location. In order to export, firms must pay a periodic fixed cost of a nominal value of f_{ex}.[25] This per-period fixed export cost is the same for urban and rural producers. Producers supplying goods to the foreign market are not exempt from the fixed production cost. Any exporting firm maximizes its profits by also serving the domestic market (as the domestic revenue r_d^m is positive for any firm remaining in the industry). Thus, the profit of an exporting firm producing in region m with productivity θ, denoted by $\pi^m(\theta)$, will be equal to $[r_d^m(\theta) + r_f^m(\theta)]/\sigma - w^m f^m - f_{ex}$. However, it will be convenient for the subsequent analysis to write it as $\pi_d^m(\theta) + \pi_f^m(\theta)$, where

$$\pi_d^m(\theta) = \frac{r_d^m(\theta)}{\sigma} - w^m f^m \tag{2.21}$$

$$\pi_f^m(\theta) = \frac{r_f^m(\theta)}{\sigma} - f_{ex} = \frac{\tau^{1-\sigma} r_d^m(\theta)}{\sigma} - f_{ex} \tag{2.22}$$

We will refer to $\pi_d^m(\theta)$ as the domestic profit because it accounts for domestic sales earnings and $\pi_f^m(\theta)$ will be the export profit because it reflects revenues from foreign sales.

In the open economy setup, there are four possible strategies available to a successful entrant: it can be an urban firm supplying goods to the domestic market, an urban exporter, a producer in the rural region that sells goods exclusively to domestic consumers, or a rural exporter. A firm will choose the strategy that yields the highest expected value of future profits. The value function of a firm with productivity θ is

[25] As explained in Melitz [2003], it is equivalent for a firm, in terms of resource expenditure, to incur a one-time fixed cost in the initial period that enables the firm to export in all periods, and to spread the fixed cost of exporting evenly over time in such a way that the discounted value of the sum of the periodic payments is equal to the one-time payment.

$v(\theta) = \max\left\{\pi_d^u(\theta)/\delta, [\pi_d^u(\theta) + \pi_f^u(\theta)]/\delta, \pi_d^r(\theta)/\delta, [\pi_d^r(\theta) + \pi_f^r(\theta)]/\delta\right\}$. The zero cutoff productivity level is determined by $\theta^* = \inf\{\theta : v(\theta) \geq 0\}$. As in the closed economy model, in the following discussion we assume that the fixed cost in the urban region is sufficiently low relative to the factor price so that low-productivity firms locate there. In particular, it warrants that $\pi^u(\theta^*) = 0$ and $\pi^r(\theta^*) < 0$. Moreover, we suppose that the fixed and variable trade costs, f_{ex} and τ, are sufficiently high so that low-productivity firms find it unprofitable to export. However, it remains ambiguous whether the first exporting firms (along the productivity line) will be urban or rural. The lowest productivity level at which it is profitable to export, or the export cutoff productivity level, is defined as

$$\theta_{ex} = \inf\left\{\theta : \theta \geq \theta^* \text{ and } \max\left\{\pi_d^u(\theta) + \pi_f^u(\theta), \pi_d^r(\theta) + \pi_f^r(\theta)\right\} \geq \max\left\{\pi_d^u(\theta), \pi_d^r(\theta)\right\}\right\} \tag{2.23}$$

In addition, the urban-rural cutoff productivity is defined as

$$\theta_{u,r} = \inf\left\{\theta : \theta \geq \theta^* \text{ and } \max\left\{\pi_d^r(\theta), \pi_d^r(\theta) + \pi_f^r(\theta)\right\} \geq \max\left\{\pi_d^u(\theta), \pi_d^u(\theta) + \pi_f^u(\theta)\right\}\right\} \tag{2.24}$$

We can distinguish between two cases: (a) $\theta_{ex} < \theta_{u,r}$; and (b) $\theta_{ex} > \theta_{u,r}$. In case (a), when moving along the productivity line to the right, one first encounters urban exporters before coming across rural firms of any kind (non-exporters or exporters). Firms with productivity levels superior to, and in the vicinity of $\theta_{u,r}$, could *a priori* be either rural non-exporters or rural exporters. The following lemma rules out the possibility to observe the former type of firms in this case.

Lemma 1. *If urban exporters operate below a given level of productivity $\theta_{u,r}$, where $\theta_{ex} < \theta_{u,r}$, then rural non-exporters cannot operate above $\theta_{u,r}$; only rural exporters can.*

Proof If rural non-exporters directly follow urban exporters, then according to (2.24), it implies that $\pi_f^u(\theta_{u,r}) > 0$, $\pi_f^r(\theta_{u,r}) < 0$, and $\pi_d^r(\theta_{u,r}) \geq \pi_d^u(\theta_{u,r}) + \pi_f^u(\theta_{u,r})$. However, we know from (2.20) that $\pi_f^r(\theta) > \pi_f^u(\theta)\ \forall\ \theta \in \Re^+$ given that $w^r < w^u$. It is true in particular for $\theta = \theta_{u,r}$. Hence, it must be the case that $\pi_f^r(\theta_{u,r}) > 0$, which contradicts the premise and implies that firms with productivities greater than $\theta_{u,r}$ are rural exporters. QED

Thus, case (a) is characterized by the succession along the productivity line of urban non-exporters, urban exporters, and rural exporters.

In case (b), when moving along the productivity line to the right, we observe rural non-exporters before coming across exporting firms of any sort (urban or rural).

Firms with productivity levels superior to, and in the vicinity of θ_{ex}, could *a priori* be either urban exporters or rural exporters. The following lemma shows that rural non-exporters cannot be superseded by urban exporters.

Lemma 2. *If rural non-exporters operate below a given level of productivity θ_{ex}, where $\theta_{ex} > \theta_{u,r}$, then urban exporters cannot operate above θ_{ex}; only rural exporters can.*

Proof If urban exporters directly follow rural non-exporters, then according to (2.23), $\pi_d^r(\theta_{ex}) > \pi_d^u(\theta_{ex})$, $\pi_d^u(\theta_{ex})+\pi_f^u(\theta_{ex}) > \pi_d^r(\theta_{ex})+\pi_f^r(\theta_{ex})$, and $\pi_d^u(\theta_{ex})+\pi_f^u(\theta_{ex}) \geq \pi_d^r(\theta_{ex})$. However, we know that $\pi_f^r(\theta) > \pi_f^u(\theta)\ \forall\ \theta \in \Re^+$ as $w^r < w^u$. This is also true for $\theta = \theta_{ex}$. Hence, it must be the case that $\pi_d^r(\theta_{ex})+\pi_f^r(\theta_{ex}) > \pi_d^u(\theta_{ex})+\pi_f^u(\theta_{ex})$, which contradicts the premise. QED

Thus, case (b) is characterized by the succession along the productivity line of urban non-exporters, rural non-exporters, and rural exporters. Note that, in the special case where $\theta_{ex} = \theta_{u,r}$, firms with productivity levels inferior to the common value of θ_{ex} and $\theta_{u,r}$ are urban non-exporters, and firms with productivities superior to this level are rural exporters. In what follows, we focus on cases (a) and (b) as they are more interesting to analyze.

2.4.2 Relationships Among the Cutoff Productivity Levels

2.4.2.1 Case (a): Urban Non-exporters, Urban Exporters, and Rural Exporters

Using the domestic revenue and profit expressions (2.19) and (2.21) we rewrite $\pi_d^u(\theta^*) = 0$ as $R(\rho P)^{\sigma-1}/\sigma = w^u f^u (w^u)^{\sigma-1}/(\theta^*)^{\sigma-1}$. In addition, $\pi_f^u(\theta_{ex}) = 0$ if and only if $(\theta_{ex})^{\sigma-1} = f_{ex}(\tau w^u)^{\sigma-1}(R(\rho P)^{\sigma-1}/\sigma)^{-1}$ (because of the export revenue and profit expressions (2.20) and (2.22)). Substituting the expression of $R(\rho P)^{\sigma-1}/\sigma$ into that of $(\theta_{ex})^{\sigma-1}$ and solving for θ_{ex} yields

$$\theta_{ex} = \eta\theta^*, \text{ where } \eta \equiv \tau\left(\frac{f_{ex}}{w^u f^u}\right)^{1/\sigma-1} \tag{2.25}$$

The condition $\eta > 1$ ensures the existence of urban non-exporters. The derivation of the relationship between θ^* and $\theta_{u,r}$ is analogous to that in the closed economy. We substitute the zero cutoff profit condition, $\pi_d^u(\theta^*) = 0$, into the urban-rural cutoff profit condition, $\pi^u(\theta_{u,r}) = \pi^r(\theta_{u,r})$, and solve for $\theta_{u,r}$. We obtain

$$\theta_{u,r} = \gamma\theta^*, \text{ where } \gamma \equiv \frac{1}{(1+\tau^{1-\sigma})^{1/(\sigma-1)}}\left[\frac{w^r f^r - w^u f^u}{w^u f^u}\left(\frac{(w^u)^{\sigma-1} - (w^r)^{\sigma-1}}{(w^r)^{\sigma-1}}\right)^{-1}\right]^{1/\sigma-1} \tag{2.26}$$

Note that the conditions $\eta > 1$ and $\gamma/\eta > 1$ together guarantee the existence of both urban non-exporters and urban exporters. The former condition means that f_{ex} must be large relative to $w^u f^u$. The second condition imposes that f_{ex} be relatively small with respect to $w^r f^r - w^u f^u$ for given values of τ, w^u, and w^r. Hence, case (a) is likely to arise when the fixed cost of exporting is relatively small compared with the difference between the fixed operating cost in the rural region and that in the urban region.

2.4.2.2 Case (b): Urban Non-exporters, Rural Non-exporters, and Rural Exporters

The relationship between θ^* and $\theta_{u,r}$ is the same as in the closed economy, that is, $\theta_{u,r} = \alpha\theta^*$, where α is defined in (2.8). The relationship between θ^* and θ_{ex} is obtained as follows: the export profit cutoff condition, $\pi_f^r(\theta_{ex}) = 0$, entails $(\theta_{ex})^{\sigma-1} = f_{ex}(\tau w^r)^{\sigma-1}(R(\rho P)^{\sigma-1}/\sigma)^{-1}$ (from (2.20) and (2.22)). Then, we substitute in this equality the same expression of $R(\rho P)^{\sigma-1}/\sigma$ as in case (a) and solve for θ_{ex}, which yields

$$\theta_{ex} = \beta\theta^*, \text{ where } \beta \equiv \tau\frac{w^r}{w^u}\left(\frac{f_{ex}}{w^u f^u}\right)^{1/\sigma-1} \tag{2.27}$$

Both urban and rural non-exporters are present in the industry provided that $\alpha > 1$ and $\beta/\alpha > 1$. The latter condition says that, unlike case (a), f_{ex} must be large relative to $w^r f^r - w^u f^u$, for given values of τ, w^u, and w^r. Thus, case (b) is likely to prevail as long as the fixed cost of exporting remains large relative to the additional fixed cost of producing in the rural region over that of operating in the urban region. The difference between the two cases is illustrated in figures 2.2 and 2.3.

2.4.3 Open Economy Equilibrium—Case (a): Urban Non-exporters, Urban Exporters, and Rural Exporters

2.4.3.1 Price Index and Average Productivities

The distribution of incumbent firms' productivity levels in equilibrium, μ, is defined as in (2.9). The probabilities that firms entering the industry and locating in the urban and rural regions would export are $q_{ex}^u = [G(\theta_{u,r}) - G(\theta_{ex})]/[G(\theta_{u,r}) - G(\theta^*)]$ and $q_{ex}^r = 1$, respectively. The masses of exporting firms in regions u and r amount to $M_f^u = q_{ex}^u M^u$ and $M_f^r = M^r$, respectively.[26] According to (3.2), the price index is

[26]The mass of varieties sold in any country, originating from its local producers and from abroad, is given by $M' = M + M_f^u + M_f^r$.

given by

$$P = \left[\int_0^{\theta_{u,r}} p_d^u(\theta)^{1-\sigma} M\mu(\theta)d\theta + \int_{\theta_{ex}}^{\theta_{u,r}} p_f^u(\theta)^{1-\sigma} M\mu(\theta)d\theta \right.$$
$$\left. + \int_{\theta_{u,r}}^{\infty} p_d^r(\theta)^{1-\sigma} M\mu(\theta)d\theta + \int_{\theta_{u,r}}^{\infty} p_f^r(\theta)^{1-\sigma} M\mu(\theta)d\theta \right]^{1/(1-\sigma)}$$

Like in the closed economy, the price index can be rewritten as a function of average productivities:

$$P = M^{1/(1-\sigma)} \left[s^u \left(p_d^u(\tilde{\theta^u}(\theta^*, \theta_{u,r}))^{1-\sigma} + q_{ex}^u p_f^u(\tilde{\theta^u}(\theta_{ex}, \theta_{u,r}))^{1-\sigma} \right) \right.$$
$$\left. + s^r \left(p_d^r(\tilde{\theta^r}(\theta_{u,r}))^{1-\sigma} + p_f^r(\tilde{\theta^r}(\theta_{u,r}))^{1-\sigma} \right) \right]^{1/(1-\sigma)} \tag{2.28}$$

where $\tilde{\theta^u}$ and $\tilde{\theta^r}$ are defined as in (2.11) and (2.12), respectively. Note that $\tilde{\theta^u}(\theta^*, \theta_{u,r})$ is the average productivity over the population of domestic urban firms, and $\tilde{\theta^u}(\theta_{ex}, \theta_{u,r})$ is the average productivity of domestic urban exporters alone.

2.4.3.2 Equilibrium Conditions and Determination

The average revenue of urban firms received from domestic sales can be expressed as $r_d^u(\tilde{\theta^u}(\theta^*, \theta_{u,r})) = r_d^u(\theta^*)(\tilde{\theta^u}(\theta^*, \theta_{u,r})/\theta^*)^{\sigma-1}$. Like in the closed economy, one can substitute the expression for $r_d^u(\tilde{\theta^u}(\theta^*, \theta_{u,r}))$ into the domestic profit function defined in (2.21) to obtain the average domestic profit level over all urban firms:

$$\bar{\pi}_d^u(\theta^*, \theta_{u,r}) = \frac{r_d^u(\theta^*)}{\sigma} \left(\frac{\tilde{\theta^u}(\theta^*, \theta_{u,r})}{\theta^*} \right)^{\sigma-1} - w^u f^u$$

The zero cutoff profit condition, $\pi_d^u(\theta^*) = 0$, which entails $r_d^u(\theta^*) = \sigma w^u f^u$, is thus equivalent to the following relationship between the average urban domestic profit and θ^*:

$$\bar{\pi}_d^u(\theta^*, \theta_{u,r}) = w^u f^u \left[\left(\frac{\tilde{\theta^u}(\theta^*, \theta_{u,r})}{\theta^*} \right)^{\sigma-1} - 1 \right]$$

where $\theta_{u,r}$ is a function of θ^* as in (2.26). Similarly, the average revenue of urban exporting firms from foreign sales is given by $r_f^u(\tilde{\theta^u}(\theta_{ex}, \theta_{u,r})) = r_f^u(\theta_{ex})(\tilde{\theta^u}(\theta_{ex}, \theta_{u,r})/\theta_{ex})^{\sigma-1}$. By substituting the expression for $r_f^u(\tilde{\theta^u}(\theta_{ex}, \theta_{u,r}))$ into the export profit function defined in (2.22), one obtains the average export profit level of all urban exporters:

$$\bar{\pi}_f^u(\theta_{ex}, \theta_{u,r}) = \frac{r_f^u(\theta_{ex})}{\sigma} \left(\frac{\tilde{\theta^u}(\theta_{ex}, \theta_{u,r})}{\theta_{ex}} \right)^{\sigma-1} - f_{ex}$$

Then, the export cutoff profit condition, $\pi_f^u(\theta_{ex}) = 0$, or, $r_f^u(\theta_{ex}) = \sigma f_{ex}$, implies the following implicit relationship between the average urban export profit and θ^*:

$$\bar{\pi}_f^u(\theta_{ex}, \theta_{u,r}) = f_{ex}\left[\left(\frac{\tilde{\theta^u}(\theta_{ex}, \theta_{u,r})}{\theta_{ex}}\right)^{\sigma-1} - 1\right]$$

where θ_{ex} is a function of θ^* as in (2.25). Given that q_{ex}^u is the fraction of exporters among urban firms, the average profit over all urban firms is

$$\bar{\pi}^u(\theta^*, \theta_{ex}, \theta_{u,r}) = w^u f^u\left[\left(\frac{\tilde{\theta^u}(\theta^*, \theta_{u,r})}{\theta^*}\right)^{\sigma-1} - 1\right] + q_{ex}^u f_{ex}\left[\left(\frac{\tilde{\theta^u}(\theta_{ex}, \theta_{u,r})}{\theta_{ex}}\right)^{\sigma-1} - 1\right] \tag{2.29}$$

The above equation is the zero cutoff profit condition in the open economy.

The urban-rural cutoff profit condition in the open economy (see Appendix A.2.1 for its derivation) also relates the average profit over all rural firms to θ^*:

$$\begin{aligned}\bar{\pi}^r(\theta^*, \theta_{ex}, \theta_{u,r}) = w^u f^u &\left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta^*}\right)^{\sigma-1} - \left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{u,r}}\right)^{\sigma-1}\right] \\ &+ w^r f^r\left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{u,r}}\right)^{\sigma-1} - 1\right] + f_{ex}\left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{ex}}\right)^{\sigma-1} - 1\right]\end{aligned} \tag{2.30}$$

The net expected value of entry is $v_e(\theta^*, \theta_{u,r}) = [G(\theta_{u,r}) - G(\theta^*)]\bar{\pi}^u/\delta + [1 - G(\theta_{u,r})]\bar{\pi}^r/\delta - f_e$. The free entry condition, that is, $v_e(\theta^*, \theta_{u,r}) = 0$, holds if and only if

$$[G(\theta_{u,r}) - G(\theta^*)]\,\bar{\pi}^u + [1 - G(\theta_{u,r})]\,\bar{\pi}^r = \delta f_e \tag{2.31}$$

Equations (2.29), (2.30), and (2.31) define the open economy equilibrium conditions. We show in Appendix A.2.2 that these three conditions determine a unique equilibrium $(\theta^*, \bar{\pi}^u, \bar{\pi}^r)$. Furthermore, the equilibrium zero cutoff productivity level determines, according to (2.25) and (2.26), unique export cutoff level, θ_{ex}, and urban-rural cutoff level, $\theta_{u,r}$, respectively.

Similarly to the closed economy, the mass of firms in region m is determined by the aggregate consumer expenditure and the average profits in the two regions:

$$M^m = q^m \frac{R}{\bar{r}} \tag{2.32}$$

where $\bar{r} = \sigma[q^u(\bar{\pi}^u + w^u f^u + q_{ex}^u f_{ex}) + q^r(\bar{\pi}^r + w^r f^r + q_{ex}^r f_{ex})]$, q_{ex}^u is the probability that an urban firm export, and $q_{ex}^r = 1$. Lastly, the equilibrium price index is determined by the masses of urban and rural firms and q_{ex}^u as in (2.28).

2.4.4 Open Economy Equilibrium—Case (b): Urban Non-exporters, Rural Non-exporters, and Rural Exporters

2.4.4.1 Price Index and Average Productivities

We now consider the case in which we observe urban non-exporters, rural non-exporters, and rural exporters. The probabilities that firms will export, conditional on their entry and location in the urban and rural regions, are $q_{ex}^u = 0$ and $q_{ex}^r = [1 - G(\theta_{ex})]/[1 - G(\theta_{u,r})]$, respectively. The masses of exporting firms are zero in region u and $M_f^r = q_{ex}^r M^r$ in region r.[27] The price index is

$$P = \left[\int_0^{\theta_{u,r}} p_d^u(\theta)^{1-\sigma} M\mu(\theta)d\theta + \int_{\theta_{u,r}}^{\infty} p_d^r(\theta)^{1-\sigma} M\mu(\theta)d\theta + \int_{\theta_{ex}}^{\infty} p_f^r(\theta)^{1-\sigma} M\mu(\theta)d\theta \right]^{1/(1-\sigma)}$$

We can also express the price index as a function of average productivities:

$$P = M^{1/(1-\sigma)} \left[s^u p_d^u(\tilde{\theta^u}(\theta^*, \theta_{u,r}))^{1-\sigma} + s^r \left(p_d^r(\tilde{\theta^r}(\theta_{u,r}))^{1-\sigma} + q_{ex}^r p_f^r(\tilde{\theta^r}(\theta_{ex}))^{1-\sigma} \right) \right]^{1/(1-\sigma)} \tag{2.33}$$

where $\tilde{\theta^r}(\theta_{ex})$ is the average productivity of domestic rural exporters alone.

2.4.4.2 Equilibrium Conditions and Determination

The zero cutoff profit condition is like (2.14) because not one urban firm exports, :

$$\bar{\pi}^u(\theta^*, \theta_{u,r}) = w^u f^u \left[\left(\frac{\tilde{\theta^u}(\theta^*, \theta_{u,r})}{\theta^*} \right)^{\sigma-1} - 1 \right] \tag{2.34}$$

where $\theta_{u,r}$ is a function of θ^* as in (2.8). The urban-rural cutoff profit condition (see Appendix A.3.1) is given by

$$\begin{aligned} \bar{\pi}^r(\theta^*, \theta_{u,r}, \theta_{ex}) = w^u f^u & \left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta^*} \right)^{\sigma-1} - \left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{u,r}} \right)^{\sigma-1} \right] \\ & + w^r f^r \left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{u,r}} \right)^{\sigma-1} - 1 \right] + q_{ex}^r f_{ex} \left[\left(\frac{\tilde{\theta^r}(\theta_{ex})}{\theta_{ex}} \right)^{\sigma-1} - 1 \right] \end{aligned} \tag{2.35}$$

The expression of the net expected value of entry is the same as in case (a), and thus the free entry condition is identical to (2.31). This condition (2.31) together

[27] The mass of varieties sold in any country equals $M' = M + M_f^r$.

with (2.34) and (2.35) characterizes the open economy equilibrium in case (b). The existence and uniqueness of the equilibrium $(\theta^*, \bar{\pi}^u, \bar{\pi}^r)$ is established in Appendix A.3.2. The equilibrium zero cutoff productivity level θ^*, in turn, identifies unique urban-rural and export cutoff levels according to (2.8) and (2.27). The expression of the mass of firms in region m is the same as in (4.8) except that, in this case, $q_{ex}^u = 0$ because no urban firm exports and $0 < q_{ex}^r < 1$. The masses of firms and q_{ex}^r determine the equilibrium price index as in (2.33).

2.5 The Impact of Trade Liberalization

In this section, we make use of the model to answer our initial query: how does an expansion in trade opportunities affect the geographic location of industries within a country? We suppose that a decline in the iceberg trade cost causes an improvement in trade opportunities and we assess its impact on the distribution of firms and aggregate production between the urban and rural regions by performing comparative statics in the open economy with respect to τ and comparing two steady-state equilibria.

2.5.1 Firm Exit, Relocation, and Export Market Entry

Case (a)—urban non-exporters, urban exporters, and rural exporters: we show in Appendix A.4.1 that θ^* is negatively related to τ. That is, a decrease in the iceberg trade cost, from τ to τ', entails an increase in the zero cutoff productivity level, from θ^* to $\theta^{*\prime}$. As a result the least productive urban firms are driven out of the industry. A decrease in τ also implies a downward shift in the export cutoff productivity level, from θ_{ex} to θ'_{ex}, which induces the most productive urban non-exporters to enter the export market. Moreover, a decrease in τ causes the urban-rural cutoff productivity level to fall, from $\theta_{u,r}$ to $\theta'_{u,r}$, which makes the most productive urban firms willing to relocate to the rural region (see Figure 2.4).

Case (b)—urban non-exporters, rural non-exporters, and rural exporters: a decrease in τ leads to an increase in θ^*, which compels the least productive urban firms to exit the industry. In contrast to case (a), the fall in τ results in an increase in the urban-rural cutoff productivity level $\theta_{u,r}$. Consequently the least productive rural non-exporters are forced to relocate to the urban region. The export cutoff productivity level, θ_{ex}, shifts downward like in case (a), which gives an incentive to the most productive rural non-exporters to start serving the foreign market (see Figure 2.5).

2.5.2 Reallocation of Market Shares and Profits

We have seen that a decrease in the iceberg trade cost affects firms' choice of location and decision to enter the foreign market. We now discuss the impact on their revenues and profits. We consider a domestic firm with productivity $\theta > \theta^*$ located in region $m \in \{u, r\}$. Let $r_d^m(\theta)$ and $(r_d^m)'(\theta)$ denote the firm's domestic revenues under the old and new trade regimes, respectively. We show in Appendix A.4.2 that $\tau' < \tau$ entails $(r_d^m)'(\theta) < r_d^m(\theta)$. In other words, a decrease in the iceberg trade cost causes a decline in the domestic sales of any urban or rural firm, in case (a) as well as in case (b). The revenue of a firm that is not productive enough to export in the new equilibrium ($\theta^{*\prime} < \theta < \theta'_{ex}$) comes from its domestic sales, that is, $r^m(\theta) = r_d^m(\theta)$. Thus, in case (a), urban non-exporters bear a loss in revenue, and so do urban and rural non-exporters in case (b). In both cases, non-exporters also accrue smaller profits.

The revenues of a firm with productivity $\theta > \theta^*$ in region m from domestic and foreign sales are $r_d^m(\theta) + r_f^m(\theta)$ under the old equilibrium and $(r_d^m)'(\theta) + (r_f^m)'(\theta)$ under the new one. It is shown in Appendix A.4.2 that $\tau' < \tau$ implies $(r_d^m)'(\theta) + (r_f^m)'(\theta) > r_d^m(\theta) + r_f^m(\theta)$. That is, a decrease in the iceberg trade cost brings about an increase in the total of domestic and foreign sales of every urban or rural firm. Thus, a firm that is sufficiently productive to export in the new equilibrium ($\theta > \theta'_{ex}$) earns a greater revenue than it did in the old trade regime, despite the contraction of its domestic sales. In case (a), the new urban exporters and the urban and rural firms that exported prior to the change in τ garner additional revenues. In case (b), the new and pre-existing rural exporters experience a growth in their sales. In both cases, the firms that were exporters in the old equilibrium ($\theta > \theta_{ex}$) obtain higher profits under the new trade regime. The change in profit of a pre-existing urban exporter in case (a) is given by

$$\Delta\pi^u(\theta) = \frac{1}{\sigma}[(r^u)'(\theta) - r^u(\theta)] = w^u f^u \theta^{\sigma-1}\left[\frac{1 + (\tau')^{1-\sigma}}{(\theta^{*\prime})^{1-\sigma}} - \frac{1 + \tau^{1-\sigma}}{(\theta^*)^{1-\sigma}}\right]$$

Similarly, the change in profit of a pre-existing rural exporter in case (a) or (b) is

$$\Delta\pi^r(\theta) = \frac{1}{\sigma}[(r^r)'(\theta) - r^r(\theta)] = \left(\frac{w^u}{w^r}\right)^{\sigma-1} w^u f^u \theta^{\sigma-1}\left[\frac{1 + (\tau')^{1-\sigma}}{(\theta^{*\prime})^{1-\sigma}} - \frac{1 + \tau^{1-\sigma}}{(\theta^*)^{1-\sigma}}\right]$$

Note that $\tau' < \tau$ implies that the term in the brackets in each of those two expressions is positive. Hence, the change in profits of pre-existing urban and rural exporters is an increasing function of their productivity levels. Moreover, the gains in profit of rural exporters relative to those of urban exporters are positively related to the factor price ratio w^u/w^r.

As the cost of trading falls, the non-exporters (urban non-exporters in case (a); rural non-exporters in case (b)) with productivity levels between θ'_{ex} and θ_{ex} enter the export market. The new exporters earn higher revenues than before, yet they also have to cover a larger fixed cost to export. Consider for instance a firm with productivity θ'_{ex} in case (a). This firm exports in the new equilibrium but its export profit, $(\pi^u_f)'(\theta'_{ex})$, is zero. As $(r^u_d)'(\theta'_{ex}) < r^u_d(\theta'_{ex})$, then its new total profit must be less than before. Thus, in case (a) (case (b)), only the most productive firms among the new urban (respectively, rural) exporters earn greater profits after the fall in τ.

In case (a), the exporters moving from the urban region to the rural region accrue even bigger profits, because they now have a lower marginal cost. The expansion of trade opportunities following a decrease in τ allows the high-productivity urban exporters to sell more units of output into the foreign market. The growth in their output creates an incentive to pay the higher fixed cost of operating in the rural location to reduce the variable cost of producing more units of output. This mechanism explains how a fall in τ leads to a downward shift in the urban-rural cutoff productivity level. In case (b), the decline in τ causes the domestic market share of the low-productivity rural non-exporters to shrink, and thus constrains them to produce smaller quantities. The contraction in their output acts as an incentive to relocate to the urban region and pay the lower fixed cost of operating there because less can be saved on variable costs if they were to produce in the rural region. Hence, $\theta_{u,r}$ shifts up.

As the representative consumer's expenditures are given by R, r^m_d/R and $(r^m_d)'/R$ represent a firm's share of the domestic market in the old and new trade equilibria. Thus, the decrease in the domestic sales of domestic firms corresponds to a decrease in their domestic market share. Foreign exporters capture a share of the domestic market away from domestic firms. As the balance of trade is nil (by symmetry), the aggregate revenue of the domestic industry is also equal to R. Then, $r^m(\theta)/R$ and $(r^m)'(\theta)/R$ are a firm's share of the domestic industry revenue, which includes the revenues from both the domestic and the foreign markets, in the old and new trade equilibria. Hence, the more productive domestic firms that export gain market shares within the domestic industry relative to the less productive firms that serve the domestic market only. The market share gains of the former comes from the growth of their foreign sales at the expenses of foreign firms. The exit of firms with productivity levels between θ^* and $\theta^{*\prime}$ (their market share becomes zero) and the relocation of the most productive urban exporters from the high factor price region to the low factor price region in case (a), or the relocation of the least productive rural non-exporters to the urban region in case (b), contribute further to the reallocation of market shares (and profits) across firms within the domestic industry.[28]

The above discussion underlines the difference betwen cases (a) and (b) regarding the relationship between the movement of firms and the reallocation of market shares between the urban and rural locations. For a more formal description of this difference, let $\varsigma^u \equiv R^u/R$ be the market share of all urban firms. Given that $R^u = M^u\bar{r}^u$ and $R = M\bar{r}$, where $\bar{r}$ is the average revenue of all firms, ς^u can be rewritten as $s^u\varrho^u$, where $s^u \equiv M^u/M$ and $\varrho^u \equiv \bar{r}^u/\bar{r}$. Differentiating ς^u with respect to τ yields

$$\frac{\partial\varsigma^u}{\partial\tau} = \frac{\partial s^u}{\partial\tau}\varrho^u + s^u\frac{\partial\varrho^u}{\partial\tau} \tag{2.36}$$

The above equation shows that two factors contribute to the change in the urban market share following a change in the iceberg trade cost. The first factor is the change in the regional distribution of firms. The second one reflects the change in the size of the typical urban firm relative to the average firm. In case (a), both terms of the derivative in (2.36) are positive. As τ decreases, θ^* goes up and $\theta_{u,r}$ goes down, which implies that the mass of urban firms falls relative to the mass of rural firms, and thus relative to the mass of all firms: $\partial s^u/\partial\tau > 0$. As the revenue of urban non-exporters decreases, the average revenue of urban firms declines relative to the average rural revenue and thus relative to the average revenue. Hence, $\partial\varrho^u/\partial\tau > 0$. In case (b), the average urban revenue also falls relative to the average revenue. But the sign of $\partial s^u/\partial\tau$ is ambiguous, unlike in case (a), because both θ^* and $\theta_{u,r}$ shift up. Whether the share of urban firms increases or decreases ultimately depends on the shape of the productivity distribution. Nevertheless, the smaller the urban firms are relative to the average firm (smaller ϱ^u) and the higher the initial share of urban firms (higher s^u), the greater the chances are that a decrease in τ leads to a fall in the market share of the urban region.[29] In summary, a decrease in τ leads to an unambiguous reallocation of market shares from the urban region to the rural region in case (a). In case (b), the reallocation of market shares may occur in the same direction under some circumstances. In conclusion, the theoretical predictions from cases (a) and (b) are consistent with the anecdotal evidence that a greater exposure

[28]The prediction of this model regarding the reallocation of market shares across domestic firms following a decrease in the iceberg trade cost is consistent with that of the Melitz [2003] model.

[29]The hypothesis of a relatively high initial share of urban firms seems particularly appropriate in the context of a leading urban region and lagging rural region with urban wages higher than rural wages.

to international trade was associated with the geographic dispersion of manufacturing firms and production away from large cities in some emerging countries like Mexico (see Hanson 1998) and Argentina (see Sanguinetti and Volpe Martincus 2009), or away from large industrial centers as in Japan for instance (Tomiura 2003).

2.6 Impact of Trade Liberalization on the Location of Colombian Industries

The open economy model showed that a decrease in the international trade cost causes a shift in the distribution of firms and production between the urban and rural regions in the domestic country, which under plausible circumstances actually leads to a reallocation of firms and production to the rural location. In this section, we analyze the Colombian data to find out how trade policy affected the distribution of manufacturing activities between metropolitan areas (large cities) and nonmetropolitan areas (medium-size to small cities and towns) from 1984 through 1991. To be exact, this analysis aims at verifying whether the observed effects of trade policy on the location of industrial activities concur with the model's predictions, not at testing the model. From 1985 to 1992, Colombia implemented a comprehensive reform of tariffs applied to manufacturing imports and removed non-tariff import barriers. Tariffs for individual industries, initially very uneven, eventually flattened out (the variability in the tariff change across industries makes possible the identification of a relationship between this variable and other industry variables) and the average tariff dropped from 55% in 1984 to 21% in 1991, and further down to 13% in 1992. However, as we do not have access to census data beyond 1991, the analysis is limited to the period 1984-91. The econometric results provide weak evidence that tariff reductions negatively affected the metro shares of plants and output from 1984 through 1991.

2.6.1 Data and Empirical Approach

We use data on Colombian manufacturing plants for 1984 and 1991 from the same source as that described in section 2.3. In 1991, 6336 plants with at least ten employees were in activity and 83.7% of them were found in metro areas.[30] Real production reached 151.9 billion 1980 pesos, metro producers accounting for 76.2% of this amount, a figure slightly lower than that in 1984. Exports rose to 17.0 billion 1980 pesos while the metro share of exports grew 11.9 percentage points, to 62.3%.

[30]Plants with less than ten employees were excluded from the statistics for 1991 to facilitate the comparison with those for 1984.

Metro establishments employed 81.3% of the manufacturing workforce numbering 490,678. As Table 2.1 shows, the average within-industry shares of plants, output, and exports remained relatively stable through the period 1984–91, rising 1.4, 0.3, and 3.5 percentage points, respectively. The output-size quotient of metro producers decreased just 0.5 percentage point while their export-size quotient increased 6 percentage points. The growth in metro export shares within industries seems to largely explain the mild upward trend in metro export share at the sector level. Nonetheless, theses variables exhibit considerable variability across industries as their standard deviation, minimum, and maximum values indicate. In some industries, metro plants contracted, at the level of the typical plant, or in aggregate at the industry level, while in others they expanded.

We estimate the impact of a trade policy change on two measures of change in the spatial distribution of industrial activities, namely, the changes in the share of plants located in metro areas and the change in the share of production originating from metro plants, at the four digit level of the ISIC, from 1984 through 1991. Specifically, we regress these two variables on the change in the import tariff for four-digit-level industries of the ISIC, from 1983 to 1990. We opt for a cross-section regression in long differences in these variables instead of a panel data approach because it may take time for adjustments in location choices to be carried into effect. This approach to assessing the effect of a shift in trade policy was previously adopted by Trefler [2004]. We use the lagged tariff difference to deal with potentially endogenous tariffs. Indeed, as pointed out by Fernandes [2007], firms might have lobbied the government to set tariff levels according to their commercial interest. The econometric model is specified as follows:

$$\Delta s^i_{84-91} = \beta_0 + \mathbf{x}'\boldsymbol{\beta}_1 + \beta_2 \Delta\tau^i_{83-90} + \beta_3 s^i_{84} + \beta_4 s^i_{84} \times \Delta\tau^i_{83-90} + (\mathbf{z}^i)'\boldsymbol{\beta}_5 + \epsilon_i \quad (2.37)$$

where the superscript i indicates the four-digit-level industry i; Δs^i_{84-91} denotes either the change in the share of metro plants or the change in the share of metro production in industry i from 1984 through 1991; $\mathbf{x}$ is a vector of indicator variables for two-digit-level industries; $\Delta\tau^i_{83-90}$ is the change in industry i's tariff from 1983 to 1990; $\mathbf{z}^i$ is a vector of industry characteristics; and ϵ_i is a disturbance term.

As relative factor endowments probably differ between metro and nonmetro regions and industry factor intensities matter for location decisions, we select the labor, capital, energy, and materials cost shares in 1984 as industry characteristics. Because the metro share of plants or of production is a censored variable (it is comprised between 0 and 1), its difference over a period of time is also censored (it assumes values between -1 and 1). To address this issue, we include the corresponding initial metro share, denoted by s^i_{84}, among the regressors, and the product between the initial metro share and the tariff change, $s^i_{84} \times \Delta\tau^i_{83-90}$.

Details about the construction of the variables are given in Appendix A.5.2. We use *ad valorem*, applied tariff rates obtained from the DNP website.[31] These tariffs presumably are averages of the nominal tariffs applied to products belonging to the four-digit-level categories.[J. Garcia-Garcia, World Bank, pers. comm.] During the period of analysis, Colombia resorted to other import protection instruments besides tariffs, like import licenses. As Fernandes [2003] claims, the levels of import protection across industries indicated by tariffs well match other measures (she found that tariffs are highly positively correlated with effective protection rates, negatively correlated with import penetration ratios, and positively related to the extent of coverage by import licenses). Thus, these tariffs should well approximate import protection in general.

Note that we look at the impact of trade liberalization on the geography of manufacturing activities in a way that differs from the approaches taken by other studies such as Hanson [1998], Tomiura [2003], Rodríguez-Pose and Sánchez-Reaza [2005], and Sanguinetti and Volpe Martincus [2009]. First, we examine the effect of trade policy changes on the shift in the geographic location of industries over a time period. Second, we consider measures of the spatial distribution of establishments and output instead of employment or/and gross product.

2.6.2 Regression Results

The Breusch-Pagan test[32] reveals the estimated errors from the OLS estimation of the share-of-plants (or plants) regression equation are correlated with those from the share-of-production (or production) model. However, as the number of observations is small, we do not use the seemingly unrelated regressions estimator (it would not produce efficient estimates). Table 2.3 contains the coefficient estimates and the t statistics computed with robust standard errors. The estimates under column heading (1) were obtained from a version of (2.37) without factor cost shares. The coefficient estimate for the tariff change is negative and statistically insignificant in both the plants and production regressions. The estimated coefficient of the initial share of metro plants is negative and significant at the 1% level in the plants regression.

[31] Tariff rates published by national custom authorities for duty administration purposes [R. Moreira, Departamento Nacional de Planeacion, pers. comm.]

[32] Command `sureg` (*depvar1 varlist1*) (*depvar2 varlist2*), `corr` in Stata

In this regression, the interaction variable between the initial share and the tariff change is positive and significant at the 10% level. In the production regression, the coefficient estimates for the initial share of metro production and for the interaction term have the same signs as in the plants regression, but they are insignificant. To compute the estimated effect of the tariff change on the growth in the share of metro plants, one must derive the estimated regression equation with respect to the tariff change, and evaluate the derivative at some value of the initial metro share. So, $\partial \widehat{\Delta s^i_{84-91}} / \partial \Delta \tau^i_{83-90} = \hat{\beta}_2 + \hat{\beta}_4 s^i_{84}$, where $\hat{\beta}_2 = -1.136$ and $\hat{\beta}_4 = 1.347$. We choose to evaluate this expression at the mean value of s^i_{84}, and at one because a significant number of industries were found exclusively in metro areas in 1984. At the mean initial metro share of plants in 1984, the derivative equals -0.0113. Thus, for an industry with the average initial metro share of plants, holding everything else constant, a 10 percentage points decrease in tariff is estimated to entail a 0.113 percentage points increase in the share of metro plants from 1984 to 1991, on average, over all industries. For an initial metro share of 1, the value of the derivative is 0.211. Hence, for an industry present only in metro areas in 1984, a 10 percentage points decrease in tariff is estimated to cause a 2.11 percentage points decline in the share of metro plants by 1991. The estimates from the regression including factor cost shares (column heading (2)) are qualitatively the same, except for the significance level of the estimated coefficient of the initial share of metro plants, which is now significant at the 5% level. Besides, the change in the share of metro plants is negatively correlated with energy intensity. The change in the share of metro production is positively affected by capital and materials intensities. The coefficient of determination is relatively high in all regressions (it varies between 0.232 and 0.436) and it is greater in the plants regressions than it is in the production regressions.

We provide alternative estimations of (2.37) to check for the robustness of those results. First, we examine the impact of tariff policy changes on the spatial distribution of plants and production at the sub-national level because the effect previously estimated may actually reflect interregional reallocations of plants and production. As Colombia became more open to trade, producers may have relocated from inland locations to the Caribbean or Pacific coasts, closer to ports. On the other hand, manufacturers initially located in coastal regions may have been unable to sustain import competition while inland producers were better protected from it. In Table 2.4, under column heading (1), we report the results from the estimation of (2.37) without the establishments from the Caribbean region (departments of Atlántico, Bolívar, Cesar, Córdoba, La Guajira, Magdalena, and Sucre). In the plant regression, the estimated coefficients of the tariff change, initial metro share, and interaction term have the same signs as before but they are statistically insignificant. Again, the estimate for the capital share is positive, and it is the only one being significant. In the production regression, the coefficient estimate for the tariff change is negative and significant at the 10% level; the estimate for the interaction variable is positive and significant at the 10% level; the estimate for the capital share is positive and significant at the 5% level; and the other estimates are insignificant. To obtain the estimated impact of the tariff change on the change in the share of metro production, we evaluate the derivative of the estimated regression function at the mean value of the initial metro share of production (for the sub-national area considered, 0.832) and at one. Hence, for a typical industry as regards the initial metro share, all else constant, a 10 percentage points tariff reduction is estimated to entail a 0.454 percentage point contraction in the share of metro production, on average, over all industries. For an initial metro share of 1, the estimated impact of a 10 percentage points tariff cut is a 1.68 percentage points decline in the share of metro production.

Second, we examine the consequences of altering the definition of the metro variable. Here, only the four cities of Colombia with the largest industrial outputs are considered to be metro. Results are under column heading (2). The signs of the coefficient estimates for the first three variables are the same as before but they are insignificant. The capital share coefficient estimate indicates that the metro share of production increased at a higher rate in capital intensive industries.

Third, we compare the estimates under column heading (2) in Table 2.3 with these obtained using the 1985 data, under heading (3) in Table 2.4. Since the 1985 data contain plants with less than ten employees, we keep them in the 1991 data. For the 1985-91 regressions, the tariff change from 1984 to 1990 replaces that from 1983 to 1990, and the factor cost shares are those in 1985. The new coefficient estimates exhibit the same signs as those of the 1984-91 estimates (except for the initial metro share in the production regression) but not one of the variables of interest is significant. The capital share coefficient estimates are positive and significant in both regressions and the energy share coefficient estimate is still negative and significant in the plant regression.

The evidence on the impact of tariff policy on the distribution of manufacturing activities between metro and nonmetro areas is relatively weak. Yet, a pattern emerges: Tariff liberalization tends to be associated with a redistribution of plants or/and output outside of metro areas in industries that were initially largely concentrated in large cities, whereas it induces a concentration of activities in metro areas in industries relatively dispersed before the liberalization of imports. In general, an increase in import competition should put the least productive manufacturers at a disadvantage and drive them out of business. Thus, we may infer from our findings that in industries concentrated in metro areas at the onset of the liberalization process, low-productivity plants were relatively localized in metro areas, which is consistent with our model. On the other hand, in industries dispersed between metro and nonmetro locations, low-productivity plants are relatively more present in nonmetro areas, which points to a different situation from that depicted by the model.

2.7 Conclusion

We presented a theory and empirical evidence on the relationship between international trade and the geographic location of industries within a country from an original perspective. A model was developed to shed light on the impact of trade liberalization on the distribution of manufacturing activities between metropolises and small urban areas. We built on the Melitz [2003] model by assuming that a country consists of two locations where firms may produce. The price of the factor of production in the rural location is lower than that in the urban location, but producers incur a higher fixed cost when operating in the former. As firms' location choice depends on their productivity level, more productive firms locate in the rural region.

In the open economy, firms jointly decide whether to export and where to locate. Depending on the value of the parameters, one of two possible patterns in firm location and export status prevails. In case (a), low-productivity firms are urban non-exporters, firms with intermediate productivity levels are urban exporters, and high-productivity firms are rural exporters. In case (b), low-productivity firms are urban non-exporters, firms in between are rural non-exporters, and high-productivity firms are rural exporters. We have shown that, in case (a), trade liberalization prompts the most productive urban exporters to relocate to the rural region. The enhancement of export opportunities translates into higher export revenues, allowing those firms to bear the higher fixed cost of producing in the rural region. Thus, trade liberalization raises the rural share of firms and production. In case (b), trade liberalization makes the least productive rural non-exporters relocate to the urban region, which tends to reduce the share of rural firms. Yet, only rural exporters benefit from lower trade costs as they gain foreign market shares, which leads to a reallocation market shares towards the rural region.

Using data from Colombian manufacturing industries for the period 1981–84, we found that metropolitan and nonmetropolitan plants differ in terms of productivity in some industries. Metro plants tend to exhibit higher labor and capital productivities, whereas nonmetro plants have a propensity for having higher energy and materials productivities. Then, we estimated the impact of a change in import tariff, over the period of trade liberalization reforms, on the distribution of plants and production between metro and nonmetro areas in Colombia. We obtained evidence that the decline in import tariffs had an effect on the change in the shares of nonmetro plants and production from 1984 through 1991. For industries relatively concentrated in metro areas before trade liberalization, the tariff decrease induced the decentralization of manufacturing activities. On the other hand, for industries relatively dispersed between metro and nonmetro areas initially, tariff liberalization led to further concentration of plants and production in metro areas.

These results reinforce the idea that policy makers should consider the implications of trade policy for the location of manufacturing activities. In particular, they should have an understanding of the consequences of trade liberalization for the sustainability of manufacturing bases in regions outside large cities, where industrial activities generate a major share of income and alternatives are scarce. Trade liberalization may be beneficial to lagging or peripheral regions as it fosters the decentralization of some industries. But other industries may become more centralized in large cities as a country becomes more exposed to trade, leading to manufacturing job losses in small urban areas. We hope this paper will spur more theoretical and empirical research to understand the interaction between regional economic disparities within countries and international trade.

Table 2.1: Geographic Location of Colombian Manufacturing Activities Within Industries, 1984–91

Variable	Mean	Std. Dev.	Min.	Max.	*N*
Level in 1984					
Share of metro plants	0.835	0.186	0.189	1	94
Share of metro output	0.841	0.216	0.003	1	94
Share of metro exports	0.792	0.325	0	1	85
Relative output size of metro plants[a]	1.012	0.272	0.009	1.667	76
Relative export size of metro plants[b]	0.948	0.469	0	2.439	70
Difference between 1984 and 1991 (% points)					
Share of metro plants	1.369	9.837	-31.434	50	94
Share of metro output	0.336	9.273	-42.582	55	94
Share of metro exports	3.546	20.381	-35.7	98.657	81
Relative output size of metro plants	-0.5	17.7	-83.5	73	69
Relative export size of metro plants	6	33	-80	129.1	66

Note: The mean and other statistics of every variable are estimated over all four-digit level industries of the ISIC.

[a] The average size of metro plants divided by the average size of all plants, in terms of real production, within an industry. Estimation is based on industries with at least one metro plant and one nonmetro plant.

[b] The average size of metro plants divided by the average size of all plants, in terms of real export sales, within an industry. Estimation is based on industries with at least one metro plant and one nonmetro plant.

Table 2.2: Sign and Statistical Significance of Productivity Differences Between Metro and Nonmetro Plants

Variable	Frequency Distribution of the t Statistics[a]		
	$t \leq -t_{0.05,n}$	$-t_{0.05,n} < t < t_{0.05,n}$	$t \geq t_{0.05,n}$
$metro_j \times l^i_{j,t}$	1	24	4
$metro_j \times k^i_{j,t}$	1	23	5
$metro_j \times e^i_{j,t}$	4	22	3
$metro_j \times m^i_{j,t}$	6	19	4
$metro_j$[b]	9	6	14

[a] Intervals of the t statistics frequency distribution are defined with respect to the t value for a mass in both tails of the probability density function of 0.05, and an industry-specific number of degrees of freedom, n. All t statistics are calculated with robust standard errors.

[b] The t statistics of *metro*'s coefficient estimate is derived from an industry-specific regression of estimated TFP on the metro variable.

Table 2.3: Impact of Tariff Policy on Metro Shares of Plants and Production From 1984 to 1991

	(1)		(2)	
	Plants	Production	Plants	Production
Tariff change, 1983-90 (in % points)	-1.136	-1.161	-0.910	-0.745
	(-1.67)	(-1.19)	(-1.55)	(-1.41)
Metro share in 1984 (in %)	-0.181***	-0.0905	-0.125**	-0.0503
	(-2.78)	(-1.26)	(-2.20)	(-1.06)
Metro share in 1984 × Tariff change[a]	1.347*	1.380	1.138*	0.949
	(1.89)	(1.26)	(1.82)	(1.60)
Labor share in 1984			12.47	59.14
			(0.39)	(1.60)
Capital share in 1984			56.39	88.88**
			(1.57)	(2.11)
Energy share in 1984			-108.7*	-47.70
			(-1.76)	(-0.69)
Materials share in 1984			0.614	43.95*
			(0.03)	(1.69)
Observations	80	80	80	80
R^2	0.355	0.232	0.436	0.347

Notes: Every regression includes a constant and a set of dummy variables for two-digit level industries.
t statistics computed with robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[a] In percentage points.

Table 2.4: Impact of Tariff Policy on Metro Shares of Plants and Production—Robustness Checks

	(1)		(2)		(3)	
	Plants	Production	Plants	Production	Plants	Production
Tariff change, 1983-90 (in % points)	-0.672	-0.562*	-0.463	-0.0409	-0.336	-0.418
	(-1.11)	(-1.68)	(-1.06)	(-0.12)	(-0.90)	(-1.00)
Metro share in 1984 (in %)	-0.0703	-0.0534	-0.171	-0.00440	-0.0825	0.0326
	(-1.04)	(-1.55)	(-1.60)	(-0.08)	(-1.44)	(0.67)
Metro share in 1984 × Tariff change[a]	0.904	0.730*	0.571	0.126	0.431	0.536
	(1.35)	(1.98)	(1.15)	(0.30)	(1.10)	(1.18)
Labor share in 1984	25.71	54.27	18.16	78.29*	34.66	50.01
	(0.83)	(1.50)	(0.38)	(1.91)	(1.42)	(1.65)
Capital share in 1984	52.56*	76.51**	80.22	135.2***	69.48**	99.81**
	(1.86)	(2.02)	(1.52)	(3.00)	(2.05)	(2.53)
Energy share in 1984	-50.98	-38.54	-88.40	35.36	-110.6**	-70.10
	(-0.94)	(-0.59)	(-0.98)	(0.37)	(-2.05)	(-1.46)
Materials share in 1984	15.77	35.09	1.815	47.82	9.459	37.42
	(0.82)	(1.42)	(0.05)	(1.36)	(0.46)	(1.60)
Observations	78	78	78	78	80	80
R^2	0.345	0.340	0.365	0.382	0.435	0.423

Notes: Every regression includes a constant and a set of dummy variables for two-digit level industries. Column headings have the following meanings: (1) Estimation excluding the Caribbean region; (2) estimation where metro areas include only Bogotá, Medellín, Cali, and Barranquilla; (3) estimation using the 1985 data in place of the 1984 data.
t statistics computed with robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[a] In percentage points.

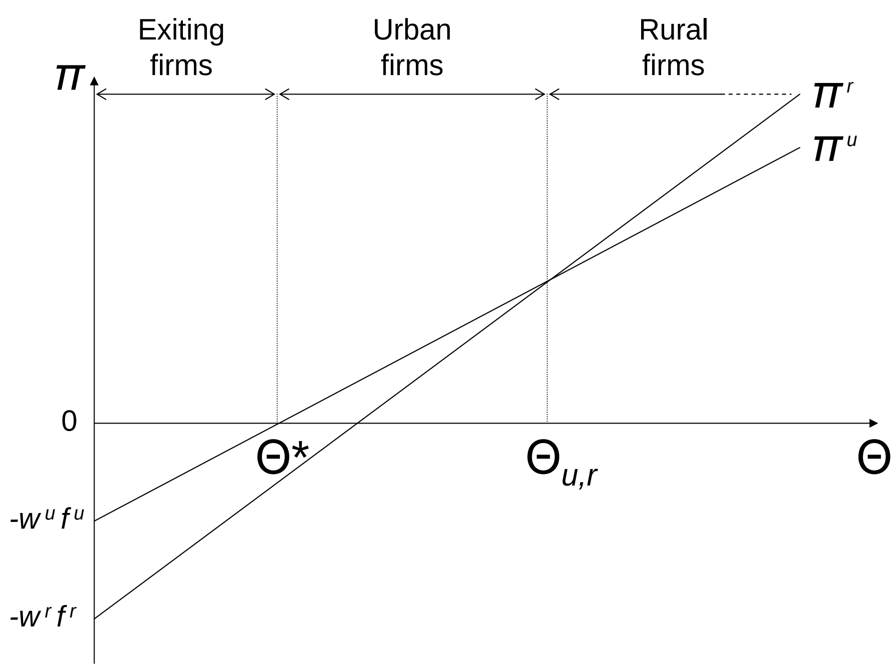

Figure 2.1: Operating Profits in the Urban and Rural Locations

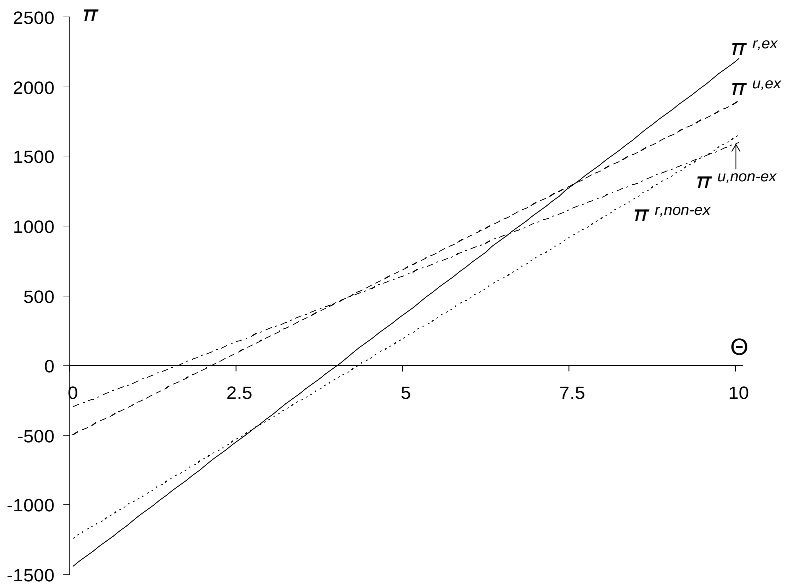

Figure 2.2: Case (a): Urban Non-Exporters, Urban Exporters, Rural Exporters

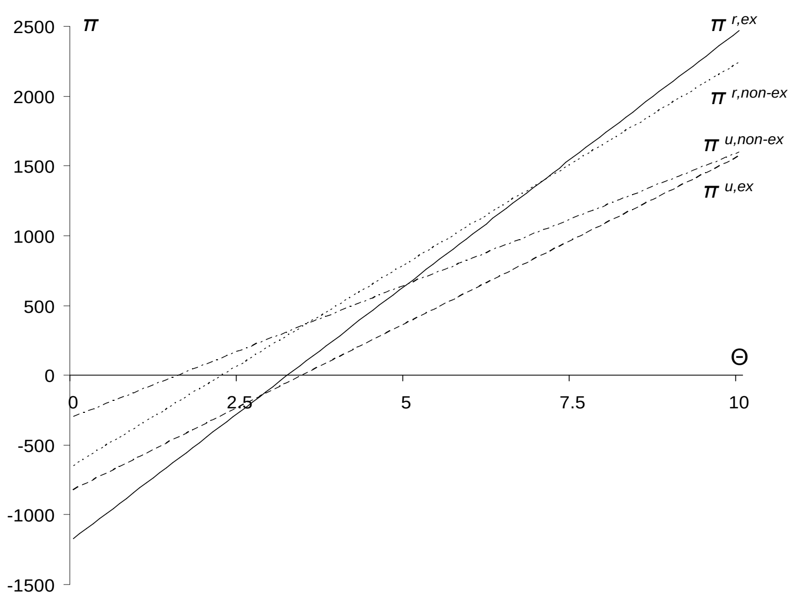

Figure 2.3: Case (b): Urban Non-Exporters, Rural Non-Exporters, Rural Exporters

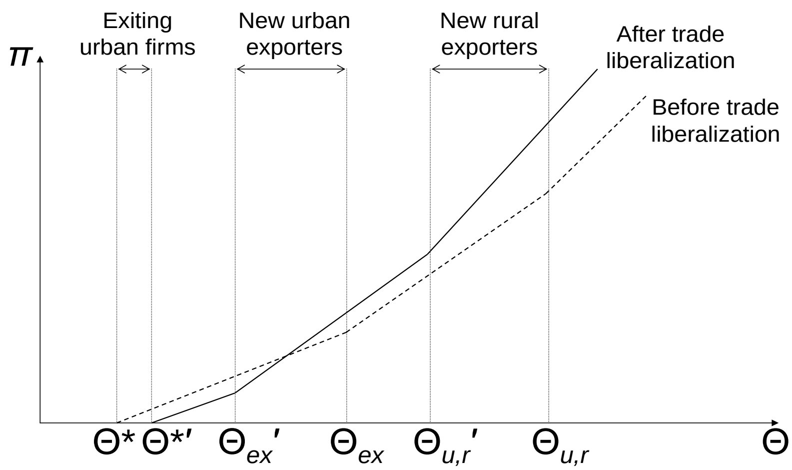

Figure 2.4: Relocation of Firms and Reallocation of Profits—Case (a)

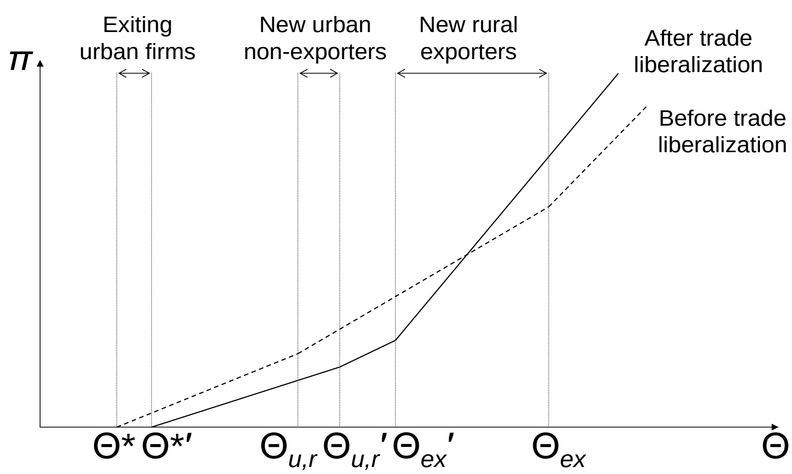

Figure 2.5: Relocation of Firms and Reallocation of Profits—Case (b)

Chapter 3 Trade Policy, Capital Mobility and Industrial Geography

3.1 Introduction

Recently, declining transport and communication costs and commercial policy reforms have substantially lowered barriers to trade between countries, which has spurred exceptional, widespread growth in international commerce. The evidence suggests that this increasing exposure to trade has not been without consequences for the economic geography of countries and that it has produced greater domestic spatial inequalities in many cases. This study offers a contribution to the theoretical literature examining the relationship between international trade and internal economic geography.

Numerous studies have looked at the link between trade openness and internal geography from a rather broad perspective. Using data from 85 countries, Ades and Glaeser [1995] uncovered a positive (negative) relationship between tariff protection (trade flows) and the concentration of population in a single, large urban area. Several studies analyzed the evolution of regional differences in income within countries. La Fuente and Vives [1995] observed that incomes were converging among European Union (EU) states while their economic integration was progressing, but also found evidence of divergence among regions within their borders. The piece by Armstrong [1995] corroborated this observation by showing that convergence among EU sub-national regions slowed down during the periods 1970–80 and 1980–90 compared with the preceding decades, and especially among regions belonging to the same country. Quah [1996] confirmed that Spain and Portugal have experienced an increase in regional income disparities in the course of the European integration (see also Esteban [1994] and Sala-i-Martin [1996] about Spain). Rodríguez-Pose and Gill [2006] argued that regional income inequalities increased in several emerging and developed countries between the 1970s and the 1990s as their manufacturing exports grew relative to their agricultural exports. A collection of empirical studies sponsored by the *Spatial Disparities in Human Development* project of the World Institute for Development Economics Research of the United Nations University supported the claim that rising trade and financial openness could partially explain increasing economic inequalities between urban and rural areas, and among regions with different assets, in emerging and developing countries [Kanbur and Venables, 2005]. For instance, Rodríguez-Pose and Sánchez-Reaza [2005] noticed that, in Mexico, a state's distance to the capital city had a negative impact on its per-capita gross domestic product growth rate between 1980 and 1985, but that it no longer affected state-level growth during the period following the inception of trade liberalization reforms, in 1985. Instead, differences in skilled labor endowment among regions caused rising spatial inequalities. In West Africa and in China, large, coastal cities have benefited the most from foreign direct investments, which has exacerbated urban-rural income inequalities [Te Welde and Morrissey, 2005, Ge, 2006].

Trade has apparently affected the spatial distribution of domestic manufacturing activities too. Hanson [1998] documented the migration of manufacturing activities from the region of Mexico's capital city towards the United States (US) border following the unilateral trade reforms of the 1980s. His analysis also highlighted the break-up of old industrial centers previously supported by input-output linkages. Tomiura [2003] provided evidence for a similar phenomenon occurring in Japan in the 1990s. Brulhart and Traeger [2005] reported that manufacturing establishments have become more evenly distributed within EU states in the recent past. In the case of Argentina, Sanguinetti and Volpe Martincus [2009] found that industries with lower tariffs were more decentralized away from the city of Buenos Aires and its surrounding region.

The literature offers theoretical explanations for the apparent relationship between changes in trade openness and changes in the geographic distribution of economic or manufacturing activities within national borders. First nature geography certainly contributes to spatial disparities within a country trading with others. Regions enjoying a better access to foreign markets simply because of geographical proximity or because they are coastal or/and have waterways may benefit more from trade policy reforms or improvements in transportation technologies. For instance, the natural advantage of Chinese coastal regions partly explains their faster growth relative to the interior regions in the recent past; proximity to the US has been a major factor behind the economic success of Mexico's border states compared with the southern states. Other explanations rely on second-nature geography, which refers to variations in the spatial density of population, producers of goods and services, and infrastructures, and to the distance among them. Krugman and Livas-Elizondo [1996] examined the link between external trade and internal geography through an extension of the core-periphery Krugman [1991] model.[1] They contended that in a relatively closed

[1] In the two-location Krugman [1991] model, manufacturing firms with scales economies prefer to locate where the market is larger to save on transport costs. The concentration of producers in one location tend to raise the wage rate of manufacturing workers there. Since these workers are mobile, they tend to migrate to that location, thus reinforcing the agglomeration effect of market size. In addition, as consumers, mobile workers are attracted to the region with a larger

economy, the tendency of the manufacturing industry to concentrate in one location, which arises from the interplay of scale economies with (domestic) trade costs, is stronger than the factor causing its dispersion, which stems from competition in the land and labor markets.[2] As the economy opens up to trade, manufacturers export a larger share of their production while consumers substitute foreign products for domestic ones; and thus the incentives for firms and consumers/workers to concentrate in one location wane. From numerical simulations, they concluded that lower external trade costs favor the spatial dispersion of industrial activities and argued that their model provides a good explanation for the shift in the location of manufacturing activities away from Mexico City in the wake of trade liberalization as well as for similar outcomes in other emerging countries. Paluzie [2001] took a slightly different approach as she maintained the assumption that there is an immobile workforce acting as a dispersion factor. Her model shows that a decrease in the cost of international trade tends to leave the symmetric equilibrium (in which manufacturing workers are evenly distributed between the two domestic regions) unstable while it makes the core-periphery outcome (the whole manufacturing workforce in one location) stable. She claimed that it provides an explanation for the deterioration of the convergence trend among Spain's regions upon its accession to the European Community. Monfort and Nicolini [2000] further extended the Krugman and Livas-Elizondo [1996] analysis by assuming that the external trading partner also comprises two regions (their model features immobile workers). Their findings corroborated those of Paluzie [2001] as their numerical simulations confirm that lower trade costs tend to render unstable the dispersion of industrial firms within countries while making the agglomeration equilibrium stable. Both articles alleged that the contradiction between their conclusion and that of Krugman and Livas-Elizondo [1996] comes from the inclusion of urban costs in the latter's model. Even though a decrease in the external trade cost weakens the agglomeration force, the peripheral region too is subject to a stronger competition from foreign firms, and thus domestic firms, in the absence of agglomeration costs, have weak incentives to disperse.

Like Monfort and Nicolini [2000], Behrens et al. [2006] and Behrens et al. [2007] analyzed the effect of external trade on internal geography using a two-country model where each country is divided into two regions. Unlike previous studies, their model

manufacturing industry as the price of the manufacturing good is lower there. Migration then leads to firm relocation (because of the market size effect) which lowers the price of manufactures again, and so on. Factors of dispersion are due to immobile workers plus the effect of competition for market shares among spatially concentrated firms.

[2]Whereas the Krugman [1991] model relies on some immobile workers as a factor of dispersion of industrial firms, Krugman and Livas-Elizondo [1996] invoke urban congestion costs.

builds on the variation of the core-periphery model presented by Ottaviano et al. [2002]. [3] The Behrens et al. [2007] model predicts the agglomeration of the manufacturing industry within each country when the interregional trade cost is low relative to the international trade cost. As the external trade cost falls, dispersion of the manufacturing sector becomes possible. Behrens et al. [2006] introduced an element of first-nature geography by assuming that a region has a better access to the foreign market than the other (in one of the two countries), to represent the advantage that a coastal or border region may have over an inland region. They found that relatively low internal or (and) external trade costs result in the agglomeration of the industry in the coastal region, while agglomeration occurs inland when these parameters are relatively high. They pointed out that because of the trade-off between access to the foreign market and exposure to foreign competition, a coastal region does not necessarily have an advantage in attracting manufacturing firms.

The purpose of this paper is to supplement this theoretical literature with a simple model, based on the Martin and Rogers [1995] footloose capital (FC) model, that yields analytical solutions (and does not require numerical simulations) while preserving the assumption of constant elasticity of substitution (CES) preferences. Thus it allows one to perform comparative statics with respect to trade policy variables, and gives the possibility to study the interaction between trade and regional policies. In addition, unlike previous studies, I make the distinction between the impact of bilateral reduction in trade costs and of unilateral trade liberalization. Like Behrens et al. [2006], I also consider the case in which the regions of a country have unequal access to the external trading partner.

Here is the outline of the paper: Section 2 presents the FC model, with few modifications from its original version. Section 3 describes the extension of the FC model to an open economy setting. In section 4, I perform comparative statics with respect to trade policy variables to analyze the impact of trade on internal geography. Concluding remarks are in section 5.

3.2 The Footloose Capital Model

Suppose the economy of a country is based on two factors and two sectors of production. A perfectly competitive sector employs labor to produce a homogeneous good

[3] In the Ottaviano et al. [2002] model, demand for the varieties of the manufacturing good is linear in price, the price elasticity is variable, and the mark-up decreases with the size of the market. Skilled labor, the factor employed in the sector with scale economies, is geographically mobile, while unskilled labor, the factor of the agricultural sector, is immobile.

with a constant returns to scale technology. This sector is referred to as agriculture. In the other sector, manufacturing, firms use a technology with scale economies to make a horizontally differentiated good. This technology requires a fixed amount of capital and a variable input of labor. The manufacturing sector is characterized by monopolistic competition among a continuum of firms.

The territory of the country is split into two regions, u and v, and is populated by a continuum of households of mass $L > 0$. A mass L^m of households ($0 < L^m < L$) consumes goods and supplies labor services in region m, where $m \in \{u, v\}$. Every household can supply one unit of labor. While households are immobile, interregional trade in goods allows them to consume the agricultural commodity and the differentiated goods produced in the region other than that where they live.

The household population is endowed with a capital stock of mass $K > 0$. The mass of capital owned by region m's residents is denoted by K^m ($0 < K^m < K$). Households living in region m can rent capital to manufacturers located in that region as well as to producers from the other region. Note that they can spend the return from their capital on goods produced in their region as well as on goods imported from the other location.

3.2.1 Demand

In each region, the preferences of households are aggregated through a representative consumer. The preferences of region m's consumer are represented by a Cobb-Douglas utility function defined over the consumption of the agricultural good and the differentiated product: [4]

$$U^m = (x_o^m)^\alpha (X^m)^{1-\alpha},$$

where x_o^m is the consumption of the agricultural good in region m, X^m is an index based on the consumption of varieties of the differentiated good, and α is the elasticity of substitution between the two (it is also the expenditure share of the agricultural good). X^m is a CES utility index defined over a continuum of potential varieties of the differentiated product:

$$X^m = \left[\int_{j \in J^m} x_m(j)^\rho dj \right]^{1/\rho},$$

where j indexes these varieties; J^m denotes the set of varieties supplied at location m; $x_m(j)$ is the quantity of variety j consumed in region m; ρ is a parameter such that

[4] In the Dixit and Stiglitz [1977] model of monopolistic competition, products in an industry, or sector, are good substitutes among themselves, but poor substitutes for other commodities in the economy. The purpose of the homogeneous good is to aggregate these commodities into one good.

$\rho = (\sigma - 1)/\sigma$, where σ is the elasticity of substitution between any pair of varieties. It is assumed that $\sigma > 1$ (varieties are substitutes), or, equivalently, $0 < \rho < 1$. The elasticity of substitution and the residual demand elasticity for any particular variety are equal provided that the set J^m has a non-zero measure. Since the varieties available in region m also are in region n, I omit the superscript of the set of varieties afterward.

The representative consumer maximizes utility subject to the budget constraint

$$p_o^m x_o^m + \int_{j \in J} p_m(j) x_m(j) dj = I^m,$$

where p_o^m and $p_m(j)$ are the prices paid by region m's households for the agricultural good and variety j, respectively, and I^m is their aggregate income (or expenditure). Hence, demand for the agricultural good is given by

$$x_o^m = \frac{\alpha I^m}{p_o^m}.$$

Demand for variety j is expressed as

$$x_m(j) = \frac{(1-\alpha) I^m}{P^m} \left(\frac{p_m(j)}{P^m} \right)^{-\sigma}, \tag{3.1}$$

where P^m is the price index of the differentiated good for region m, which is defined as

$$P^m = \left[\int_{j \in J} p_m(j)^{1-\sigma} dj \right]^{1/(1-\sigma)}. \tag{3.2}$$

The expression of the indirect utility function, which measures the representative consumer's welfare, is

$$V^m = \frac{I^m}{(p_o^m)^\alpha (P^m)^{1-\alpha}}.$$

3.2.2 Production

Units of the agricultural good and of labor are defined so that producing one unit of the good requires one unit of labor. Therefore, perfect competition ensures that the agricultural good's price in region m equals the wage rate there, w^m, that is, $p_o^m = w^m$. Interregional trade in this good is assumed to be free.[5] As a result, the agricultural price in region u is equated with that in region v (the common price is denoted by p_o). As long as both regions produce the agricultural commodity, free

[5] This assumption is standard in the new trade theory and new economic geography (NEG) literatures.

trade indirectly equalizes wages between the two locations.[6] Finally, setting the wage rate to one, as a normalization, implies that the agricultural price is also one, that is, $w^u = w^v = p_o = 1$. This makes the agricultural good the numéraire.

In the manufacturing sector, each firm produces only one variety of the differentiated product; and every variety is made by a single firm (there are no scope economies). Manufacturers use a common technology, which requires one unit of capital and a variable quantity of labor (as varieties are identical in all respects, the index j is omitted in what follows). The cost function of a producer located in region m, c^m, is specified as

$$c^m(y) = r^m + \beta y,$$

where r^m denotes the rental rate of capital at location m, β is the marginal cost, in terms of labor, and y denotes output. Provided that $r^m > 0$, the average cost, $c^m(y)/y$, decreases with output. Because each firm requires one unit of capital to operate, the mass of manufacturing firms in the economy is the same as the mass of capital, K. Likewise, the mass of firms located in region m, M^m, is the same as the capital mass invested there.

Manufacturers can deliver goods to households residing in the same region as that where they are located (their local market) without bearing any trade cost. However, they must incur a per-unit trade cost to serve the other region's market (their non-local market). This cost takes the form of an iceberg trade cost: producers must ship τ units of output, where $\tau > 1$, to deliver one unit of it to the non-local market. In practice, the cost of trading between regions would essentially involve transport costs and other transaction costs related to the geographic distance factor. As the trade cost effectively shifts the marginal cost of serving the market, firms maximize the profit from sales in the local market independently of their price and output decision in the non-local market.

For a firm in region m, the variable profit earned from local sales (gross of the fixed cost) is

$$\pi_m^m = (p_m^m - \beta)\, y_m^m, \qquad (3.3)$$

where p_m^m is the price charged to region m's households and y_m^m the volume of output sold to them. The firm faces the residual demand (4.1). The profit-maximizing price is determined by substituting (4.1) into (3.3) and obtaining the maximum of the

[6] For agricultural production to take place in both regions, no location must have enough labor to match the country-wide demand for the agricultural commodity. In other words, this condition must hold: $\alpha(I^u + I^v) > \max\{p_o L^u, p_o L^v\}$.

resulting expression with respect to p_m^m. Thus, the local price is

$$p_m^m = \frac{\beta}{\rho}. \tag{3.4}$$

As a consequence of the assumption of CES utility, firms charge a mark-up over their marginal cost that is independent of the mass of firms in the market. By substituting the demand function (4.1) and the price formula (3.4) into (3.3), one obtains the maximal value of the local variable profit:

$$\pi_m^m = \frac{s_m^m}{\sigma},$$

where s_m^m, the local sales revenue, is given by

$$s_m^m = \frac{(1-\alpha)I^m\left(P^m\rho\right)^{\sigma-1}}{\beta^{\sigma-1}}.$$

Similarly, the variable profit earned from non-local sales is

$$\pi_n^m = \left(p_n^m - \tau\beta\right)y_n^m,$$

where p_n^m is the price charged to region n's households and y_m^m is the physical volume of sales. The optimal non-local price is $p_n^m = \tau\beta/\rho$ and thus the maximum variable profit derived from non-local sales is $\pi_n^m = s_n^m/\sigma$, where s_n^m stands for the non-local sales revenue,

$$s_n^m = \frac{(1-\alpha)I^n\left(P^n\rho\right)^{\sigma-1}}{(\tau\beta)^{\sigma-1}}.$$

In the FC model, the ratio of the non-local price to the local price is just equal to the trade cost. The total of the local and non-local variable profits, net of the fixed cost, is $\pi^m = \pi_m^m + \pi_n^m - r^m$, or

$$\pi^m = \frac{(1-\alpha)\rho^{\sigma-1}}{\sigma\beta^{\sigma-1}}\left[I^m\left(P^m\right)^{\sigma-1} + \tau^{1-\sigma}I^n\left(P^n\right)^{\sigma-1}\right] - r^m. \tag{3.5}$$

By using the price index (3.2) and the price formulas, it is rewritten as follows:

$$\pi^m = \frac{(1-\alpha)I}{\sigma K}\left[\frac{s_I^m}{s_M^m + \tau^{1-\sigma}s_M^n} + \frac{\tau^{1-\sigma}s_I^n}{\tau^{1-\sigma}s_M^m + s_M^n}\right] - r^m, \tag{3.6}$$

where $I = I^u + I^v$ is the domestic income, $s_I^m \equiv I^m/I$ is the income share of region m, and $s_M^m \equiv M^m/K$ designates the share of manufacturing firms sited in region m.

In the context of the FC model, it is appropriate to distinguish between two equilibrium concepts. The first type of equilibrium takes the regional distribution of capital—and thus of manufacturing firms—as exogenous. This equilibrium concept is intended to represent the economy in the short run, when capital is immobile. The second, the long-run equilibrium, allows households to circulate capital between regions in search of the highest return. The next two subsections characterize these equilibria. The equilibrium conditions are derived only for the differentiated product market. As long as this market clears, by Walras' law, the agricultural market clears too.

3.2.3 Short-Run Equilibrium

3.2.3.1 Free Entry Condition

In the short run, in each region, the supply of capital is inelastic. Since firms must meet a fixed capital requirement, free entry into the manufacturing industry raises the rental rate of capital up to the point where it equals the typical firm's profit.[7] Thus, free entry drives profits to zero. From (3.6), this implies the following equality:

$$r^m = \frac{(1-\alpha)I}{\sigma K}\left[\frac{s_I^m}{s_M^m + \tau^{1-\sigma}s_M^n} + \frac{\tau^{1-\sigma}s_I^n}{\tau^{1-\sigma}s_M^m + s_M^n}\right]. \tag{3.7}$$

The return to the mobile factor depends not only on the regional distribution of expenditures, but also on the spatial allocation of capital. Thus, in the short run, the return to capital may differ between the two regions even if these are symmetric with regard to expenditure levels. The capital rental rate differential between the two regions is given by

$$r^m - r^n = A\left[\left(1+\tau^{1-\sigma}\right)\left(s_I^m - \frac{1}{2}\right) - \left(1-\tau^{1-\sigma}\right)\left(s_M^m - \frac{1}{2}\right)\right], \tag{3.8}$$

where $A \equiv [(1-\alpha)I/(\sigma K)](1-\tau^{1-\sigma})/[(s_M^m + \tau^{1-\sigma}s_M^n)(\tau^{1-\sigma}s_M^m + s_M^n)]$. Equation (3.8) renders explicit the determinants of the relative rental rate in region m, that is, the attractiveness to investors of region m compared with that of region n. The sign of $r^m - r^n$ is the same as the sign of the expression inside square brackets. The first term of that expression represents the influence of region m's relative market size; all else equal, the larger it is, the greater the rental rate differential. As will be shown later, a region with a larger capital endowment exhibits a higher income. This, in turn, commands a positive rental rate differential, which in the long run attracts capital and induces an expansion in the mass of manufacturing firms. Hence, the local market size acts as an agglomeration force for capital and the manufacturing

[7] One may view household as potential "entrepreneurs" willing to enter the industry by incurring the fixed cost of capital. They would bid the capital rental rate up until the expected profit from entry is zero.

industry. On the other hand, the relative mass of manufacturing firms, in the second term, has a negative effect on the rental rate differential; all else constant, the bigger the mass of firms in region m, the less attractive the region is to capital owners. Intuitively, when rival firms are numerous, a firm can only capture a small share of the market. Then, competition among firms for market shares acts as a dispersion force for the manufacturing industry. In addition, a decrease in the trade cost or in the elasticity of substitution strengthens the agglomeration effect of regional market size and weakens the dispersion force due to regional competition.

The capital rental rate should be expressed as a function of the regional shares of capital, labor, and manufacturing firms only to fully characterize the short-run equilibrium. For that, one must first express domestic and regional incomes as functions of the labor and capital endowments.

3.2.3.2 Regional Income Distribution

Domestic income is the sum of labor and capital incomes; that is, $I = R + L$, where $R = M^u r^u + M^v r^v$ and $L = L^u + L^v$. Let $S^m \equiv \int (s_u^m + s_v^m)$ denote the aggregate revenue of manufacturing firms located in region m and $S = S^u + S^v$ the revenue of the domestic manufacturing industry. The free entry condition (equation (3.7)) implies that the aggregate return to the capital invested in region m, namely, $M^m r^m$, amounts to the aggregate variable profit of that region's manufacturers, that is, $M^m r^m = S^m/\sigma$; and so, $R = S/\sigma$. Furthermore, $S = (1-\alpha)I$ since households spend a fraction $1-\alpha$ of their income on the differentiated good. Hence, $R = (1-\alpha)I/\sigma$ and $I = \sigma L/(\sigma - 1 + \alpha)$.

Similarly, the income of region m is $I^m = R^m + L^m$, where $R^m = K_u^m r^u + K_v^m r^v$, K_n^m being the mass of capital from region m invested in region n. Whereas regional labor income is given, finding capital income is more difficult. Indeed, since in the short run the rental rate of capital may diverge between regions, one must know how much of region m's capital is invested in each region to calculate the income of region m's investors. This hypothesis resolves that issue: The fraction of the capital invested in region n that is owned by region m's households equals region m's share of the domestic capital endowment. That is,

$$\frac{K_n^m}{M^n} = \frac{K^m}{K} \; \forall (m,n) \in \{u,v\} \times \{u,v\}.$$

Accordingly, region m's capital income can be expressed as $R^m = s_K^m R = s_K^m (1-\alpha)I/\sigma$, where $s_K^m \equiv K^m/K$ is region m's capital endowment share.

Eventually, region m's income/expenditure share is given by

$$s_I^m = \frac{1-\alpha}{\sigma} s_K^m + \frac{\sigma - 1 + \alpha}{\sigma} s_L^m. \tag{3.9}$$

In this expression, region m's income share is a weighted average of its capital and labor endowment shares. And if regions u and v are symmetric with respect to their capital and labor endowments ($s_K^m = s_L^m = 1/2$), then income is evenly split between them. Besides, the weight of the capital share is greater the larger the share of expenditures allocated to the manufacturing good and the smaller the elasticity of substitution between varieties.

The regional distribution of income is independent of the geographic allocation of firms. Indeed, the assumption made about the composition, with respect to its origin, of the capital invested in a region entails $R^m/K^m = R/K = (1-\alpha)I/(\sigma K)$. In other words, the average rental rate of capital received by region m's households is just the average rental rate for the economy as a whole, $\bar{r}$. This is true in particular when r^u and r^v assume a common value, r, in which case $r = \bar{r}$. However, as shown in the next subsection, the regional distribution of expenditures affects the spatial distribution of firms in the long run.

3.2.4 Long-Run Equilibrium

In the short run, the return to capital may vary from one region to the other. But in the long run, the mobility of capital allows investors to seek the highest return on their assets (in terms of the numéraire) by reallocating capital between regions. The economy reaches a long-run equilibrium when the returns to capital in regions u and v are equalized so that no investor has an incentive to transfer its capital, or when capital is wholly concentrated in a region while the latter still offers a higher return. A formal definition of potential equilibria with capital mobility is given below.

Definition 1 (Equilibria with capital mobility)**.** Let $\Delta r \equiv r^u - r^v$ be the capital return differential between locations u and v. A regional distribution of capital investment, or manufacturing firms, (M^u, M^v), is an equilibrium if

$$-1^{l\{M^u=0\}} \Delta r \geq 0, \tag{3.10}$$

where l is the indicator operator (it takes the value one if its statement if true and zero otherwise) and the equality holds if $0 < s_M^u < 1$.

Two types of equilibrium may occur. First, an interior equilibrium corresponds to a situation where no region has zero manufacturing firms, that is, $0 < s_M^u < 0$, and capital is remunerated at the same rate in both regions. An agglomerated, or core-periphery, equilibrium occurs when either $s_M^u = 0$ or $s_M^u = 1$ and the return to capital is unequal between regions.

NEG models usually describe factor movement as follows: The product market adjusts instantaneously upon the transfer of factor from one region to the other. The owners of the mobile factor make investment decisions according to a "myopic" adjustment process, which in this case means that the incentive for investors to move capital around is given by the current differential in rental rate between regions (since investors spend capital income in the region where they reside no matter where capital is utilized). Thus, capital is assumed to move between locations according to the following equation:

$$\frac{ds_M^u}{dt} = \Delta r s_M^u s_M^v.$$

First, I characterize the interior long-run equilibrium by solving (3.10) for the geographical distribution of the mobile factor. The condition $r^u = r^v$ holds if and only if

$$s_M^m - \frac{1}{2} = \frac{1+\tau^{1-\sigma}}{1-\tau^{1-\sigma}} \left(s_I^m - \frac{1}{2} \right). \tag{3.11}$$

For this solution to be valid, its value must be comprised between zero and one. Depending on the values taken by τ, σ, and s_I^m, it could be smaller than zero, or greater than one. Second, from here, the core-periphery equilibria are easily solved for. If the expression above is smaller (greater) than zero (one), then s_M^m is zero (one), which means that the manufacturing industry is present only in region m (n).

By subtituting (3.9) into (3.11), one obtains

$$s_M^m - \frac{1}{2} = \frac{1+\tau^{1-\sigma}}{1-\tau^{1-\sigma}} \left[\frac{1-\alpha}{\sigma} \left(s_K^m - \frac{1}{2} \right) + \frac{\sigma - 1 + \alpha}{\sigma} \left(s_L^m - \frac{1}{2} \right) \right],$$

where the share of the manufacturing industry located in region m is expressed as a function of the primitives of the model only. Furthermore, using (3.11), one can express the relative manufacturing share of region u as a function of its relative income, as $s_M^u - s_M^v = T(s_I^u - s_I^v)$, where

$$T^a \equiv \frac{1+\tau^{1-\sigma}}{1-\tau^{1-\sigma}}. \tag{3.12}$$

Because $T^a > 1$, a difference in income between regions u and v results in a disproportionate divergence in the mass of manufacturing firms. This result is analog to the home-market effect (HME) in models of international trade. Moreover, this effect is negatively related to τ and σ.

The long-run closed economy equilibrium is illustrated in Figure 3.1. That depiction has two main components: The first, line (I), represents the regional distribution of income as in (3.9). The second, line (M), pictures the relationship between $s^u_M - s^v_M$ and $(s^u_I - s^v_I)$ drawn from (3.11). The point of intersection of both lines, at the origin of the plan, corresponds to an equilibrium where both income and manufacturers are evenly distributed between regions u and v. A shift in the income distribution towards u, as represented by (I'), leads to a new equilibrium, E'. The regional market effect implies that (M) is steeper that the 45^o-degree line. Hence, at E', the change in the relative income of region u is matched by a disproportionate change in its relative share of manufacturing firms.

3.3 Open Economy Model

3.3.1 Open Economy Setup

In this section, the standard FC model (with a two-region economy) is extended by adding a third location, the foreign country, with which the two-region domestic country exchanges the agricultural and manufacturing goods. Capital is mobile between countries. The foreign country has a household population of mass $L^f > 0$ and is endowed with a capital stock of mass $K^f > 0$.

Preferences and technologies are the same in both countries. Therefore, the foreign representative consumer's demand for variety j is given by $x^f(j) = \left[(1-\alpha)I^f/P^f\right] \times \left(p_f(j)/P^f\right)^{-\sigma}$, where I^f is the foreign income, P^f is the foreign price index, and $p_f(j)$ is the price of variety j in the foreign country. The set of varieties available to foreign households is the same as the set of varieties consumed by domestic households, J. Let r^f and M^f denote the rental rate of capital in the foreign country and the mass of capital invested there (that is, the mass of foreign manufacturers), respectively. The world's masses of labor and capital (or manufacturers) are $L = L^u + L^v + L^f$ and $K = K^u + K^v + K^f = M^u + M^v + M^f$, respectively.

Domestic and foreign agricultural firms can export output at no cost. Thus, the agricultural price is the same in the three locations. On condition that the agricultural good is produced everywhere, wages are equalized across regions and countries.[8] Again, units of the commodity and of labor are chosen so that $w^u = w^v = w^f = p_o = 1$. Domestic manufacturing firms have to pay for an iceberg trade cost to ship goods to the foreign country. The foreign market may or may

[8] Like in the closed economy, for that condition to hold, no location must have enough labor to fulfill the world's demand for the agricultural good.

not be equally accessible to both domestic regions alike. For example, a country may comprise a coastal region, with a port providing convenient access to overseas markets, and a landlocked region (or a region lacking a good transportation infrastructure). To represent this kind of situation, I decompose the export cost into internal and external trade costs. Precisely, for region m's firms, the cost of exporting is $\tau_d^m \tau_f^d$, where $\tau_d^m \geq 1$ and $\tau_f^d > \tau$ are the internal and external components, respectively. The external component of the trade cost not only comprises freight costs and other costs associated with a transaction at distance, but it also captures the effect of commercial policy, that is, customs duties, administrative barriers to trade, and other non-tariff barriers such as regulatory standards. The internal component take into account the transport and distribution costs of getting goods to the shipping point.Additionally, $\max\{\tau_d^u/\tau_d^v, \tau_d^v/\tau_d^u\} \leq \tau$, or, in other words, the difference in foreign market access between regions u and v is no greater than the interregional trade cost. Conversely, to allow for asymmetry in export market access, foreign manufacturers must ship $\tau_d^f \tau_d^m$ units of output, where $\tau_d^f > \tau$, to deliver one unit to region m. The external trade costs τ_d^f and τ_f^d may be unequal if countries impose different policy barriers to trade.

The variable profit that a firm located in region m earns from domestic (local and non-local) sales, gross of the capital outlay, is given by the first term on the right-hand side of (3.5). Similarly to the pricing strategy in the domestic market, the firm sells its product in the foreign market at a price of $p_f^m = \tau_d^m \tau_f^d \beta/\rho$. Thus, its profit from export sales is $\pi_f^m = s_f^m/\sigma$, where s_f^m is given by

$$s_f^m = \frac{(1-\alpha) I^f \left(P^f \rho\right)^{\sigma-1}}{(\tau_d^m \tau_f^d \beta)^{\sigma-1}}.$$

The total profit of the firm, net of the fixed cost, is $\pi^m = \pi_m^m + \pi_n^m + \pi_f^m - r^m$, or, explicitly,

$$\pi^m = \frac{(1-\alpha)\rho^{\sigma-1}}{\sigma\beta^{\sigma-1}} \left[I^m \left(P^m\right)^{\sigma-1} + \tau^{1-\sigma} I^n \left(P^n\right)^{\sigma-1} + (\tau_d^m \tau_f^d)^{1-\sigma} I^f \left(P^f\right)^{\sigma-1} \right] - r^m.$$

By substituting in the price indices and the mark-up pricing formulas, it is rewritten as

$$\pi^m = \frac{(1-\alpha) I}{\sigma K} \left[\frac{s_I^m}{d^m} + \frac{\tau^{1-\sigma} s_I^n}{d^n} + \frac{(\tau_d^m \tau_f^d)^{1-\sigma} s_I^f}{d^f} \right] - r^m,$$

where $I = I^u + I^v + I^f$ is the world income; $s_I^m \equiv I^m/I$ and $s_I^f \equiv I^f/I$ are the income shares of region m and the foreign country; $d^m \equiv s_M^m + \tau^{1-\sigma} s_M^n + (\tau_d^f \tau_d^m)^{1-\sigma} s_M^f$ and $d^f \equiv (\tau_d^m \tau_f^d)^{1-\sigma} s_M^m + (\tau_d^n \tau_f^d)^{1-\sigma} s_M^n + s_M^f$; and $s_M^m \equiv M^m/K$ and $s_M^f \equiv M^f/K$ are the shares of manufacturers in region m and in the foreign country.

Likewise, the profit of a foreign firm derived from sales in its home country and in the domestic country is expressed as

$$\pi^f = \frac{(1-\alpha)I}{\sigma K}\left[\frac{(\tau_d^f \tau_d^m)^{1-\sigma} s_I^m}{d^m} + \frac{(\tau_d^f \tau_d^n)^{1-\sigma} s_I^n}{d^n} + \frac{s_I^f}{d^f}\right] - r^f.$$

3.3.2 Short-Run Equilibrium

Again, given a capital allocation across the three locations of the world, free entry in the manufacturing industry entails zero profits:

$$r^m = \frac{(1-\alpha)I}{\sigma K}\left[\frac{s_I^m}{d^m} + \frac{\tau^{1-\sigma} s_I^n}{d^n} + \frac{(\tau_d^m \tau_f^d)^{1-\sigma} s_I^f}{d^f}\right] \tag{3.13}$$

and

$$r^f = \frac{(1-\alpha)I}{\sigma K}\left[\frac{(\tau_d^f \tau_d^m)^{1-\sigma} s_I^m}{d^m} + \frac{(\tau_d^f \tau_d^n)^{1-\sigma} s_I^n}{d^n} + \frac{s_I^f}{d^f}\right]. \tag{3.14}$$

The free entry condition entails that the world capital income is $R = (1-\alpha)I/\sigma$ and the world income $I = \sigma L/(\sigma - 1 + \alpha)$. As in the closed economy model, to determine regional capital income, it must be assumed that the fraction of the capital invested in any of the three locations that is owned by region m's households, or by country f's households, equals region m's share, or country f's share, respectively, of the world capital stock. That is, $K_n^m/M^n = K^m/K \; \forall (m,n) \in \{u,v\} \times \{u,v\}$, $K_m^f/M^m = K^f/K \; \forall m \in \{u,v\}$, and $K_f^f/M^f = K^f/K$. This assumption implies that region m's capital income and foreign capital income are given by $R^m = s_K^m R = s_K^m(1-\alpha)I/\sigma$ and $R^f = s_K^f R = s_K^f(1-\alpha)I/\sigma$, respectively, where $s_K^f \equiv K^f/K$ is the foreign capital endowment share. In addition, it means that the average capital income perceived by households living in any of the three locations equals the world average capital income, $\bar{r}$, as $R^m/K^m = R/K = (1-\alpha)I/(\sigma K)$ and $R^f/K^f = R/K = (1-\alpha)I/(\sigma K)$. As before, these equalities hold when r^u, r^v, and r^f take the same value, r, which is then equal to $\bar{r}$, or $(1-\alpha)I/(\sigma K)$.

3.3.3 Long-Run Equilibrium

In the long run, capital is reallocated across regions and countries so that its return is everywhere the same. The definition of the long-run equilibria closely follows that in the closed economy.

Definition 2 (Equilibria with capital mobility)**.** Let $\Delta^l r \equiv r^l - \bar{r}$ be the differential between the capital return at location l, $l \in \{u, v, f\}$, and the average capital rental rate worldwide. A world distribution of capital investment (M^u, M^v, M^f) is an equilibrium if

$$-1^{l\{M^l=0\}}\Delta^l r \geq 0 \ \forall l \in \{u, v, f\}, \tag{3.15}$$

where l is the indicator operator and the equality holds if $0 < s_M^l < 1$.

To begin with, I consider the interior equilibrium, where the manufacturing industry is present in all three locations in the long run. This equilibrium is defined by a system of three equations, that is, $r^l = \bar{r} = (1-\alpha)I/(\sigma K)$ for all $l \in \{u, v, f\}$. Using the free entry conditions (4.14) and (4.15), this system is rewritten as

$$\begin{cases} q^m + \tau^{1-\sigma} q^n + \left(\tau_d^m \tau_f^d\right)^{1-\sigma} q^f = 1 \\ \tau^{1-\sigma} q^m + q^n + \left(\tau_d^n \tau_f^d\right)^{1-\sigma} q^f = 1 \\ \left(\tau_d^f \tau_d^m\right)^{1-\sigma} q^m + \left(\tau_d^f \tau_d^n\right)^{1-\sigma} q^n + q^f = 1 \end{cases}$$

where $q^l \equiv s_I^l / d^l$. The equilibrium manufacturing shares s_M^u, s_M^v, and s_M^f solve this system of equations. The solution for the long-run share of manufacturers in region m is given by

$$s_M^m = \left[\left[\left(\tau_d^f \tau_d^n\right)^{1-\sigma} - 1 \right] s_I^m \left(q^m\right)^{-1} + \left[\tau^{1-\sigma} - \left(\tau_d^f \tau_d^m\right)^{1-\sigma} \right] s_I^n \left(q^n\right)^{-1} + \left(\tau_d^f \tau_d^m\right)^{1-\sigma} \right.$$
$$\left. - \left(\tau_d^f \tau_d^n \tau\right)^{1-\sigma} \right] \left[\left(1 - \tau^{1-\sigma}\right) \left[\left(\tau_d^f \tau_d^m\right)^{1-\sigma} + \left(\tau_d^f \tau_d^n\right)^{1-\sigma} - 1 - \tau^{1-\sigma} \right] \right]^{-1}, \tag{3.16}$$

where q^m stands for

$$q^m \equiv \left[\left(\tau_f^d \tau_d^f\right)^{1-\sigma} \left[\left(\tau_d^n\right)^{2(1-\sigma)} - \left(\tau_d^m \tau_d^n\right)^{1-\sigma} \right] + \left(\tau_f^d\right)^{1-\sigma} \left[\left(\tau_d^m\right)^{1-\sigma} - \left(\tau_d^n \tau\right)^{1-\sigma} \right] + \tau^{1-\sigma} \right.$$
$$\left. -1 \right] \left[\left(\tau_f^d \tau_d^f\right)^{1-\sigma} \left[\left(\tau_d^m\right)^{2(1-\sigma)} + \left(\tau_d^n\right)^{2(1-\sigma)} - 2\left(\tau_d^m \tau_d^n \tau\right)^{1-\sigma} \right] + \tau^{2(1-\sigma)} - 1 \right]^{-1}.$$

Equation (3.16) is valid as long as the value of the right-hand side is comprised between zero and one. For some values of the elasticity of substitution, the income shares, and the trade costs, it may actually be less than zero or greater than one. If it is less (greater) than zero (one), then s_M^m is zero (one), which means that none of (all) the manufacturing firms are located in region m. Note that this solution is consistent with that obtained by Baldwin et al. from a three-country model with additional restrictions on the parameters, because if $\tau_d^m = \tau_d^n = 1$ and $\tau_f^d = \tau_d^f = \tau$, then $s_M^l - 1/3 = [(1 + 2\tau^{1-\sigma})/(1 - \tau^{1-\sigma})](s_I^l - 1/3)$.

Since this analysis focuses on the impact of external trade on the distribution of the manufacturing industry between regions u and v, the variable of interest is the difference between the share of firms located in region u and the share of firms located in region v, that is, $s_M^u - s_M^v$. In the next section, I perform comparative statics on this variable with respect to the external trade costs.

3.4 Impact of Trade Liberalization

I shall examine the impact of trade liberalization under different circumstances. First, I assess the effect of a bilateral decline in trade costs when domestic regions are identical with regard to their access to the foreign market, although they differ in terms of market size (case of symmetrical geography). Then, I consider the consequences of unilateral trade liberalization by the domestic country. Third, I evaluate the impact of bilateral trade liberalization when a domestic region has an advantage over the other in accessing the foreign market (case of asymmetrical geography).

3.4.1 Symmetrical Geography

3.4.1.1 Bilateral Trade Liberalization

Suppose that $\tau_d^u = \tau_d^v = 1$ and $\tau_f^d = \tau_d^f = \tau_{ex}$. However, s_I^u may or may not be equal to s_I^v. Then, according to (3.16), the share of manufacturing firms located in region m is

$$s_M^m = \left[\left[1 + \tau^{1-\sigma} - 2\tau_{ex}^{2(1-\sigma)}\right] \left[(1 - \tau_{ex}^{1-\sigma})s_I^m + (\tau_{ex}^{1-\sigma} - \tau^{1-\sigma})s_I^n\right] \right.$$
$$\left. \times \left[(1 - \tau^{1-\sigma})(1 - \tau_{ex}^{1-\sigma})\right]^{-1} - \tau_{ex}^{1-\sigma} \right] \left[1 + \tau^{1-\sigma} - 2\tau_{ex}^{1-\sigma}\right]^{-1}.$$

So therefore, $s_M^u - s_M^v = T(s_I^u - s_I^v)$, where

$$T \equiv \frac{1 + \tau^{1-\sigma} - 2\tau_{ex}^{2(1-\sigma)}}{(1 - \tau^{1-\sigma})(1 - \tau_{ex}^{1-\sigma})}.$$

T can also be expressed as

$$T = \frac{1 + \tau_{ex}^{1-\sigma}}{1 - \tau^{1-\sigma}} + \frac{\tau^{1-\sigma} - \tau_{ex}^{2(1-\sigma)}}{(1 - \tau^{1-\sigma})(1 - \tau_{ex}^{1-\sigma})} > 1. \tag{3.17}$$

One can readily check that the first and second terms on the right-hand side of (3.17) are respectively greater than one and zero (since $\sigma > 1$ and $\tau_{ex} > \tau > 1$) and thus that their sum is greater than one. Moreover, the second term is greater than $(\tau^{1-\sigma} - \tau_{ex}^{1-\sigma})/(1-\tau^{1-\sigma})$. Hence, T is also greater than T^a (cf. equation (3.12)). This means that the existence of an external trading partner makes the regional market effect stronger. An uneven distribution of income between domestic regions gives rise to a more imbalanced spatial distribution of the manufacturing industry than in the closed economy.

The derivative of T with respect to τ_{ex} indicates the direction of the impact of bilateral trade liberalization on the spatial distribution of manufacturing firms:

$$\frac{\partial T}{\partial \tau_{ex}} = -(\sigma - 1)\frac{\tau_{ex}^{-\sigma}\left(1 + \tau^{1-\sigma} - 4\tau_{ex}^{1-\sigma} + 2\tau_{ex}^{2(1-\sigma)}\right)}{\left(1 - \tau^{1-\sigma}\right)\left(1 - \tau_{ex}^{1-\sigma}\right)^2}.$$

This derivative is positive if $1+\tau^{1-\sigma} - 4\tau_{ex}^{1-\sigma} + 2\tau_{ex}^{2(1-\sigma)} < 0$; that is, if $\underline{\tau}_{ex} < \tau_{ex} < \overline{\tau}_{ex}$, where

$$\underline{\tau}_{ex} \equiv \left[1 + \left(\frac{1-\tau^{1-\sigma}}{2}\right)^{1/2}\right]^{1/(1-\sigma)} \text{ and } \overline{\tau}_{ex} \equiv \left[1 - \left(\frac{1-\tau^{1-\sigma}}{2}\right)^{1/2}\right]^{1/(1-\sigma)}.$$

Note that $\lim_{\tau \to 1} \underline{\tau}_{ex} = 1$, $\lim_{\tau \to +\infty} \underline{\tau}_{ex} = \left[1 + (1/2)^{1/2}\right]^{1/(1-\sigma)} < 1$, and $\underline{\tau}_{ex}$ is monotonically decreasing with τ. Hence, given that $\tau_{ex} > \tau > 1$, it is always the case that $\tau_{ex} > \underline{\tau}_{ex}$. On the other hand, $\lim_{\tau \to 1} \overline{\tau}_{ex} = 1$ and $\lim_{\tau \to +\infty} \overline{\tau}_{ex} = \left[1 - (1/2)^{1/2}\right]^{1/(1-\sigma)} > 1$. Moreover, $\overline{\tau}_{ex}$ increases with τ although it is concave in τ. Thus, when τ is relatively small, it is possible for τ_{ex} to be smaller than $\overline{\tau}_{ex}$ while still being greater than τ. If it is the case, then the derivative of T is positive. Otherwise, when τ_{ex} is greater than $\overline{\tau}_{ex}$, the derivative of T is negative.

The impact of bilateral trade liberalization depends on the initial magnitude of trade costs. It dampens down the regional market effect when both interregional and international trade costs are relatively low. As the domestic economy becomes more open to trade, domestic producers export a larger share of their output abroad, domestic households shift their expenditures toward foreign goods, and thus the influence of the relative size of regions declines. Whether the domestic economy is large relative to the foreign economy, in which case the former attracts capital investment/firms from abroad, or small, the agglomeration force wanes relative to the dispersion force and a spatial reallocation of capital and firms occurrs in favor of the smaller region. This case matches the finding of Krugman and Livas-Elizondo [1996]. Figure 3.1 depicts the equilibrium in the open economy with symmetrical geography and the impact of bilateral trade liberalization when barriers to international trade are relatively low. The absolute value of the relative regional income share is at most equal to the domestic income share, s_I^d. Likewise, the absolute value of the relative regional industry share is at most equal to the domestic share of the manufacturing industry, s_M^d. The broken and solid lines (M) and (M') represent the relationship between the relative regional income and manufacturing shares before and after trade liberalization, respectively. For any level of interregional trade cost, these two lines would be steeper than the line in autarky. The Figure shows that a decline in the external trade cost induces a reallocation of manufacturers from the large region, u, toward the small region, v. Moreover, the relative regional expenditure share at which manufacturing industries are full agglomerated in region u rises from $T^{-1}s_M^d$ to $(T')^{-1}(s_M^d)'$.

Then again, a decrease in the external trade cost amplifies the regional market effect when this cost (or trade costs in general) is (are) high. Under these circumstances, a greater exposure to trade leads to a stronger agglomeration force in the domestic country. This effect would be described in Figure 3.1 by a counterclockwise rotation of line (M). This case well agrees with the conclusions of Monfort and Nicolini [2000] and Paluzie [2001].

3.4.1.2 Unilateral Trade Liberalization

As before, let $\tau_d^u = \tau_d^v = 1$, but now, $\tau_f^d \neq \tau_d^f$. In this case, the difference in the share of manufacturing firms between regions u and v is given by $s_M^u - s_M^v = T(s_I^u - s_I^v)$, where

$$T \equiv \frac{1 + \tau^{1-\sigma} - 2\left(\tau_f^d \tau_d^f\right)^{1-\sigma}}{\left(1 - \tau^{1-\sigma}\right)\left(1 - \left(\tau_f^d\right)^{1-\sigma}\right)}. \tag{3.18}$$

Like T in (3.17), that in (3.18) is greater than T^a.

The derivative of T with respect to τ_d^f gives the direction of the impact of unilateral trade liberalization on the distribution of manufacturing firms between regions u and v:

$$\frac{\partial T}{\partial \tau_d^f} = (\sigma - 1)\frac{2\left(\tau_f^d\right)^{1-\sigma}\left(\tau_d^f\right)^{-\sigma}}{\left(1 - \tau^{1-\sigma}\right)\left(1 - \left(\tau_f^d\right)^{1-\sigma}\right)}.$$

Clearly, this derivative is positive for all values of the trade cost parameters (as long as $\tau > 1$, $\tau_f^d > \tau$, and $\tau_d^f > \tau$). Hence, unilateral trade liberalization reduces the regional market effect. A reduction in the cost of importing foreign goods lowers their price and leads domestic consumers to reallocate their expenditures toward these goods.

As a result, the agglomeration force declines relative to the dispersion force and, if $s_I^u > s_I^v$, that induces a relocation of producers from region u toward region v.

3.4.2 Asymmetrical Geography

In the asymmetrical case, let u be the coastal region and v the inland region. To make the analysis straightforward, I assume that $\tau_d^u = 1$, $\tau_d^v = \tau$, $\tau_f^d = \tau_d^f = \tau_{ex}$, and that the regions have the same size, $s_I^u = s_I^v = s_I^d/2$, where s_I^d denotes the income share of the domestic country. The expression of the difference in firm share between u and v is

$$s_M^u - s_M^v = \frac{\tau_{ex}^{1-\sigma}}{1-\tau_{ex}^{1-\sigma}} \left[\frac{3 + \tau^{2(1-\sigma)} - 4\left(\tau\tau_{ex}\right)^{1-\sigma}}{\left(1-\tau^{1-\sigma}\right)\left[1-\left(\tau\tau_{ex}\right)^{1-\sigma}\right]} \frac{s_I^d}{2} - 1 \right].$$

Since the coefficient of $s_I^d/2$ is positive (because $\tau_{ex} > \tau > 1$), s_M^u is greater than s_M^v if and only if $s_I^d > \underline{s}_I^d \equiv [2(1-\tau^{1-\sigma})(1-(\tau\tau_{ex})^{1-\sigma})]/[3+\tau^{2(1-\sigma)} - 4(\tau\tau_{ex})^{1-\sigma}]$. In other words, the income share of the domestic country must be relatively large for the majority of the domestic manufacturing sector to be located in the coastal region. The derivative of $\underline{s}_I^d$ with respect to τ_{ex} is given by

$$\frac{\partial \underline{s}_I^d}{\partial \tau_{ex}} = -2(\sigma-1)\frac{\tau^{1-\sigma}\left(1-\tau^{1-\sigma}\right)\left(1-\tau^{2(1-\sigma)}\right)\tau_{ex}^{-\sigma}}{\left[3+\tau^{2(1-\sigma)} - 4\left(\tau\tau_{ex}\right)^{1-\sigma}\right]^2} < 0.$$

This derivative is negative, which indicates that, as the domestic and foreign economies become more integrated through trade, $\underline{s}_I^d$ rises. Thus, in the course of trade liberalization, a shift in the spatial distribution of domestic manufacturers may occur, resulting in the majority of them to be located in the landlocked region.

To interpret this finding in an informal way, one can argue that, in a relatively large domestic country, manufacturers prefer to locate in the coastal region, u, which gives them an advantage to serve the markets of all three locations, including a large local market, while being subject to a moderate foreign competition, since the foreign country is relatively small and thus has a small industry. Conversely, in a relatively small domestic country, the coastal location gives direct access to only a small local market but is exposed to a strong foreign competition, since the other country boasts a large industry. Thus, the majority of domestic manufacturers retreat in the inland region, v, where they are sheltered from foreign competition.

This finding highlights the fact that, insofar as internal geography is concerned, a greater exposure to external trade can yield different outcomes for different countries. It is consistent with the model of Behrens et al. [2006], which predicts that remoteness from the foreign market does not necessarily constitute a disadvantage, since it may make it costlier to export (and import) goods, but then again it may offer protection from foreign competitors.[9]

3.5 Conclusion

A simple model of new economic geography, the footloose capital model (FC), was implemented to analyze the impact of external trade on the internal geography of manufacturing activities. This analysis demonstrated the following propositions: (1) when the domestic country is relatively poorly integrated through trade with the foreign country, bilateral trade liberalization tends to increase the disproportion in the mass of the manufacturing industry between domestic regions of different sizes; however, when barriers to international trade are low, it induces a reallocation of manufacturing firms in favor of the small region; (2) unilateral liberalization unambiguously reduces the interregional disparity in the size of the manufacturing industry; (3) and when one region has an advantage over the other in serving the foreign market and importing foreign goods, the effect of bilateral liberalization depends on the size of the domestic country; if it is relatively small, trade liberalization may cause a shift of the manufacturing industry from the coastal to the inland region; but if the country is large, the largest share of the industry remains in the coastal region.

According to this approach, the location of firms is driven by demand side factors. The extent of the market differs across locations, because of differences in endowments of capital and labor, and due to the fact that the geographic concentration of firms determines the degree of competition for market shares. This analytical framework differs from that of chapter one of this dissertation as the latter differentiates regions with regard to production costs. Behind the assumption of spatial differences in production costs lies the idea that the level of economic development is heterogeneous across locations, in terms of urbanization, presence of industries supporting manufacturing activities (business services for intance), transaction costs, and infrastructure. These disparities have implications for the supply-side factors of the location decision.

The open economy version of the FC model may be useful to study the interaction between commercial and regional policies. For instance, one may want to determine optimal tax and fiscal policies with the objective to mitigate the effect of trade liberalization on the geographical reallocation of firms. Specifically, the implementation of different levels of subsidization of the capital investment cost or income tax could

[9] In their model with linear demand functions and without income effects, the magnitude of the external trade cost relative to the internal transport cost is the determinant factor.

prevent firm relocation so as to avoid the loss of employment in a small or landlocked region.

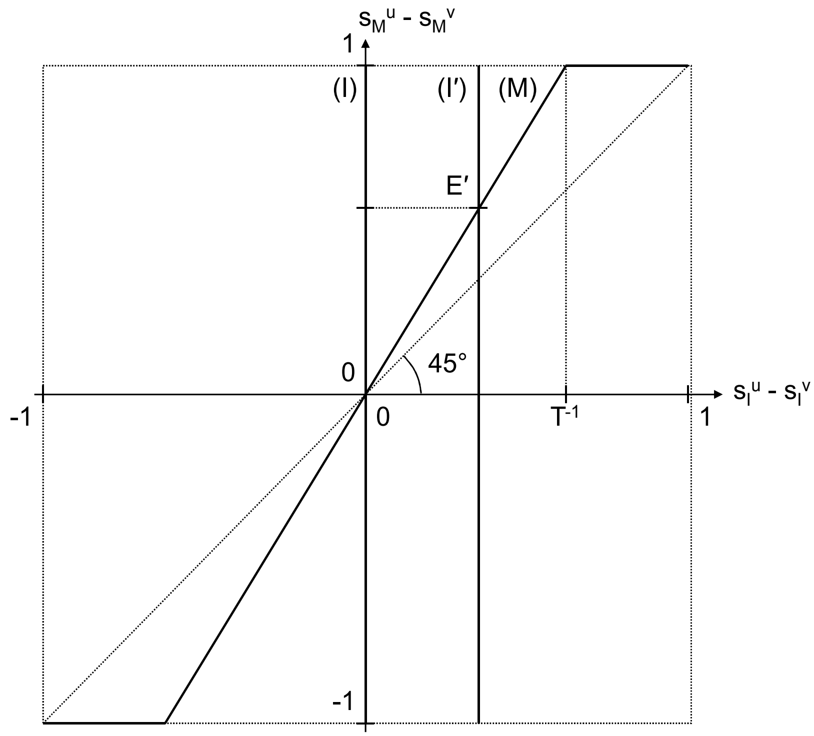

Figure 3.1: Closed Economy Equilibrium

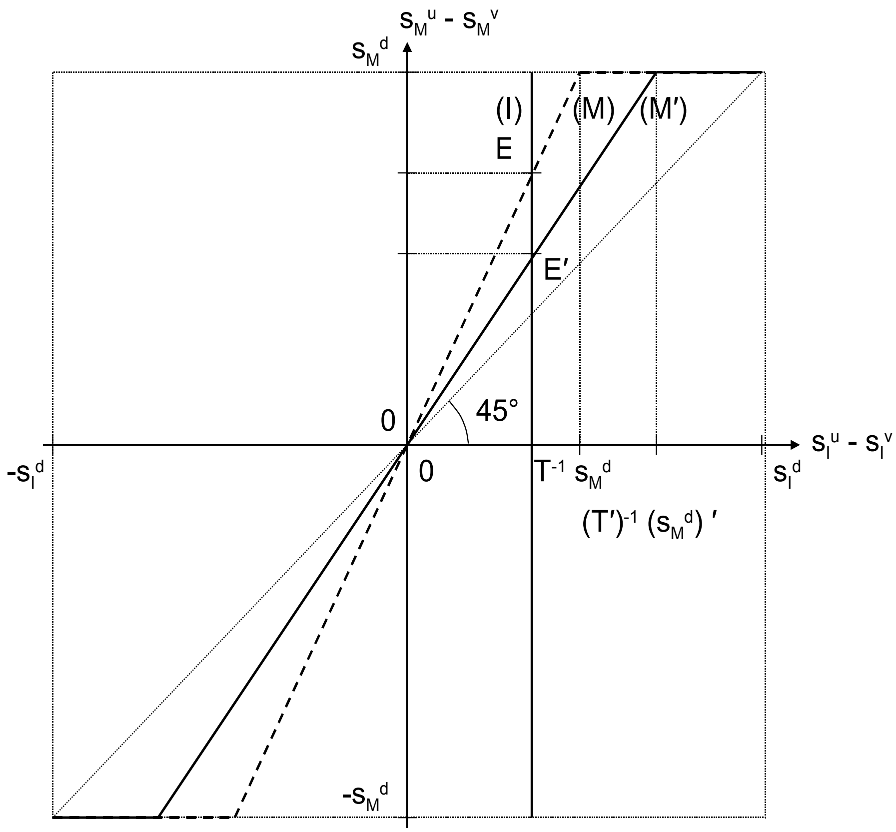

Figure 3.2: Open Economy Equilibrium—Case of Symmetrical Geography

Chapter 4 The Impact of Trade Liberalization on Productivity Across Regions

4.1 Introduction

A number of theoretical and empirical studies establish a link between international trade and economic geography or the location of manufacturing activities within countries. This paper focuses on a related aspect of the link between trade and geography by examining how trade liberalization affects industry productivity across regions.

A critical issue in the debate about the benefits from commercial policy reform has revolved around its effect on productivity growth in the manufacturing sector. The trade policy reforms implemented by several countries of Latin America and other emerging economies in the recent past have provided researchers with opportunities to assess the impact of an increase in trade exposure on productivity at the levels of manufacturing establishments and industries. Empirical studies showed that falling import barriers and improvements in export market access favor the international diffusion of productivity-enhancing technologies such as technologies embodied in capital goods and intermediate inputs, encourages some plants to invest in more productive technologies (Ederington and McCalman [2007]) and skill upgrading, and induces within-industry reallocations of market shares and the exit of the least productive firms (Pavcnik [2002]).

Although these studies shed light on the many ways in which a greater trade openness affects the performance of plants and industries, they did not examine how producers' response to a change in a trade policy would differ across regions within countries. Yet, the countries that have been studied, like many others, exhibit an uneven distribution of economic activities over their territory. They feature regional differences in terms of population density, urbanization, infrastructure quality, and proximity to ports and borders, which imply differences in market access. Furthermore, other empirical studies revealed that the size of a geographic market has significant implications for the mean, variance, and other characteristics of the distribution of firm- or establishment-level variables such as price, output, employment, and productivity. For instance, Campbell and Hopenhayn [2005] found that U.S. retail establishments located in larger regional markets command larger sales and employment. Syverson [2004, 2007] showed that plants producing cement are larger and more productive and charge lower prices, on average, in larger U.S. regional markets. In addition, he found that the least productive establishments in larger markets are more productive than the least productive ones in smaller markets. Such evidence raises the question whether regional economies with different geographic attributes within a country respond in different ways to a shift in trade openness.

This paper builds on the Melitz and Ottaviano [2008] model of trade with monopolistic competition and heterogeneous firms to look into the link between trade openness, regional differences in market access, and regional difference productivity. In Melitz and Ottaviano [2008], firms are endowed with different productivity levels (that is, marginal costs), as in Melitz [2003], which induces a self-selection of firms in the domestic and export markets. But the demand for the varieties of the differentiated good are linear in their own price, as in Ottaviano et al. [2002]. In this context, more productive firms can charge lower prices and firms must impose smaller mark-ups when the market is larger and the number of competitors is greater. Low-production firms cannot survive in this environment and they exit the industry. Thus, market size affects the equilibrium distribution of productivity levels; larger markets exhibit higher average productivity and larger firms (in terms of output or sales), and a more variety. In Melitz and Ottaviano [2008], the impact of trade liberalization on firm selection in an industry occurs through product market competition (not labor market competition as in the Melitz [2003] model with a constant price elasticity of demand) and the size of the domestic economy matters for the response of aggregate industry performance to a shift in exposure to trade with a foreign country. This paper applies the Melitz and Ottaviano [2008] model by considering a domestic economy made of two locations as in Krugman and Livas-Elizondo [1996] and trading with another country. The domestic population is distributed between these two locations in some proportions. This framework allows one to examine how trade liberalization affects relative regional aggregate industry performance given an allocation of the population between the two regions. In the empirical section, I use data from Colombian manufacturing establishments to assess the dependence of aggregate industry productivity growth on tariff protection, the size of the regional market, and the geographical situation of regions.

The remainder of the paper is organized a follows. Section 2 presents a closed economy model with two regions. Section 3 works out the open economy model. Section 4 investigates the impact of trade liberalization on aggregate industry performance across regions. Section 5 concludes.

4.2 Closed Economy Model

The economy of a country consists of two sectors employing labor as the sole factor of production. One sector produces a homogeneous good under constant returns to scale and perfect competition. In the other sector, the manufacturing industry, firms makes a horizontally differentiated good using an increasing returns to scale technology. This industry is characterized by monopolistic competition among a continuum of firms.

The country is populated by a continuum of households and comprises two regions, u and v. A mass of households equal to $L^m > 0$ consumes goods and supply labor services in region m, where $m \in \{u, v\}$ labels either region. Every household supplies one unit of labor. While households are tied to their region, the commerce of goods allows them to consume the agricultural commodity and the differentiated product originating from the other region.

4.2.1 Demand

Households have the same preferences, in both locations. The preferences of household k from region m, where $k \in [0, L^m]$ indexes households, are described by a quasi-linear utility function defined as the sum of the consumption of the homogeneous good[1] and a quadratic sub-utility function defined over a continuum of potential varieties of the differentiated product: [2]

$$u^k = x_o^k + \alpha \int_{j \in J_m} x^k(j)dj - \frac{\beta}{2} \int_{j \in J_m} \left(x^k(j)\right)^2 dj - \frac{\chi}{2} \left[\int_{j \in J_m} x^k(j)dj\right]^2,$$

where x_o^k is household k's consumption of the homogeneous good; j indexes potential varieties and J_m denotes the set of varieties available at location m; $x^k(j)$ is the quantity of variety j consumed; and the parameters α, β, and χ are strictly positive numbers. The parameters α and χ measure the desirability of the differentiated product relative to that of the homogeneous good. A greater value of α (χ) means higher (lower) demand for the differentiated product relative to the demand for the homogeneous good. The coefficient β parameterizes the preference for variety in the consumption of the differentiated product. The higher β, the less substitutable the varieties are between themselves. Under certain conditions, variations in income may not affect the demand for the differentiated good since the utility function is quasi-linear.

[1] The Dixit and Stiglitz [1977] model of monopolistic competition views products in an industry as well substitutable between themselves although being poor alternatives for goods of other industries. The role of the homogeneous good is to aggregate these other goods into one.

[2] This form of utility function was introduced by Dixit [1979].

Household k maximizes utility subject to the budget constraint

$$p_{o,m}x_o^k + \int_{j\in J_m} p_m(j)x^k(j)dj = p_{o,m}e_o + w^m,$$

where $p_{o,m}$ and $p_m(j)$ are the prices of the homogeneous good and variety j in region m, respectively, and e_o is the quantity of the homogeneous good with which every household is endowed.[3] Let the price of the homogeneous good be normalized to one, making this good the numéraire. Thus, household k's inverse demand for variety j is given by

$$p_m(j) = \alpha - \beta x^k(j) - \chi \int_{j\in J_m} x^k(j)dj.$$

The demand is obtained by integrating the inverse demand over the set J_m, solving for $\int_{j\in J_m} x^k(j)dj$, and substituting the expression of the latter into the inverse demand equation. From there, one obtains the market demand for variety j in region m by multiplying the household's demand by L^m, as follows:

$$x^m(j) = L^m\left(a^m - bp_m(j) + c^m\bar{p}^m\right), \tag{4.1}$$

where $a^m \equiv \alpha/(\beta + M_m\chi)$, $b \equiv 1/\beta$, and $c^m \equiv M_m\chi/\big(\beta(\beta + M_m\chi)\big)$; M_m is a measure of J_m, or the mass of manufacturing products supplied in region m; and $\bar{p}^m = \int_{j\in J_m} p_m(j)dj/M_m$ is the average price of the differentiated good at location m. Note that $x^m(j) > 0$ if and only if $p_m(j) < \hat{p}^m \equiv (a^m + c^m\bar{p}^m)/b$. Hence, if the price of variety j is greater than or equal to $\hat{p}^m$, the threshold price, then there is no demand for this variety. The price elasticity of demand is given by $\big(\hat{p}^m/p_m(j)-1\big)^{-1}$. Thus, the threshold price and the elasticity of demand increase with the own price and decrease with the average price; they also vary with the number of competing products.

The indirect utility function quantifies household welfare:

$$v^k = e_o + w^m + \frac{M_m}{2}\left(\frac{a^m}{\alpha}\left(\alpha - \bar{p}^m\right)^2 + b\left(\sigma_p^m\right)^2\right), \tag{4.2}$$

where w^m denotes the wage rate in region m and $\left(\sigma_p^m\right)^2 = \int_{j\in J_m}\big(p_m(j)-\bar{p}^m\big)^2 dj/M_m$ is the variance of manufacturing prices at this location. According to the inverse demand function, any variety consumed by households is bought at a price less than α, which entails $\bar{p}^m < \alpha$. Hence, all else constant, household welfare increases when the average

[3] As long as the homogeneous good endowment is sufficiently large, the consumption of it in equilibrium is non-zero. Otherwise, since the marginal utility of both the homogeneous good and the differentiated product is bounded, demand for either of the two may be zero.

price falls. Additionally, welfare increases with the price variance as households may substitute the numéraire and less expensive varieties for more expensive ones to a greater extent. Greater variety in the differentiated product also improves welfare.

4.2.2 Production

If one assumes that the production of one unit of the homogeneous good requires one unit of labor (as the definition of units is arbitrary), then perfect competition entails that the good's price in region m equals the wage rate there. Free interregional trade in the homogeneous good equates the price in region u to that in v. Furthermore, as long as both regions produce this good, free trade indirectly equalizes wages between locations. Under these circumstances, since the homogeneous is also the numéraire, $w^u = w^v = 1$.

The production of manufactures exhibits scale economies as entry into the manufacturing industry requires a fixed outlay. However, once a firm joined that industry, it can produce a variety of the differentiated product with a constant returns to scale technology. The cost function of incumbent firm j located in region m is given by

$$l(j) = \frac{y(j)}{\theta(j)},$$

where $y(j)$ denotes firm j's output and $\theta(j)$ parameterizes its productivity level ($\theta(j) > 0$). In the absence of scope economies, every firm produces only one variety;[4] and every variety originates from a single firm.[5] Since all varieties enter the utility function in a symmetric way and all firms are identical in every respect besides the level of productivity, let the θ variable index both of them in what comes next.

Firms can deliver output to households living in the same region (the local market) without incurring any trade cost. However, they must bear a per-unit trade cost to serve dwellers of the other region (the non-local market). This cost takes the form of an iceberg trade cost, so that firms must ship τ units of output, where $\tau > 1$, to deliver one unit of it to the other region's market. In effect, this cost shifts the schedule of the marginal cost of supplying the market; so firms maximize the profit

[4]Intuitively, a multi-product firm would impose a mark-up higher than that set by a single-product manufacturer, since varieties are substitutes for each other. In this situation, single-product competitors would undercut the multi-product firm. Hence, firms do not actually have an incentive to produce more than one variety.

[5]If several firms produced a variety, then they would behave as an oligopoly. For instance, if there were two firms producing the same variety, then, whether Cournot or Bertrand competition prevailed, both would earn a profit smaller than that obtained by a firm enjoying a monopoly for that product.

from sales in their region independently of their profit-maximizing price and output decision for sales in the other region.

The operating profit that a firm from region m with productivity θ earns from sales in its local market is

$$\pi_m^m(\theta) = \left(p_m^m(\theta) - \frac{1}{\theta}\right) y_m^m(\theta), \tag{4.3}$$

where p_m^m is the price charged to region m's households and y_m^m is the output quantity (the superscript and subscript indicate the origin and destination location(s), respectively). The optimal price is determined by substituting the demand equation (4.1) into the local profit function (4.3) and then maximizing the resulting expression with respect to $p_m^m(\theta)$. One must proceed by assuming that, from the perspective of a single firm, the mass of firms serving the local market and the average price are exogenous. This assumption is justified by the fact that there is a continuum of firms and one firm alone does not influence those aggregate variables (that is characteristic of monopolistic competition). Formally, the derivative of (4.1) with respect to $p_m^m(\theta)$ is just equal to $-L^m b$. Hence, in the local market, the profit-maximizing price is

$$p_m^m(\theta) = \frac{1}{2}\left(\hat{p}^m + \frac{1}{\theta}\right). \tag{4.4}$$

The optimal output level is obtained by substituting the price rule (4.4) into (4.1):

$$y_m^m(\theta) = \frac{L^m b}{2}\left(\hat{p}^m - \frac{1}{\theta}\right). \tag{4.5}$$

By substituting the optimal output (4.5) and (4.4) into (4.3), one obtains the maximal local profit level:

$$\pi_m^m(\theta) = \frac{L^m b}{4}\left(\hat{p}^m - \frac{1}{\theta}\right)^2. \tag{4.6}$$

Similarly, the operating profit derived from sales in the non-local market is

$$\pi_n^m(\theta) = \left(p_n^m(\theta) - \frac{\tau}{\theta}\right) y_n^m(\theta),$$

where $(m, n) \in \{u, v\}$, p_n^m is the price charged to region n's households, and y_n^m is the sales volume. The optimal price and volume for sales in the non-local market and the maximal value of the non-local profit are given by

$$p_n^m(\theta) = \frac{1}{2}\left(\hat{p}^n + \frac{\tau}{\theta}\right), \; y_n^m(\theta) = \frac{L^n b}{2}\left(\hat{p}^n - \frac{\tau}{\theta}\right), \text{ and } \pi_n^m(\theta) = \frac{L^n b}{4}\left(\hat{p}^n - \frac{\tau}{\theta}\right)^2.$$

In both the local and non-local markets, the price decreases with the productivity level of the firm. However, both the absolute and relative margins and output increase with productivity. Hence, a firm with a higher productivity level earns a greater profit. In addition, the optimal price is unambiguously increasing with α and the regional average price level. All else equal, firms net higher profits from sales in a larger region, in terms of population, although the size of the regional market may also affect the regional average price level and the mass of firms supplying that market, in equilibrium. This makes its overall effect on profits *a priori* unpredictable.

4.2.3 Firm Entry and Exit

Consider a dynamic economy with an infinite time horizon. Production and consumption take place in every time period, while preferences and technology are constant over time. In every time period, in each location, some firms enter the manufacturing industry and some exit. Entrants originate from a continuum of identical, potential firms, or "entrepreneurs". To enter the industry, these entrepreneurs have to incur a sunk cost of f_e units of local labor. The determination of the firm-specific productivity parameter follows the Hopenhayn-Melitz modeling of heterogeneous firms (see Melitz 2003). Upon paying the entry fee, an entrepreneur from region m randomly draws a productivity level θ from a distribution given by a continuous p.d.f. g, with support (θ_0, ∞) and c.d.f. G. Once the entrepreneur—now a firm—learned its productivity level, she decides whether to produce the good in the current and future periods or to opt out of the industry without producing if its productivity level is too low.[6] Explicitly, the firm remains in the industry if it is sufficiently productive to sell output in the local market at a price lower than the local threshold price but higher than its marginal cost, that is, if $1/\theta \leq p_m^m(\theta) \leq \hat{p}^m$. Hence, the lowest productivity level among firms remaining in the industry, or zero-cutoff productivity level, is determined by $\theta^m = \inf\{\theta : \theta \geq \theta_0 \text{ and } p_m^m(\theta) \leq \hat{p}^m\}$, which entails $p_m^m(\theta^m) = \hat{p}^m$, or $\hat{p}^m = 1/\theta^m$ (see (4.4)).[7]

Similarly, the firm participates in the non-local market if it can sell output in region n at a price lower then that region's threshold price but still higher than its marginal cost, that is, if $\tau/\theta \leq p_n^m(\theta) \leq \hat{p}^n$. Thus, the lowest productivity level among

[6] Several empirical studies (for instance, dunne et al. 1989 and Baily et al. 1992) find that newer plants are more likely to shut down than older establishments. This evidence suggests that entrants may face greater risk or uncertainty about the demand for their product or their costs of production and marketing, while they must incur sunk costs as they enter the industry. In this model, low-productivity firms exit the industry before they even start to produce.

[7] The presence of a fixed cost of production is unnecessary to induce firm selection based on productivity. Low-productivity firms cannot lower their price below the threshold without making a negative profit. Thus, they have no incentive to stay in the industry upon entry. A similar argument holds for the participation in the non-local market.

firms serving both the local and non-local markets, or non-local-cutoff productivity level, is given by $\theta_n^m = \inf\{\theta : \theta \geq \theta_0 \text{ and } p_n^m(\theta) \leq \hat{p}^n\}$, which implies $p_n^m(\theta_n^m) = \hat{p}^n$, or $\hat{p}^n = \tau/\theta_n^m = 1/\theta^n$ (from the expression of the non-local price).

Firms are forever sited in the region where they entered the industry. Although their lifespan can be infinite incumbent firms face the possibility of forced liquidation with probability δ in every period.[8]

Like Helpman et al. [2004], Melitz and Ottaviano [2008], and others, I assume that the distribution of the productivity level takes the form of a Pareto distribution. Let the random variable Θ stands for the level of productivity. The distribution of Θ is Pareto with minimum value θ_0 and shape parameter κ if it is defined as

$$G(\theta) = P(\Theta < \theta) = 1 - \left(\frac{\theta_0}{\theta}\right)^{\kappa}, \text{ where } \theta \in [\theta_0, +\infty) \text{ and } \kappa \in \mathbb{R}_+.$$

The Pareto distribution is skewed to the right, that is, most of the probability mass lies to the right of the modal value. A smaller value of κ indicates a more uniform distribution. However, the variance of Θ is finite if and only if $\kappa > 2$ (par). If the distribution is truncated to the left, it remains a Pareto distribution with shape parameter κ. As noted by Chaney [2008], the primary purpose of this assumption is to allow one to conveniently solve a model with heterogeneous firms. Nevertheless, studies have found that the Pareto distribution is a fairly good description of the observed size distribution of firms (see Axtell [2001] and Luttmer [2007] for evidence from the United States) and total factor productivity distribution (see Gatto et al. [2006] for evidence from manufacturing industries in Europe).

The exogenous distribution of productivity levels, g, along with the probability of actual entry in the industry in location m, $1 - G(\theta^m)$, determine the equilibrium productivity distribution. As the probability that a firm be constrained to cease its activities, δ, is independent of its productivity level, the exit of incumbents does not alter the endogenous productivity distribution. Thus, this distribution is defined by the p.d.f. g conditional on actual entry:

$$\mu^m(\theta) = \begin{cases} \frac{g(\theta)}{1-G(\theta^m)} = \kappa\frac{(\theta^m)^\kappa}{\theta^{\kappa+1}} & \text{if } \theta \geq \theta^m, \\ 0 & \text{otherwise.} \end{cases}$$

The decision of entrants to remain in the industry or to leave it determines the zero-cutoff productivity level, θ^m, which then shapes the equilibrium distribution of productivity levels and thus affects the average productivity level. This analysis is

[8]Involuntary exit may be interpreted as the result of an adverse shock caused by, for instance, unforeseen changes in market conditions.

concerned with steady-state equilibria, for which aggregate variables are constant over time.

Using the Pareto distribution, the mean productivity of region m manufacturers can be expressed as

$$\bar{\theta}^m = \frac{\kappa}{\kappa - 1}\theta^m$$

and the average price of manufacturing products sold in this region as

$$\bar{p}^m = \frac{2\kappa + 1}{2(\kappa + 1)}\frac{1}{\theta^m}. \tag{4.7}$$

Regional average productivity and price are increasing and decreasing functions of the zero-cutoff productivity level, respectively.

4.2.4 Equilibrium

4.2.4.1 Cutoff Profit Conditions

Since π_m^m monotonically increases with productivity, the least productive firm in activity must make a zero profit, or $\pi_m^m(\theta^m) = 0$. This is the zero-cutoff profit condition and it is equivalent to $\hat{p}^m = 1/\theta^m$ (see (4.6)). This condition is the same as that derived from the definition of the zero-cutoff productivity level. Since the firm with productivity θ^m charges the threshold price in the local market, the demand for its product is nonexistent and thus it neither earns nor lose anything. The zero-cutoff profit condition determines the mass of manufacturing firms (from the definition of $\hat{p}^m$, a^m, b, and c^m):

$$M_m = \frac{\beta}{\chi}\frac{\alpha - 1/\theta^m}{1/\theta^m - \bar{p}^m}.$$

Substituting the expression of $\bar{p}^m$ previously derived into that of M_m yields

$$M_m = 2\frac{\beta}{\chi}(\kappa + 1)\left(\alpha\theta^m - 1\right). \tag{4.8}$$

Thus, there is a positive relationship between the zero-cutoff productivity level and the mass of firms participating in region m's market.

The non-local profit of the firm with the lowest level of productivity among those firms shipping goods to both the local and non-local market must be zero too. The non-local-cutoff profit condition, $\pi_n^m(\theta_n^m) = 0$, is equivalent to $\hat{p}^n = \tau/\theta_n^m$, which is like the condition obtained fron the definition of the non-local-cutoff productivity level. The non-local-cutoff profit condition for region n together with the zero-cutoff profit condition for region m imply the following relationship between the cutoff productivity levels θ^m and θ_m^n:

$$\theta_m^n = \tau\theta^m. \tag{4.9}$$

Hence, to ship goods to region m, manufacturers from region n must be at least τ times more productive than the producers from the destination region having the lowest productivity level.

The analysis relies on the assumption that households consume a non-negative amount of numéraire. For this to be the case, they must spend less than all of their income on the manufacturing good. In other words, the aggregate revenue of regions m and n's firms from sales in the former must be strictly smaller than the aggregate income of region m's households: $M_m \bar{p}^m \bar{y}_m < L^m(1+e_o)$, where $\bar{y}_m$ denotes the average volume of sales in region m's market, over all firms serving this market and $\bar{y}_m = (L^m b/2)[1/(\kappa+1)](1/\theta^m)$. That condition is equivalent to this one: $(\alpha\theta^m - 1)/(\chi(\theta^m)^2) < [2(\kappa+1)/(2\kappa+1)](1+e_o)$, which says that α (χ) must be relatively small (large), or that the preference for differentiated should be sufficiently mild.

4.2.4.2 Free Entry Condition

The present value of the average profit flow of firms from region m is $\sum_{t=0}^{\infty}(1-\delta)^t \bar{\pi}^m = \bar{\pi}^m/\delta$, where $\bar{\pi}^m$ is calculated as follows:

$$\bar{\pi}^m = \frac{1}{M^m}\left[\int_{\theta^m}^{\infty} \pi_m^m(\theta) M^m \mu^m(\theta) d\theta + \int_{\theta_n^m}^{\infty} \pi_n^m(\theta) M^m \mu^m(\theta) d\theta\right].$$

Thus, in region m, the expected value of future profits net of entry cost, or, value of entry in the industry, v_e, is given by

$$v_e(\theta^m, \theta_n^m) = [1 - G(\theta^m)]\,\frac{\bar{\pi}^m}{\delta} - f_e.$$

No firm would have an incentive to enter the industry if v_e were taking a negative value. In this case, region m would specialize in the production of the numéraire (that does not occur provided that α be sufficiently large). Otherwise, given the infinite pool of potential firms, free entry keeps the value of entry from exceeding zero in equilibrium. Therefore, free entry imposes the following condition:

$$\int_{\theta^m}^{\infty} \pi_m^m(\theta) g(\theta) d\theta + \int_{\theta_n^m}^{\infty} \pi_n^m(\theta) g(\theta) d\theta = \delta f_e.$$

With g taking the form of the Pareto distribution, one can express the free entry condition as

$$\frac{L^m}{(\theta^m)^{\kappa+2}} + \frac{L^n \tau^2}{(\theta_n^m)^{\kappa+2}} = c,$$

where $c \equiv 2\delta f_e \beta(\kappa+1)(\kappa+2)/\theta_0^\kappa$. Since there are two locations, there is a system of two free entry conditions. Using the relationship among the cutoff productivity levels (4.9), this system of equations ($(m,n) \in \{u,v\}$) is rewritten as

$$\frac{L^m}{(\theta^m)^{\kappa+2}} + \frac{L^n}{\tau^\kappa (\theta^n)^{\kappa+2}} = c. \tag{4.10}$$

4.2.4.3 Equilibrium Determination

The zero-cutoff productivity levels in regions u and v solving the system of equations (4.10) are

$$\theta^m = \left[\frac{L^m}{c}\left(1+\tau^{-\kappa}\right)\right]^{1/(\kappa+2)}. \tag{4.11}$$

For selection to happen when firms enter the industry, θ_0 must be relatively small so that $\theta^m \geq \theta_0$. There is a positive relationship between regional market size and the minimum productivity level. Furthermore, greater interregional integration (a lower trade cost) entails a higher zero-cutoff productivity level. Higher barriers to entry (higher f_e) lowers the minimum productivity level as they reduce competition in the market. When varieties are less substitutable between themselves, this also reduces competition and thus lowers θ^m. A higher liquidation rate and a more skewed productivity distribution have also the effect of reducing competition and lowering θ^m.

Stability condition At both locations, the mass of suppliers of the manufacturing good must be constant over time in the steady-state equilibrium. Since a mass δM_m of suppliers exits region m's market in every period, there must also be masses M_e^m and M_e^n attempting to enter the industry in region m and n, respectively, so that $[1-G(\theta^m)]M_e^m + [1-G(\theta_m^n)]M_e^n = \delta M_m$.

4.2.4.4 Welfare Analysis

Given the chosen form of the productivity distribution, the indirect utility function (4.2) is rewritten as

$$v^k(L^m) = e_0 + 1 + \frac{1}{2\chi}\left(\alpha - \frac{1}{\theta^m}\right)\left(\alpha - \frac{\kappa+1}{\kappa+2}\frac{1}{\theta^m}\right) \tag{4.12}$$

Each factor of the third term is positive since $p_m(j) < \alpha$ for all $j \in J_m$, as was mentioned earlier. Thus, an increase in the zero-cutoff productivity level raises welfare. It does so by increasing product variety and lowering the average price despite decreasing a^m and the price variance.

Differences in market size between regions results in disparities in aggregate industry performance and welfare. As the minimum level of productivity is an increasing function of market size, a larger region exhibits a higher average productivity level, a lower average price, more entrants and incumbents, and product variety, and higher welfare. In addition, the dispersion and variance of productivity levels and prices is smaller.

As the market expands, sales and profits expectations improve, which tend to raise the value of entering the industry and results in a larger number of entrants. At the same time, firms may want to lower their price as they have a chance to sell more units of output. As a consequence, low-productivity firms cannot survive this more competitive environment, which raises the zero-cutoff productivity level, lowers the expected profit conditional on productivity, and keeps the value of entry at zero.

The zero-cutoff productivity level of region m firm does not depend on the size of region n. As explained in Melitz and Ottaviano [2008], the neutral effect of the other region's size on the θ^m actually reflects factors acting in opposite directions. For producers in a relatively small region m, local competition is soft while revenues from sales in the larger region n can be sizeable. Then again, they face an intense competition in the non-local market (where margins are low) and also in their own region from non-local manufacturers.

4.3 Open Economy Model

Up until now, the economy functioned in autarky. Next, I present an extension of that model to the case of an open economy allowing me to examine the impact of trade on regional differences in aggregate industry performance. The domestic country (d) is allowed to trade the numéraire and the differentiated good with a foreign country (f). To make the model simpler, suppose the foreign country consists of a single location. The mass of foreign households equals L^f. These households share the characteristics of domestic households. Thus, in country f, the market demand for variety j, $j \in J_f$, is given by

$$x^f(j) = L^f \left(a^f - b p_f(j) + c^f \bar{p}^f \right),$$

where $a^f \equiv \alpha/(\beta + M_f \chi)$, $b \equiv 1/\beta$, and $c^f \equiv M_f \chi / \big(\beta(\beta + M_f \chi)\big)$; M_m designates the mass of manufacturing products supplied at location f; and $\bar{p}^f = \int_{j \in J_f} p_f(j) dj / M_f$ is the average price of the differentiated good in the foreign country.

In both sectors, foreign producers employ the same technologies as those used in the domestic country. International trade in the numéraire is free and therefore it equalizes prices and wages across regions and countries ($w^u = w^v = w^f = 1$ provided that production of this good takes place in all locations).

Manufacturers from both domestic locations incur an iceberg trade cost to export their output. However, producers from one region may not enjoy the same access to the foreign market as those from the other region. One region may be on the coast and may have a port while the other may be landlocked; or one region may be closer to the main commercial partner. To capture the role of geography, that trade cost is decomposed into internal and external trade costs. Manufacturers from region m must ship $\tau_d^m \tau_f^d$ units of output, where $\tau_d^m \geq 1$ and $\tau_f^d > \tau$, to supply one unit to the foreign market. Besides, $\max\{\tau_d^u/\tau_d^v, \tau_d^v/\tau_d^u\} \leq \tau$, that is, the difference in access to the foreign market between regions u and v is no greater than the cost of trading between them. In the opposite direction, a shipment from the foreign country must be $\tau_d^f \tau_d^m$ times larger than the quantity intended to be sold in region m. Anyway, so as not to clutter expressions, I let $\tau_f^m \equiv \tau_d^m \tau_f^d$ and $\tau_m^f \equiv \tau_d^f \tau_d^m$, keeping in mind that international trade costs can be broken down into two tiers.

The operating profit derived from sales in the foreign market by a firm from region m with productivity θ is

$$\pi_f^m(\theta) = \left(p_f^m(\theta) - \frac{\tau_f^m}{\theta} \right) y_f^m(\theta),$$

where p_f^m is the price charged to foreign households and y_f^m is the export volume. The optimal price and export volume for sales in the foreign market and the maximal value of the foreign profit are given by

$$p_f^m(\theta) = \frac{1}{2}\left(\hat{p}^f + \frac{\tau_f^m}{\theta}\right),\ y_f^m(\theta) = \frac{L^f b}{2}\left(\hat{p}^f - \frac{\tau_f^m}{\theta}\right),\ \text{and}\ \pi_f^m(\theta) = \frac{L^f b}{4}\left(\hat{p}^f - \frac{\tau_f^m}{\theta}\right)^2,$$

where $\hat{p}^f$ is the threshold price in country f. The price, output, and profit functions for foreign firms are obtained in a similar manner.

4.3.1 Firm Entry and Export Market Participation

A firm from region m makes its way into the foreign if it can sell output at a price lower than foreign threshold price but still higher than the marginal cost, that is, if $\tau_f^m/\theta < p_f^m(\theta) < \hat{p}^f$. Hence, the minimum productivity level to export, or export-cutoff productivity level, is defined by $\theta_f^m = \inf\{\theta : \theta \geq \theta_0 \text{ and } p_f^m(\theta) \leq \hat{p}^f\}$, which entails $p_f^m(\theta_f^m) = \hat{p}^f$, or $\hat{p}^f = \tau_f^m/\theta_f^m$ (from the export price expression).

4.3.2 Equilibrium

The export profit of a firm from region m with the lowest level of productivity among exporters must equal zero. The export-cutoff profit condition, $\pi_f^m(\theta_f^m) = 0$, is also equivalent to $\hat{p}^f = \tau_f^m/\theta_f^m$. The export-cutoff profit condition for country f and the zero-cutoff profit condition for region m together imply the following relationship between the cutoff productivity levels θ^m and θ_m^f:

$$\theta_m^f = \tau_m^f \theta^m \tag{4.13}$$

Likewise, $\theta_f^m = \tau_f^m \theta^f$. Thus, the least productive exporters from country f (region m) must be at least τ_m^f (τ_f^m) times more productive than producers with the zero-cutoff productivity level in region m (country f). A direct consequence of this relationship between zero- and export-cutoff productivity levels is that the productivity distribution of foreign exporters adjusted for the cost of exporting to region m matches that of region m producers. Hence, the expressions of the average price and the mass of firms present in region m's market in the open economy are also given by (4.7) and (4.8). The welfare function (4.12) applies to the open economy as well.

The free entry condition for location m is stated as

$$\int_{\theta^m}^{\infty} \pi_m^m(\theta)g(\theta)d\theta + \int_{\theta_n^m}^{\infty} \pi_n^m(\theta)g(\theta)d\theta + \int_{\theta_f^m}^{\infty} \pi_f^m(\theta)g(\theta)d\theta = \delta f_e.$$

Using the Pareto distribution and the relationships among the cutoff productivity levels, it is rewritten as

$$\frac{L^m}{(\theta^m)^{\kappa+2}} + \frac{L^n}{\tau^\kappa (\theta^n)^{\kappa+2}} + \frac{L^f}{(\tau_f^m)^\kappa (\theta^f)^{\kappa+2}} = c. \tag{4.14}$$

The free entry condition for the foreign country is

$$\int_{\theta_m^f}^{\infty} \pi_m^f(\theta)g(\theta)d\theta + \int_{\theta_n^f}^{\infty} \pi_n^f(\theta)g(\theta)d\theta + \int_{\theta^f}^{\infty} \pi_f^f(\theta)g(\theta)d\theta = \delta f_e,$$

or, equivalently,

$$\frac{L^u}{(\tau_u^f)^\kappa (\theta^u)^{\kappa+2}} + \frac{L^v}{(\tau_v^f)^\kappa (\theta^v)^{\kappa+2}} + \frac{L^f}{(\theta^f)^{\kappa+2}} = c. \tag{4.15}$$

The equilibrium cutoff productivity levels θ^u, θ^v, and θ^f solve the system of equations (4.14) and (4.15). The equilibrium value of θ^u is given by

$$\theta^m = \left[\frac{L^m}{c} T\right]^{1/(\kappa+2)}, \tag{4.16}$$

where

$$T \equiv \left[\tau^{-2\kappa} - \tau^{-\kappa}\left[\left(\tau_f^m\right)^{-\kappa}\left(\tau_n^f\right)^{-\kappa} + \left(\tau_m^f\right)^{-\kappa}\left(\tau_f^n\right)^{-\kappa}\right] + \left(\tau_f^m\right)^{-\kappa}\left(\tau_m^f\right)^{-\kappa} + \left(\tau_f^n\right)^{-\kappa}\left(\tau_n^f\right)^{-\kappa} - 1\right]\left[\tau^{-\kappa}\left[1 - \left(\tau_f^n\right)^{-\kappa}\right] + \left(\tau_f^m\right)^{-\kappa} + \left(\tau_n^f\right)^{-\kappa}\left[\left(\tau_f^n\right)^{-\kappa} - \left(\tau_f^m\right)^{-\kappa}\right] - 1\right]^{-1}. \quad (4.17)$$

The minimum value of the Pareto distribution must be sufficiently small so that $\theta^m \geq \theta_0$.

4.4 The Impact of Trade Liberalization

This section considers the regional impact of trade liberalization under various circumstances: (1) bilateral decrease in trade costs when domestic regions have an equal same access to the foreign market (case of symmetrical geography), (2) unilateral trade liberalization by the domestic country, and (3) bilateral trade liberalization when a domestic region is closer to the foreign country than the other (case of asymmetrical geography).

4.4.1 Symmetrical Geography

4.4.1.1 Bilateral Trade Liberalization

Suppose that $\tau_d^u = \tau_d^v = 1$ and $\tau_f^d = \tau_d^f = \tau_{ex}$ (then, $\tau_f^u = \tau_f^v = \tau_u^f = \tau_v^f = \tau_{ex}$). Therefore, the zero-cutoff productivity level for region m, (4.16), becomes

$$\theta^m = \left[\frac{L^m}{c}\frac{1 + \tau^{-\kappa} - 2\tau_{ex}^{-2\kappa}}{1 - \tau_{ex}^{-\kappa}}\right]^{1/(\kappa+2)}.$$

To compare θ^m in the open economy with its value in the closed economy, one may look at the difference between T and $1 + \tau^{-\kappa}$. In this particular case,

$$T - (1 + \tau^{-\kappa}) = \frac{\tau_{ex}^{-\kappa}\left[1 + \tau^{-\kappa} - 2\tau_{ex}^{-\kappa}\right]}{1 - \tau_{ex}^{-\kappa}} > 0.$$

Since this difference is positive, θ^m is higher in the open economy. Low-productivity firms were viable in the closed economy but they cannot sustain the exposure to import competition which operates through an increase in the number of competitors and a downward shift in the average price and thus which tends to raise the price elasticity of demand. As they can no longer sell output at a price above their marginal cost, they exit the industry. Since, the zero-cutoff productivity level goes up, trade also causes growth in regional average productivity. In addition, welfare rises as imports bring about more variety in the consumption of manufacturing good.

In the open economy, the direction of the impact of trade liberalization on zero-cutoff productivity level is given by the sign of its derivative with respect to τ^{ex}:

$$\frac{\partial \theta^m}{\partial \tau_{ex}} = -\frac{\kappa}{\kappa+2}\left[\frac{L^m}{c}\right]^{1/(\kappa+2)} \tau_{ex}^{-\kappa-1}\left[2-\frac{1-\tau^{-\kappa}}{(1-\tau_{ex}^{-\kappa})^2}\right]\left[\frac{1+\tau^{-\kappa}-2\tau_{ex}^{-2\kappa}}{1-\tau_{ex}^{-\kappa}}\right]^{-\frac{\kappa+1}{\kappa+2}}.$$

This expression is negative if $2-(1-\tau^{-\kappa})/(1-\tau_{ex}^{-\kappa})^2 > 0$; that is, if $\tau_{ex} > \tilde{\tau}_{ex}$, where

$$\tilde{\tau}_{ex} \equiv \left[1-\left(\frac{1-\tau^{-\kappa}}{2}\right)^{1/2}\right]^{-1/\kappa}.$$

Note that $\lim_{\tau\to 1}\tilde{\tau}_{ex} = 1$ and $\lim_{\tau\to+\infty}\tilde{\tau}_{ex} = \left[1-(1/2)^{1/2}\right]^{1/(1-\sigma)} > 1$. Besides, $\tilde{\tau}_{ex}$ is monotonically increasing and concave in τ. Thus, when τ is relatively small, it is possible for τ_{ex} to be smaller than $\tilde{\tau}_{ex}$ while still being greater than τ. In that case, the derivative of θ^m with respect to τ^{ex} is positive. Otherwise, when τ_{ex} is greater than $\tilde{\tau}_{ex}$, the derivative is negative.

The direction of the impact of bilateral trade liberalization varies according to the relative magnitude of the international trade cost. When it is high relative to the interregional trade cost ($\tau_{ex} > \tilde{\tau}_{ex}$), trade liberalization results in an increase in regional zero-cutoff productivity levels and thus in regional average productivity. As barriers to external trade fall, the mass of firms competing in regional markets expands, which raises the zero-cutoff productivity levels and leads the least productive firms to exit. Furthermore, since regional market size enters the derivative of θ^m multiplicatively, the increase in the zero-cutoff productivity level is increasing in L^m. Hence, a larger region would experience greater growth in average productivity following trade liberalization. Since a larger regional market is more competitive, only more productive foreign exporters can penetrate this market upon trade liberalization, which makes selection even harsher.

When the gap between the international and interregional trade costs is relatively low ($\tau_{ex} < \tilde{\tau}_{ex}$), trade liberalization leads to a decrease in regional zero-cutoff productivity levels and in regional average productivity.[9] This outcome arises because as the international trade cost falls further, firms from region m lose the advantage in serving region n's market they held over foreign firms, while they do not gain a better access to the foreign market than foreign firms do in the opposite direction. As a consequence, there is a relocation of manufacturers firms from the domestic

[9] Like in Melitz and Ottaviano [2008], one must assume that the international cost remains above a minimum level. Otherwise, as the trade cost fell below this level, the number of entrants at the domestic locations would tend toward zero and so therefore these locations would specialize in the production of the numéraire.

country to the foreign country (through a decrease and increase in the rates of entry in these two countries, respectively) with an ensuing decline in regional zero-cutoff productivity levels and average productivity in the domestic country. Similarly as before, a larger region would undergo a steeper fall in average productivity following trade liberalization, since more firms would relocate away from this region and it would become less competitive.

4.4.1.2 Unilateral Trade Liberalization

Now assume that $\tau_f^d \neq \tau_d^f$ while, as before, $\tau_d^u = \tau_d^v = 1$. In this case, the zero-cutoff productivity level is

$$\theta^m = \left[\frac{L^m}{c} \frac{1 + \tau^{-\kappa} - 2 \left(\tau_f^d\right)^{-\kappa} \left(\tau_d^f\right)^{-\kappa}}{1 - \left(\tau_f^d\right)^{-\kappa}} \right]^{1/(\kappa+2)}.$$

Here, the difference between T and $1 + \tau^{-\kappa}$ is

$$T - (1 + \tau^{-\kappa}) = \frac{\left(\tau_f^d\right)^{-\kappa} \left[1 + \tau^{-\kappa} - 2\left(\tau_d^f\right)^{-\kappa}\right]}{1 - \left(\tau_f^d\right)^{-\kappa}} > 0.$$

Thus, θ^m is higher in the open economy, even if international trade costs are asymmetric. Note however that if it were costlier for the domestic country to export than it were for the foreign country, then θ^m would be smaller than in a situation of symmetric costs. In that case, the foreign country draws firms as it gives access to a larger market. Hence, the foreign market is more competitive and the zero-cutoff and average productivity levels are higher there.

When international trade are asymmetric, the direction of the impact of unilateral trade liberalization by the domestic country on the zero-cutoff productivity level of region m is given by the sign of its derivative with respect to τ_d^f:

$$\frac{\partial \theta^m}{\partial \tau_d^f} = \frac{2\kappa}{\kappa+2} \left[\frac{L^m}{c}\right]^{1/(\kappa+2)} \left(\tau_f^d\right)^{-\kappa} \left(\tau_d^f\right)^{-\kappa-1} \frac{\left[1 + \tau^{-\kappa} - 2\left(\tau_f^d\right)^{-\kappa}\left(\tau_d^f\right)^{-\kappa}\right]^{-\frac{\kappa+1}{\kappa+2}}}{\left[1 - \left(\tau_f^d\right)^{-\kappa}\right]^{1/(\kappa+2)}} > 0.$$

Hence, unilateral trade liberalization causes a decrease in regional zero-cutoff productivity levels and in regional average productivity. As the cost if exporting goods to the domestic country falls, the advantage of being located in the foreign country increases as it gives access to a larger market. There is a shift in the location of the manufacturing industry from the domestic country to the foreign country (through a decrease and increase in the rates of entry in these two countries, respectively), and as the degree of competition both domestic regions drops, region average productivity falls along. Like in the case of bilateral trade liberalization, a larger market size exacerbate these changes and in particular the larger domestic region stands to lose more in terms of average productivity (and welfare) from unilateral liberalization.

4.4.2 Asymmetrical Geography

Suppose that region u is closer to the foreign market than v because, for instance, it has a direct access to a commercial port. Accordingly, I assume that $\tau_d^u = 1$, $\tau_d^v = \tau$, $\tau_f^d = \tau_d^f = \tau_{ex}$, and that the regions have the same size, $L^u = L^v = L^d/2$, where L^d denotes the mass of domestic households. The expression of the zero-cutoff productivity level for region u is

$$\theta^u = \left[\frac{L^d}{2c} \frac{(1+\tau^{-\kappa})(1+\tau_{ex}^{-\kappa})}{1-(\tau\tau_{ex})^{-\kappa}} \right]^{1/(\kappa+2)} .$$

and for region v it is

$$\theta^v = \left[\frac{L^d}{2c} (1+\tau^{-\kappa}) \right]^{1/(\kappa+2)} .$$

If one compares these zero-cutoff productivity levels to that of a region with the same size in the closed economy, then θ^u is found to be larger and θ^v the same. Region u is like a hub among the three locations and thus it is a competitve market in the open economy since more domestic firms locate there to gain access the foreign market. Region v however is isolated from the international market and thus is not affected by the exposure of the domestic country to trade. For the same reason, trade liberalization does not further induce changes in this region, unlike the coastal region. The derivative of θ^u with respect to τ_{ex} is

$$\frac{\partial \theta^u}{\partial \tau_{ex}} = -\frac{\kappa}{\kappa+2} \left[\frac{L^d}{2c} \right]^{1/(\kappa+2)} \frac{(1+\tau^{-\kappa})^{\frac{\kappa+3}{\kappa+2}} \tau_{ex}^{-\kappa-1}}{[1-(\tau\tau_{ex})^{-\kappa}]^2} \left[\frac{(1+\tau^{-\kappa})(1+\tau_{ex}^{-\kappa})}{1-(\tau\tau_{ex})^{-\kappa}} \right]^{-\frac{\kappa+1}{\kappa+2}} < 0.$$

So, since a decrease in the external trade cost entails an increase in θ^u, bilateral trade liberalization encourages further entry and intensifies competition in region m's market. Consequently, the coastal region experience an increase in average productivity while the aggregate productivity level of region v is left unchanged.

The case of unilateral liberalization with asymmetrical geography is not formally treated as the analytical expressions of θ^u and its derivative are complicated, however, one may reason intuitively to deduce its impact. For a decrease in τ_d^f, the foreign country would get a better access to the domestic regions, the coastal region in particular, while nothing would change for the domestic regions in terms of export market access. Thus, one would expect the zero-cutoff productivity level in region u to drop as manufacturers would relocate to the foreign country and the former become a less competitive market. Thus, region u would undergo a fall in average productivity and welfare.

4.4.3 Summary

In the model, the direction of the impact of trade liberalization on aggregate industry productivity across regions depends on the reciprocity of the shift in exposure to international trade, the physical and economic geography of the domestic country, and also, in one case, on the initial level of the external trade cost. When domestic regions are of different sizes, a bilateral decrease in the trade cost raises the aggregate productivity differential between the large region and the small one; but it has the reverse effect when the cost of trade between countries is already low. Unilateral trade liberalization reduces the difference in productivity between regions. When regional markets have the same size but regions face different costs of access to the foreign market, bilateral trade liberalization results in higher productivity growth in the region with better access to the foreign country. Under the same geographical circumstances, unilateral trade liberalization should lower aggregate productivity in this region. Thus, the impact of trade liberalization is contingent upon different factors, but since these factors are observable in practice, one could control for them in an empirical analysis. In that sense, the model provides guidance for an empirical investigation as to which variables to consider and what spatial heterogeneity in the response to trade to expect; without suggesting a unique hypothesis. The following section considers the case in Colombia to look at whether different regions were affected by trade liberalization in different ways.

4.5 Empirical Evidence

This section investigates the effects of unilateral trade liberalization on aggregate industry productivity across regions of Colombia using establishment-level data from the manufacturing sector. Colombia has been highlighted before as a good case study to examine the impact of trade policy reforms on productivity growth (see Fernandes [2007] and Ederington and McCalman [2007] for instance). High and irregular tariffs and other stringent measures of trade restriction protected Colombia's manufacturing sector from import competition in the post-War period, but this country started to reduce its trade barriers in the late 1970s as it was planning to access the GATT. While Colombia became a contracting party of the GATT in 1981, unfavorable macroeconomic conditions delayed further trade liberalization until the mid-1980s. Eventually, between 1985 and 1992, Colombia substantially reduced its tariffs across-the-board and dismantled other import barriers. By the end of that short period, it achieved relatively low tariffs, on a par across products. The uneven physical and economic geography of Colombia also makes it an interesting case to explore the variation in the impact of exposure to trade on productivity across regions within a country. The Colombian territory is relatively large, exhibits an unequal geographical distribution of economic activities, and regional economies have been relatively isolated from each other in the past due to geographical barriers to trade and a fragmented society.

4.5.1 Econometric Method

The empirical analysis of the marginal impact of tariff policy and its interaction with region-level geographic characteristics on aggregate industry productivity is based on a linear panel-data, fixed-effects regression model. In this model, the regressand is a measure of aggregate productivity for an industry-region pair and the set of regressors comprises the *ad valorem* import tariff for the industry, a measure of the extent of the region's market, and two terms of interaction, between the tariff covariate and a geographical indicator for the region and between tariff and size of the geoggraphical market. The empirical model is specified as follows:

$$pr_{i,t}^m = \gamma_t + \beta_1 \tau_{i,t-1} + \beta_2 \text{Coastal}^m \times \tau_{i,t-1} + \beta_3 \text{GDP}_t^m + \beta_4 \text{GDP}_t^m \times \tau_{i,t-1} + \alpha_i^m + \epsilon_{i,t}^m, \quad (4.18)$$

where $pr_{i,t}^m$ denotes an aggregate measure of productivity for industry i in region m in year t; γ_t is a non-stochastic time fixed effect; $\tau_{i,t-1}$ is the tariff for industry i in year $t-1$; the indicator Coastal takes the value one if region m borders an ocean, zero otherwise; GDP_t^m is the gross domestic product of region m in year t; α_i^m is an industry-region-specific, time-invariant stochastic error term, that is, an industry-region fixed effect; and $\epsilon_{i,t}^m$ $(0, \sigma_\epsilon^2)$ is an i.i.d. disturbance.

The fixed-effects (FE) model permits regressors to be endogenous provided they are correlated with α_i^m but not with $\epsilon_{i,t}^m$. For instance, unobserved, time-invariant characteristics of an industry in a particular region may simultaneously influence both the industry aggregate productivity and the tariff level, while not affecting the consistency of the estimator. One may consider an alternative specification of model (4.18) in which two separate sets of fixed effects, one for industries and one for regions, replace the set of industry-region pairs fixed effects. In this case, $\alpha_i^m \equiv \alpha_i + \alpha^m$. This specification of the fixed-effects allows unobserved industry and region characteristics to influence productivity and regressors in a mutually independent way. Additionally, year fixed effects control for macroeconomic factors shifting aggregate productivity in the same way across industry-region pairs.

The use of the lagged tariff has two purposes. First, it serves to address the issue of endogeneity of contemporaneous tariffs. Indeed, there may be a systematic dependence of tariff protection on productivity as industries with many unproductive firms would press policy-makers to raise tariff barriers. Fernandes [2007] reports a set of facts suggesting that this wasn't the case for Colombia. Second, as explained by Tybout [1992], the ambiguity surrounding the trade policy agenda of the Colombian government at that time could have caused producers to adjust to changes in tariffs with a lag.

The following two sub-sections explain how productivity is measured at the plant level and how it is subsequently aggregated into an industry-level productivity measure to be used as the explained variable of regression model (4.18).

4.5.2 Establishment-level Productivity Estimation

4.5.2.1 Estimation Method

The estimator employed to measure Colombian manufacturing establishments' total factor productivity (TFP) is that of Levinsohn and Petrin [2003]. This estimation method assumes that profit-maximizing plants in an industry produce a homogeneous good using a Cobb-Douglas technology[10] and that the factors explaining the variation in productivity across plants are Hicks-neutral. The econometric model of the production function of plant j in industry i is specified as

$$y_{i,t}^j = \beta_0 + \beta_k k_{i,t}^j + \beta_l l_{i,t}^j + \beta_e e_{i,t}^j + \beta_m m_{i,t}^j + \omega_{i,t}^j + \eta_{i,t}^j, \tag{4.19}$$

where $y_{i,t}^j$ is the logarithm of plant j's output in year t; $k_{i,t}^j$ the log of its capital stock (the state variable); $l_{i,t}^j$, $e_{i,t}^j$, and $m_{i,t}^j$ are the logs of the labor, energy, and intermediate inputs utilized by plant j (the variable inputs), respectively. The parameter β_0 reflects a Hicks-neutral productivity factor common to all plants in the industry and constant over time.

Output is also determined by a plant-year-specific stochastic error term, $\omega_{i,t}^j + \eta_{i,t}^j$. The first term, $\omega_{i,t}^j$, is a state variable in the plant's objective function, it determines

[10] This estimator can be applied to other production functions as well.

the demand for inputs and the decision to exit, and it cannot be observed directly by someone else than the plant's decision maker. Since $\omega^i_{j,t}$ is contemporaneously correlated with inputs quantities, an estimator of the production function parameters that would not take into account the correlation between inputs and $\omega^j_{i,t}$ (the ordinary least squares (OLS) estimator for instance) would produce biased and inconsistent coefficient and TFP estimates. This problem is commonly referred to as the simultaneity (or endogeneity) bias.[11] The second component of the disturbance term, $\eta^i_{j,t}$, embodies random productivity shocks occuring after the choice of inputs is made as well as measurement errors. Although it could be serially correlated, it does not affect the producer's profit-maximization problem.

Olley and Pakes [1996] first showed that investment can serve as a proxy for the unobservable productivity shock. This method, however, is unsuitable in many cases: it requires strictly positive values for investment in every time period but plant-level data often contain a large number of zero values. As alleged by Levinsohn and Petrin [2003], adjustment costs tend to prevent plants from adjusting capital to productivity innovations in a continuous manner. Levinsohn and Petrin [2003] present an alternative method using intermediate inputs as proxy (intermediate inputs are the inputs which are typically substracted from revenue to compute value added). They assume that demand for intermediate inputs depends on the plant's state variables:[12]

$$m^j_{i,t} = m_{i,t}(k^j_{i,t}, \omega^j_{i,t}).$$

If $m_{i,t}$ is monotonically increasing in $\omega^j_{i,t}$ conditional on $k^j_{i,t}$ (see conditions in Appendix A of Levinsohn and Petrin [2003] for when this is true), the intermediate inputs demand function can be inverted to obtain productivity as a function of inputs consumption and capital (which are observable):

$$\omega^j_{i,t} = \omega_{i,t}(k^j_{i,t}, m^j_{i,t}).$$

Furthermore, productivity is assumed to follow a first-order Markov process:

$$\omega_t = \mathrm{E}\left[\omega_t|\omega_{t-1}\right] + \xi_t,$$

[11] The first formal analysis of the issue of correlation between the unobservable, transmitted firm-specific productivity level and the choice of inputs is due to Marschak and Andrews [1944]. Griliches and Mairesse [1998] retrace the debate about this issue. Establishments experiencing positive productivity shocks are expected to respond by expanding output, which necessitates more inputs. On the contrary, negative productivity shocks cause firms to scale back output and thus consume less inputs. Moreover, one would expect the most easily adjustable inputs to be more correlated with contemporaneous productivity $\omega^j_{i,t}$ (see Marschak and Andrews [1944] and Griliches and Mairesse [1998] for more detailed expositions).

[12] The intermediate inputs demand function is the same for all plants in an industry.

where ξ_t is a stochastic error term that is uncorrelated with $k^j_{i,t}$ but that may be correlated with $l^j_{i,t}$ and $e^j_{i,t}$. Unless it is necessary, the industry subscript is omitted from now on. The expression of the production function becomes

$$y^j_t = \beta_l l^j_t + \beta_e e^j_t + \phi_t(k^j_t, m^j_t) + \eta^j_t, \tag{4.20}$$

where

$$\phi_t(k^j_t, m^j_t) = \beta_0 + \beta_k k^j_t + \beta_m m^j_t + \omega_t(k^j_t, m^j_t).$$

In the first stage of the procedure, consistent estimates of β_l, β_e, and ϕ_t (denoted by $\hat{\beta}_l$, $\hat{\beta}_e$, and $\hat{\phi}_t$, respectively) are obtained with the OLS estimator after substitution of a third-order polynomial in k^j_t and m^j_t for ϕ in (4.20).

The second stage consists in estimating the parameters β_k and β_m. For any numbers β^*_k and β^*_m, a predicted value of ω^j_t (up to a scalar) is given by

$$\hat{\omega}^j_t = \hat{\phi}_t - \beta^*_k k^j_t - \beta^*_m m^j_t,$$

where $\hat{\phi}_t = \hat{y}_t - \hat{\beta}_l l^j_t - \hat{\beta}_e e^j_t$. A pooled OLS regression of $\hat{\omega}^j_t$ on a third-order polynomial in $\hat{\omega}^j_{t-1}$ yields a consistent estimate of $\mathrm{E}[\omega^j_t|\omega^j_{t-1}]$, $\mathrm{E}[\widehat{\omega^j_t|\omega^j_{t-1}}]$. Now, for any numbers β^*_k and β^*_m, one can obtain a predicted value of the error term $\eta^j_t + \xi^j_t$:

$$\widehat{\eta^j_t + \xi^j_t} = y^j_t - \hat{\beta}_l l^j_t - \hat{\beta}_e e^j_t - \beta^*_k k^j_t - \beta^*_m m^j_t - \mathrm{E}[\widehat{\omega^j_t|\omega^j_{t-1}}].$$

The parameter β_k is identified from the assumption that yesterday's investment is independent of today's unexpected shift in productivity, ξ^j_t, which implies the moment condition

$$\mathrm{E}[\eta^j_t + \xi^j_t | k^j_t] = 0.$$

Similarly, the parameter β_m is identified from the assumption that yesterday's utilization of intermediate inputs is independent of ξ_t, that is,

$$\mathrm{E}[\eta^j_t + \xi^j_t | m^j_t] = 0.$$

Thus, the generalized method of moments estimator of β_k and β_m is given by

$$\min_{\beta^*_k, \beta^*_m} \left[\sum_{t,j} \left(\widehat{\eta^j_t + \xi^j_t} \right) k^j_t \right]^2 + \left[\sum_{t,j} \left(\widehat{\eta^j_t + \xi^j_t} \right) m^j_t \right]^2$$

Additional, over-identifying moment conditions using lagged capital, lagged labor, lagged energy, and intermediate inputs lagged twice serve to improve efficiency. Finally, plant j's TFP in year t is calculated as $\exp(\omega^j_t) = \exp(y^j_t - \hat{\beta}_k k^j_t - \hat{\beta}_l l^j_t -$

$\hat{\beta}_e e_t^j - \hat{\beta}_m m_t^j)$. Later, when the notation ω_t^j is used, it refers to the exponential of the predicted productivity.

The estimation of production functions at the plant level entails additional thorny issues beyond the simultaneity problem. These issues are related to the entry and exit of plants (see below), imperfect competition in the output and inputs markets, multi-product plants. When the output or inputs markets are imperfectly competitive, using deflated sales or value added to measure output and expenditures on inputs without controlling for plant-level prices leads to an omitted variable bias. Fernandes [2007] notes that the estimation of a production function with deflated revenue as a measure of physical production and deflated expenditures on raw materials, energy, and capital as measures of physical utilization entails that predicted TFP values are close to the revenue per unit input bundle and thus combine actual efficiency and price-cost mark-ups. But as long as these mark-ups increase with efficiency, such TFP measures are still correlated with plant efficiency. A similar problem noted by Syverson [2004] arises when output demand and inputs supply conditions vary across locations. Besides, if plants produce multiple products with different production technologies, then, failure to estimate the production function at the appropriate product level, rather than at the plant level, introduces a bias in the standard (OLS) TFP estimate.

4.5.2.2 Production Data

I use an unbalanced panel data set of Colombian manufacturing establishments for the period 1983-1991. The *Departamento Administrativo Nacional de Estadística* (DANE, Colombia's statistical agency) collected the data through the annual Colombian census of the manufacturing sector.[13] Plants are categorized in industries with a four digit code defined in the International Standard Industrial Classification (ISIC). This data set contains information about plant characteristics such as employment, capital stock, energy consumption, raw materials and intermediate goods utilized, the value of production, and so forth.

The production function (4.19) is estimated by industry at the 3-digit level, over the period 1981-1991, and with raw materials consumption as a proxy for the transmitted productivity shocks since this item is positive in all years for most plants in the data set. The variables are constructed as follows:

[13] Mark Roberts provided the data. See Roberts [1996] for a comprehensive description of the data.

- Output, $y^j_{i,t}$: the natural logarithm (log) of the real value of production of an establishment.
- Capital, $k^j_{i,t}$: the log of the deflated total book value of fixed assets excluding land.
- Labor, $l^j_{i,t}$: the log of total employment.
- Materials, $m^j_{i,t}$: the log of the deflated value of raw materials and inputs consumption (calculated as raw materials purchases minus the net increase in inventories).
- Energy, $e^j_{i,t}$: the log of the deflated value of energy consumption plus purchases of fuels and lubricants consumed by the establishment.

Like most plant-level data sets, the Colombia data contains plants entering, exiting, or switching industries over the data collection period. Both theoretical models and the empirical evidence suggest that successful entry, growth, and liquidation are, to a large extent, related to productivity. The fact that the presence of plants in a sample dataset or their absence from it may be systematically related to an unobservable variable like productivity implies that an estimator that does not account for the selection of plants may be biased (selection bias). For instance, Olley and Pakes [1996] find that the self-selection of plants shutting down results in a downward bias of the coefficient estimate of capital obtained from the OLS estimator. However, they warn against excluding plants with incomplete time series of data as this can cause further sample selection issues. [14]

4.5.3 Aggregate Productivity Measurement

A variable like the zero-cutoff productivity level in the model is evidently not directly observable in an industry. In spite of this, one may look at industry-level measures of productivity that correspond, in the model, to aggregate variables influenced by the zero-cutoff productivity level. According to the theory, they should reflect the unobservable minimum productivity level to survive in an industry and thus the effect of competition.

The study of the ready-mixed concrete industry by Syverson [2004] considers the following measures of central tendency of the TFP distribution in a geographical

[14] They design a method to correct for the selection bias (in addition to the simultaneity bias), but they report that this correction does not significantly alter their results.

market: weighted average TFP (where weights are producers' output shares) and median TFP. He also examines productivity dispersion using an interquartile TFP range and the minimum productivity level as measured by the tenth percentile of the TFP distribution (to mitigate the inluence of outliers). In addition, he looks at two measures of average plant size, average output and producer-demand ratio. All of these characteristics of the empirical productivity distribution may shed light on the consequences of a greater exposure to import competition for plant selection and market shares reallocations in Colombia. However, this study does not focus on just one industry and the data available present limitations regarding the identification of geographical markets and the sample size within spatial units. That is why I shall only utilize measures of central tendency of the productivity distribution (weighted-average and median TFP).

The output-weighted average productivity for industry i and region m in year t, denoted by $\breve{\omega}_{i,t}^{m}$, is calculated in the following manner (the subscript i and superscript m are omitted for clarity):

$$\breve{\omega}_t = \sum_{j=1}^{N_t} s_t^j \omega_t^j,$$

where s_t^j is plant j's output share in industry i, and ω_t^j is the estimated TFP of plant j. Then, the weighted average productivity is decomposed as in Olley and Pakes [1996]:

$$\begin{aligned}\breve{\omega}_t &= \sum_{j=1}^{N_t} \left(\bar{s}_t + \Delta s_t^j\right)\left(\bar{\omega}_t + \Delta\omega_t^j\right) \\ &= N_t \bar{s}_t \bar{\omega}_t + \sum_{j=1}^{N_t} \Delta s_t^j \Delta\omega_t^j \\ &= \bar{\omega}_t + \sigma(s_t^j, \omega_t^j)\end{aligned}$$

where $\Delta s_t^j \equiv s_t^j - \bar{s}_t$, $\Delta\omega_t^j \equiv \omega_t^j - \bar{\omega}_t$, $\bar{s}_t$ and $\bar{\omega}_t$ denotes the (unweighted) mean output share and TFP, respectively, and $\sigma(s_t^j, \omega_t^j)$ is the sample covariance between output share and plant productivity. Thus, the output-weighted average productivity is the sum of plants' productivities mean and the covariance between their output share and their productivity. The covariance between output and productivity measures the degree to which more productive plants produce more output. The growth in output-weighted average productivity has two components: the increase in average productivity over plants and the reallocation of market shares toward more productive plants.[15]

[15]Other methods of decomposition account for the contribution of firm entry and exit to output-

In practice, an increase in exposure to trade generally affects productivity at the plant level (see Fernandes [2007] and Ederington and McCalman [2007] about Colombia), as producers may adopt more productive technologies, employ better inputs, and improve management. But in the model presented above, technology is constant over time and thus trade liberalization does not affect firm-level productivity. Instead, the model predicts that as low-productivity firms exit the industry, a reallocation of market shares in favor of high-producticity firms occurs. The empirical analysis focuses on this effect and for this reason the allocative efficiency component of aggregate industry productivity $\breve{\omega}_t$ is retained as a dependent variable in regression model (4.18). The other variable of interest is the median TFP because it is expected to go up as the least productive plants shut down.

Next I describe the data on tariffs and the features of Colombia's physical and economic geography relevant to the empirical analysis.

4.5.4 Tariff Data and Geography

4.5.4.1 Tariff Data

The tariff data are *ad valorem*, applied tariff rates[16] at the 4-digit level of the ISIC. They are simple, arithmetic averages of the nominal tariffs for the products belonging to the 4-digit level industries. They were obtained from Jorge Garcia, from the World Bank, and the DANE website. Since the times series of tariffs is incomplete, I use industry data for the years 1983, 1985, 1987, 1989, and 1991 in conjunction with tariff data for 1982, 1984, 1985 (in place of 1986), 1988, and 1990. Tariffs are also missing for some products for the period of the analysis either because they were not imported or because that information was kept confidential.[17]

As mentioned earlier, there were other measures of import protection imposed by the Colombian government like import licenses. Fernandes [2003] reports that the level of protection indicated by tariffs is consistent with other measures available

weighted average productivity growth (see Roberts [1996], Aw et al. [2001]). As Fernandes [2007] reports that firm turnover contributes little to aggregate productivity growth in Colombia, I do not use these alternative decompositions.

[16] Tariff rates published by national custom authorities for duty administration purposes ([R. Moreira, Departamento Nacional de Planeacion, pers. comm.])

[17] The industries for which tariffs are not available at the ISIC 4-digit level, for the period covered by the empirical analysis, are the following: Dietetic products (ISIC code 3123), Cotton products (3216), Wool products (3217), Synthetic fiber textiles (3218), Clothing, excluding shoes (3221), Miscellaneous chemical products (3528), Lead and zinc (3721), Tin and nickel (3722), Precious metal (3723), Plumbing and heating products (3814), Other machinery (3826, 3827), and Miscellaneous manufacturing industries (3904).

(tariffs are found to be highly positively correlated with effective protection rates, negatively correlated with import penetration ratios, and positively related to the extent of coverage of import licenses). Tariffs should be expected to provide a good proxy for import restrictions in general.

4.5.4.2 Colombia's Geography

The territory of Colombia, with an area of about 1,140,000 km^2, is relatively large. This country is situated in the northwestern part of South America. Its neighboring countries are Venezuela and Brazil to the east, Ecuador and Peru to the south, and Panama to the northwest. Colombia is also bounded by the Pacific Ocean to the west, and the Atlantic Ocean, through the Caribbean Sea, to the north (see map in Figure 4.1).

Colombia is usually divided into five regions: the Andean highlands, which consist of the three Andean ranges (*Cordillera* (chain) *Occidental*, *Cordillera Central*, and *Cordillera Oriental*, or, alternatively, West Range, Centre Range, and East Range); the Pacific lowlands; the Caribbean lowlands; the plains (*Los Llanos*) lying to the east of the Andes Mountains; and the Amazon region to the southeast (these last two regions comprise most of the country's land area but are sparsely populated). The Andean highlands extend from the southwest, at the Ecuador border, toward the northeast, over the Caribbean region and Venezuela. They are separated by two narrow stretches of valley lowlands, traversed by rivers running from south to north. The Magdalena River in the valley separating the *cordilleras Central* and *Oriental* is navigable from the Caribbean Sea to the city of Neiva and constitutes a major way of transportation. The Cauca Valley sits between the *cordilleras Occidental* and *Central*. The capital city of Bogotá, sited in the *Cordillera Oriental*, in the basin of Cundinamarca, rises at 2,650 meters above sea level.

The geographic distribution of Colombia's population and economic activity is very uneven. About three-fourth of the population inhabits the basins and plateaus of the highlands and the valley lowlands in the center of the country and the remainder lives in the Caribbean lowlands. The region of Bogotá is the core population and industrial center of the country. The basins and plateaus to the north of Bogotá are also densely populated and very active in the industrial sector. The cities of Medellín, located at the northern tip of the *Cordillera Central*, Cali, in the middle of the *Cordillera Occidental*, and Barranquilla, on the Caribbean coast, form secondary core population and industrial centers. Cali enjoys an easy access to the Pacific coast and the port of Buenaventura through a pass in the *cordillera*. Although the Andean region is much larger than the Caribbean region in demographic and economic terms, the latter has a significant share of the economic activity and most of the country's foreign trade transits through Barranquilla, Cartagena, Santa Marta, and other ports cities located along the Caribbean coastline. A large part of the country's agricultural production takes place inland from these cities.

In Colombia, regional disparities in income and economic development have historically been relatively large. A number of studies lend support to the hypothesis of rising economic disparities among regions during the second half of the twentieth century (for instance, see Birchenall and Murcia [1996], Bonet and Meisel [1999], Meisel [1993], and Rocha and Vivas [1998]). In particular, it seems that a relative economic decline of the Caribbean and North-central regions was taking place during that period, whereas the gross domestic product share of Bogotá's economy was rising, from 15 percent in 1960 to 20 percent by 2000 (Bonet and Meisel [1999]). Meanwhile, the shares of GDP of the West-central, South-central, and Pacific regions remained about constant.

This study adopts a definition of regions that parallels the conventional partition of Colombia's territory. The fact that the country's topography imposes significant transport costs for interregional trade well matches the hypothesis of the theoretical model that regional markets are imperfectly integrated. Specifically, I identify four regions: the Caribbean, Central, Inland, and Pacific regions. Central encompasses the northern parts of the *cordilleras Occidental* and *Central*, which contains Medellín; Inland coincides with the eastern Andean lowlands, which includes Bogotá; the Pacific region includes Calí. These regions are defined according to the departments they comprise,[18] as follows:

- Caribbean region: Atlántico, Bolívar, Cesar, Córdoba, La Guajira, Magdalena, and Sucre
- Central region: Antioquia, Caldas, Chocó[19], Quindío, and Risaralda
- Inland region: Boyacá, Caquetá, Cundinamarca, Huila, Intendencia de Casanare, Meta, Norte de Santander, Santafé de Bogotá, D.C., Santander, and Tolima
- Pacific: Cauca, Nariño, and Valle del Cauca

[18]Departments are the primary administrative units in Colombia. Since the Colombian manufacturing census reports the department when a plant is located, it is possible to isolate plant-level data by region.

[19]The department of Chocó is usually associated with the Pacific region, which makes sense from a physical geography perspective. Yet, the transport infrastructure (roads especially) establish a stronger link with the department of Antioquia and the rest of the Central region.

An alternative definition of regions used for the regression analysis joins together the Central and Pacific regions into a Greater Pacific region.

Regions are characterized by their GDP as a proxy for demand density. GDP data at the department level are obtained from DANE [1995] and are aggregated at the level of regions as defined above. Figure 4.2 shows the level and evolution of GDP for the four regions of Colombia.

4.5.5 Results

The results from the estimation of regression model (4.18) across four regions (Caribbean, Central, Inland, and Pacific) are reported in Table 4.1. The model was estimated with both output-TFP covariance and median TFP as dependent variables. In addition, for comparison, it was estimated with mean TFP as a dependent variable, the first term of the decomposition of output-weighted average productivity. For each type of dependent variable, two specifications were estimated, one with industry-region pair fixed effects, labeled by "Ind.-reg. FE," and one with both industry and region fixed effects, denoted by "Ind.&reg. FE." White's robust standard errors should be used if the disturbance $\epsilon_{i,t}^{m}$ is heteroscedastic. However, it may also be aucorrelated within industries, regions, or even industry-region pairs, in which case one should rely on cluster-robust standard errors for proper inference. Here, the cluster-robust standard errors are defined with respect to industry-region units of observation. This correction of standard errors still presumes that the number of observations is large and the $\eta_{i,t}^{m}$'s are independent across industry-region pairs. Thus, for each fixed-effects specification (that is, Ind.-reg. FE or Ind.&reg. FE), estimated coefficients are reported with both types of corrected standard errors.[20]

In the output-TFP covariance regression using industry-region fixed effects, none of the coefficients are significant. With both industry and region (separate) fixed effects, the coefficient estimate for the lagged tariff is positive and significant at the 5 percent level (based on the robust standard error). Thus, the tariff rate is estimated to have a positive effect on the covariance of output and productivity, all else constant. In other words, tariff liberalization would decrease the efficiency of market shares allocation within an industry. This result seems counter-intuitive but it is consistent with the theoretical outcome in the case of unilateral trade liberalization. It is worth mentioning that Fernandes [2007] does not obtain conclusive results with regard to

[20] It is also likely that errors are either correlated across regions or across industries (or both). However, there is no available command in the software STATA to correct for such types of correlation since industries and regions do not nest each other.

the effect of the tariff rate on the output-productivity covariance (her analysis is at the 3-digit level and the tariff coefficient she estimates is negative but insignificant at conventional levels). The estimated coefficients of the tariff rate's interaction with the coastal indicator and of regional GDP are insignificant. The estimated coefficient for the interaction of tariff with regional GDP is negative and significant at the 1 percent level (with the robust standard error). According to this estimate, an industry located in a larger geographic market and facing lower tariff protection experience a greater reallocation of market shares toward more productive plants. With the cluster-robust standard error, the same coefficient remains significant only at the 10 percent level.

In the median TFP regression, both fixed-effects specifications yield a negative and significant coefficient estimate (based on robust standard errors) for the regional GDP variable, suggesting that an industry in a larger geographic market has more low productivity plants. This result is inconsistent with the model in general since a larger market should have a higher zero-cutoff productivity level and thus a higher median productivity level.

Table 4.2 reports estimates based on a partition of Colombia into three regions (Caribbean, Greater Pacific, and Inland). For output-TFP covariance Ind.&reg. FE regression (7), the estimated coefficient of the tariff-GDP interaction term is still negative and significant (at the 10 percent level, with robust standard errors). The main difference concerns the coastal interaction term that is now positive and significant at the 1 percent level (with robust standard errors). This result means that a lower tariff for an industry located closer to a coast is estimated to entail lower productivity through reallocation of market shares toward more productive plants. This estimate matches the predicted impact of unilateral trade liberalization. The results for the median TFP regression are similar to those obtained in the four-region case.

4.6 Conclusion

Theoretical and empirical studies of the implication of trade policy for plant- and industry-level productivity usually assume that production takes place in a single location and thus that all producers face the same exposure to domestic and foreign competition. Yet, regional disparities in demand density or proximity to external trading partners are ubiquitous and they matter (see Syverson [2004] for evidence about the role of geographic market size in determining the plant productivity distribution within an industry; see Limao and Venables [2001] about the influence of coastal access on economic performance). These variations in regional geographic characteristics within a country may interact with trade policy in determining the aggregate performance of industries.

This paper has presented an extension of the Melitz and Ottaviano [2008] model that incorporates such regional differences and generates predictions regarding the impact of trade liberalization on aggregate industry productivity across regions exhibiting different market sizes and degrees of exposure to foreign trade. This model provides guidance for the empirical analysis of regional disparities within an open economy. This framework of analysis was implemented to study the effect of trade liberalization on Colombia's regions.

The evidence on the influence of tariff protection on aggregate industry productivity across regions, in terms of market shares allocation efficiency, is weak. However, it seems that regional market size and foreign market access, through proximity to ports, play a role in determining the impact of tariff liberalization on industry performance.

Table 4.1: Impact of Tariff Policy and Geography on Industry Productivity in Colombia, 1983–91—Four Regions

	Mean TFP				Output-TFP covariance				Median TFP			
	Ind.-reg. FE		Ind.&reg. FE		Ind.-reg. FE		Ind.&reg. FE		Ind.-reg. FE		Ind.&reg. FE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Robust	Cluster	Robust	Cluster	Robust	Cluster	Robust	Cluster	Robust	Cluster	Robust	Cluster
Lagged tariff	7.183	7.183	8.297	8.297	25.59	25.59	57.04**	57.04	-13.59	-13.59	-12.28	-12.28
	(0.55)	(0.30)	(0.94)	(0.70)	(0.85)	(0.60)	(1.97)	(1.60)	(-0.53)	(-0.30)	(-0.74)	(-0.55)
Coastal × Lagged tariff	4.166	4.166	0.310	0.310	-15.82	-15.82	-5.309	-5.309	15.34	15.34	3.972	3.972
	(1.00)	(0.56)	(0.14)	(0.10)	(-1.10)	(-0.79)	(-0.81)	(-0.53)	(1.29)	(0.73)	(1.10)	(0.66)
Regional GDP	-12.90**	-12.90	-12.10**	-12.10	20.94	20.94	20.03	20.03	-26.58**	-26.58	-24.75**	-24.75
	(-2.24)	(-1.26)	(-2.30)	(-1.35)	(1.45)	(1.01)	(1.16)	(1.19)	(-2.05)	(-1.14)	(-2.24)	(-1.27)
GDP × Lagged tariff	-0.0456	-0.0456	-0.0434	-0.0434	-0.0701	-0.0701	-0.332***	-0.332*	0.0116	0.0116	0.0381	0.0381
	(-0.75)	(-0.47)	(-0.86)	(-1.35)	(-0.65)	(-0.47)	(-2.69)	(-1.75)	(0.11)	(0.07)	(0.46)	(0.70)
Observations	1463	1463	1463	1463	1464	1464	1464	1464	1463	1463	1463	1463
Adjusted R^2	0.001	0.001	0.885	0.885	0.004	0.004	0.502	0.502	0.004	0.004	0.704	0.704

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes:

Every regression is estimated using the years 1983, 1985, 1987, 1989, and 1991 and comprises fixed effects for all but one of these years. Regions are Caribbean, Central, Inland, and Pacific.

Table 4.2: Impact of Tariff Policy and Geography on Industry Productivity in Colombia, 1983–91—Three Regions

	Mean TFP				Output-TFP covariance				Median TFP			
	Ind.-reg. FE		Ind.&reg. FE		Ind.-reg. FE		Ind.&reg. FE		Ind.-reg. FE		Ind.&reg. FE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Robust	Cluster	Robust	Cluster	Robust	Cluster	Robust	Cluster	Robust	Cluster	Robust	Cluster
Lagged tariff	10.77	10.77	7.598	7.598	38.83	38.83	62.35	62.35	-26.16	-26.16	-21.61	-21.61
	(0.39)	(0.23)	(0.49)	(0.37)	(0.78)	(0.54)	(1.26)	(1.01)	(-0.53)	(-0.30)	(-0.72)	(-0.54)
Coastal × Lagged tariff	7.173	7.173	4.525	4.525	-17.42	-17.42	22.52***	22.52	21.14	21.14	3.966	3.966
	(1.23)	(0.77)	(1.18)	(1.03)	(-1.11)	(-0.81)	(2.63)	(1.46)	(1.36)	(0.82)	(0.72)	(0.49)
Regional GDP	-22.81**	-22.81	-21.91**	-21.91	27.55*	27.55	22.97	22.97	-38.91**	-38.91	-35.82**	-35.82
	(-1.98)	(-1.16)	(-2.01)	(-1.22)	(1.80)	(1.19)	(1.19)	(1.19)	(-2.15)	(-1.24)	(-2.30)	(-1.33)
GDP × Lagged tariff	-0.0705	-0.0705	-0.0452	-0.0452	-0.0811	-0.0811	-0.325*	-0.325	0.0161	0.0161	0.0484	0.0484
	(-0.61)	(-0.40)	(-0.60)	(-0.84)	(-0.53)	(-0.38)	(-1.92)	(-1.31)	(0.09)	(0.06)	(0.39)	(0.52)
Observations	1113	1113	1113	1113	1114	1114	1114	1114	1113	1113	1113	1113
Adjusted R^2	0.005	0.005	0.884	0.884	0.004	0.004	0.494	0.494	0.006	0.006	0.699	0.699

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes:

Every regression is estimated using the years 1983, 1985, 1987, 1989, and 1991 and comprises fixed effects for all but one of these years.

Regions are Caribbean, Greater Pacific, and Inland.

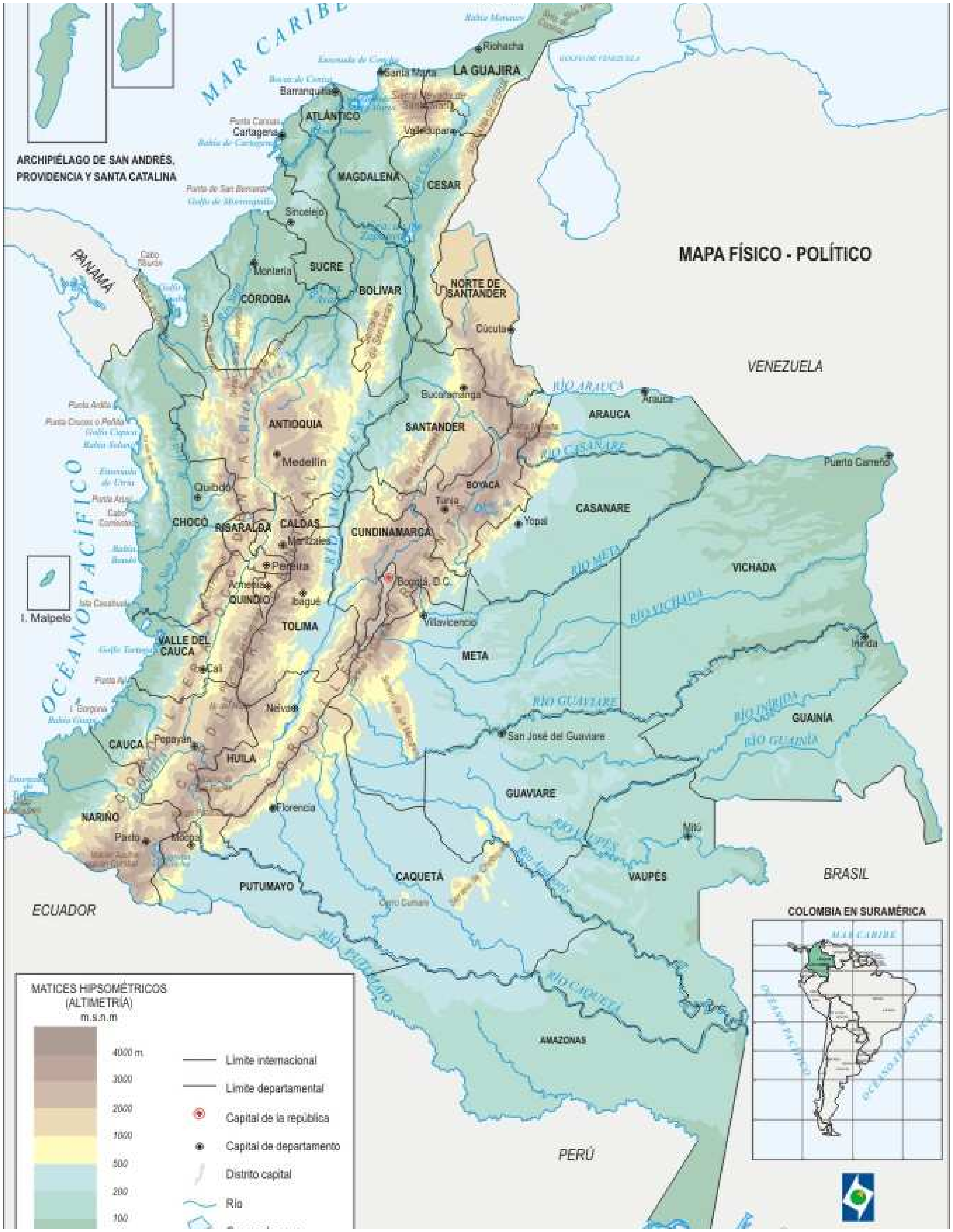

Figure 4.1: Map of Colombia

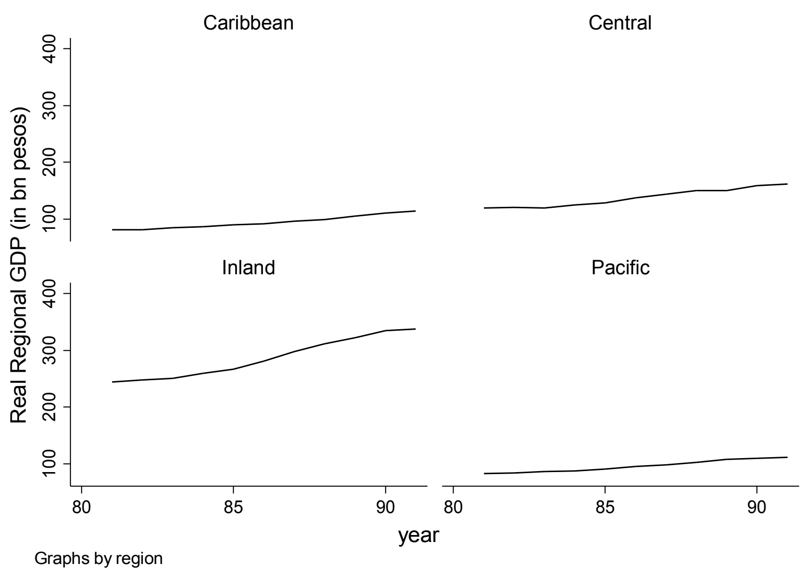

Figure 4.2: Regional Gross Domestic Product in Colombia, 1981–91

Appendix A Trade Liberalization and the Geographic Location of Industries

A.1 Closed Economy Equilibrium

A.1.1 Derivation of the Average Rural Profit

The equality $r^u(\theta_{u,r})/r^u(\theta^*) = (\theta_{u,r}/\theta^*)^{\sigma-1}$ (see (2.13)) and the zero cutoff profit condition, $\pi^u(\theta^*) = 0$, together entail $r^u(\theta_{u,r}) = \sigma w^u f^u(\theta_{u,r}/\theta^*)^{\sigma-1}$. In addition, the urban-rural cutoff profit condition, $\pi^u(\theta_{u,r}) = \pi^r(\theta_{u,r})$, can be rewritten as $r^r(\theta_{u,r}) = r^u(\theta_{u,r}) + \sigma(w^r f^r - w^u f^u)$. Substituting the expression of $r^u(\theta_{u,r})$ into the latter equation gives

$$r^r(\theta_{u,r}) = \sigma\left[w^u f^u\left[\left(\frac{\theta_{u,r}}{\theta^*}\right)^{\sigma-1} - 1\right] + w^r f^r\right]$$

By analogy to the derivation of the average urban profit, the average rural profit is given by $\pi^r(\tilde{\theta^r}(\theta_{u,r})) = [\tilde{\theta^r}(\theta_{u,r})/\theta_{u,r}]^{\sigma-1} r^r(\theta_{u,r})/\sigma - w^r f^r$ (using (2.5) and (2.13)). Substituting the expression of $r^r(\theta_{u,r})$ into that of $\pi^r(\tilde{\theta^r}(\theta_{u,r}))$ yields

$$\pi^r(\tilde{\theta^r}(\theta_{u,r})) = w^u f^u\left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta^*}\right)^{\sigma-1} - \left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{u,r}}\right)^{\sigma-1}\right] + w^r f^r\left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{u,r}}\right)^{\sigma-1} - 1\right]$$

A.1.2 Existence and Uniqueness of the Equilibrium Zero Cutoff Productivity Level

We show that the zero cutoff profit condition (2.14), the urban-rural cutoff profit condition (2.15), and the free entry condition (3.7) determine a unique cutoff productivity level θ^*. To do so we prove that there is a unique value of θ, θ^*, that satisfies the equilibrium condition

$$[G(\alpha\theta) - G(\theta)]\,\bar{\pi}^u(\theta,\alpha\theta) + [1 - G(\alpha\theta)]\,\bar{\pi}^r(\theta,\alpha\theta) = \delta f_e \tag{A.1}$$

where $\bar{\pi}^u(\theta,\alpha\theta) = w^u f^u k^u(\theta,\alpha\theta)$ and $\bar{\pi}^r(\theta,\alpha\theta) = w^u f^u k_1^r(\theta,\alpha\theta) + w^r f^r k_2^r(\alpha\theta)$; $k^u(\theta,\alpha\theta) = [\tilde{\theta^u}(\theta,\alpha\theta)/\theta]^{\sigma-1} - 1$, $k_1^r(\theta,\alpha\theta) = [\tilde{\theta^r}(\alpha\theta)/\theta]^{\sigma-1} - [\tilde{\theta^r}(\alpha\theta)/\alpha\theta]^{\sigma-1}$, and $k_2^r(\alpha\theta) = [\tilde{\theta^r}(\alpha\theta)/\alpha\theta]^{\sigma-1} - 1$; $\tilde{\theta^u}(\theta,\alpha\theta) = \left[[1/(G(\alpha\theta) - G(\theta))]\int_\theta^{\alpha\theta}\xi^{\sigma-1}g(\xi)d\xi\right]^{1/(\sigma-1)}$ and $\tilde{\theta^r}(\alpha\theta) = \left[[1/(1-G(\alpha\theta))]\int_{\alpha\theta}^{\infty}\xi^{\sigma-1}g(\xi)d\xi\right]^{1/(\sigma-1)}$. A sufficient condition for the existence and uniqueness of the solution is that the left-hand side of equation (A.1) be monotonically decreasing on $(0,\infty)$, tending towards infinity for values of θ near zero, and approaching zero for infinitely large values of θ.

The derivatives of k^u, k^r_1, and k^r_2 with respect to θ, denoted by $(k^u)'$, $(k^r_1)'$, and $(k^r_2)'$, respectively are given by:

$$
\begin{aligned}
(k^u)'(\theta,\alpha\theta) &= \frac{\alpha^\sigma g(\alpha\theta)-g(\theta)}{G(\alpha\theta)-G(\theta)} - [k^u(\theta,\alpha\theta)+1]\left[\frac{\alpha g(\alpha\theta)-g(\theta)}{G(\alpha\theta)-G(\theta)}+\frac{\sigma-1}{\theta}\right]\\
(k^r_1)'(\theta,\alpha\theta) &= -(\alpha^{\sigma-1}-1)\frac{\alpha g(\alpha\theta)}{1-G(\alpha\theta)} + k^r_1(\theta,\alpha\theta)\left[\frac{\alpha g(\alpha\theta)}{1-G(\alpha\theta)}-\frac{\sigma-1}{\theta}\right]\\
(k^r_2)'(\alpha\theta) &= -\frac{\alpha g(\alpha\theta)}{1-G(\alpha\theta)} + [k^r_2(\alpha\theta)+1]\left[\frac{\alpha g(\alpha\theta)}{1-G(\alpha\theta)}-\frac{\sigma-1}{\theta}\right]
\end{aligned}
$$

Let j^u and j^r denote functions of θ defined as $j^u(\theta,\alpha\theta) = [G(\alpha\theta)-G(\theta)]\bar{\pi}^u(\theta,\alpha\theta)$ and $j^r(\theta,\alpha\theta) = [1-G(\alpha\theta)]\bar{\pi}^r(\theta,\alpha\theta)$. The derivatives of j^u and j^r with respect to θ (denoted $(j^u)'$ and $(j^r)'$, respectively) are

$$
\begin{aligned}
(j^u)'(\theta,\alpha\theta) &= w^u f^u \alpha g(\alpha\theta)(\alpha^{\sigma-1}-1) - [G(\alpha\theta)-G(\theta)]w^u f^u [k^u(\theta,\alpha\theta)+1]\frac{\sigma-1}{\theta}\\
(j^r)'(\theta,\alpha\theta) &= -\, w^u f^u \alpha g(\alpha\theta)(\alpha^{\sigma-1}-1)\\
&\quad - [1-G(\alpha\theta)]\big[w^u f^u k^r_1(\theta,\alpha\theta) + w^r f^r [k^r_2(\alpha\theta)+1]\big]\frac{\sigma-1}{\theta}
\end{aligned}
$$

Let j be the sum of j^u and j^r. Thus, $j(\theta,\alpha\theta)$ is the left-hand side of (A.1). The derivative of j with respect to θ is given by $j'(\theta,\alpha\theta) = (j^u)'(\theta,\alpha\theta) + (j^r)'(\theta,\alpha\theta)$, that is,

$$
\begin{aligned}
j'(\theta,\alpha\theta) = -\Big[&[G(\alpha\theta)-G(\theta)]w^u f^u[k^u(\theta,\alpha\theta)+1]\\
&+ [1-G(\alpha\theta)]\big[w^u f^u k^r_1(\theta,\alpha\theta) + w^r f^r[k^r_2(\alpha\theta)+1]\big]\Big]\frac{\sigma-1}{\theta} < 0 \quad \text{(A.2)}
\end{aligned}
$$

The elasticity of j with respect to θ is:

$$
\begin{aligned}
&\frac{j'(\theta,\alpha\theta)\theta}{j(\theta,\alpha\theta)} = -(\sigma-1)\\
&\times\frac{[G(\alpha\theta)-G(\theta)]w^u f^u[k^u(\theta,\alpha\theta)+1] + [1-G(\alpha\theta)]\big[w^u f^u k^r_1(\theta,\alpha\theta) + w^r f^r[k^r_2(\alpha\theta)+1]\big]}{[G(\alpha\theta)-G(\theta)]k^u(\theta,\alpha\theta) + [1-G(\alpha\theta)]\big[w^u f^u k^r_1(\theta,\alpha\theta) + w^r f^r k^r_2(\alpha\theta)\big]}\\
&< -(\sigma-1)
\end{aligned}
$$

This elasticity is less than $-(\sigma-1)$ because the fraction on the right-hand side is greater than one and thus it is strictly negative. We also know that j is nonnegative. Hence, j must be falling to zero as θ goes to infinity. In addition, $\lim_{\theta\to 0} k^u(\theta) = \infty$, $\lim_{\theta\to 0} k^r_1(\theta) = \infty$, and $\lim_{\theta\to 0} k^r_2(\theta) = \infty$. Hence, we have $\lim_{\theta\to 0} j(\theta) = \infty$. Therefore, j is monotonically decreasing from infinity to zero on $(0,\infty)$.

A.2 Open Economy Equilibrium—Case (a)

A.2.1 Derivation of the Average Rural Profit

Using the equality $r_d^u(\theta_{u,r})/r_d^u(\theta^*) = (\theta_{u,r}/\theta^*)^{\sigma-1}$ (see (2.13)) and the zero cutoff profit condition, $\pi_d^u(\theta^*) = 0$, one obtains $r_d^u(\theta_{u,r}) = \sigma w^u f^u (\theta_{u,r}/\theta^*)^{\sigma-1}$. Similarly, substituting the export cutoff condition, $\pi_f^u(\theta_{ex}) = 0$, into the equality $r_f^u(\theta_{u,r})/r_f^u(\theta_{ex}) = (\theta_{u,r}/\theta_{ex})^{\sigma-1}$ gives $r_f^u(\theta_{u,r}) = \sigma f_{ex}(\theta_{u,r}/\theta_{ex})^{\sigma-1}$. The urban-rural cutoff profit condition, $\pi^u(\theta_{u,r}) = \pi^r(\theta_{u,r})$, now restated as $[r_d^u(\theta_{u,r}) + r_f^u(\theta_{u,r})]/\sigma - w^u f^u = r^r(\theta_{u,r})/\sigma - w^r f^r$, after rearranging the terms entails $r^r(\theta_{u,r}) = r_d^u(\theta_{u,r}) + r_f^u(\theta_{u,r}) + \sigma(w^r f^r - w^u f^u)$. Substituting the expressions for $r_d^u(\theta_{u,r})$ and $r_f^u(\theta_{u,r})$ into the one for $r^r(\theta_{u,r})$ yields

$$r^r(\theta_{u,r}) = \sigma \left[w^u f^u \left[\left(\frac{\theta_{u,r}}{\theta^*} \right)^{\sigma-1} - 1 \right] + f_{ex} \left(\frac{\theta_{u,r}}{\theta_{ex}} \right)^{\sigma-1} + w^r f^r \right]$$

The expression of the average rural profit in the open economy is derived by substituting the expression of $r^r(\theta_{u,r})$ into $\pi^r(\tilde{\theta^r}(\theta_{u,r})) = [\tilde{\theta^r}(\theta_{u,r})/\theta_{u,r}]^{\sigma-1} r^r(\theta_{u,r})/\sigma - w^r f^r - f_{ex}$, which eventually gives

$$\begin{aligned}\bar{\pi}^r(\theta^*, \theta_{ex}, \theta_{u,r}) = w^u f^u & \left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta^*} \right)^{\sigma-1} - \left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{u,r}} \right)^{\sigma-1} \right] \\ & + w^r f^r \left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{u,r}} \right)^{\sigma-1} - 1 \right] + f_{ex} \left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{ex}} \right)^{\sigma-1} - 1 \right]\end{aligned}$$

A.2.2 Existence and Uniqueness of the Equilibrium Cutoff Productivity Level

We show that the zero cutoff profit condition (2.29), the urban-rural cutoff profit condition (2.30), and the free entry condition (2.31) define a unique cutoff productivity level θ^* by proving that there is a unique value of θ, θ^*, satisfying the equilibrium condition

$$[G(\gamma\theta) - G(\theta)]\, \bar{\pi}^u(\theta, \eta\theta, \gamma\theta) + [1 - G(\gamma\theta)]\, \bar{\pi}^r(\theta, \eta\theta, \gamma\theta) = \delta f_e \qquad \text{(A.3)}$$

$$\text{where } \bar{\pi}^u(\theta, \eta\theta, \gamma\theta) = w^u f^u k^u(\theta, \gamma\theta) + \frac{G(\gamma\theta) - G(\eta\theta)}{G(\gamma\theta) - G(\theta)} f_{ex} k^u(\eta\theta, \gamma\theta)$$

$$\text{and } \bar{\pi}^r(\theta, \eta\theta, \gamma\theta) = w^u f^u k_1^r(\theta, \gamma\theta) + w^r f^r k_2^r(\gamma\theta) + f_{ex} k_3^r(\eta\theta, \gamma\theta)$$

k^u, k_1^r, k_2^r, $\tilde{\theta^u}$, and $\tilde{\theta^r}$ are defined as previously; $k_3^r(\eta\theta, \gamma\theta) = [\tilde{\theta^r}(\gamma\theta)/\eta\theta]^{\sigma-1} - 1$. A sufficient condition for the existence and uniqueness of the solution is that the left-hand side of equation (A.3) be monotonically decreasing on the interval $(0, \infty)$, going towards infinity for values of θ close to zero, and tending towards zero for infinitely large values of θ.

Let j_f^u and j_f^r be functions of θ defined as $j_f^u(\eta\theta,\gamma\theta) = [G(\gamma\theta)-G(\eta\theta)]f_{ex}k^u(\eta\theta,\gamma\theta)$ and $j_f^r(\eta\theta,\gamma\theta) = [1-G(\gamma\theta)]f_{ex}k_3^r(\eta\theta,\gamma\theta)$. The derivatives of j_f^u and j_f^r with respect to θ are given by

$$(j_f^u)'(\eta\theta,\gamma\theta) = f_{ex}\gamma g(\gamma\theta)\left[\left(\frac{\gamma}{\eta}\right)^{\sigma-1} - 1\right] - [G(\gamma\theta)-G(\eta\theta)]f_{ex}[k^u(\eta\theta,\gamma\theta)+1]\frac{\sigma-1}{\theta}$$

$$(j_f^r)'(\eta\theta,\gamma\theta) = -f_{ex}\gamma g(\gamma\theta)\left[\left(\frac{\gamma}{\eta}\right)^{\sigma-1} - 1\right] - [1-G(\gamma\theta)]f_{ex}[k_3^r(\eta\theta,\gamma\theta)+1]\frac{\sigma-1}{\theta}$$

We define j_f as $j_f^u + j_f^r$; thus $j(\theta,\gamma\theta) + j_f(\eta\theta,\gamma\theta)$ is the left-hand side of (A.3). The derivative of j_f with respect to θ (denoted j_f') is given by $j_f'(\eta\theta,\gamma\theta) = (j_f^u)'(\eta\theta,\gamma\theta) + (j_f^r)'(\eta\theta,\gamma\theta)$, that is,

$$\begin{aligned} j_f'(\eta\theta,\gamma\theta) = -f_{ex}\big[&[G(\gamma\theta)-G(\eta\theta)][k^u(\eta\theta,\gamma\theta)+1] \\ &+ [1-G(\gamma\theta)][k_3^r(\eta\theta,\gamma\theta)+1]\big]\frac{\sigma-1}{\theta} < 0 \end{aligned} \quad \text{(A.4)}$$

Thus, the elasticity of j_f with respect to θ is given by

$$\begin{aligned} \frac{j_f'(\eta\theta,\gamma\theta)\theta}{j_f(\eta\theta,\gamma\theta)} &= -(\sigma-1) \\ &\times \frac{[G(\gamma\theta)-G(\eta\theta)][k^u(\eta\theta,\gamma\theta)+1] + [1-G(\gamma\theta)][k_3^r(\eta\theta,\gamma\theta)+1]}{[G(\gamma\theta)-G(\eta\theta)]k^u(\eta\theta,\gamma\theta) + [1-G(\gamma\theta)]k_3^r(\eta\theta,\gamma\theta)} < -(\sigma-1) \end{aligned}$$

As the fraction on the right-hand side is greater than one, the elasticity of j_f with respect to θ is less than $-(\sigma-1)$, and as a result it is strictly negative. Therefore, as θ goes to infinity, j_f must be decreasing towards zero. In addition, $\lim_{\theta\to 0} k^u(\eta\theta,\gamma\theta) = \infty$ and $\lim_{\theta\to 0} k_3^r(\eta\theta,\gamma\theta) = \infty$. Thus, we also have $\lim_{\theta\to 0} j_f(\eta\theta,\gamma\theta) = \infty$. Hence, j_f is monotonically decreasing from infinity to zero on $(0,\infty)$. As γ does not depend on θ, we know from the appendix A.1.2 that j is monotonically decreasing from infinity to zero on $(0,\infty)$. Therefore, $j + j_f$ is also monotonically decreasing from infinity to zero on $(0,\infty)$. Thus, equation (A.3) determines a unique cutoff level θ^*.

A.3 Open Economy Equilibrium—Case (b)

A.3.1 Derivation of the Average Rural Profit

Note that in this case, the expression of $r_d^r(\theta_{u,r})$ is identical to that derived in (A.1.1) for $r^r(\theta_{u,r})$ (because below $\theta_{u,r}$ urban firms do not export and immediately above

$\theta_{u,r}$ rural firms do not export either). Thus, $\bar{\pi}^r_d$, the average rural domestic profit, is now defined as $\bar{\pi}^r$ in (2.15). In addition, the average rural export profit is given by $\pi^r_f(\tilde{\theta^r}(\theta_{ex})) = [\tilde{\theta^r}(\theta_{ex})/\theta_{ex}]^{\sigma-1} r^r_f(\theta_{ex})/\sigma - f_{ex}$. Substituting the export cutoff condition, $\pi^r_f(\theta_{ex}) = 0$ (that is $r^r_f(\theta_{ex}) = \sigma f_{ex}$), into the last expression yields

$$\bar{\pi}^r_f(\theta_{ex}) = f_{ex}\left[\left(\frac{\tilde{\theta^r}(\theta_{ex})}{\theta_{ex}}\right)^{\sigma-1} - 1\right]$$

Thus, the average profit over all rural firms, $\bar{\pi}^r(\theta^*, \theta_{u,r}, \theta_{ex}) = \bar{\pi}^r_d(\theta^*, \theta_{u,r}) + \bar{\pi}^r_f(\theta_{ex})$, is

$$\bar{\pi}^r(\theta^*, \theta_{u,r}, \theta_{ex}) = w^u f^u \left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta^*}\right)^{\sigma-1} - \left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{u,r}}\right)^{\sigma-1}\right] + w^r f^r \left[\left(\frac{\tilde{\theta^r}(\theta_{u,r})}{\theta_{u,r}}\right)^{\sigma-1} - 1\right] + q^r_{ex} f_{ex}\left[\left(\frac{\tilde{\theta^r}(\theta_{ex})}{\theta_{ex}}\right)^{\sigma-1} - 1\right]$$

A.3.2 Existence and Uniqueness of the Equilibrium Cutoff Productivity Level

As in case (a) we want to show that the zero cutoff profit condition (2.34), the urban-rural cutoff profit condition (2.35), and the free entry condition (2.31) identify a unique cutoff productivity level θ^*.

$$[G(\alpha\theta) - G(\theta)]\,\bar{\pi}^u(\theta, \alpha\theta) + [1 - G(\alpha\theta)]\,\bar{\pi}^r(\theta, \alpha\theta, \beta\theta) = \delta f_e \tag{A.5}$$
$$\text{where } \bar{\pi}^u(\theta, \alpha\theta) = w^u f^u k^u(\theta, \alpha\theta)$$
$$\text{and } \bar{\pi}^r(\theta, \alpha\theta, \beta\theta) = w^u f^u k^r_1(\theta, \alpha\theta) + w^r f^r k^r_2(\alpha\theta) + f_{ex} k^r_4(\beta\theta)$$

In addition, $k^r_4(\beta\theta) = [\tilde{\theta^r}(\beta\theta)/\beta\theta]^{\sigma-1} - 1$. A sufficient condition for the existence and uniqueness of the equilibrium is that the left-hand side of equation (A.5) is monotonically decreasing on $(0, \infty)$. In other words, it should be tending towards infinity for values of θ near zero and approaching zero when θ tends towards infinity.

We define j^r_f as $j^r_f(\beta\theta) = [1 - G(\beta\theta)] f_{ex} k^r_4(\beta\theta)$. Recall that $j(\theta, \alpha\theta) + j^r_f(\beta\theta)$ is the left-hand side of (A.5). The derivative of j^r_f with respect to θ (denoted $(j^r_f)'$) is given by

$$(j^r_f)'(\beta\theta) = -f_{ex}[1 - G(\beta\theta)][k^r_4(\beta\theta) + 1]\frac{\sigma - 1}{\theta}$$

The elasticity of $j^r_f(\beta\theta)$ with respect to θ is

$$\frac{(j^r_f)'(\beta\theta)\theta}{j^r_f(\beta\theta)} = -(\sigma - 1)\frac{k^r_4(\beta\theta) + 1}{k^r_4(\beta\theta)} < -(\sigma - 1)$$

The fraction on the right-hand side is greater than one, so the elasticity of j_f^r with respect to θ is less than $-(\sigma - 1)$. Thus, it is strictly negative. As θ approaches infinity, j_f^r must be going towards zero. We also have that $\lim_{\theta \to 0} k_4^r(\beta\theta) = \infty$ and $\lim_{\theta \to 0} j_f^r(\beta\theta) = \infty$. Therefore, j_f^r is monotonically decreasing from infinity to zero on the interval $(0, \infty)$. We showed in appendix A.1.2 that j is monotonically decreasing from infinity to zero on $(0, \infty)$. Thus, $j + j_f^r$ is also monotonically decreasing from infinity to zero on $(0, \infty)$, and hence equation (A.5) determines a unique cutoff level θ^*.

A.4 The Impact of Trade Liberalization

A.4.1 Shifts in the Cutoff Productivity Levels

A.4.1.1 Case (a): Urban Non-exporters, Urban Exporters, and Rural Exporters

We derive comparative statics of the zero cutoff, export cutoff, and urban-rural cutoff productivity levels (θ^*, θ_{ex}, and $\theta_{u,r}$) for a change in the variable trade cost, τ. Recall (from appendix A.2.2) that θ^*, $\theta^* \in (0, \infty)$, is the equilibrium zero cutoff productivity level if and only if

$$j(\theta^*, \theta_{u,r}) + j_f(\theta_{ex}, \theta_{u,r}) = \delta f_e \tag{A.6}$$

where θ_{ex} and $\theta_{u,r}$ are implicitly defined as functions of θ^* as in (2.25) and (2.26). Differentiating (A.6) with respect to τ gives:

$$\frac{\partial j(\cdot)}{\partial \theta^*}\frac{\partial \theta^*}{\partial \tau} + \frac{\partial j(\cdot)}{\partial \theta_{u,r}}\frac{\partial \theta_{u,r}}{\partial \tau} + \frac{\partial j_f(\cdot)}{\partial \theta_{ex}}\frac{\partial \theta_{ex}}{\partial \tau} + \frac{\partial j_f(\cdot)}{\partial \theta_{u,r}}\frac{\partial \theta_{u,r}}{\partial \tau} = 0$$

Then, using the fact that $\partial j(\cdot)/\partial \theta_{u,r} = (1/\gamma)\partial j(\cdot)/\partial \theta^*$, $\partial j_f(\cdot)/\partial \theta_{ex} = (1/\eta)\partial j_f(\cdot)/\partial \theta^*$, and $\partial j_f(\cdot)/\partial \theta_{u,r} = (1/\gamma)\partial j_f(\cdot)/\partial \theta^*$, by substituting in $\partial \theta_{ex}/\partial \tau = (\theta_{ex}/\theta^*)\partial \theta^*/\partial \tau + \theta_{ex}/\tau$ and $\theta_{u,r}/\partial \tau = (\theta_{u,r}/\theta^*)\partial \theta^*/\partial \tau + \theta_{u,r}/(\tau + \tau^\sigma)$ and rearranging the terms, one obtains

$$\frac{\partial \theta^*}{\partial \tau} = -\frac{\theta^*}{\tau}\frac{1}{2(1+\tau^{\sigma-1})}\frac{j'(\theta^*, \gamma\theta^*) + (2+\tau^{\sigma-1})j_f'(\eta\theta^*, \gamma\theta^*)}{j'(\theta^*, \gamma\theta^*) + j_f'(\eta\theta^*, \gamma\theta^*)} < 0 \tag{A.7}$$

Given that $j'(\theta, \gamma\theta) < 0$ and $j_f'(\eta\theta, \gamma\theta) < 0 \;\forall\; \theta \in (0, \infty)$. Substituting (A.7) into $\partial \theta_{ex}/\partial \tau = (\theta_{ex}/\theta^*)\partial \theta^*/\partial \tau + \theta_{ex}/\tau$ yields

$$\frac{\partial \theta_{ex}}{\partial \tau} = \frac{\theta_{ex}}{\tau}\frac{1}{2(1+\tau^{\sigma-1})}\frac{(1+2\tau^{\sigma-1})j'(\theta^*, \gamma\theta^*) + \tau^{\sigma-1}j_f'(\eta\theta^*, \gamma\theta^*)}{j'(\theta^*, \gamma\theta^*) + j_f'(\eta\theta^*, \gamma\theta^*)} > 0 \tag{A.8}$$

Then, substituting (A.7) into $\partial\theta_{u,r}/\partial\tau = (\theta_{u,r}/\theta^*)\partial\theta^*/\partial\tau + \theta_{u,r}/(\tau+\tau^\sigma)$ gives

$$\frac{\partial\theta_{u,r}}{\partial\tau} = \frac{\theta_{u,r}}{\tau}\frac{1}{2(1+\tau^{\sigma-1})}\frac{j'(\theta^*,\gamma\theta^*) - \tau^{\sigma-1}j_f'(\eta\theta^*,\gamma\theta^*)}{j'(\theta^*,\gamma\theta^*) + j_f'(\eta\theta^*,\gamma\theta^*)} \tag{A.9}$$

To determine the sign of $\partial\theta_{u,r}/\partial\tau$ we have to know the sign of the $j'(\theta^*,\gamma\theta^*) - \tau^{\sigma-1}j_f'(\eta\theta^*,\gamma\theta^*)$. Using (A.2), (A.4), the expressions for k^u, k_1^r, k_2^r, k_3^r, and (2.25), we obtain

$$\begin{aligned}
\frac{j'(\theta^*,\gamma\theta^*)}{j_f'(\eta\theta^*,\gamma\theta^*)} &= \frac{w^u f^u}{f_{ex}} \\
&\times \frac{[G(\gamma\theta^*)-G(\theta^*)][k^u(\theta^*,\gamma\theta^*)+1] + [1-G(\gamma\theta^*)](k_1^r(\theta^*,\gamma\theta^*) + \frac{w^r f^r}{w^u f^u}[k_2^r(\gamma\theta^*)+1])}{[G(\gamma\theta^*)-G(\eta\theta^*)][k^u(\eta\theta^*,\gamma\theta^*)+1] + [1-G(\gamma\theta^*)][k_3^r(\eta\theta^*,\gamma\theta^*)+1]} \\
&= \frac{\tau^{\sigma-1}}{\eta^{\sigma-1}}\frac{\frac{\int_{\theta^*}^{\gamma\theta^*}\xi^{\sigma-1}g(\xi)d\xi}{(\theta^*)^{\sigma-1}} + \frac{\int_{\gamma\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi}{(\theta^*)^{\sigma-1}} - \frac{\int_{\gamma\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi}{(\gamma\theta^*)^{\sigma-1}} + \frac{w^r f^r}{w^u f^u}\frac{\int_{\gamma\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi}{(\gamma\theta^*)^{\sigma-1}}}{\frac{\int_{\eta\theta^*}^{\gamma\theta^*}\xi^{\sigma-1}g(\xi)d\xi}{(\eta\theta^*)^{\sigma-1}} + \frac{\int_{\gamma\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi}{(\eta\theta^*)^{\sigma-1}}} \\
&= \tau^{\sigma-1}\left[\int_{\theta^*}^{\gamma\theta^*}\xi^{\sigma-1}g(\xi)d\xi + \int_{\gamma\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi - \gamma^{1-\sigma}\int_{\gamma\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi\right. \\
&\left. + \gamma^{1-\sigma}\frac{w^r f^r}{w^u f^u}\int_{\gamma\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi\right]\left[\int_{\eta\theta^*}^{\gamma\theta^*}\xi^{\sigma-1}g(\xi)d\xi + \int_{\gamma\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi\right]^{-1} \\
&= \tau^{\sigma-1}\frac{\int_{\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi + \gamma^{1-\sigma}\int_{\gamma\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi\left(\frac{w^r f^r}{w^u f^u}-1\right)}{\int_{\eta\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi} > \tau^{\sigma-1}
\end{aligned}$$

The inequality follows from the fact that $\int_{\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi > \int_{\eta\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi$ as $\eta > 1$, $w^r f^r/w^u f^u > 1$, and therefore the fraction is greater than one. Hence, $\partial\theta_{u,r}^*/\partial\tau > 0$.

A.4.1.2 Case (b): Urban Non-exporters, Rural Non-exporters, and Rural Exporters

We now obtain comparative statics of the cutoff productivity levels for a change in τ from the equilibrium condition

$$j(\theta^*, \theta_{u,r}) + j_f^r(\theta_{ex}) = \delta f_e \tag{A.10}$$

where $\theta_{u,r}$ and θ_{ex} are implicitly defined as functions of θ^* in (2.8) and (2.27). The derivative of (A.10) with respect to τ is:

$$\frac{\partial j(\cdot)}{\partial\theta^*}\frac{\partial\theta^*}{\partial\tau} + \frac{\partial j(\cdot)}{\partial\theta_{u,r}}\frac{\partial\theta_{u,r}}{\partial\tau} + \frac{\partial j_f^r(\cdot)}{\partial\theta_{ex}}\frac{\partial\theta_{ex}}{\partial\tau} = 0$$

Substituting in the previous equation $\partial j(\cdot)/\partial\theta_{u,r} = (1/\alpha)\partial j(\cdot)/\partial\theta^*$, $\partial j_f^r(\cdot)/\partial\theta_{ex} = (1/\beta)\partial j_f^r(\cdot)/\partial\theta^*$, $\partial\theta_{u,r}/\partial\tau = (\theta_{u,r}/\theta^*)\partial\theta^*/\partial\tau$, and $\partial\theta_{ex}/\partial\tau = (\theta_{ex}/\theta^*)\partial\theta^*/\partial\tau + \theta_{ex}/\tau$ and rearranging the terms we obtain:

$$\frac{\partial\theta^*}{\partial\tau} = -\frac{\theta^*}{\tau}\frac{(j_f^r)'(\beta\theta^*)}{2j'(\theta^*,\alpha\theta^*) + (j_f^r)'(\beta\theta^*)} < 0 \tag{A.11}$$

This inequality holds because $j'(\theta,\alpha\theta) < 0$ and $(j_f^r)'(\beta\theta) < 0$, $\forall\theta \in (0,\infty)$. Substituting equation (A.11) into $\partial\theta_{ex}/\partial\tau = (\theta_{ex}/\theta^*)\partial\theta^*/\partial\tau + \theta_{ex}/\tau$ yields

$$\frac{\partial\theta_{ex}}{\partial\tau} = \frac{\theta_{ex}}{\tau}\frac{2j'(\theta^*,\alpha\theta^*)}{2j'(\theta^*,\alpha\theta^*) + (j_f^r)'(\beta\theta^*)} > 0 \tag{A.12}$$

Like $\partial\theta^*/\partial\tau$, the derivative of $\theta_{u,r}$ with respect to τ is negative.

A.4.2 Reallocation of Market Shares

A.4.2.1 Case (a): Urban Non-exporters, Urban Exporters, and Rural Exporters

For all $\theta \in [\theta^*,\infty)$, $r_d^u(\theta)/r_d^u(\theta^*) = (\theta/\theta^*)^{\sigma-1}$ (using (2.19)). As $\pi_d^u(\theta^*) = 0$ entails $r_d^u(\theta^*) = \sigma w^u f^u$, then $r_d^u(\theta,\theta^*) = \sigma w^u f^u(\theta/\theta^*)^{\sigma-1}$. We know that $\partial\theta^*/\partial\tau < 0$. Hence, $\partial r_d^u(\theta,\theta^*)/\partial\tau > 0$. Similarly, for all $\theta \in [\theta^*,\infty)$, $r_d^r(\theta)/r_d^r(\theta^*) = (\theta/\theta^*)^{\sigma-1}$. Given that $r_d^r(\theta^*)/r_d^u(\theta^*) = (w^u/w^r)^{\sigma-1}$, it follows that $r_d^r(\theta,\theta^*) = (w^u/w^r)^{\sigma-1}r_d^u(\theta,\theta^*)$. Thus, $\partial r_d^r(\theta,\theta^*)/\partial\tau > 0$. Let $r_d^m(\theta) \equiv r_d^m(\theta,\theta^*)$ and $(r_d^m)'(\theta) \equiv r_d^m(\theta,\theta^{*\prime})$, for all $m \in \{u,r\}$. Consequently $\tau' < \tau$ entails $(r_d^m)'(\theta) < r_d^m(\theta)$.

For all $\theta \in [\theta^*,\infty)$, $r_d^m(\theta)+r_f^m(\theta) = (1+\tau^{1-\sigma})r_d^m(\theta)$ (from (2.19) and (2.20)). Thus, the revenue from domestic and foreign sales of a firm in the urban region is given by $(1+\tau^{1-\sigma})\sigma w^u f^u(\theta/\theta^*)^{\sigma-1}$. Likewise, the total of the domestic and foreign revenues of a rural firm is $(1+\tau^{1-\sigma})(w^u/w^r)^{\sigma-1}\sigma w^u f^u(\theta/\theta^*)^{\sigma-1}$. Therefore, the derivative with respect to τ of the sum of the domestic and export revenues has the same sign as the derivative of $(1+\tau^{1-\sigma})/(\theta^*)^{\sigma-1}$, regardless of whether the firm is urban or rural.

$$\begin{aligned}\frac{\partial(1+\tau^{1-\sigma})/(\theta^*)^{\sigma-1}}{\partial\tau} &= (\sigma-1)\frac{1+\tau^{1-\sigma}}{(\theta^*)^{\sigma-1}}\left[-\frac{\tau^{-\sigma}}{1+\tau^{1-\sigma}} - \frac{\partial\theta^*}{\partial\tau}\frac{1}{\theta^*}\right] \\ &= (\sigma-1)\frac{1+\tau^{1-\sigma}}{(\theta^*)^{\sigma-1}\tau}\left[-(1+\tau^{\sigma-1})^{-1} - \frac{\partial\theta^*}{\partial\tau}\frac{\tau}{\theta^*}\right]\end{aligned}$$

Using (A.7), the second term in the brackets can be rewritten as

$$-\frac{\partial\theta^*}{\partial\tau}\frac{\tau}{\theta^*} = \frac{1}{2}\left(1+\tau^{\sigma-1}\right)^{-1} + \frac{1}{2}\left[1 + \frac{j'(\theta^*,\gamma\theta^*)}{j_f'(\eta\theta^*,\gamma\theta^*)}\right]^{-1} < (1+\tau^{\sigma-1})^{-1}$$

This inequality follows from the fact that $j'(\theta^*, \gamma\theta^*)/j'_f(\eta\theta^*, \gamma\theta^*) > \tau^{\sigma-1}$ as it is shown in appendix A.4.1. Hence, $\partial\left[(1+\tau^{1-\sigma})/(\theta^*)^{\sigma-1}\right]/\partial\tau < 0$ and $\partial\left[r_d^m(\theta) + r_f^m(\theta)\right]/\partial\tau < 0$ and $\tau' < \tau$ entails $(r_d^m)'(\theta) + (r_f^m)'(\theta) > r_d^m(\theta) + r_f^m(\theta)$.

A.4.2.2 Case (b): Urban Non-exporters, Rural Non-exporters, and Rural Exporters

The expressions of the urban and rural domestic revenues in case (b) are the same as in case (a). Thus, for any firm with productivity $\theta > \theta^*$ in location $m \in \{u, r\}$, $r_d^m(\theta)$ is positively related to τ. The derivative of the total revenue from domestic and export sales with respect to τ also has the same sign as the derivative of $(1+\tau^{1-\sigma})/(\theta^*)^{\sigma-1}$. However, because the expression of $(\partial\theta^*/\partial\tau)\tau/\theta^*$ differs from that in case (a), a proof specific to case (b) must be provided.

$$-\frac{\partial\theta^*}{\partial\tau}\frac{\tau}{\theta^*} = \left[1 + 2\frac{j'(\theta^*, \alpha\theta^*)}{(j_f^r)'(\beta\theta^*)}\right]^{-1} < (1+\tau^{\sigma-1})^{-1}$$

The above inequality is justified as follows:

$$\frac{j'(\theta^*, \alpha\theta^*)}{(j_f^r)'(\beta\theta^*)} = \frac{w^u f^u}{f_{ex}}$$

$$\times \frac{[G(\alpha\theta^*) - G(\theta^*)][k^u(\theta^*, \alpha\theta^*) + 1] + [1 - G(\alpha\theta^*)](k_1^r(\theta^*, \alpha\theta^*) + \frac{w^r f^r}{w^u f^u}[k_2^r(\alpha\theta^*) + 1])}{[1 - G(\beta\theta^*)][k_4^r(\beta\theta^*) + 1]}$$

$$= \frac{\tau^{\sigma-1}}{\beta^{\sigma-1}}\left(\frac{w^r}{w^u}\right)^{\sigma-1} \frac{\frac{\int_{\theta^*}^{\alpha\theta^*}\xi^{\sigma-1}g(\xi)d\xi}{(\theta^*)^{\sigma-1}} + \frac{\int_{\alpha\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi}{(\theta^*)^{\sigma-1}} - \frac{\int_{\alpha\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi}{(\alpha\theta^*)^{\sigma-1}} + \frac{w^r f^r}{w^u f^u}\frac{\int_{\alpha\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi}{(\alpha\theta^*)^{\sigma-1}}}{\frac{\int_{\beta\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi}{(\beta\theta^*)^{\sigma-1}}}$$

$$= \tau^{\sigma-1}\left(\frac{w^r}{w^u}\right)^{\sigma-1}\left[\int_{\theta^*}^{\alpha\theta^*}\xi^{\sigma-1}g(\xi)d\xi + \int_{\alpha\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi - \alpha^{1-\sigma}\int_{\alpha\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi\right.$$

$$\left. + \alpha^{1-\sigma}\frac{w^r f^r}{w^u f^u}\int_{\alpha\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi\right]\left[\int_{\beta\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi\right]^{-1}$$

$$= \tau^{\sigma-1}\left(\frac{w^r}{w^u}\right)^{\sigma-1}\frac{\int_{\theta^*}^{\alpha\theta^*}\xi^{\sigma-1}g(\xi)d\xi + \int_{\alpha\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi + \alpha^{1-\sigma}\int_{\alpha\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi\left(\frac{w^r f^r}{w^u f^u} - 1\right)}{\int_{\beta\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi}$$

Substituting the expression of α from (2.8) into the above equation we obtain

$$\frac{j'(\theta^*, \alpha\theta^*)}{(j_f^r)'(\beta\theta^*)} = \tau^{\sigma-1}\left[\frac{\left(\frac{w^r}{w^u}\right)^{\sigma-1}\int_{\theta^*}^{\alpha\theta^*}\xi^{\sigma-1}g(\xi)d\xi}{\int_{\beta\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi} + \frac{\int_{\alpha\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi}{\int_{\beta\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi}\right] > \tau^{\sigma-1}$$

as $(w^r/w^u)^{\sigma-1}\int_{\theta^*}^{\alpha\theta^*}\xi^{\sigma-1}g(\xi)d\xi / \int_{\beta\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi > 0$, and $\int_{\alpha\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi > \int_{\beta\theta^*}^{\infty}\xi^{\sigma-1}g(\xi)d\xi$ as $\alpha < \beta$ by assumption. Hence $r_d^m(\theta) + r_f^m(\theta)$ is negatively related to τ.

A.5 Data Description

A.5.1 Data for the Estimation of the Production Functions

The variables in regression model (2.18) are defined as follows:
Output, $y^i_{j,t}$: natural logarithm (log) of the real value of production. The real value of production is obtained by dividing the nominal value of production by an output price index calculated at the ISIC three-digit level and expressed in 1980 pesos. **?** explains why the use of the production value is preferable to that of a value added measure for this particular data set.
Labor, $l^i_{j,t}$: log of total employment.
Capital stock, $k^i_{j,t}$: log of the total book value of fixed assets excluding land, deflated by the same price index as that used to deflate the value of production.
Energy, $e^i_{j,t}$: log of the deflated value of energy consumption plus purchases of fuels and lubricants.
Materials, $m^i_{j,t}$: log of the deflated value of raw materials and intermediate inputs utilization.
Metro, *metro*: a dummy variable equal to one for a plant located in a metro area and zero otherwise. The algorithm matching plants across years implies that a plant relocating from one area to another is assigned a new identifier. Thus, the metro variable should be constant over time, unless the definition of metro areas changed or data entry errors were committed. To remove any ambiguity regarding the interpretation of the interaction terms' coefficients, should *metro* vary over time, we retain its value in the first year a plant appears in the records (during the 1981-1984 period) for the subsequent years (it happens in only a few cases).

A.5.2 Data for the Analysis of Trade Liberalization

The variables in regression model (2.37) are constructed as follows:
Change in the share of metro plants, Δs^i_{84-91}: difference between the share of metro plants in 1991 and that in 1984 within industry i, at the four-digit level of the ISIC.
Change in the share of metro production, Δs^i_{84-91}: difference between the share of production from metro plants in 1991 and that in 1984 within industry i, at the four-digit level of the ISIC.
Tariff change, $\Delta\tau^i_{83-90}$: difference in industry i's tariff between 1990 and 1983.
Labor cost share in 1984: average share of the labor cost in total factor cost (sum of labor cost, total book value of fixed assets excluding land, value of energy consumption plus purchases of fuels and lubricants, and value of raw materials and intermediate

inputs consumption) in 1984, over all plants in industry i.

Capital share in 1984: average share of the total book value of fixed assets excluding land in total factor cost in 1984.

Energy share in 1984: average share of energy consumption value (including purchases of fuels and lubricants) in total factor cost in 1984.

Materials share in 1984: average share of the raw materials and intermediate inputs cost in total factor cost in 1984.

References for Chapter 1

A.F. Ades and E.L Glaeser. Trade and circuses: Explaining urban giants. *Quarterly Journal of Economics*, 110:195–227, 1 1995.

E.J. Bartelsman and M. Doms. Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic Literature*, 38, 2000.

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

International Bank for Reconstruction and Development/The World Bank. Reshaping economic geography. World Development Report, 2009.

P. K. Krugman and R. Livas-Elizondo. Trade policy and the third world metropolis. *Journal of Development Economics*, 49:137–150, 1 1996.

P.K. Krugman. Scale economies, product differentiation, and the pattern of trade. *American Economic Review*, 70:950–959, 5 1980.

P.K. Krugman. Increasing returns and economic geography. *Journal of Political Economy*, 99:483–499, 3 1991.

P. Martin and C.A. Rogers. Industrial location and public infrastructure. *Journal of International Economics*, 39, 3 1995.

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

J.D Sachs and A. Warner. Economic reform and the process of global integration. *Brookings Papers on Economic Activity*, 26, 1995.

References for Chapter 2

D. Ackerberg, C.L. Benkard, S. Berry, and A. Pakes. Econometric tools for analyzing market outcomes. In J.J. Heckman and E.E. Leamer, editors, *Handbook of Econometrics*, volume 6. Elsevier, Amsterdam, 2007.

P. Antràs and E. Helpman. Global outsourcing. *Journal of Political Economy*, 112: 552–580, 3 2004.

K. Behrens, C. Gaigné, G.I.P. Ottaviano, and J.-F. Thisse. Is remoteness a locational disadvantage? *Journal of Economic Geography*, 6, 2006.

K. Behrens, C. Gaigné, G.I.P. Ottaviano, and J.-F. Thisse. Countries, regions and trade: On the welfare impacts of economic integration. *European Economic Review*, 51, 2007.

A. B. Bernard, S. Redding, and P. K. Schott. Factor price equality and the economies of the United States. Unpublished manuscript, November 2005a.

A. B. Bernard, R. Robertson, and P. K. Schott. Is Mexico a lumpy country? Unpublished manuscript, June 2005b.

P. Bustos. The impact of trade on technology and skill upgrading: Evidence from Argentina. Unpublished manuscript, November 2005.

J. Cavailhès, C. Gaigné, T. Tabuchi, and J.-F. Thisse. Trade and the structure of cities. *Journal of Urban Economics*, 62, 2007.

S. K. Clerides, S. Lach, and J. R. Tybout. Is learning by exporting important? micro-dynamic evidence from Colombia, Mexico, and Morocco. *Quarterly Journal of Economics*, 113:903–947, 3 1998.

DANE. Conceptos básicos. Available from the DANE website http://www.dane.gov.co. Accessed May 2008.

Dirección de Desarrollo Empresarial DNP. Aranceles nominales (1974 - 2007) por Clasificación Industrial Internacional Uniforme (ciiu rev. 2) [data file]. Available from the DNP website, http://www.dnp.gov.co. Accessed May 2008.

J. Ederington and P. McCalman. Endogenous firm heterogeneity and the dynamics of trade liberalization. *Journal of International Economics*, 74:422–440, 2 2008.

R. A. Erickson. The filtering-down process: Industrial location in a nonmetropolitan area. *Professional Geographer*, 28:254–260, 3 1976.

A. M. Fernandes. Trade policy, trade volumes and plant-level productivity in Colombian manufacturing industries. Policy Research Working Paper 3064, World Bank, 2003.

A. M. Fernandes. Trade policy, trade volumes and plant-level productivity in Colombian manufacturing industries. *Journal of International Economics*, 71:52–71, 1 2007.

J. L. Findeis and L. Jensen. Employment opportunities in rural areas: Implications for poverty in a changing policy environment. *American Journal of Agricultural Economics*, 80:1000–1007, 0 1998.

C. Forman, A. Goldfarb, and S. Greenstein. How did location affect adoption of the commercial Internet? global village vs. urban leadership. *Journal of Urban Economics*, 58:389–420, 0 2005.

H. F. Gale, D. A. McGranahan, R. Teixeira, and E. Greenberg. Rural competitiveness: Results of the 1996 Rural Manufacturing Survey. Agricultural Economic Report 776, U.S. Department of Agriculture, Economic Research Service, 1999.

E. L. Glaeser and D. C. Maré. Cities and skills. *Journal of Labor Economics*, 19: 316–342, 2 2001.

P. K. Goldberg and N. Pavcnik. Trade, wages, and the political economy of trade protection: Evidence from the Colombian trade reforms. *Journal of nternational Economics*, 66:75–105, 1 2005.

G. H. Hanson. Regional adjustment to trade liberalization. *Regional Science and Urban Economics*, 28:419–444, 4 1998.

R. L. Holden and R. Wertheimer II. Differences in living costs between urban and rural areas, final report. Technical report, Urban Institute, 1980.

T. J. Holmes and J. J. Stevens. Geographic concentration and establishment scale. *Review of Economics and Statistics*, 84, 2002.

T. J. Holmes and J. J. Stevens. Geographic concentration and establishment size: Analysis in an alternative economic geography model. *Journal of Economic Geography*, 4:227–250, 3 2004.

D. Keeble, P. L. Owens, and C. Thompson. The urban-rural manufacturing shift in the European Community. *Urban Studies*, 20:405–418, 0 1983.

P. K. Krugman and R. Livas-Elizondo. Trade policy and the third world metropolis. *Journal of Development Economics*, 49:137–150, 1 1996.

P.K. Krugman. Increasing returns and economic geography. *Journal of Political Economy*, 99:483–499, 3 1991.

J. O. Lanjouw and P. Lanjouw. Rural non-farm employment: A survey. Policy Research Working Paper 1463, The World Bank, 1995.

J. Levinsohn and A. Petrin. Estimating production function using inputs to control for unobservables. *Review of Economic Studies*, 70:317–341, 2 2003.

R. S. Mack and P. V. Schaeffer. Nonmetropolitan manufacturing in the United States and product cycle theory: A review of literature. *Journal of Planning Literature*, 8:124–139, 2 1993.

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

P. Monfort and R. Nicolini. Regional convergence and international integration. *Journal of Urban Economics*, 48, 2000.

S. Olley and A. Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64:1263–1297, 6 1996.

E. Paluzie. Trade policies and regional inequalities. *Papers in Regional Science*, 80, 2001.

N. Pavcnik. Trade liberalization, exit, and productivity improvement: Evidence from Chilean plants. *Review of Economic Studies*, 69:245–276, 1 2002.

E. Phimister, E. Vera-Toscano, and A. Weersink. Female participation and labor market attachment in rural Canada. *American Journal of Agricultural Economics*, 84:210–221, 0 2002.

J. Roback. Wages, rents, and the quality of life. *Journal of Political Economy*, 90: 1257–1278, 0 1982.

M. J. Roberts. Colombia, 1977–85: Producer turnover, margins, and trade exposure. In M. J. Roberts and J. R. Tybout, editors, *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity, and Market Structure*. Oxford University Press, New York, NY, 1996.

M. J. Roberts and J. R. Tybout. The decision to export in Colombia: An empirical model of entry with sunk costs. *American Economic Review*, 87:545–564, 4 1997.

A. Rodríguez-Pose and J. Sánchez-Reaza. Economic polarization through trade: Trade liberalization and regional growth in Mexico. In *Spatial Inequality and Development*. Oxford University Press, New York, NY, 2005.

D. Roth. Thinking about rural manufacturing: A brief history. *Rural America*, 15: 12–19, 1 2000.

P. Sanguinetti and C. Volpe Martincus. Tariffs and manufacturing location in argentina. *Regional Science and Urban Economics*, 39, 2009.

E. Tomiura. Changing economic geography and vertical linkages in Japan. *Journal of the Japanese and International Economies*, 17:561–581, 4 2003.

D. Trefler. The long and short of the Canada-U.S. Free Trade Agreement. *American Economic Review*, 94:870–895, 4 2004.

E. Vera-Toscano, E. Phimister, and A. Weersink. Panel estimate of the Canadian rural/urban women's wage gap. *American Journal of Agricultural Economics*, 86: 1138–1151, 4 2004.

References for Chapter 3

A.F. Ades and E.L Glaeser. Trade and circuses: Explaining urban giants. *Quarterly Journal of Economics*, 110:195–227, 1 1995.

H.W. Armstrong. Convergence among regions of the European Union, 1950–1990. *Papers in Regional Science*, 74:143–152, 2 1995.

R. Baldwin, R. Forslid, P. Martin, G.I.P. Ottaviano, and F. Robert-Nicoud. chapter 14, page 333.

K. Behrens, C. Gaigné, G.I.P. Ottaviano, and J.-F. Thisse. Is remoteness a locational disadvantage? *Journal of Economic Geography*, 6, 2006.

K. Behrens, C. Gaigné, G.I.P. Ottaviano, and J.-F. Thisse. Countries, regions and trade: On the welfare impacts of economic integration. *European Economic Review*, 51, 2007.

M. Brulhart and R. Traeger. An account of geographic concentration patterns in europe. *Regional Science and Urban Economics*, 35, 2005.

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

J.M. Esteban. La desigualdad interregional en Europa y en España: Descripcióne y análisis. In *Crecimiento y convergencia regional en España y Europa*. Instituto de Análisis Económico, Barcelona, 1994.

Y. Ge. Regional inequality, industry agglomeration and foreign trade. Research Paper 105, United Nations University-World Institute for Development Economics, 2006.

G. H. Hanson. Regional adjustment to trade liberalization. *Regional Science and Urban Economics*, 28:419–444, 4 1998.

R. Kanbur and A.J. Venables. Rising spatial disparities and development. Policy Brief 3, World Institute for Development Economics Research of the United Nations University, 2005.

P.K. Krugman. Increasing returns and economic geography. *Journal of Political Economy*, 99:483–499, 3 1991.

P.K. Krugman and R. Livas-Elizondo. Trade policy and the third world metropolis. *Journal of Development Economics*, 49:137–150, 1 1996.

A. La Fuente and X. Vives. Infrastructure and education as instruments of regional policy: Evidence from Spain. *Economic Policy*, 20, 1995.

P. Martin and C.A. Rogers. Industrial location and public infrastructure. *Journal of International Economics*, 39, 3 1995.

P. Monfort and R. Nicolini. Regional convergence and international integration. *Journal of Urban Economics*, 48, 2000.

G.I.P. Ottaviano, T. Tabuchi, and J.-F. Thisse. Agglomeration and trade revisited. *International Economic Review*, 43:409–435, 2 2002.

E. Paluzie. Trade policy and regional inequalities. *Papers in Regional Science*, 80, 2001.

D.T. Quah. Empirics for economic growth and convergence. *European Economic Review*, 40, 1996.

A. Rodríguez-Pose and N. Gill. How does trade affect regional disparities? *World Development*, 34:1201–1222, 7 2006.

A. Rodríguez-Pose and J. Sánchez-Reaza. Economic polarization through trade: Trade liberalization and regional growth in Mexico. In *Spatial Inequality and Development*. Oxford University Press, Oxford, 2005.

X. Sala-i-Martin. Regional cohesion: Evidence and theories of regional growth and convergence. *European Economic Review*, 40, 1996.

P. Sanguinetti and C. Volpe Martincus. Tariffs and manufacturing location in argentina. *Regional Science and Urban Economics*, 39, 2009.

D.W. Te Welde and O. Morrissey. Spatial inequality for manufacturing wages in five african countries. In *Spatial Inequality and Development*. Oxford University Press, Oxford, 2005.

E. Tomiura. Changing economic geography and vertical linkages in Japan. *Journal of the Japanese and International Economies*, 17:561–581, 4 2003.

References for Chapter 4

B. Aw, X. Chen, and M. Roberts. Firm-level evidence on productivity differentials and turnover in Taiwanese manufacturing. *Journal of Development Economics*, 66, 2001.

R. L. Axtell. Zipf distribution of U.S. firm sizes. *Science*, 293, 2001.

M. N. Baily, C. Hulten, D. Campbell, T. Bresnahan, and R. E. Caves. Productivity dynamics in manufacturing plants. *Brookings Papers on Economic Activity. Microeconomics*, 1992.

J. Birchenall and G. Murcia. Convergencia regional: Una revisión del caso colombiano. *Desarrollo y Societad*, 40, 1996.

J. A. Bonet and A. Meisel. La convergencia regional en Colombia: Une visión de largo plazo, 1926 – 1995. *Coyuntura Económica*, 29:69–106, 1 1999.

J. Campbell and H. Hopenhayn. Market size matters. *Journal of Industrial Economics*, 53, 2005.

T. Chaney. Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review*, 98:1707–1721, 4 2008.

DANE. *Estadísticas básicas departamentales de Colombia 1980-1992*. División de Ediciones del Departamento Administrativo Nacional de Estadística, Santafé de Bogotá, D.C., 1995.

A. K. Dixit. A model of duopoly suggesting a theory of entry barriers. *Bell Journal of Economics*, 10:20–32, 1 1979.

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

T. dunne, M. J. Roberts, and L. Samuelson. The growth and failure of U.S. manufacturing plants. *Quarterly Journal of Economics*, 104, 1989.

J. Ederington and P. McCalman. The impact of trade liberalization on productivity within and across industries: Theory and evidence. University of Kentucky and University of California-Santa Cruz, July 2007.

A. M. Fernandes. Trade policy, trade volumes and plant-level productivity in Colombian manufacturing industries. Policy Research Working Paper 3064, World Bank, 2003.

A. M. Fernandes. Trade policy, trade volumes and plant-level productivity in Colombian manufacturing industries. *Journal of International Economics*, 71:52–71, 1 2007.

M. Del Gatto, G. Mion, and G. I. P. Ottaviano. Trade integration, firm selection and the costs of non-europe. Discussion paper 5730, CEPR, 2006.

Z. Griliches and J. Mairesse. Production functions: The search for identification. In *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*. Cambridge University Press, Cambridge, MA, 1998.

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

P. K. Krugman and R. Livas-Elizondo. Trade policy and the third world metropolis. *Journal of Development Economics*, 49:137–150, 1 1996.

J. Levinsohn and A. Petrin. Estimating production function using inputs to control for unobservables. *Review of Economic Studies*, 70:317–341, 2 2003.

N. Limao and A. J. Venables. Infrastructure, geographical disadvantage, transport costs, and trade. *World Bank Economic Review*, 15, 2001.

E. G. J. Luttmer. Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics*, 122:1103–1144, 3 2007.

J. Marschak and W. H. Andrews. Random simultaneous equations and the theory of production. *Econometrica*, 12, 1944.

A. Meisel. Polarización or convergencia? a propósito de Cárdenas, Pontón y Trujillo. *Coyuntura Económica*, 23:153–160, 2 1993.

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

M.J. Melitz and G.I.P. Ottaviano. Market size, trade, and productivity. *Review of Economic Studies*, 75, 2008.

S. Olley and A. Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64:1263–1297, 6 1996.

G. I. P. Ottaviano, T. Tabuchi, and J.-F. Thisse. Agglomeration and trade revisited. *International Economic Review*, 43:409–435, 2 2002.

N. Pavcnik. Trade liberalization, exit, and productivity improvement: Evidence from Chilean plants. *Review of Economic Studies*, 69:245–276, 1 2002.

M. J. Roberts. Colombia, 1977-85: Producer turnover, margins, and trade exposure. In M. J. Roberts and J. R. Tybout, editors, *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity, and Market Structure*. Oxford University Press, New York, NY, 1996.

R. Rocha and A. Vivas. Crecimiento regional en Colombia: Persiste la desigualdad? *Revista de Economía de Rosario*, 1:67–108, 1 1998.

C. Syverson. Market structure and productivity: A concrete example. *Journal of Political Economy*, 112, 2004.

C. Syverson. Prices, spatial competition, and heterogeneous producers: An empirical test. *Journal of Industrial Economics*, 55:197–222, 2 2007.

J. Tybout. Linking trade and productivity: new research directions. *World Bank Economic Review*, 6, 1992.

Vita

Fabien Tondel

Born February 9, 1980 in Nice, Alpes-Maritimes, France

EDUCATION

University of Kentucky, Lexington

Master of Science in Economics, 2006

Master of Science in Agricultural Economics, 2004

Ingénieur, Etablissement National d'Enseignement Supérieur Agronomique de Dijon (National School of Higher Agronomy Studies of Dijon), France, 2003

Baccalaureat général, 1998

PAST EMPLOYMENT

Graduate Research and Teaching Assistant, Department of Agricultural Economics, University of Kentucky, 2002-2009

Consultant, Von Allmen Center for Entrepreneurship, University of Kentucky, 2004

AWARDS

Kentucky Opportunity Fellowship, 2007-2009