

ESSAYS ON PATTERNS OF INTERNATIONAL TRADE

by

Isao Kamata

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2008

Doctoral Committee:

Professor Alan V. Deardorff, Chair
Emeritus Professor Robert M. Stern
Assistant Professor Jagadeesh Sivadasan
Assistant Professor Juan Carlos Hallak, Universidad de San Andrés

© Isao Kamata

All rights reserved
2008

To My Mom

ACKNOWLEDGEMENTS

I am extremely grateful to all the people who have directly or indirectly supported me to complete my studies for the doctoral degree. Without these individuals, including those whom I cannot list here, I could not have completed this challenging task.

I am deeply indebted to the four members of my dissertation committee. First and foremost, I would like to express my deep gratitude to Alan Deardorff, the chair, for his wise counsel, timely guidance, and continuous encouragement. Throughout the seven years of my doctoral study, he has been a wonderful advisor and mentor with great patience and endless amount of support. Special gratitude is also extended to Juan Carlos Hallak for his insightful suggestions regarding my work, from the overall picture to the details. He guided me to pursue research in international trade, and has shown me what good research is and what a good researcher is like. Many thanks are also due to Jagadeesh Sivadasan for his invaluable comments, which have always been helpful in finding ways to make my papers more “punchy.” My heartfelt appreciation also goes out to Bob Stern for his considerate advice and generous support throughout my dissertation research. I am honored to have him in my dissertation committee.

I would like to offer my sincere thanks to Gary Saxonhouse, who was initially part of my dissertation committee, for his very supportive advice and mentoring in the early stages of my study for this dissertation. Although he passed away in 2006, I

continued to be greatly supported from the resources and wisdom that he left to me. Throughout the process of completing this dissertation, the memory of his gentle smile kept encouraging me.

I owe a great deal to instructors of the English Language Institute at the University of Michigan (UM) for their help in writing this doctoral thesis. In particular, I would like to thank Chris Feak for her generosity in providing timely and continuous reviews of my writing along with stimulating feedback, which has also helped me re-organize my thoughts. I would also like to thank the UM Center for Japanese Studies for the fellowships supporting the last year of my dissertation work.

I would like to acknowledge Na Yang, my co-author of the preliminary work found in Chapter III of this dissertation. I greatly benefited from extensive discussions of the topic with her. It was a very interesting and wonderful experience to work with such a nice and productive peer with strengths different from my own.

I am very thankful to my fellow classmates for inspiring discussions and enjoyable conversations during my study in the degree program. In particular, I would like to extend my special thanks to my former officemates Phacharaphot Nuntramas and Jose Berrospide for friendship throughout the time we shared in the program. Even though they have graduated, they have continuously given me hearty advice and cheer during my final and most important stage of completion.

In the end, I thank my family in Japan, especially my mother, for her firm faith in and ceaseless support of me in this venture of my life.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	viii
LIST OF TABLES	x
CHAPTER	
I. Introduction	1
II. Comparative Advantage, Firm Heterogeneity, and Selection of Exporters	4
2.1 Introduction	4
2.2 The Model	10
2.2.1 Consumption	10
2.2.2 Production	11
2.2.3 Entry and Equilibrium in Autarky	13
2.2.4 Export	15
2.2.5 Factor Prices	19
2.2.6 Probability of Exporting	20
2.3 Data	23
2.3.1 Chilean Data	24
2.3.2 Colombian Data	25
2.3.3 Indian Data	25
2.3.4 U.S. Data	25
2.3.5 Factor Endowment Data	26
2.4 Empirical Analysis	27
2.4.1 Individual Country Analysis	27
2.4.2 Pooled Analysis	32
2.4.3 Robustness Check of Pooled Analysis	35
2.4.4 Factor Prices	36
2.5 Conclusion	41
2.6 Appendix A	42
2.6.1 Derivation of Equation (2.12)	42

2.6.2	Derivation of Equations (2.17) & (2.18)	43
2.6.3	Derivation of Equation (2.21)	44
2.6.4	Relative Factor Prices under Costly Trade	45
2.6.5	Proof of Proposition	48
2.7	Appendix B	52
2.8	Appendix C	56
2.9	Appendix D	57

III. Explaining Export Varieties: The Unexplored Role of Comparative Advantage 78

3.1	Introduction	78
3.2	The Model	81
3.2.1	Basic Framework	81
3.2.2	Consumption	82
3.2.3	Production and Export	82
3.2.4	Zero Profit	84
3.2.5	Entry and Equilibrium under Costly Trade	85
3.2.6	Mass of Firms and Export Varieties	86
3.3	The Data	90
3.4	Empirical Tests	93
3.4.1	Measuring Exported Varieties	93
3.4.2	Regressions for Aggregate North and South	94
3.4.3	Pooled Regression for Dependent Parameter Specification	96
3.4.4	Alternative Measure of Export Varieties	98
3.5	Conclusion	99
3.6	Appendix	100
3.6.1	Derivation of Equations (3.2.20)	100
3.6.2	Proof of Proposition	101

IV. Revisiting the Revisited: An Alternative Test of the Monopolistic Competition Model of International Trade 125

4.1	Introduction	125
4.2	Monopolistic Competition Model and Volume of Trade	130
4.2.1	Aggregate volume of trade	132
4.2.2	Volume of Trade in the Differentiated Sectors	133
4.2.3	Discussion on Production Structure-adjusted Size Similarity	135
4.3	Empirical Approaches to Estimate Volume-of-Trade Equations	136
4.3.1	OLS Estimation of Log-linearized Form	137
4.3.2	Non-linear Model for Zero-trade Observations: Poisson Quasi-maximum Likelihood Estimation	139
4.3.3	Tobit Estimation of Log-linearized Form	141
4.4	The Data	142
4.4.1	Industry/commodity classifications for the production data and trade data	144
4.4.2	Zero-trade Observations	146
4.5	Empirical Results	146
4.5.1	OLS Estimation of Log-linear Form	146
4.5.2	Alternative Estimation of the Log-linear Form: Tobit	149

4.5.3	Poisson Quasi-maximum Likelihood (PQML) Estimation	150
4.5.4	Comparison of Four Approaches to Estimation	151
4.5.5	Summary and Discussion	151
4.5.6	World GDP Share of Trading Countries	153
4.5.7	Robustness Check: Alternative Groupings of Differentiated Sectors	154
4.6	Conclusion	156
4.7	Appendix A	158
4.7.1	Hummels and Levinsohn (1995)	158
4.7.2	Debaere (2005)	159
4.7.3	Countries in the Data	160
V.	Conclusion	172
	BIBLIOGRAPHY	174

LIST OF FIGURES

Figure

2.1	“Gap” between Productivity Cutoffs in Costly-Trade Equilibrium (for Skilled-labor Abundant Country)	75
2.2	Fraction of Exporters among All Active Firms vs Industry Skill Intensity: Chile	76
2.3	Fraction of Exporters among All Active Firms vs Industry Skill Intensity: Colombia ..	76
2.4	Fraction of Exporters among All Active Firms vs Industry Skill Intensity: India	77
2.5	Fraction of Exporters among All Active Firms vs Industry Skill Intensity: the United States	77
3.1.1	Number of Exporters to the U.S. in Each Manufacturing Industry; Year 1990	116
3.1.2	Number of Exporters to the U.S. in Each Manufacturing Industry; Year 1995	117
3.1.3	Number of Exporters to the U.S. in Each Manufacturing Industry; Year 2000	118
3.2.1	Number of Exporters vs Industry Skilled-labor Intensity in the U.S. Manufacturing Imports; Year 1990	119
3.2.2	Number of Exporters vs Industry Skilled-labor Intensity in the U.S. Manufacturing Imports; Year 1995	119
3.2.3	Number of Exporters vs Industry Skilled-labor Intensity in the U.S. Manufacturing Imports; Year 2000	120
3.3.1	Number of Varieties vs Industry Skilled-labor Intensity in the U.S. Manufacturing Imports; Year 1990	121
3.3.2	Number of Varieties vs Industry Skilled-labor Intensity in the U.S. Manufacturing Imports; Year 1995	121
3.3.3	Number of Varieties vs Industry Skilled-labor Intensity in the U.S. Manufacturing Imports; Year 2000	122

3.4.1	Individual Exporter Country Regression for 1990: Slope Coefficient vs Skill Abundance of the Country	123
3.4.2	Individual Exporter Country Regression for 1995: Slope Coefficient vs Skill Abundance of the Country	123
3.4.3	Individual Exporter Country Regression for 2000: Slope Coefficient vs Skill Abundance of the Country	124
4.1	Volume of Bilateral Trade per Production vs Size Similarity Index; for OECD Countries (in logarithm; mean-differenced)	171
4.2	Volume of Bilateral Trade per Production vs Size Similarity Index; for Non-OECD Countries (in logarithm; mean-differenced)	171

LIST OF TABLES

Tables

2.1	Fractions of Exporting Firms in Manufacturing Industries (2-digit U.S. SIC): Chile, Colombia, India, and the United States	61
2.2a	Human Capital to Labor Ratio (H/L)	62
2.2b	Tertiary Education Completion in Total Population over Age 15	62
2.3	Individual Country Regression: Chile	63
2.4	Individual Country Regression: Colombia	63
2.5	Individual Country Regression: India	64
2.6	Comparison of Estimated Coefficients for Industry Skill Intensity among Three Countries (with 25 three-digit ISIC manufacturing industries)	64
2.7	Individual Country Regression: The United States	65
2.8	Comparison of Estimated Coefficients for Industry Skill Intensity among Four Countries (with 17 two-digit U.S. SIC manufacturing industries)	65
2.9	Concordance from Three-digit ISIC (revision 2) to Two-digit 1987 U.S. SIC	66
2.10	Cost Share of Skilled Labor ($skill_i$) in 17 Two-digit U.S. SIC Manufacturing Industries	67
2.11	Regression for Four Countries and 17 Two-digit U.S. SIC Manufacturing Industries ...	67
2.12	Alternative Skilled-labor Cost Share in 17 Two-digit U.S. SIC Manufacturing Industries: Wage Share of Workers with One or More Years of College Education	68
2.13	Alternative Relative Skilled-labor Abundance (S/U) _c in Four Countries: Percentage of Population with Tertiary Education Attainment	68
2.14	Regression for Four Countries and 17 Two-digit U.S. SIC Manufacturing Industries with Alternative Measures of Industry Skill Intensity and Country Skill Abundance	68

2.15	Relative Wage of Skilled to Unskilled Labor (s/w)	69
2.16	Regression with Relative Factor Price (Skilled-to-Unskilled Wage Ratio)	69
2.17	Relative Wage of Skilled to Unskilled Labor (s/w), Measured Based on Same Occupations	70
2.18	Regression with Relative Factor Price (Skilled-to-Unskilled Wage Ratio) Measured Based on Same Occupations	70
2.C.1	Country-Industry-Year Pooled Regression (1): Without Dummies	71
2.C.2	Country-Industry-Year Pooled Regression (2): With Industry Dummies	71
2.C.3	Country-Industry-Year Pooled Regression (3): With Country Dummies	72
2.C.4	Country-Industry-Year Pooled Regression (4): With Year Dummies	72
2.C.5	Country-Industry-Year Pooled Regression (5): With Industry and Country Dummies ..	73
2.C.6	Country-Industry-Year Pooled Regression (6): With Industry and Year Dummies	73
2.C.7	Country-Industry-Year Pooled Regression (7): With Country and Year Dummies	74
2.C.8	Country-Industry-Year Pooled Regression (8): With Industry, Country, and Year Dummies	74
3.1	U.S. Imports and Varieties	105
3.2	Country List (115 countries)	106
3.3	Factor Abundance of Countries: Skilled Labor (S) to Unskilled Labor (U) Ratio	107
3.4.1	Relative Skilled-labor (S) and Unskilled-labor (U) Intensity of Manufacturing Industries: for 4-digit U.S. SIC Industries, Year 1992	108
3.4.2	Relative Skilled-labor (S) and Unskilled-labor (U) Intensity of Manufacturing Industries: for 6-digit NAICS Industries, Year 1997	109
3.4.3	Relative Skilled-labor (S) and Unskilled-labor (U) Intensity of Manufacturing Industries: for 6-digit NAICS Industries, Year 2002	110
3.5	List of Countries in Aggregate North and South	111
3.6	Skilled-to-Unskilled Labor Ratios (S/U) of North and South	112
3.7	Regressions for Aggregate North and South	113
3.8	Pooled Regressions for Individual Exporters	114

3.9	Pooled Regressions using Alternative Measure of Export Varieties	115
4.1	List of Countries and Years in Data	162
4.2	Result of OLS Estimation, with Positive-Trade Observations	163
4.3	Result of OLS Estimation, with All Observations	163
4.4	Result of Tobit Estimation	164
4.5	Results of Poisson Quasi-maximum Likelihood (PQML) Estimation	164
4.6	Akaike Information Criteria (AIC) of Estimated Models	165
4.7	Alternative Groupings of Differentiated Sectors	166
4.8.1	Result of OLS Estimation for Alternative Sector Grouping (1), with Positive-Trade Observations	167
4.8.2	Result of OLS Estimation for Alternative Sector Grouping (1), with All Observations	167
4.8.3	Result of Tobit Estimation for Alternative Sector Grouping (1)	168
4.8.4	Results of Poisson Quasi-maximum Likelihood (PQML) Estimation for Alternative Sector Grouping (1)	168
4.9.1	Result of OLS Estimation for Alternative Sector Grouping (2), with Positive-Trade Observations	169
4.9.2	Result of OLS Estimation for Alternative Sector Grouping (2), with All Observations	169
4.9.3	Result of Tobit Estimation for Alternative Sector Grouping (2)	170
4.9.4	Results of Poisson Quasi-maximum Likelihood (PQML) Estimation for Alternative Sector Grouping (2)	170

CHAPTER I

Introduction

This dissertation theoretically and empirically studies the roles of comparative advantage, monopolistic competition, and firm-level heterogeneity in international trade in various aspects.

The following Chapter II investigates cross-country and cross-industry variation in the fractions of exporters among domestic firms. The paper presents a model of an economy in which countries are asymmetrically endowed with two production factors, industries vary in the relative intensity of the use of these factors, and firms differ in productivity level. The model predicts that the shares of exporting firms in the number of domestic producers are ranked in order of the industry's relative intensity of the factor with which the country is relatively well-endowed. This quasi-Heckscher-Ohlin prediction is empirically tested using data from the manufacturing censuses of Chile, Colombia, India, and the United States. The result of the analysis shows that the correlation between the exporter fractions and industry skill intensities is larger (more positive) for a country with higher skilled-labor abundance. This result is evidence of the theoretical prediction and demonstrates the role of comparative advantage in exporter selection.

The third chapter is based on a co-authored paper with Na Yang. The paper examines how factor proportions determine the number of product varieties, or the extensive margin, in exports of countries. The model of a two-factor, two-country and multi-industry economy with productivity-heterogeneous firms that is introduced in the second chapter suggests that countries export more varieties in industries in which the countries have a comparative advantage. This theoretical prediction is confirmed by empirical tests that use disaggregated data on U.S. imports. The tests show that relatively (un)skilled-labor abundant countries tend to export more varieties in more (un)skilled-labor intensive industries. This chapter provides both a theoretical foundation and empirical evidence for the importance of factor proportions in explaining the pattern of product varieties in exports.

Chapter IV proposes an alternative test of the monopolistic competition model of international trade, based on its implication of a positive correlation between the volume of trade and the similarity among trading countries in the size of the economy. In the existing literature this implication has been tested for aggregate trade, which includes the sectors that are not characterized by product differentiation. In contrast, this paper focuses on trade of differentiated products, which are the sectors that the monopolistic competition model is designed to describe. The amended prediction is tested with disaggregated data on manufacturing trade and production, using various estimation procedures including a non-linear method to handle zero-trade observations. The result from this alternative approach demonstrates that (i) trade in the differentiated sectors among OECD countries is well described by the monopolistic competition model; but (ii) for non-OECD countries the predicted relationship between trade and country size

similarity is more pronounced in the *non*-differentiated sectors, implying that trade flows among non-rich countries may be driven or crucially influenced by forces other than horizontal product differentiation.

CHAPTER II

Comparative Advantage, Firm Heterogeneity, and Selection of Exporters

2.1 Introduction

Some firms export but others do not. Since Bernard and Jensen (1995) pointed this out for the United States, this fact has been confirmed for various countries and industries. The fraction of exporters among domestic firms in a country, however, varies widely across industries. For instance, as shown in Table 2.1, in the United States, 49% of firms in the electric equipment industry export, while only 13% export in the stone, clay, and glass products industry,¹ even though the total number of firms is almost the same in the two industries. A difference in the fraction of exporters among domestic firms can also be seen across countries in the same industry. For example, 54% of Indian firms in the apparel industry are exporters, while only 12% of American firms in this same industry export. At the same time, the share of exporters in the electric equipment

¹ The data are for the year 1992. In 1992, the share of exporters in all firms was 22% in the manufacturing industry in total. Therefore, the exporter share in the electric equipment industry (U.S. Standard Industry Classification code 36) was more than the double of the exporter share in all the manufacturing sectors, while the share in the stone, clay, and glass products industry (U.S. SIC code 32) was about 60% of that in the whole manufacturing industry.

industry is 17% in India, but 49% in the United States.² These examples illustrate that cross-industry variation in exporter fraction (i.e., in what industries firms are more likely to export) does not follow the same pattern for all countries.

This difference in the likelihood of domestic firms being exporters in different industries and countries indicates that country-based and industry-based influences must be involved. These affect individual firms' decision to export or not. To date, however, empirical studies have focused on firm-level determinants generating heterogeneity in export behavior among firms. Little work has investigated how the fraction of exporters among domestic firms differs across industries and countries and what generates these differences. This paper explains this cross-industry and cross-country variation in the exporter fraction from the perspective of comparative advantage, in particular comparative advantage in terms of factor proportion. Although other potential country-specific or industry-specific determinants of the selection of exporters can be considered, the empirical analysis in this paper shows that the observed patterns of the exporter fraction can be well explained by comparative advantage, or countries' relative factor abundance and industries' relative factor intensity.

The influence of factor proportion-based comparative advantage on difference in firm-level export decision has been theoretically examined by Bernard, Redding and Schott (2007). They incorporate the model by Melitz (2003), which has provided a theoretical benchmark explaining the empirical regularity of self-selection of exporters (i.e., firms that are the most productive in a domestic market become exporters), into a two-country, two-factor and two-industry framework. To derive a prediction describing

² The data for India are for the fiscal year 1997/98 (April 1997-March 1998). In this year, the share of exporters in all the manufacturing firms was 14% in India.

an empirical relationship between the exporter fraction and factor proportion, this paper extends the model by Bernard, Redding and Schott to a multi-industry framework. That is, this paper considers an economy that comprises two countries differing in the relative abundance of two production factors (skilled and unskilled labor) and a large number of industries differing in the relative intensity of the two production factors. In these two countries each industry is populated with a continuum of firms differing in total factor productivity. Two threshold levels of firms' productivity, one of which divides domestic producers from "exiters" and the other divides these domestic producers into exporters and non-exporters, are created through monopolistic competition and costly international trade. However, the impact of international trade on the two productivity cutoffs is asymmetric across industries, due to the difference in factor proportion. Keener competition among firms seeking larger potential export profits raises the domestic-production productivity cutoff more in comparative-advantage industries, while the cutoff for exporting is relatively lower in these industries due to the comparative advantage over foreign competitors. This impact of trade on the two productivity cutoffs is more pronounced with the strength of comparative advantage; as a result, the "gap" between the two productivity cutoffs, which is measured as the ratio of the export cutoff to the domestic-production cutoff, is the largest in the industry with the lowest relative intensity of the factor with which the country is relatively well-endowed, and the smallest in the industry with the highest relative intensity of that factor. This ratio of the two productivity cutoffs determines the ex post fraction of exporters among domestic producers (the smaller the gap, the larger the fraction). Therefore, if all other conditions are equal between countries and among industries, in the relatively more skilled-labor

abundant country, the exporter fraction rises with an industry's relative skilled-labor intensity, and vice versa.

Empirically, this theoretical prediction is examined as a correlation between the fraction of exporters among domestic firms and the relative skill intensity of industries. That is, the correlation should be larger (i.e., more positive or less negative) for a country with higher relative skilled-labor abundance, compared to less skilled-labor abundant countries. This empirical prediction is tested using data from the manufacturing censuses of Chile, Colombia, India, and the United States. These four countries represent a variety of country groups in terms of relative skill abundance. The results of estimation for individual countries present that the correlation between the exporter fraction and industry skill intensity in fact differs across countries, and the values of correlation coefficients estimated for the four countries follow the order of the countries' skilled-labor abundance; i.e., the correlation is of the largest positive for the United States, and declines for Chile, Colombia, and towards a negative value for India. This relationship between the countries' skill abundance and the correlation between the exporter fraction and industry skill intensity is more formally tested using pooled data for these four countries and 17 manufacturing industries classified according to the two-digit U.S. Standard Industrial Classification (SIC). The result confirms that the correlation between the exporter fraction and industry skill intensity rises (toward positive) with the relative skill abundance of a country.³ This result is robust across alternative measures of country skill abundance and industry skill intensity. The estimation using relative factor price (the ratio of skilled-labor wage to unskilled-labor wage) as another measure of comparative

³ This result can also be interpreted from the cross-country point of view in the following way: the correlation between the exporter fraction and country relative skill abundance is larger, or more positive, in a more skill-intensive industry.

advantage also supports the quasi-Heckscher-Ohlin prediction about the exporter fraction.

The finding of quasi-Heckscher-Ohlin effect on the exporter fraction in this paper has an additional implication for international trade. In the representative- (or symmetric-)firm framework by Romalis (2004), the industrial composition of a country's exports in terms of the number of firms (or product varieties) is symmetric to the industrial composition of the country's domestic production.⁴ In other words, a country has larger shares of world total exporters in comparative-advantage industries because the country has larger shares of world total producers. This is not necessarily the case in the current model. Since a country's comparative advantage also affects the mechanism of exporter selection, it may be that despite a small share of producers in the world, a country's share of exporters is large in a comparative-advantage industry. Some examples are found in Table 2.1. In the apparel industry, for instance, the number of Indian domestic firms is the double of the number of Chilean firms; however, the number of Indian exporters is eight-fold that of Chilean exporters in that industry. That is, the effect of comparative advantage on the number of firms (or the extensive margin) can be magnified through exporter selection.⁵

This study adds to the literature in two ways. First, this paper is the first to empirically investigate cross-country and cross-industry asymmetry in the (self-)selection of exporters. Since Bernard and Jensen (1995), a great number of empirical studies have investigated differences between exporters and non-exporters focusing on a single

⁴ More accurately, the number of domestic producers is the same as the number of exporters since the model does not have the mechanism of the selection of exporters (all domestic firms export).

⁵ The effect of comparative advantage can be even more magnified in the volume of exports. For example, in the apparel industry the volume of Indian exports is more than 50 times greater than the volume of Chilean exports. (In contrast, in Romalis' model the share in the volume of exports is also symmetric to the share in the number of firms.) Although this paper does not directly address this issue, the present model has the potential for explaining this magnification of the effect of comparative advantage in export volume as a result of differences in relative productivity among exporters in different industries.

country, some of which further narrow their focuses on a single industry (see Greenaway and Kneller (2007), Lopez (2005), and Wagner (2007) that are extensive surveys of this literature). This paper takes one step back in focus and addresses the issue of firm-level heterogeneity in export behavior from a cross-country and cross-industry perspective. In addition, this paper adds to few theoretical studies on firm-level heterogeneity in export decision that takes into account asymmetry of countries such as Falvey et al. (2004) and Melitz and Ottaviano (2008), or of both countries and industries such as Bernard, Redding and Schott (2007).

Secondly, this paper empirically demonstrates the effect of the factor proportion-based comparative advantage on another dimension of international trade, i.e., the fraction of exporting firms among domestic firms. The Heckscher-Ohlin framework, or the factor proportion theory, has been empirically tested for the specialization patterns of countries' net trade flows (e.g., Baldwin (1971), Harkness (1978), and Stern and Maskus (1981); also see the survey by Deardorff (1984)), production (Harrigan and Zakrajsek (2000) and Fitzgerald and Hallak (2004)), and the relative volume of trade (Romalis, 2004).⁶ The next chapter III of this dissertation demonstrates that the factor-proportion framework also provides a prediction about the relative product variety in countries' exports. This paper demonstrates that the (quasi-) Heckscher-Ohlin framework also explains the patterns of the exporter fractions. In addition, this paper extends the model by Bernard, Redding & Schott to a multi-industry framework, which is analogous to the work by Dornbusch, Fischer and Samuelson (1980) that extends the standard Heckscher-Ohlin model with perfect competition, and Romalis' (2004) extension of the monopolistic

⁶ Another large branch of the literature is empirical tests of the Heckscher-Ohlin-Vanek model of the factor contents of trade.

competition model by Helpmand and Krugman (1985).

The organization of the rest of the paper is as follows. The next section presents the economic model and derives the prediction of the cross-industry pattern of exporter selection. The third section describes the data that are used in the empirical analysis, which is demonstrated in Section 2.4. The concluding section discusses the results and implications.

2.2 The Model

This paper adopts the model by Bernard, Redding and Schott (2007) and extends it to the framework of two countries, two factors and multiple industries. The modeled economy comprises two countries, Home (H) and Foreign (F); two factors, skilled labor (S) and unskilled labor (U); and $N (>2)$ industries. Within each industry there is a continuum of firms that are heterogeneous in productivity. Countries differ in factor endowments: Home is relatively abundant in skilled labor, and Foreign is relatively abundant in unskilled labor; i.e., $\frac{\bar{S}^H}{\bar{U}^H} > \frac{\bar{S}^F}{\bar{U}^F}$ where \bar{S}^H (\bar{U}^H) and \bar{S}^F (\bar{U}^F) denote the total inelastic supply of (un)skilled labor in Home and Foreign, respectively.

2.2.1 Consumption

The representative consumer possesses Cobb-Douglas preferences over $N > 2$ industries that are described by the following first-tier utility function:

$$U = C_1^{\alpha_1} C_2^{\alpha_2} \dots C_N^{\alpha_N}, \quad \sum_{i=1}^N \alpha_i = 1 \quad (2.1)$$

where C_i represents the consumption index for Industry $i = 1, \dots, N$. The representative consumer consumes all the available product varieties within each industry, and the

industry-wise consumption index C_i takes the following CES (or Dixit-Stiglitz) form:

$$C_i = \left[\int_{\omega \in \Omega_i} q_{\omega}^{\rho} d\omega \right]^{\frac{1}{\rho}} \quad (2.2)$$

where ω indexes product varieties within an industry, Ω_i denotes a set of available varieties in Industry i , and q_{ω} represents the quantity of each variety consumed.

Accordingly, the price index P_i over individual varieties of products in Industry i is defined as:

$$P_i = \left[\int_{\omega \in \Omega_i} p_{i,\omega}^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}} \quad (2.3)$$

where $\sigma = \frac{1}{1-\rho} > 1$ is the constant elasticity of substitution across varieties.

2.2.2 Production

Each firm produces a unique variety of products. A firm's total cost of production is the sum of fixed costs and variable costs. The fixed costs are the same for all firms in an industry within a country,⁷ but the variable costs vary across firms according to the difference in their productivity $\phi \in (0, \infty)$. The cost function for Firm ω in Industry i in each country is:

$$\begin{aligned} \Gamma_{i,\omega}^H &= \left[f_i + \frac{q_{i,\omega}}{\phi_{i,\omega}} \right] \cdot (s^H)^{\beta_i} (w^H)^{1-\beta_i} \\ \Gamma_{i,\omega}^F &= \left[f_i + \frac{q_{i,\omega}}{\phi_{i,\omega}} \right] \cdot (s^F)^{\beta_i} (w^F)^{1-\beta_i} \end{aligned} \quad (2.4)$$

where s is the wage for skilled labor, w is the wage for unskilled labor, and the

⁷ As shown in Equation (2.4), since the fixed costs also depend on the prices of two production factors, the fixed costs is in general different between the two countries due to the difference in factor prices.

superscripts H and F denote Home and Foreign, respectively. The industries are ranked according to the Cobb-Douglas cost share of skilled labor (β_i), such that the industry indexed with a large number for i has a larger skilled-labor cost share:

$0 < \beta_1 < \beta_2 \dots < \beta_{N-1} < \beta_N < 1$. Within an industry, the cost share of each factor does not differ across countries or across firms. Note that the factor intensity of Industry i is also ranked using the rank of β_i ,⁸ and thus β_i can be regarded as an indirect index of industry factor intensity.

In what follows, I present equations and expressions for Home, unless otherwise noted. The equations for Foreign are symmetric.

With the Dixit-Stiglitz preferences, the optimal price of a firm's product variety equals a constant markup ($1/\rho$) over the marginal cost of production.

$$p_{i,\omega}^H(\phi_{i,\omega}) = \frac{(s^H)^{\beta_i} (w^H)^{1-\beta_i}}{\rho \phi_{i,\omega}} \quad (2.5)$$

Revenue of each firm from its domestic (Home) sales thus takes the following form:

$$r_{i,\omega}^H(\phi_{i,\omega}) = \alpha_i Y^H \left(\frac{(s^H)^{\beta_i} (w^H)^{1-\beta_i}}{\rho \phi_{i,\omega} P_i^H} \right)^{1-\sigma} \quad (2.6)$$

where Y^H is the total national income of Home. The profit of each firm is equal to revenue minus production costs, which is as follows:

$$\pi_{i,\omega}^H(\phi_{i,\omega}) = \frac{r_{i,\omega}^H(\phi_{i,\omega})}{\sigma} - f_i (s^H)^{\beta_i} (w^H)^{1-\beta_i} \quad (2.7)$$

⁸ Since the equilibrium relative factor intensity in each industry is

$$\frac{S_i}{S_i + U_i} = \frac{\beta_i}{\beta_i + (1 - \beta_i) \cdot (s/w)},$$

and therefore for any relative wage $s/w > 0$, $S_i/(S_i + U_i)$ is larger for a larger β_i .

2.2.3 Entry and Equilibrium in Autarky

To describe the general idea with simpler expressions, I first describe the equilibrium in an autarkic economy. To enter the domestic market, each firm incurs a sunk entry cost. Firms discover their productivity after the entry. The productivity parameter ϕ is randomly drawn from a distribution $G(\phi)$, which is common across countries. The entry cost also depends upon the prices of the two input factors, and takes the following form:

$$f_{ei}(s^H)^{\beta_i}(w^H)^{1-\beta_i}, \quad f_{ei} > 0 \quad (2.8)$$

In other words, the Cobb-Douglas cost share of each factor in an industry commonly affects the sunk entry cost as well.

After paying the sunk entry cost (and realizing a productivity level), a firm must earn at least zero profit to remain and produce in the market. In other words, if the firm observes that its productivity is too low to earn a positive profit, it will shut down and exit. The minimum productivity requirement, or the productivity cutoff, for domestic production ϕ_i^* is thus determined by the following zero-profit condition:

$$r_i^H(\phi_i^{*H}) = \sigma_i f_{ei}(s^H)^{\beta_i}(w^H)^{1-\beta_i} \quad (2.9)$$

In Industry i , all the firms whose productivity is higher than or equal to ϕ_i^{*H} will continue operation, while less productive firms will exit.

The value of each firm is determined as the present discount value of the future profit flows, which is expressed as follows:

$$v_{i,\omega}^H(\phi_{i,\omega}) = \max\left\{0, \sum_{t=0}^{\infty} (1-\delta)^t \pi_{i,\omega}^H(\phi_{i,\omega})\right\} = \max\left\{0, \frac{\pi_{i,\omega}^H(\phi_{i,\omega})}{\delta}\right\} \quad (2.10)$$

where $\delta < 1$ is an exogenous probability of firm death in each period. In the long run

equilibrium, the expected value of entry, $V_{i,\omega}$, will equal the sunk entry cost for each firm in each industry. Since the expected value of entry is the expected value of the firm (or future profit stream) conditional on the *ex ante* probability of successful entry, the free-entry condition is as follows:

$$V_{i,\omega}^H = [1 - G(\phi_i^{*H})] \frac{\bar{\pi}_i^H}{\delta} = f_{ei} (s^H)^{\beta_i} (w^H)^{1-\beta_i} \quad (2.11)$$

where $\bar{\pi}_i^H$ represents the per-period expected future profit for the firm successfully entering into the market in Industry i . That is, $\bar{\pi}_i^H \equiv \pi_i^H(\bar{\phi}_i^\lambda)$ where $\bar{\phi}_i^\lambda$ is the average productivity of the successful entrees in the industry.⁹

In the case of an autarkic economy, by combining the zero profit condition (2.9) and the free entry condition (2.11), the following equation to determine the cutoff-level productivity ϕ_i^{*H} is derived:

$$\frac{f_i}{\delta} \int_{\phi_i^{*H}}^{\infty} \left[\left(\frac{\phi}{\phi_i^{*H}} \right)^{\sigma-1} - 1 \right] g(\phi) d\phi = f_{ei} \quad (2.12)$$

where $g(\cdot) = G'(\cdot)$ is the common density function of productivity ϕ .¹⁰ The left-hand side of Equation (2.12) monotonically decreases as the value of ϕ_i^{*H} increases, and thus a unique value of ϕ_i^{*H} is identified since the right-hand side of the equation is constant.

⁹ The average productivity of the successfully entering firms is determined by the *ex post* distribution of the productivities defined with the zero-profit cutoff productivity level: i.e.;

$$\bar{\phi}_i^\lambda = \bar{\phi}(\phi_i^{*\lambda}) = \left[\frac{1}{1 - G(\phi_i^{*\lambda})} \int_{\phi_i^{*\lambda}}^{\infty} \phi^{\sigma-1} g(\phi) d\phi \right]^{\frac{1}{\sigma-1}}$$

where $g(\cdot) = G'(\cdot)$ is a density function of productivity ϕ .

¹⁰ See Appendix A for the derivation of Equation (2.12).

2.2.4 Export

The main interest of this paper is a trading equilibrium, and I now analyze the decisions of the firms when a country is open to trade with the other country.

For each firm to export, it must incur per-year fixed costs for export, which depend on the domestic factor prices and industry factor intensity, as the fixed costs for domestic production and the sunk entry cost do. Specifically, the per-year fixed costs for export are described as $f_{xi} (s^H)^{\beta_i} (w^H)^{1-\beta_i}$, $f_{xi} > 0$. In addition, international trade is subject to variable “iceberg” shipping costs such that only a proportion $1/\tau_i$ ($\tau_i > 1$) of the shipped quantity of products reaches the other country. The variable costs are assumed to be symmetric between the two countries.

The optimal export price of the product of Firm ω in Home in Industry i ($p_{xi,\omega}^H$) is equal to the constant markup ($1/\rho$) over the marginal production cost inclusive of the iceberg transportation costs. That is;

$$p_{xi,\omega}^H(\phi) \equiv \tau_i \cdot p_{i,\omega}^H(\phi) = \frac{\tau_i (s^H)^{\beta_i} (w^H)^{1-\beta_i}}{\rho \phi_{i,\omega}} \quad (2.13)$$

Accordingly, Firm ω 's revenue from export to the Foreign market is:

$$r_{xi,\omega}^H(\phi) = \alpha_i Y^F \left(\frac{\tau_i (s^H)^{\beta_i} (w^H)^{1-\beta_i}}{\rho \phi_{i,\omega} P_i^F} \right)^{1-\sigma}$$

Firms produce either to serve only the domestic market or to serve both domestic and foreign markets, depending on their productivity.¹¹ Therefore, the total revenue of each firm is now as follows:

$$r_{i,\omega,total}^H(\phi) = r_{i,\omega}^H(\phi) \quad \text{if the firm serves only the domestic market;}$$

¹¹ To preview the result, if they are sufficiently productive, domestic producers can also export.

$$r_{i,\omega,total}^H(\phi) = r_{i,\omega}^H(\phi) + r_{xi,\omega}^H(\phi) \quad \text{if the firm also exports.}$$

As in the closed economy case, the zero-profit condition and the free-entry condition jointly identify the productivity cutoff at which additional profits from exporting are zero. The profit of each firm now consists of two parts:

$$\pi_{i,\omega,total}^H(\phi) = \pi_{i,\omega}^H(\phi) + \max\{0, \pi_{xi,\omega}^H(\phi)\} \quad (2.14)$$

$$\text{where } \pi_{i,\omega}^H(\phi) = \frac{r_{i,\omega}^H(\phi)}{\sigma} - f_i(s^H)^{\beta_i} (w^H)^{1-\beta_i};$$

$$\pi_{xi,\omega}^H(\phi) = \frac{r_{xi,\omega}^H(\phi)}{\sigma} - f_{xi}(s^H)^{\beta_i} (w^H)^{1-\beta_i}.$$

Accordingly, the zero-profit condition is two-fold, which consists of the following two equations:

Zero-profit condition for domestic production, which involves the domestic producer productivity cutoff ϕ_i^{*H} :

$$r_i^H(\phi_i^{*H}) = \sigma f_i(s^H)^{\beta_i} (w^H)^{1-\beta_i} \quad (2.15)$$

Zero-profit condition for export, which involves the exporter productivity cutoff

ϕ_{xi}^{*H} :

$$r_{xi}^H(\phi_{xi}^{*H}) = \sigma f_{xi}(s^H)^{\beta_i} (w^H)^{1-\beta_i} \quad (2.16)$$

Equations (2.15) and (2.16) jointly determine the relationship between the two cutoffs ϕ_i^* and ϕ_{xi}^* for each country, as follows:

$$\phi_{xi}^{*H} = \Lambda_i^H \cdot \phi_i^{*H} \quad \text{for Home} \quad (2.17)$$

$$\phi_{xi}^{*F} = \Lambda_i^F \cdot \phi_i^{*F} \quad \text{for Foreign} \quad (2.18)$$

where $\Lambda_i^H = \tau_i \left(\frac{P_i^H}{P_i^F} \right) \left(\frac{Y^H}{Y^F} \cdot \frac{f_{xi}}{f_i} \right)^{\frac{1}{\sigma-1}}$ and $\Lambda_i^F = \tau_i \left(\frac{P_i^F}{P_i^H} \right) \left(\frac{Y^F}{Y^H} \cdot \frac{f_{xi}}{f_i} \right)^{\frac{1}{\sigma-1}}$,¹² and P_i^H and P_i^F

are the industry price indexes in Home and Foreign, respectively. Because empirical studies have shown that exporting firms tend to be more productive than non-exporters, I focus on the case where the productivity cutoff for export is higher than that for domestic production: i.e., $\phi_{xi}^{*H(F)} > \phi_i^{*H(F)}$, or $\Lambda_i^{H(F)} > 1$. This would be the case when the fixed costs for export is sufficiently higher than fixed costs for (domestic) production ($f_{xi} > f_i$), and/or the variable trade costs (τ_i) are sufficiently large. In this case, only a portion of firms that successfully enter the domestic market can export, i.e., selection of exporters occurs.¹³ Of all the firms in Home that draw a random productivity in return for the sunk entry cost, a fraction of $G(\phi_i^{*H})$ will exit because their revenues can not cover the fixed costs for domestic production. A fraction $G(\phi_{xi}^{*H}) - G(\phi_i^{*H})$ of the firms will serve only the Home domestic market because they will not be able to cover the higher fixed costs for export. Only the remaining firms (the fraction of $1 - G(\phi_{xi}^{*H})$), which are the most productive, will be exporters.

The free-entry condition also comprises two parts: the expected future profit stream from the domestic market, and the expected future profit from the export market multiplied by the probability of being an exporter conditional on the firm successfully entering and staying the domestic market. The value (or the expected total future profit) of Firm ω is:

¹² See Appendix A for the derivation of Equations (2.17) and (2.18).

¹³ $\Lambda_i^H > 1$ will also hold when the industry price index in Home is higher than that in Foreign ($P_i^H > P_i^F$), and/or the Home economy is larger than Foreign ($Y^H > Y^F$). However, this also implies that Λ_i^H could be less than one if Home price index is sufficiently lower than that of Foreign, and/or the Home economy is sufficiently smaller than the Foreign economy.

$$v_{i,\omega}^H(\phi_{i,\omega}) = \max\left\{0, \frac{\pi_{i,\omega}^H(\phi_{i,\omega})}{\delta}\right\} + \chi_i^H \cdot \max\left\{0, \frac{\pi_{xi,\omega}^H(\phi_{i,\omega})}{\delta}\right\} \quad (2.19)$$

where $\chi_i^H \equiv \frac{1-G(\phi_i^{*H})}{1-G(\phi_i^{*H})}$ is the probability of exporting conditional on the firm successfully entering and producing in the domestic market. Hence, the free-entry condition with costly international trade is that the *ex ante* expected value of initial entry equals the sunk entry cost:

$$V_i^H = \frac{1-G(\phi_i^{*H})}{\delta} [\bar{\pi}_i^H + \chi_i^H \bar{\pi}_{xi}^H] = f_{ei} (s^H)^{\beta_i} (w^H)^{1-\beta_i} \quad (2.20)$$

where $\bar{\pi}_i^H \equiv \pi_i^H(\bar{\phi}_i^H)$ is the per-period profit of the average domestic producer from the domestic sales, and $\bar{\pi}_{ix}^H \equiv \pi_{xi}^H(\bar{\phi}_{xi}^H)$ is the per-period profit of the average exporter from export sales.¹⁴

Combining this free-entry condition (2.20) with the zero-profit condition (2.15) and (2.16) yields the following equation¹⁵:

$$\frac{f_i}{\delta} \int_{\phi_i^{*H}}^{\infty} \left[\left(\frac{\phi}{\phi_i^{*H}} \right)^{\sigma-1} - 1 \right] g(\phi) d\phi + \frac{f_{xi}}{\delta} \int_{\phi_{ix}^{*H}}^{\infty} \left[\left(\frac{\phi}{\phi_{xi}^{*H}} \right)^{\sigma-1} - 1 \right] g(\phi) d\phi = f_{ei} \quad (2.21)$$

The first term of the left-hand side of this equation is monotonically decreasing in ϕ_i^{*H} ,

¹⁴ The average productivity level of the group of domestically-producing firms is defined with the cutoff productivity for domestic production in each country, as follows:

$$\bar{\phi}_i^{\lambda}(\phi_i^{*\lambda}) = \left[\frac{1}{1-G(\phi_i^{*\lambda})} \int_{\phi_i^{*\lambda}}^{\infty} \phi^{\sigma-1} g(\phi) d\phi \right]^{\frac{1}{\sigma-1}}$$

Similarly, the average productivity level of the group of exporters is defined with the cutoff productivity for export:

$$\bar{\phi}_{xi}^{\lambda}(\phi_{xi}^{*\lambda}) = \left[\frac{1}{1-G(\phi_{xi}^{*\lambda})} \int_{\phi_{xi}^{*\lambda}}^{\infty} \phi^{\sigma-1} g(\phi) d\phi \right]^{\frac{1}{\sigma-1}}$$

¹⁵ The derivation of Equation (2.21) is shown in the Appendix A.

and the second term is monotonically decreasing in ϕ_{xi}^{*H} . Since ϕ_{xi}^{*H} increases as ϕ_i^{*H} increases (from Equations (2.17) and (2.18), $\phi_{xi}^{*H(F)} = \Lambda_i^{H(F)} \cdot \phi_i^{*H(F)}$, $\Lambda_i^{H(F)} > 1$), the whole of the left-hand side of the equation monotonically decreases as the value of ϕ_i^{*H} increases. With the right-hand side being constant, this Equation (2.21) solves for the unique value of the domestic production cutoff ϕ_i^{*H} and accordingly the export cutoff ϕ_{xi}^{*H} .

2.2.5 Factor Prices

Because of fixed and variable trade costs, factor price equalization (FPE) fails. However, the relative prices of two factors will converge partially such that equilibrium relative factor prices will fall between their autarky and free trade levels. In autarky, the wage for skilled labor relative to that for the unskilled is lower in the skill-abundant Home. Opening the country to costly trade will result in an increase in the relative reward for the abundant factor in each country (i.e., s/w will rise in the Home and w/s will rise (or s/w will fall) in the Foreign), which will decrease the difference in relative factor prices between the two countries. That is;

$$\left(\frac{s^H}{w^H} \right)^A \left(\frac{s^H}{w^H} \right)^{CT} \left(\frac{s^H}{w^H} \right)^{FT}$$

$$\left(\frac{s^F}{w^F} \right)^A \left(\frac{s^F}{w^F} \right)^{CT} \left(\frac{s^F}{w^F} \right)^{FT}$$

where A , CT , and FT indicate autarky, costly trade, and free trade, respectively.¹⁶ The right-hand side (the third term) of the inequality above will be equal to one when free

¹⁶ See Appendix A for demonstration for the equilibrium factor prices.

trade leads to factor price equalization (FPE).¹⁷

This difference in equilibrium relative factor reward implies that the impacts of costly trade will differ across countries and industries due to factor proportion-based comparative advantage. The profits derived from exporting will also vary across countries, across industries, and across heterogeneous firms.

2.2.6 Probability of Exporting

Having analyzed both firm-level production heterogeneity and country-level factor prices in equilibrium, I unite them to analyze the determinants of a firm's exporting status. The *ex ante* probability for a domestic producer to be an exporter is determined by the two productivity cutoffs: ϕ_i^* for domestic production and ϕ_{xi}^* for export. That is, as previously defined, the probability is expressed as follows:

$$\chi_i^H \equiv \frac{1 - G(\phi_{xi}^{*H})}{1 - G(\phi_i^{*H})} < 1. \quad ^{18}$$

In the equilibrium, this probability equals the *ex post* fraction of exporting firms in all the domestically-producing firms. That is, denoting the mass of the continuum of actively-producing firms by M_i and that of exporting firms by M_{xi} ,

$$M_{xi}^H / M_i^H = \chi_i^H \quad (2.22)$$

The concern of this paper is documenting the determinants of the cross-industry patterns of this probability of a domestic producer being an exporter. Before deriving a prediction, I introduce the following assumption on the distribution for firm productivity:

¹⁷ It can be shown that a free-trade equilibrium with FPE exists in this model economy. The author can provide the proof upon request.

¹⁸ The inequality follows $\phi_{xi}^{*H} > \phi_i^{*H}$ as Equation (2.17) (Equation (2.18) for Foreign) shows.

$$\text{Assumption: } G(\phi_i) = 1 - \left(\frac{\underline{\phi}_i}{\phi_i} \right)^k \text{ for } i = 1, 2, \dots, N; k > 2\sigma$$

That is, I assume that the *ex ante* distribution of firm productivity is a Pareto distribution.¹⁹ $\underline{\phi}_i$ is the minimum value for productivity drawn in Industry i ($\phi_i \in [\underline{\phi}_i, +\infty)$), and k is a shape parameter that indicates the dispersion of productivity distribution, which is assumed to be common across industries. I assume $k > 2\sigma$ for the variances of both drawn productivities and sizes of firms (measured as domestic sales) to be finite.²⁰

Now the following proposition is derived:

Proposition: Suppose $f_i = f, f_{xi} = f_x$, and $\tau_i = \tau$ for any $i = 1, 2, \dots, N$.

Then, if $\frac{\bar{S}^H}{\bar{U}^H} > \frac{\bar{S}^F}{\bar{U}^F}$ and $\beta_1 < \beta_2 < \dots < \beta_N, \chi_1^H < \chi_2^H < \dots < \chi_N^H$ and

$$\chi_1^F > \chi_2^F > \dots > \chi_N^F.$$

Proof: See Appendix A.

This proposition implies that if fixed costs for production and export differ across industries only due to the cross-industry variation of the cost shares of two factors,²¹ and the “iceberg” shipping costs are also the same for all industries, then the *ex ante* probability for a domestic producer to be an exporter, which is equal to the *ex post*

¹⁹ See Chaney (2008) for references evidencing that a Pareto distribution well approximates the observed distribution of the sizes of the U.S. firms. A Pareto distribution is also used frequently for the distribution of firm productivity in this type of models: for example, Helpman, Melitz and Yeaple (2004), Ghironi and Melitz (2005), Melitz and Ottaviano (2008), and Bernard, Redding and Schott (2007).

²⁰ For the variance of drawn productivity to be finite, it must be that $k > 2$. For the variance of the domestic sales of firms to be finite, $k > 2(\sigma-1)$. For these two conditions for k to be satisfied for any $\sigma > 1$, I assume $k > 2\sigma$.

²¹ Recall that both production fixed costs and export fixed costs depend on factor prices:

$f_i (s^H)^{\beta_i} (w^H)^{1-\beta_i}$ and $f_{xi} (s^H)^{\beta_i} (w^H)^{1-\beta_i}$. The cost shares of the two factors in these fixed costs differ across industries. Therefore, even though the parameters f and f_x are the same for all industries, the fixed costs still vary across industries.

fraction of exporting firms to all domestic firms, will be higher in an industry that uses more intensively a production factor with which a country is relatively well-endowed. That is, if other things are equal, for the (un)skilled labor-abundant Home (Foreign) country, a larger fraction of firms that are serving the domestic market will export as an industry is more (un)skilled-labor intensive.

The key determinant of the *ex ante* probability of a domestic producer being an exporter is the “gap” between the two productivity cutoffs—the minimum productivity level for domestic production and the minimum for export. The gap between the two cutoffs, which is measured as the ratio of the export productivity cutoff to the domestic production productivity cutoff (ϕ_{xi}^* / ϕ_i^*), thus decide the *ex post* fraction of exporters among active domestic firms. The predicted ranking of the exporter fractions that the proposition states is generated from the (reversed) ranking of the “gap” between the two productivity cutoffs in equilibrium. In other words, the “gap” is smaller in an industry with a stronger degree of comparative advantage (i.e., an industry that more intensively uses a factor with which the country is relatively well-endowed), as depicted in Figure 2.1. While the “gap” is defined as the ratio of the two productivity cutoffs, Figure 2.1 expresses the “gap” as a distance between the two cutoffs on a line (one can understand that the productivity levels in the figure are shown in a log scale).²² The mechanism that generates this cross-industry ranking of the productivity “gaps” is intuitively explained by competition in domestic factor markets. Consider the case for the skilled-labor abundant Home. When the country opens up to costly international trade, the potential profit will rise for firms with high productivity, as well as for new entrants that can

²² Figure 2.1 also normalizes the autarky productivity cutoffs for domestic production in all industries for an illustrative purpose.

possibly draw a high level of productivity, due to the additional sales opportunity in the foreign market. This will raise the domestic labor demand and thus increase domestic wages. The domestic wage increase results in the increase in production costs, and due to this production cost increase, all domestic firms will require a higher level of productivity for survival, thus raising the productivity cutoff for domestic production. This increase in the domestic-production productivity cutoff will occur in all industries, while the productivity increase will be more pronounced in more skill-intensive industries. The reason for this is that the increase in potential profit from export is larger in more skill-intensive (i.e., comparative-advantage) industries, and the increase in firms' factor demands will thus be larger in more skill-intensive industries. This results in a larger increase in the demand for skilled labor than in the unskilled-labor demand, and thus the relative price of skilled labor to the unskilled will rise. As a result, the increase in production costs will be larger and accordingly the rise in the minimum productivity level required for domestic production will be greater if the industry is more skill intensive. At the same time, for the "survivor" firms, exporting will be easier in more skill-intensive industries because of the country's comparative advantage. In costly-trade equilibrium, the relative price of skilled labor will be lower in the Home than the Foreign, and this relative factor-price advantage will be more pronounced if the industry is more skill intensive. This results in a lower minimum productivity level for exporting in that industry.

2.3 Data

The model predicts, as the Proposition in the last section states, that a larger fraction of firms that are active (i.e., producing and selling their products) in the domestic

market will become exporters in an industry in which the country's comparative advantage is stronger. This implies that, for each country, the ratio of exporters to all active firms in an industry can be ranked according to the industry's intensity of the use of the factor that is better-endowed in the country relative to the rest of the world. To test this prediction empirically, I need information on how many firms in each industry are active and how many out of those active firms export to other countries. I also need information on the factor intensity of each industry, as well as the factor abundance of countries in which the firms locate. In the current study, I use four countries for which at least limited data are available: Chile, Colombia, India, and the United States.

2.3.1 Chilean Data

For Chile, I employ a firm-level dataset from an annual manufacturing census conducted by the national statistical institute of the country (Instituto Nacional de Estadísticas: INE), which was compiled and documented in English through a World Bank project.²³ The manufacturing census dataset covers all establishments with ten or more employees. The dataset contains various kinds of information on each establishment including employment by type and the values of sales and exports. The dataset also contains the code of the industry that each firm belongs to, which is according to Chile's national classification of economic activities (Clasificador de Actividades Economicas, CIIU²⁴) at the four-digit level. Although the dataset covers the years 1979 through 1996, it has the export value of each firm only for 1990 through 1996. The dataset thus enables the calculation of the ratio of exporters to active firms and the skilled-labor intensity for various manufacturing industries for 1990 through 1996.

²³ See Roberts and Tybout (1996) for the summary of the project funded by the World Bank.

²⁴ CIIU is indeed equivalent to the International Standard Industry Classification (ISIC), Revision 2.

2.3.2 Colombian Data

The dataset for Colombia that I use in this study is from an annual census of manufacturing by Colombia's national statistical office (Departamento Administrativo Nacional de Estadística: DANE), which was also compiled as a part of the World Bank project. The Colombian manufacturing census contains a variety of information on manufacturing plants with ten or more employees. The dataset covers the years 1981 through 1991. The industry code is also available according to the International Standard Industry Classification (ISIC, Revision 2) at the four-digit level. The dataset allows for the calculation of the fraction of exporters among active firms and the skilled-labor intensity for each manufacturing industry for 1981 through 1991.

2.3.3 Indian Data

Information on India is originally sourced from the Annual Survey of Industries (ASI) conducted by the Ministry of Statistics and Programme Implementation of India. The survey covers all industrial units that are registered under the Factories Act with more than 20 employees. This paper uses data that are aggregated for each of the four-digit ISIC (Revision 2) industries from the original unit level data, which include the following industry-level variables: the numbers of industrial units, exporting units, skilled workers, and unskilled workers. The data are for a single year of the Indian fiscal year 1997/98 (the period from April 1, 1997 through March 31, 1998).²⁵

2.3.4 U.S. Data

For the United States, although I do not have equally detailed firm-level data as those for the previously-mentioned three countries, I have collected the necessary data

²⁵ I thank Jagadeesh Sivadasan for providing the aggregated data for India. Also see Sivadasan (2007) for the description of the original ASI data.

from published sources. The U.S. Bureau of the Census (1992) provides the numbers of establishments and exporting establishments in each of twenty manufacturing industries classified according to the 1987 U.S. Standard Industry Classification (SIC) at the two-digit level, for the year 1992. Data on the number of production workers and the total employment in each two-digit U.S. SIC industry are available from the publications on the 1992 U.S. Census of Manufactures. These are used to measure the skilled-labor intensity of each manufacturing industry.

2.3.5 Factor Endowment Data

For the information on how well these four countries are endowed with (un)skilled labor (i.e., the skilled-to-unskilled labor ratio: S/U), I employ the ratio of human capital to labor ratio provided by Hall and Jones (1999) as my benchmark measure. They measure human capital per worker in a country using the average years of schooling in the population of the country and return-to-schooling estimates.^{26, 27} I also use data on educational attainment reported by Barro and Lee (2000) as an alternative measure of the countries' relative skill abundance for robustness checks. The details of this alternative measure are described in a later section.

Table 2.2a shows the relative skilled-labor abundance of the four countries according to the Hall and Jones' measure of human capital per worker. The summary

²⁶ More specifically, they measure human capital per worker in a country (c) as $H_c / L_c = e^{\phi(E_c)}$, where E_c is the average years of schooling in the country's population measured by Barro and Lee (1993), and $\phi'(E)$ is an estimated return to schooling in a Mincerian wage regression that is reported by Psacharopoulos (1994). They apply a piecewise linear specification to $\phi(E)$ assuming that the return to schooling differs for each segment of year length of schooling. See Hall and Jones (1999) for more details.

²⁷ I also use another measure of human capital per worker relative to the United States reported by Klenow and Rodriguez-Clare (1997). Their way of estimating the human capital to labor ratios for countries is similar but slightly different from the method applied by Hall and Jones (1999). See Klenow and Rodriguez-Clare (1997) for the details. My empirical results, however, do not change between these two measures, and thus I report only the results with Hall & Jones' measure in the following part of this paper.

statistics of the variable for 127 countries covered in their data are also presented for reference. All the numbers are based on the statistics weighted by the amount of labor reported in the source data. Similarly, Table 2.2b reports the relative skill abundance of the four countries according to Barro and Lee’s measure of tertiary education completion in the total population over age 15, with the population-weighted summary statistics for 103 countries covered in their data. Both tables show that these four countries represent diverse groups of countries in terms of skill-labor abundance. That is, the United States is very skill rich, Chile is moderately skill abundant, Colombia is about the “middle,”²⁸ and India is skill scarce or unskilled-labor abundant.

2.4 Empirical Analysis

2.4.1 Individual Country Analysis

The theoretical two-country two-factor model suggests that if a country is more (un)skilled-labor abundant relative to the rest of the world, the fraction of exporting firms among all active firms in that country will be higher in more (un)skilled-labor intensive industries. To test this prediction, I apply the following empirical model for each individual country:

$$ex_share_i = \gamma + \theta \cdot skill_i + \varepsilon_i \quad (2.23)$$

where ex_share_i = (number of exporters) / (number of active firms) in Industry i , and $skill_i$ is the skilled-labor intensity of the industry.²⁹ The theoretical prediction is that the coefficient for the skilled-labor intensity (θ) will be larger (i.e., more positive or less

²⁸ According to Hall and Jones’ data, Colombia is a slightly less skill abundant country compared to the median and the mean of the 127 countries, while the country falls between the median and mean of the 103 countries in Barro and Lee’s data.

²⁹ The measure of skilled-labor intensity is described for each country later in this subsection. Because the categories of workers in the data are different for each country, the definition of the industry skill intensity is not exactly the same for all four countries.

negative) for the country with higher skilled-labor abundance. In the rest of this subsection, I test this prediction by estimating the regression equation (2.23) using the data for each of the four sample countries: Chile, Colombia, India, and the United States, and compare the coefficient θ estimated for each country.

Chile

For Chile, using the dataset described in the previous section, I compute each variable in Equation (2.23) for 25 manufacturing industries classified according to the three-digit ISIC.³⁰ Exporters are defined as firms with positive values of exports, and active firms are firms with positive total sales of goods. The skilled-labor intensity of each industry $skill_i$ is the share of skilled workers^{31, 32} in all workers employed in each industry.³³ For Chile, as well as for Colombia in the following subsection, data are averaged over periods for cross-industry estimation; this is done to achieve a fair comparison of estimation results with those for India and the United States for which data are available only for a single year. Hence, for Chile, the average values of the variables

³⁰ The three-digit ISIC lists 28 manufacturing industries. I exclude the following four industries from the estimation: 314 (tobacco products), 353 (petroleum refineries), 354 (miscellaneous petroleum and coal products), and 390 (other manufacturing products). The last category is excluded because of its miscellaneous status. The first three categories are excluded because these industries are extremely concentrated (Carree *et al.*, 2000). In the data used in the present study, the number of active (domestic) firms is significantly lower in these industries compared to other three-digit industries in Chile, Colombia, and India. (Although industries are classified differently in the U.S. data, the number of firms in the U.S. data is also extremely small in industry categories corresponding to these three ISIC industries.) Alvarez and Lopez (2006) and Bergoeing and Repetto (2006), which use Chilean firm-level data, also exclude the tobacco and petroleum industries from their analyses because these industries “are organized as monopolies, operating with very few plants” (Bergoeing & Repetto, 2006).

³¹ The categories of workers in the dataset are more detailed. I define skilled workers as the total of owners, executives, white-collar administrative workers, and white-collar production workers. Unskilled workers are the rest of the workers employed; i.e., blue-collar production and non-production workers, workers at home, and salespersons in commission. Therefore, the unskilled-labor intensity of each industry equals $1 - skill_i$.

³² The proposition presented in the second section of this paper is based on the Cobb-Douglas cost share of skilled workers in an industry (β_i). However, as explained in Footnote 10, with an equilibrium relative wage (s/w) in a country, the ranking of industry skill intensity measured by the physical unit of labor ($S_i/(S_i+U_i)$) corresponds to the ranking of β_i .

³³ The total employment in each industry, as well as the number of (un)skilled workers, is computed by aggregating for each industry the numbers of workers (in each category) hired by the firms in the dataset.

over the years 1990 through 1996 are used for the analysis.³⁴

The result of the estimation by OLS is presented in Table 2.3, and the plot of the fractions of exporters against the industry skill intensities is shown with the fitted line in Figure 2.2. In Chile the correlation between the exporter fractions and the industry skill intensities is positive ($\hat{\theta} = 0.634$) and significant.

Colombia

Variables for estimating Equation (2.23) are computed from the previously-mentioned Colombian dataset for the 25 industries classified according to the three-digit ISIC. The value of each variable is the average from 1981 to 1991.³⁵ The definitions of an exporter and an active firm are the same as those in the Chilean case. The industry skilled-labor intensity is the share of workers in the categories of owners, management, skilled workers, and local and foreign technicians in all workers employed in each industry.³⁶

Table 2.4 shows the estimation result, and Figure 2.3 plots the fractions of exporters against the skilled-labor intensities of the 25 three-digit industries. The correlation between the exporter fractions and the industry skill intensities is not significant in Colombia ($\hat{\theta} = -0.091$).

India

For India, I compute the variables for a regression from the data described in Section 2.3 for the 25 manufacturing industries classified according to the three-digit

³⁴ I also estimated the coefficient theta using whole panel data for Chile (with time dummies, standard errors clustered by industry), and obtained virtually the same result for each country.

³⁵ The averaged data are used for the same reason as described for Chile in the previous subsection. Estimation with whole panel data does not alter the result, however.

³⁶ Unskilled labor is thus the sum of workers in other categories in the data (i.e., unskilled workers and apprentices).

ISIC. Exporter fraction in each industry is measured as the number of exporting industrial units divided by the number of all units in the industry. The skilled-labor intensity of each industry is the share of non-production workers in all employees in the industry. The variables are for the single fiscal year 1997/98.

Table 2.5 presents the result of the estimation. In India, the correlation between the exporter fractions and the industry skill intensities is negative ($\hat{\theta} = -0.865$) and fairly significant. Figure 2.4 plots the exporter fractions vs the skilled-labor intensities of the 25 manufacturing industries and shows the fitted line together.

Table 2.6 summarizes the results of these individual country regressions for the three countries in order to compare the estimates of the coefficient for the industry skill intensity. For all three countries, the data are for the common 25 manufacturing industries classified according to the three-digit ISIC. The correlation estimated by each individual country regression is larger, or more positive, for a country with higher skill abundance, as the theoretical model suggests. The coefficient estimate is negative for India while it is positive for Chile, and the estimate for Colombia falls in between.

United States

Unlike for the other three countries, for the United States, industries are not classified according to the ISIC in the available data, but are classified according to the 1987 U.S. SIC at the two-digit level. I thus estimate Equation (2.23) for the United States with 17 manufacturing industries.³⁷ The exporter fraction (ex_share_i) is the number of exporting establishments divided by the number of (all) establishments in each industry.

³⁷ There are 20 two-digit U.S. SIC manufacturing industries, but as for the other three countries, for the United States I exclude tobacco, petroleum and coal, and miscellaneous industries from estimation. The following three categories in the two-digit U.S. SIC corresponds to these three industries and thus are excluded: 21 (tobacco products), 29 (petroleum and coal products), and 39 (misc. manufacturing industries).

The industry skilled-labor intensity is measured as the share of non-production workers in all workers employed in each industry.³⁸ The variables are for the single year 1992.

The result of the estimation with the 17 manufacturing industries by OLS is presented in Table 2.7. Figure 2.5 displays the plot of the fractions of exporters against the industry skill intensities along with the fitted line. The correlation between the exporter fractions and the industry skill intensities is positive ($\hat{\theta} = 0.587$) and fairly significant in the United States.

Comparing Four Countries

The coefficient estimates for the industry skill intensity (θ) for the United States is not directly comparable with the estimates for the other three countries due to the difference in industry classification. For the cross-country comparison of the coefficient, I compute for Chile, Colombia, and India the exporter fractions and skill intensities in the two-digit U.S. SIC industries, using the concordance between the three-digit ISIC and the two-digit U.S. SIC, which is presented in Table 2.9. The re-classified data are used to re-estimate Equation (2.23) for each of these three countries with the 17 two-digit U.S. SIC manufacturing industries. Table 2.8 compares the coefficients estimated for Chile, Colombia, India, and the United States.³⁹ The table shows that the relative sizes of the coefficients correspond to the relative skill abundance of these countries. That is, the correlation between the exporter fractions and industry skill intensities is the largest

³⁸ More specifically, since in the Census of Manufactures the number of production workers is available for each manufacturing industry, I first calculate *unskilled-labor* intensity that is defined as the share of production workers in the total employment, and then compute skilled-labor intensity as one minus unskilled-labor intensity.

³⁹ For Chile and India, the coefficient estimates with the 17 two-digit U.S. SIC industries are not significant whereas the estimates with the 25 three-digit ISIC industries are significant at least at the 10% level. This can be explained by the aggregation of the industries. The aggregation of the 25 manufacturing industries to the 17 reduces the variance of the exporter fractions, which makes the size of the slope coefficient smaller. The aggregation also reduces the variance of the skill intensities among industries, which makes the standard error of the estimate larger.

(most positive) for the most skill abundant United States, the second for Chile, the third for Colombia, and the smallest (most negative) for the least skill abundant India. This result is consistent with the case for the 25 three-digit ISIC industries shown in Table 2.6, and also with the theoretical prediction.

Impact of Sample Truncation

As described in the previous section, manufacturing census data for these countries exclude small firms whose employment is below the threshold level.⁴⁰ This omission of small firms might cause bias in estimation, in particular if the fraction of exporters among domestic firms in a sample systematically overestimate or underestimate the fraction in a population in relation to industry skill intensities. This possibility of estimation bias is examined in Appendix B; however, the exclusion of small firms should not affect the result of the present empirical analysis.

2.4.2 Pooled Analysis

The predicted relationship among the relative factor abundance of countries, relative factor intensities of industries, and the ratio of exporters to all active firms has been confirmed by individual country regressions in the previous section. This quasi-Heckscher-Ohlin prediction can be tested more formally using the following empirical model:

$$ex_share_{ic} = \gamma_c + \Pi_c \cdot skill_i + \varepsilon_{ic} \quad (2.24)$$

where i indexes an industry and c indexes a country. The coefficient Π_c for the industry skilled-labor intensity, as well as the constant term γ_c , being indexed by c means that

⁴⁰ The U.S. Census of Manufactures also omits small firms that are excused from filing reports. See 1992 Census of Manufactures General Summary (MC92-S-1, pp. VII-IX) for the details of company coverage in the census.

these parameters differ across countries. In particular, the theoretical prediction is that the slope coefficient Π_c will be larger (more positive or less negative) for a country with a higher relative skilled-labor abundance, and smaller (less positive or more negative) for a country with a lower relative skilled-labor abundance (or a higher relative unskilled-labor abundance). To capture this correlation between the coefficient Π_c and the skill abundance of a country, the following structure is imposed:

$$\Pi_c = \Pi((S/U)_c) = \theta_1 + \theta_2 \cdot \log(S/U)_c \quad (2.25)$$

where $(S/U)_c$ is the skilled-labor to unskilled-labor ratio of Country c .⁴¹ By substituting (2.25) for (2.24), the following equation is derived:

$$ex_share_{ic} = \gamma_c + \theta_1 \cdot skill_i + \theta_2 \cdot skill_i \cdot \log(S/U)_c + \varepsilon_{ic} \quad (2.26)$$

This equation is estimated with pooled data for the four countries (Chile, Colombia, India, and the United States) and the 17 manufacturing industries classified according to the two-digit 1987 U.S. SIC. The pooled data allows for the inclusion of industry dummies in estimation to control for the effects of industry-specific factors other than the skill intensity, such as fixed and variable costs for export. (Recall that in Section 2.2 the theoretical model derives quasi-Heckscher-Ohlin prediction when the (factor-price-adjusted) fixed costs for production, fixed costs for exporting, and variable shipping costs are the same across industries. These costs, however, are generally not the same.⁴²)

Hence, the equation is estimated in the following form:

⁴¹ The advantage of the log-scaled measure of relative factor abundance is that the size (absolute value) of the coefficient θ_2 will be invariant to which of S or U is the denominator of the measure.

⁴² For this reason, it is ideal to control for these industry-specific costs in individual regressions in the previous subsection. However, industry-specific dummies cannot be used since the observations in individual country data are unique for each industry. In addition, no relevant measures of these costs are available. Nevertheless, the result of the individual country analysis is valid as far as these industry-specific costs are symmetric or invariant across countries.

$$ex_share_{ic} = \theta_2 \cdot skill_i \cdot \log(S/U)_c + \gamma_c + \eta_i + \varepsilon_{ic} \quad (2.26.2)$$

where γ_c and η_i are series of industry-specific and country-specific intercepts, respectively.⁴³

The fraction of exporters among active firms in each industry (ex_share_{ic}) is obtained from the data for each individual country. For Chile, the variable is of the average over the years 1990 through 1996; for Colombia, the variable is of the average over 1981 through 1991; for India, the variable is for the fiscal year 1997/98; and for the United States, the variable is for the year 1992. The skilled-labor to unskilled-labor ratio in each country ($(S/U)_c$) is measured as the human capital to labor ratio reported by Hall and Jones (1999). The variable $skill_i$ is now defined as the (Cobb-Douglas) cost share of skilled labor in each industry, which is assumed to be common across countries in the theoretical model. The cost-share measure, rather than the skill-intensity measure based on the physical amount of labor, is chosen because while the Cobb-Douglas cost share is the same for all countries, the employment-based intensity of each type of labor will differ across countries in general due to the difference in relative wage (s/w) in the costly-trade equilibrium.⁴⁴ This common cost-share variable is measured as wage payments to non-production workers as the share in the total annual payroll in each industry, using the data in the 1992 U.S. Census of Manufactures. The skilled-labor cost shares in the 17 two-digit U.S. SIC manufacturing industries are shown in Table 2.10.

The result of the estimation of Equation (2.26.2) is presented in Table 2.11. The positive and significant (at the 4% level) estimate $\hat{\theta}_2$ suggests that correlation between

⁴³ Note that the first term ($\theta_1 \cdot skill_i$) of Equation (2.26) is dropped from the estimation due to the inclusion of industry-specific dummies, because the industry skill intensity is unique for each industry.

⁴⁴ See Footnote 8.

the exporter fractions and industry skill intensities (defined by the cost shares) is larger (or more positive) for a country with a higher relative skill abundance. This result of the pooled-data analysis confirms the result of the individual country analysis, and thus supports the theoretical prediction.⁴⁵

2.4.3 Robustness Check of Pooled Analysis

Although non-production workers or white-collar workers are frequently used in empirical studies to represent skilled labor, these are crude proxies. A more desirable measure is to categorize workers according to their (potential) skill levels such as educational attainment. To check the robustness of the result of the pooled regression in the previous subsection, I employ a measure of industry skill intensity proposed by Morrow (2008). He uses data from the March U.S. Current Population Survey for the years 1988-92 that contain information on incomes and educational attainment of workers employed in various industries. He computes the Cobb-Douglas cost share of skilled workers using the share of employees in each educational category and the wage levels of workers in an educational category relative to the wage level of workers in other categories estimated from a Mincerian wage regression. While he reports the information for the three-digit ISIC manufacturing industries, for the current study, I use this information to compute the skilled-labor cost shares for the two-digit U.S. SIC industries. I define skilled labor as workers with one or more years of college education. The details of the computation are described in Appendix D. The obtained skilled-labor cost shares in the 17 two-digit U.S. SIC manufacturing industries are listed in Table 2.12.

Because the industry skill intensity is now measured based on the educational attainment of workers, for consistency I also employ an educational attainment-based

⁴⁵ See Appendix C for further empirical exercise.

measure for the countries' relative abundance of skilled labor, $(S/U)_c$. Specifically, I use the percentage of the population that has attained tertiary education reported by Barro and Lee (2000). The percentages of tertiary education attainment for Chile, Colombia, India, and the United States are shown in Table 2.13.⁴⁶

Using these alternative measures of industry skill intensity and country skill abundance, Equation (2.26.2) is re-estimated with data for the four countries and 17 manufacturing industries. The result is presented in Table 2.14. A positive coefficient is estimated ($\hat{\theta}_2 = 0.277$) at the 5% level of significance (p-value = 0.018), which indicates that the estimated relationship among the factor abundance of countries, the skill intensities of industries, and the exporter fractions is robust across different measures of country factor abundance and industry skill intensity.⁴⁷

2.4.4 Factor Prices

In the current model, the mechanism that determines the cross-industry patterns of exporter fraction operates based on the relative prices of the two factors. As described in Section 2.2, in the costly-trade equilibrium, the relative wage is not equalized between countries. This relative wage difference creates the variation in the exporter fractions between comparative-advantage industries and comparative-disadvantage industries. To confirm this mechanism, in this subsection, I estimate the exporter fraction equation using wage data.

The source of information on wages is the Occupational Wages around the World

⁴⁶ Since the exporter fractions are measured in different periods for each country, the educational attainment data for different periods are employed for each country to have the periods of the two variables being consistent. See Table 2.13 for the data periods for each country.

⁴⁷ The equation is also estimated using the alternative measure only for either industry skill intensity or for country skill abundance, maintaining the benchmark measure for the other variable. In any case, the estimated coefficient is positive and significant at the 5% level or more.

(OWW) Database, an NBER dataset provided by Freeman and Oostendorp (2000). This is a comprehensive dataset of wages of various occupations in a large number of countries. The occupational wage data in the OWW Database are derived from the “October Inquiry,” which is a wage survey conducted by the International Labour Organization (ILO).⁴⁸ However, the ILO’s data, which are based on reports from national governments, involve problems such as inconsistency in wage format (e.g., weekly or monthly, minimum or average), missing data, and erroneous records. Therefore, in the OWW Database, the original October Inquiry data have been cleaned and standardized by calibration.⁴⁹ The OWW Database thus provides comparable wage data for a large number of occupations in many countries. In the current paper, I use an updated version of the OWW database by Oostendorp (2005) that covers 161 occupations (in various industries including services and the government sector) in 137 countries for the years 1983 through 2003.⁵⁰ The numbers of occupations and years covered in the database significantly vary across countries, however. For the four countries examined in the current study, wage data are available for the following years and numbers of occupations: for Chile, 89 to 134 occupations for 1984-86; for Colombia, 41 to 124 occupations for 1988-90; for India, 13 to 93 occupations for 1985-2000; and for the United States, 11 to 152 occupations for 1984-2002.

The skilled-labor wage relative to unskilled-labor wage, s/w , needs to be measured for each country. Since both numbers and types of occupations reported in the

⁴⁸ The original ILO data are available at <http://laborsta.ilo.org/>. Also see Freeman and Oostendorp (2000) for the description of ILO’s October Inquiry.

⁴⁹ See Freeman and Oostendorp (2000) and Oostendorp (2005) for the details of the procedure of data calibration and standardization.

⁵⁰ The database is available at <http://www.nber.org/oww/>. Specifically, I employ the data on wages with country-specific and uniform calibration and lexicographic weighting (the variable “x3wl”) in the database.

dataset vary across countries and years, I calculate the relative wage (s/w) in the following three ways: (i) taking the ratio of the highest occupational wage to the lowest (whatever these occupations are); (ii) dividing occupations into ten groups by wage deciles and taking the ratio of the mean wage in the highest-wage group to the mean wage in the lowest-wage group; and (iii) taking the ratio of the 90th percentile wage to the 10th percentile wage. After computing these three relative wage measures for each of the four countries for each year, I select the values of the three variables for the following period to match the wage data period to that of the exporter fraction for each country: for Chile, the averages over 1984-86;⁵¹ for Colombia, the averages over 1988-90;⁵² for India, the averages of 1997 and 1998; and for the United States, the year 1992. Table 2.15 lists the values of the relative skilled-to-unskilled wage in the three measures for each country.⁵³ The table shows that the relative wage reflects the relative skilled-labor abundance of the countries, except for Chile and Colombia. Between these two “medium” countries, the relative positions in terms of comparative advantage are reversed when they are measured by the relative wage.

The quasi-Heckscher-Ohlin prediction that the fraction of exporters among active domestic firms is higher in comparative-advantage industries for each country is tested using these three measures of the factor price ratio. The same empirical model as in the previous subsections is applied, but the skilled-to-unskilled labor ratio (S/U) is now replaced with the skilled-to-unskilled wage ratio (s/w), and thus the regression equation is

⁵¹ The periods do not match the ones for the exporter fractions (1990-96), but these are only periods for which the wage data are available for Chile.

⁵² The periods do not completely match the ones for the exporter fractions (1981-91), but these are only periods for which the wage data are available for Colombia.

⁵³ Note that these three relative wage measures are based on different sets of occupations for different countries. In the following subsection, I estimate the same equation using alternative relative wage measures that are based on the same set of occupations for all four countries.

as follows:

$$ex_share_{ic} = \psi \cdot skill_i \cdot \log(s/w)_c + \gamma_c + \eta_i + \varepsilon_{ic} \quad (2.27)$$

η_i and γ_c are industry- and country-specific intercepts, respectively, as in Equation (2.26.2). The same 17 two-digit U.S. SIC manufacturing industries are used for estimation. As in Subsection 2.4.2, the industry skilled-labor cost share ($skill_i$) is measured as the wages to non-production workers as the share in the total payroll (in Table 2.10). Note that since the relative price of skilled labor (s/w) is lower in a more skill-abundant country, the model expects ψ to be *negative*.

The results of the estimation of Equation (2.27) using the three relative wage measures are presented in Table 2.16. With any wage measure, the estimate of ψ is negative and significant at the 5% level (p-value is between 0.019 and 0.040).⁵⁴ This indicates that the correlation between the exporter fraction and industry skill intensity (measured as the cost share) is larger in a country where the skilled labor is relatively cheaper, which is consistent with the prediction.

Relative Wage Measures Based on Same Occupations

Since the three measures of relative wage (s/w) are based on a different set of occupations for each country, the measured cross-country variation in the relative wage might be simply due to the difference of occupation composition. For instance, the large gap between the measured wages of skilled and unskilled workers in India may not reflect the relative unskilled-labor abundance of the country, but instead may be due to the fact that data for India contain extremely well-paid occupations that are not covered in data for other countries. To address this potential measurement issue, in this

⁵⁴ The equation is also estimated using the measure of skilled-labor cost share based on workers' educational attainment (following Morrow, 2008). The results do not change: with each relative wage measure, a similar size of the coefficient is estimated at least at the 5% level of significance.

subsubsection, I construct relative wage measures that are adjusted for occupation composition, and estimate Equation (2.27) with those alternative measures to check the robustness of the results presented above.

I first select a set of occupations that is in common across four countries. The occupation set consists of 25 occupations selected based on the availability of wage data in the OWW2 dataset. These 25 occupations are observed for Chile in 1985-86, for Colombia in 1988,⁵⁵ for India in 1997-98, and for the United States in 1992.⁵⁶ Based on this common occupation set, I compute the following three measures of skilled-to-unskilled wage ratio, which are similar to the relative wage measures used for the preceding estimation: (i) the ratio of the highest wage to the lowest wage in each country (maximum to minimum wage ratio); (ii) the ratio of the mean wage of the three highest-wage occupations to the mean wage of the three lowest-wage occupations (the top10% to the bottom10% mean wage ratio); and (iii) the ratio of the 90th percentile wage to the 10th percentile wage (the top 90th to the bottom 10th percentile wage ratio). The values of the three relative wage measures for each country are shown in Table 2.17.⁵⁷

These three alternative wage measures are used to re-estimate Equation (2.27) for the four countries for the 17 two-digit U.S. SIC industries. The results are presented in Table 2.18, and are very similar to the results of the preceding estimation in Table 2.16. These results imply that (i) the relative wage measures used for the first estimation,

⁵⁵ The availability of wage data in the dataset is the most limited for Colombia among the four countries. In addition, sets of data-available occupations for Colombia substantially differ across years. Therefore, to maximize the number of occupations in a common set to all four countries, I select a single year 1988 for Colombia. Chilean wage data in 1984 are also omitted for the same reason.

⁵⁶ These 25 industries are, in the occupation codes in the ILO October Inquiry: 30, 36, 59, 65, 67, 70, 82, 84, 85, 88, 90, 91, 95, 96, 97, 99, 100, 129, 130, 131, 133, 134, 135, 140, and 141. These occupations include a computer programmer, which is the best-paid occupation among these occupations in all four countries, and a laborer, which is one of the least-paid occupations in every country.

⁵⁷ Note that the order reversal between Chile and Colombia is also observed in these alternative wage measures.

although they are not adjusted for occupation composition, are well reflecting the countries' relative factor abundance, and (ii) the relative wage-driven mechanism of the theoretical model is empirically supported.

2.5 Conclusion

This paper investigates what patterns of the fractions of exporters among domestic firms emerge when countries and industries are asymmetric. The model of a two-country, two-factor and many-industry economy with productivity-heterogeneous firms, which is an extension of the two-factor, two-country and two-industry framework by Bernard, Redding and Schott (2007), suggests that a country's comparative advantage in terms of relative factor abundance explains the cross-industry and cross-country patterns of exporter selection. That is, the probability that a domestic producer will be an exporter is higher in the country's comparative-advantage industries. Furthermore, the fractions of exporters among domestic firms can be ranked according to the order of the industries' intensity of a production factor with which the country is relatively well-endowed. This quasi-Heckscher-Ohlin prediction about exporter fractions is empirically tested using data for manufacturing firms in Chile, Colombia, India, and the United States. The result of the analysis confirms the quasi-Heckscher-Ohlin pattern: the correlation between exporter fractions and the skill intensity of industries is larger, or more positive, for a country with higher skilled-labor abundance. This empirical finding is robust across alternative measures of both industry relative factor intensity and country relative factor abundance.

By empirically demonstrating the effect of factor proportion on the selection of exporters in different industries and countries, this paper highlights one role of

comparative advantage that has not been adequately explored. The result of this study implies that, through the exporter selection, the influence of a country's comparative advantage can be more pronounced in the industrial composition of the country's export than in that of the country's domestic production, at least in terms of the extensive margin.

The empirical analysis in this paper is in fact based on data for a limited number of countries. Having more countries in a sample would be desirable to more strongly confirm the theoretical quasi-Heckscher-Ohlin prediction. Nevertheless, the four sample countries in this paper represent a variety of country groups in terms of relative skilled-labor abundance, and the empirical result clearly demonstrates a comparative advantage-driven variation in the patterns of exporter selection among countries.

2.6 Appendix A

2.6.1 Derivation of Equation (2.12)

Note, from Equation (2.6), that the ratio of the revenues of two firms with different productivities is expressed with the ratio of those firms' productivities, such as follows:

$$\frac{r_i(\phi')}{r_i(\phi)} = \left(\frac{\phi'}{\phi}\right)^{\sigma-1} \Leftrightarrow r_i(\phi') = \left(\frac{\phi'}{\phi}\right)^{\sigma-1} r_i(\phi) \quad (2.A.1)$$

Using this relationship, as well as Equation (2.7) for an individual firm's profit and Equation (2.9) for the revenue of the firm with the cutoff-level productivity, the free-entry condition (2.11) becomes as follows:

$$\begin{aligned}
[1 - G(\phi_i^{*\lambda})] \frac{\bar{\pi}_i^\lambda}{\delta} &= f_{ei}(s^\lambda)^{\beta_i} (w^\lambda)^{1-\beta_i} \\
\Leftrightarrow [1 - G(\phi_i^{*\lambda})] \frac{1}{\delta} \left\{ \frac{\bar{r}_i^\lambda}{\sigma} - f_i(s^\lambda)^{\beta_i} (w^\lambda)^{1-\beta_i} \right\} &= f_{ei}(s^\lambda)^{\beta_i} (w^\lambda)^{1-\beta_i} \\
\Leftrightarrow [1 - G(\phi_i^{*\lambda})] \frac{1}{\delta} \left\{ \frac{1}{\sigma} \left(\frac{\bar{\phi}_i}{\phi_i^{*\lambda}} \right)^{\sigma-1} \cdot r_i^\lambda(\phi_i^{*\lambda}) - f_i(s^\lambda)^{\beta_i} (w^\lambda)^{1-\beta_i} \right\} &= f_{ei}(s^\lambda)^{\beta_i} (w^\lambda)^{1-\beta_i} \\
\Leftrightarrow [1 - G(\phi_i^{*\lambda})] \frac{1}{\delta} \left\{ \left(\frac{\bar{\phi}_i}{\phi_i^{*\lambda}} \right)^{\sigma-1} - 1 \right\} f_i(s^\lambda)^{\beta_i} (w^\lambda)^{1-\beta_i} &= f_{ei}(s^\lambda)^{\beta_i} (w^\lambda)^{1-\beta_i} \\
\Leftrightarrow [1 - G(\phi_i^{*\lambda})] \frac{f_i}{\delta} \left\{ \frac{1}{1 - G(\phi_i^{*\lambda})} \int_{\phi_i^{*\lambda}}^{\infty} \left(\frac{\phi}{\phi_i^{*\lambda}} \right)^{\sigma-1} g(\phi) d\phi - 1 \right\} &= f_{ei}
\end{aligned}$$

and thus Equation (2.12) follows.

2.6.2 Derivation of Equations (2.17) & (2.18)

Here I derive only Equation (2.17) for Home. Equation (2.18) for Foreign is derived analogously.

From Equation (2.13) for the optimal pricing of exported product, the revenue of an individual firm earned from the overseas market (export) is as follows:

$$r_{xi,\omega}^H(\phi) = \alpha_i Y^F \left(\frac{\tau_i (s^H)^{\beta_i} (w^H)^{1-\beta_i}}{\rho P_i^F \phi} \right)^{1-\sigma} \quad (2.A.2)$$

From this and Equation (2.6), the ratio of the revenue earned by an exporter and that earned by a domestic producer in Home country is expressed as follows:

$$\frac{r_{xi}^H(\phi)}{r_i^H(\phi)} = \tau_i^{1-\sigma} \left(\frac{P_i^F}{P_i^H} \right)^{\sigma-1} \left(\frac{Y^F}{Y^H} \right) \quad (2.A.3)$$

Equation (2.A.1) can be modified to the following equation, which implies that the ratio of two firms' productivities is a function of the ratio of the revenues that two firms earn *in the same (domestic) market*:

$$\frac{\phi'}{\phi} = \left(\frac{r_i(\phi')}{r_i(\phi)} \right)^{\frac{1}{\sigma-1}} \quad (2.A.4)$$

Using Equations (2.A.3) and (2.A.4), we can express the ratio of the productivity cutoff for exporting to the cutoff for domestic production in Home as follows:

$$\begin{aligned} \frac{\phi_{xi}^{*H}}{\phi_i^{*H}} &= \left(\frac{r_i^H(\phi_{xi}^{*H})}{r_i^H(\phi_i^{*H})} \right)^{\frac{1}{\sigma-1}} \\ &= \left(\frac{r_{xi}^H(\phi_{xi}^{*H}) \cdot \tau_i^{\sigma-1} \left(\frac{P_i^H}{P_i^F} \right)^{\sigma-1} \left(\frac{Y^H}{Y^F} \right)}{r_i^H(\phi_i^{*H})} \right)^{\frac{1}{\sigma-1}} \\ &= \tau_i \left(\frac{r_{xi}^H(\phi_{xi}^{*H})}{r_i^H(\phi_i^{*H})} \right)^{\frac{1}{\sigma-1}} \cdot \left(\frac{P_i^H}{P_i^F} \right) \left(\frac{Y^H}{Y^F} \right)^{\frac{1}{\sigma-1}} \\ &= \tau_i \left(\frac{f_{xi}}{f_i} \right)^{\frac{1}{\sigma-1}} \left(\frac{P_i^H}{P_i^F} \right) \left(\frac{Y^H}{Y^F} \right)^{\frac{1}{\sigma-1}} \end{aligned}$$

The last equality is from the zero-profit condition in the domestic market (2.6) and the zero-profit condition in the export market (2.15). Equation (2.17) thus follows by defining the right-hand side of the last line of the equation above as Λ_i^H .

2.6.3 Derivation of Equation (2.21)

Note that, from Equation (2.7);

$$\begin{aligned} \bar{\pi}_i^H &= \frac{\bar{r}_i^H}{\sigma} - f_i (s^H)^{\beta_i} (w^H)^{1-\beta_i} \\ &= \frac{1}{\sigma} \left(\frac{\bar{\phi}_i^H}{\phi_i^{*H}} \right)^{\sigma-1} \cdot r_i^{*H} - f_i (s^H)^{\beta_i} (w^H)^{1-\beta_i} \\ &= \left[\left(\frac{\bar{\phi}_i^H}{\phi_i^{*H}} \right)^{\sigma-1} - 1 \right] \cdot f_i (s^H)^{\beta_i} (w^H)^{1-\beta_i} \end{aligned}$$

The second equality is derived using (2.A.1), and the third equality is from Equation (2.15). Analogously;

$$\bar{\pi}_{xi}^H = \left[\left(\frac{\bar{\phi}_{xi}^H}{\phi_{xi}^{*H}} \right)^{\sigma-1} - 1 \right] \cdot f_{xi} (s^H)^{\beta_i} (w^H)^{1-\beta_i} .$$

Substituting these equations for the average profit levels, as well as the average

productivity levels defined as $\bar{\phi}_i^\lambda (\phi_i^{*\lambda}) = \left[\frac{1}{1-G(\phi_i^{*\lambda})} \int_{\phi_i^{*\lambda}}^{\infty} \phi^{\sigma-1} g(\phi) d\phi \right]^{\frac{1}{\sigma-1}}$ and

$\bar{\phi}_{xi}^\lambda (\phi_{xi}^{*\lambda}) = \left[\frac{1}{1-G(\phi_{xi}^{*\lambda})} \int_{\phi_{xi}^{*\lambda}}^{\infty} \phi^{\sigma-1} g(\phi) d\phi \right]^{\frac{1}{\sigma-1}}$, yields:

$$\begin{aligned} & \left[\frac{f_i}{\delta} \int_{\phi_i^{*H}}^{\infty} \left\{ \left(\frac{\phi}{\phi_i^{*H}} \right)^{\sigma-1} - 1 \right\} g(\phi) d\phi \right] \cdot (s^H)^{\beta_i} (w^H)^{1-\beta_i} \\ & + \left[\frac{f_{xi}}{\delta} \int_{\phi_{xi}^{*H}}^{\infty} \left\{ \left(\frac{\phi}{\phi_{xi}^{*H}} \right)^{\sigma-1} - 1 \right\} g(\phi) d\phi \right] \cdot (s^H)^{\beta_i} (w^H)^{1-\beta_i} \\ & = f_{ei} (s^H)^{\beta_i} (w^H)^{1-\beta_i} , \end{aligned}$$

which Equation (2.21) follows by canceling out the term $(s^H)^{\beta_i} (w^H)^{1-\beta_i}$ on the both sides.

2.6.4 Relative Factor Prices under Costly Trade

Here I demonstrate that in equilibrium the relative prices of the two production factors (S and U) is not equalized in our framework of costly trade. The wage for skilled labor relative to the that for unskilled labor will be lower in Home, where skilled labor is

relatively more abundant, than in Foreign; i.e., $\frac{s^H}{w^H} < \frac{s^F}{w^F}$.

First, note that autarky and free trade are the two extreme cases, or limits, of the costly trade. That is, the former is the limit with infinitely large trade costs ($f_{xi} \rightarrow \infty, \tau_i \rightarrow \infty$), and the latter is the limit with no additional costs for trade ($f_{xi} \rightarrow 0, \tau_i \rightarrow 1$). The equilibrium relative factor price under costly trade will fall in the range between those in these two limit cases (i.e., $\frac{s^H}{w^H} < \frac{s^F}{w^F}$). I will thus show how the relative factor prices in the two countries will be in these two limit cases.

Autarky

Since the production function (2.4) has a Cobb-Douglas form, the optimal allocation of the two factors in each industry is such that the total payment to each factor is proportional to the total revenue, which equals the total expenditure, in the industry. That is, WLOG in Home,

$$S_i^H = \left(\frac{\beta_i}{s^H} \right) R_i^H = \left(\frac{\beta_i}{s^H} \right) \alpha_i Y^H \quad (2.A.5)$$

$$L_i^H = \left(\frac{1 - \beta_i}{w^H} \right) R_i^H = \left(\frac{1 - \beta_i}{w^H} \right) \alpha_i Y^H \quad (2.A.6)$$

where R_i^H is the total revenue in Industry i in Home, which is equal to the total industry expenditure in equilibrium. The industry expenditure is proportional to the national income due to the Cobb-Douglas utility function (2.1) (i.e., $R_i^H = \alpha_i Y^H$).

Inelastic supply of each factor equals the sum of that factor allocated to each industry, that is;

$$\bar{S}^H = \sum_i S_i^H = \frac{Y^H}{s^H} \sum_i \alpha_i \beta_i \Leftrightarrow s^H \bar{S}^H = Y^H \sum_i \alpha_i \beta_i \quad (2.A.7)$$

$$\bar{U}^H = \sum_i U_i^H = \frac{Y^H}{w^H} \sum_i \alpha_i (1 - \beta_i) \Leftrightarrow w^H \bar{U}^H = Y^H \sum_i \alpha_i (1 - \beta_i) \quad (2.A.8)$$

Dividing (2.A.7) by (2.A.8) in both sides yields the following equation for Home:

$$\frac{s^H \bar{S}^H}{w^H \bar{U}^H} = \frac{\sum_i \alpha_i \beta_i}{\sum_i \alpha_i (1 - \beta_i)} \Leftrightarrow \frac{s^H}{w^H} = \left(\frac{\sum_i \alpha_i \beta_i}{\sum_i \alpha_i (1 - \beta_i)} \right) \cdot \left(\frac{\bar{U}^H}{\bar{S}^H} \right) \quad (2.A.9)$$

Analogously, for Foreign,

$$\frac{s^H}{w^H} = \left(\frac{\sum_i \alpha_i \beta_i}{\sum_i \alpha_i (1 - \beta_i)} \right) \cdot \left(\frac{\bar{U}^H}{\bar{S}^H} \right) \quad (2.A.9')$$

Since consumers share the identical preference and the Cobb-Douglas cost share of each production factor is common across countries within each industry (i.e., the parameters α_i and β_i are common across countries), the first term of the product in the right-hand side of Equations (2.A.9) and (2.A.9') is the same for both countries. Hence, the relative factor price $\frac{s}{w}$ in each country is determined by the ratio of the two factors

that the country is endowed with, $\frac{\bar{S}}{\bar{U}}$. Since $\frac{\bar{S}^H}{\bar{U}^H} > \frac{\bar{S}^F}{\bar{U}^F}$ by assumption, (2.A.9) and

(2.A.9') imply that $\frac{s^H}{w^H} < \frac{s^F}{w^F}$ in the autarky equilibrium.

Free Trade

Here I focus on the case with FPE. We can identify the equilibrium relative factor price with FPE by solving for the problem of the integrated world economy, which is characterized by Equations (2.A.5) through (2.A.9) in the autarky case described above, but omitting the country script. The common relative factor price $\frac{s}{w}$ is determined by the

world relative factor supply $\frac{\bar{S}}{\bar{U}} = \frac{\bar{S}^H + \bar{S}^F}{\bar{U}^H + \bar{U}^F}$. Hence, in the free-trade equilibrium with

$$\text{FPE, } \frac{s^H}{w^H} = \frac{s^F}{w^F}. \quad 58$$

2.6.5 Proof of Proposition

WLOG, in this proof I focus on the skill-abundant Home country.

The probability of a domestic producer being an exporter, χ_i^H , is determined by the ratio between the two productivity cutoffs, the one for domestic production ϕ_i^{*H} and the one for exporting ϕ_{xi}^{*H} . Since, as in Equation (2.17), the ratio between these two

productivity cutoffs $\left(\frac{\phi_{xi}^{*H}}{\phi_i^{*H}} = \Lambda_i^H = \tau_i \left(\frac{P_i^H}{P_i^F} \right) \left(\frac{Y^H}{Y^F} \cdot \frac{f_{xi}}{f_i} \right)^{\frac{1}{\sigma-1}} \right)$ depends upon the Home and

Foreign industry price indexes (P_i^H and P_i^F), I take the following proof strategy:

- (i) I first show that the relative industry price index (Home to Foreign) is smaller for an industry in which the skill-abundant Home has stronger comparative

advantage (i.e., $\frac{P_i^H}{P_i^F} < \frac{P_j^H}{P_j^F}$ for $i \neq j$ such that $\beta_i > \beta_j$);

- (ii) I next demonstrate that (i) implies that the ratio between the two productivity cutoffs is smaller in an industry in which the country has stronger comparative

advantage (i.e., $\frac{\phi_{xi}^{*H}}{\phi_i^{*H}} < \frac{\phi_{xj}^{*H}}{\phi_j^{*H}}$ for $i \neq j$ such that $\beta_i > \beta_j$);

and

⁵⁸ We can show that there exist the optimal allocations of the two factors to each industry in each country with FPE, although the allocations are not unique (Melvin's indeterminacy). The authors can provide the proof upon request.

(iii) I then use the results in (ii) and the relationship between the relative factor prices in the two countries in equilibrium, which has been derived in Subsection 2.6.4, to compare across industries the probability of the Home active firms to be an exporter, χ_i^H and χ_j^H .

(i) Relative industry price index in two countries:

To demonstrate that $\frac{P_i^H}{P_i^F} < \frac{P_j^H}{P_j^F}$ for $i \neq j$ such that $\beta_i > \beta_j$, here I apply a similar

logic to the one that I have used in above-mentioned 2.6.4 to show the relative factor

prices in the costly-trade equilibrium ($\frac{S^H}{W^H} < \frac{S^F}{W^F}$). The relative industry price index in

the costly-trade equilibrium is as follows:

$$\frac{P_i^H}{P_i^F} = \frac{[M_i^H (p_i^H (\bar{\phi}_i^H))^{1-\sigma} + \chi_i^F \cdot M_i^F \cdot \tau_i^{1-\sigma} (p_i^F (\bar{\phi}_{xi}^F))^{1-\sigma}]^{\frac{1}{1-\sigma}}}{[M_i^F (p_i^F (\bar{\phi}_i^F))^{1-\sigma} + \chi_i^H \cdot M_i^H \cdot \tau_i^{1-\sigma} (p_i^H (\bar{\phi}_{xi}^H))^{1-\sigma}]^{\frac{1}{1-\sigma}}} \quad (2.A.10)$$

Since the autarky equilibrium and the free-trade FPE equilibrium are the two extreme or limit cases, the relative price index in the costly trade equilibrium falls between the one in the autarky equilibrium and the one in the free-trade FPE equilibrium.

In autarky, which is characterized by $\tau_i = \infty$ and $f_{xi} = \infty$, no firms will be exporters ($\chi_i = 0$ in each country). Therefore, Equation (2.A.10) is now as follows:

$$\frac{P_i^H}{P_i^F} = \frac{(M_i^H)^{\frac{1}{1-\sigma}} \cdot p_i^H (\bar{\phi}_i^H)}{(M_i^F)^{\frac{1}{1-\sigma}} \cdot p_i^F (\bar{\phi}_i^F)} = \left(\frac{M_i^H}{M_i^F} \right)^{\frac{1}{1-\sigma}} \left(\frac{p_i^H (\bar{\phi}_i^H)}{p_i^F (\bar{\phi}_i^F)} \right) \quad (2.A.11)$$

Since $M_i = R_i / r_i (\bar{\phi}_i)$ and $R_i = \alpha_i Y$ for each country in the autarky equilibrium,

Equation (2.A.11) yields the following equation;

$$\frac{P_i^H}{P_i^F} = \left(\frac{Y^H}{Y^F} \right)^{\frac{1}{1-\sigma}} \left(\frac{r_i^F(\bar{\phi}_i^F)}{r_i^H(\bar{\phi}_i^H)} \right)^{\frac{1}{1-\sigma}} \left(\frac{p_i^H(\bar{\phi}_i^H)}{p_i^F(\bar{\phi}_i^F)} \right) \quad (2.A.12)$$

Note that the optimal pricing Equation (2.5) implies that the ratio of the prices charged by two firms with different productivity *in the same market* can be expressed as the ratio of the two productivities, i.e.;

$$p_i(\phi') = \left(\frac{\phi'}{\phi} \right) \cdot p_i(\phi)$$

Using this equation and Equation (A.1), as well as the optimal pricing (2.5) and the zero-profit condition (2.9), Equation (2.A.12) can be expressed as follows:

$$\frac{P_i^H}{P_i^F} = \left(\frac{Y^H}{Y^F} \right)^{\frac{1}{1-\sigma}} \frac{\left(\frac{\bar{\phi}_i^F}{\phi_i^{*F}} \right)^{\sigma-1} (s^F)^{\beta_i} (w^F)^{1-\beta_i}}{\left(\frac{\bar{\phi}_i^H}{\phi_i^{*H}} \right)^{\sigma-1} (s^H)^{\beta_i} (w^H)^{1-\beta_i}} \left(\frac{\bar{\phi}_i^H}{\phi_i^{*H}} \right)^{\sigma-1} \frac{(s^H)^{\beta_i} (w^H)^{1-\beta_i}}{\phi_i^{*H}}}{\left(\frac{\bar{\phi}_i^F}{\phi_i^{*F}} \right)^{\sigma-1} (s^F)^{\beta_i} (w^F)^{1-\beta_i}} \frac{\phi_i^{*F}}{\phi_i^{*F}} \quad (2.A.13)$$

Note that the productivity cutoff for each country, ϕ_i^* , is determined by the free-entry condition (2.12), which is common for the two countries. Therefore, $\phi_i^{*H} = \phi_i^{*F}$, and accordingly, $\bar{\phi}_i^H = \bar{\phi}_i^F$ since the productivity distribution is also common across

countries. These imply that $\frac{\bar{\phi}_i^H}{\phi_i^{*H}} = \frac{\bar{\phi}_i^F}{\phi_i^{*F}}$. Hence, from Equation (2.A.13) we obtain the

following:

$$\begin{aligned} \frac{P_i^H}{P_i^F} &= \left(\frac{Y^H}{Y^F} \right)^{\frac{1}{1-\sigma}} \cdot \left\{ \left(\frac{s^F}{s^H} \right)^{\beta_i} \left(\frac{w^F}{w^H} \right)^{1-\beta_i} \right\}^{\frac{1}{1-\sigma}} \\ \Leftrightarrow \frac{P_i^H}{P_i^F} &= \left(\frac{Y^H}{Y^F} \right)^{\frac{1}{1-\sigma}} \cdot \left\{ \left(\frac{s^H / w^H}{s^F / w^F} \right)^{\beta_i} \left(\frac{w^H}{w^F} \right) \right\}^{\frac{\sigma}{\sigma-1}} \end{aligned} \quad (2.A.14)$$

Analogously;

$$\frac{P_j^H}{P_j^F} = \left(\frac{Y^H}{Y^F} \right)^{\frac{1}{1-\sigma}} \cdot \left\{ \left(\frac{s^H / w^H}{s^F / w^F} \right)^{\beta_j} \left(\frac{w^H}{w^F} \right) \right\}^{\frac{\sigma}{\sigma-1}} \quad (2.A.15)$$

It has been shown in Subsection 2.6.4 that $\frac{s^H}{w^H} < \frac{s^F}{w^F}$ in autarky. Therefore, since $\beta_i > \beta_j$,

it follows that $\frac{P_i^H}{P_i^F} < \frac{P_j^H}{P_j^F}$ in the autarky equilibrium.

Next, consider the free-trade equilibrium, which is characterized by $\tau_i = 1$ and $f_{xi} = 0$. Since all domestically active firms will export, $\chi_i^\lambda = 1$ in each country λ . Furthermore, with FPE, firms in the two countries will charge the same price for both domestic sales and export if their productivities are the same, $p_i^H(\bar{\phi}_i^H) = p_{xi}^H(\bar{\phi}_{xi}^H) = p_i^F(\bar{\phi}_i^F) = p_{xi}^F(\bar{\phi}_{xi}^F)$ (the average productivity is the same across countries since it is determined by the common free-entry condition (2.12)). Hence, Equation (2.A.10) yields:

$$\frac{P_i^H}{P_i^F} = \frac{(M_i^H + M_i^F)^{\frac{1}{1-\sigma}}}{(M_i^H + M_i^F)^{\frac{1}{1-\sigma}}} = 1$$

Therefore, under costly trade, which is the intermediate case of the two extremes shown

above, $\frac{P_i^H}{P_i^F} < \frac{P_j^H}{P_j^F}$ for $i \neq j$ such that $\beta_i > \beta_j$ in equilibrium.

(ii) Ratio between the export cutoff productivity and the domestic production cutoff productivity:

From Equations (2.17) and (2.18);

$$\frac{\phi_{xi}^{*H}}{\phi_i^{*H}} = \Lambda_i^H = \tau_i \left(\frac{P_i^H}{P_i^F} \right) \left(\frac{Y^H}{Y^F} \cdot \frac{f_{xi}}{f_i} \right)^{\frac{1}{\sigma-1}}$$

$$\frac{\phi_{xj}^{*H}}{\phi_j^{*H}} = \Lambda_j^H = \tau_j \left(\frac{P_j^H}{P_j^F} \right) \left(\frac{Y^H}{Y^F} \cdot \frac{f_{xj}}{f_j} \right)^{\frac{1}{\sigma-1}}$$

Suppose $\tau_i = \tau_j = \tau$, $f_i = f_j = f$, and $f_{xi} = f_{xj} = f_x$. Then, from the result in (i) above, these two equations imply that:

$$\text{If } \beta_i > \beta_j, \text{ then } (1 <) \Lambda_i^H < \Lambda_j^H \Leftrightarrow (1 <) \frac{\phi_{xi}^{*H}}{\phi_i^{*H}} < \frac{\phi_{xj}^{*H}}{\phi_j^{*H}}. \quad (2.A.16)$$

(iii) Cross-industry comparison of the probability of exporting:

According to Assumption, $G(\phi_i) = 1 - \left(\frac{\phi_i}{\phi_i} \right)^k$, $k > 2\sigma$. Then,

$$\chi_i^H = \frac{1 - G(\phi_{xi}^{*H})}{1 - G(\phi_i^{*H})} = \frac{(\phi_i / \phi_{xi}^{*H})^k}{(\phi_i / \phi_i^{*H})^k} = \left(\frac{\phi_i^{*H}}{\phi_{xi}^{*H}} \right)^k = \left(\frac{\phi_{xi}^{*H}}{\phi_i^{*H}} \right)^{-k}, \text{ and}$$

$$\chi_j^H = \frac{1 - G(\phi_{xj}^{*H})}{1 - G(\phi_j^{*H})} = \left(\frac{\phi_{xj}^{*H}}{\phi_j^{*H}} \right)^{-k}.$$

Therefore, from (2.A.16),

$$\text{If } \beta_i > \beta_j, \text{ then } \frac{\phi_{xi}^{*H}}{\phi_i^{*H}} < \frac{\phi_{xj}^{*H}}{\phi_j^{*H}} \Leftrightarrow \left(\frac{\phi_{xi}^{*H}}{\phi_i^{*H}} \right)^{-k} > \left(\frac{\phi_{xj}^{*H}}{\phi_j^{*H}} \right)^{-k} \Leftrightarrow \chi_i^H > \chi_j^H.$$

Since this holds for any industry pair i and j ($i, j = 1, 2, \dots, N$) that satisfies $\beta_i > \beta_j$, the Proposition thus follows. ■

2.7 Appendix B

This appendix examines the potential impacts of sample truncation in manufacturing census data on the result of the empirical analysis in this paper. As described in Section 2.3, the manufacturing census of each country omits firms whose employment is less than the threshold level (ten employees for the Chile and Colombia,

and 20 employees for India; the U.S. census also excludes small firms: see the General Summary of the U.S. Census of Manufactures for the details). The omission of small firms might potentially cause the overestimation of exporter fraction, since the number of domestic firms in the data, which is the denominator of the fraction, does not count such small firms.⁵⁹ This overestimation of the exporter fraction might occur for all industries, but if the degree of the overestimation would differ systematically in relation to the factor intensity (or more specifically, the skill intensity) of the industries, it could result in the biased estimation of a correlation between the exporter fractions and the industry skill intensities. In what follows, I examine whether such estimation bias could crucially affect, or mislead, the result of the empirical findings presented in Section 2.4.

From Theory

The theoretical model presented in Section 2.2 suggests that the minimum productivity level required for domestic production is lower in comparative-disadvantage industries and higher in comparative-advantage industries. This implies that comparative-*disadvantage* industries contain more small-size firms (in terms of employment) compared to comparative-advantage industries.⁶⁰ Therefore, if the sample of firms is truncated at the same threshold employment level for all industries, the number of small domestic firms omitted from the sample is larger in comparative-*disadvantage* industries, and thus the exporter fraction is more overestimated in the sample for comparative-

⁵⁹ The omission of small firms might also affect the numerator of the fraction, if some of these small firms export. However, Bernard and Jensen (1995) and other studies have found that exporters are significantly larger than non-exporters in terms of employment size, and it is expected that, even though the small firms include exporters, the fraction of exporters in these small firms is (significantly) smaller than the exporter fraction in all firms. The exporter fraction observed in the census data might thus still overestimate the fraction in the population.

⁶⁰ The present model implies that a firm with a lower productivity level is smaller in both sales size and employment size. In addition, with the assumption of a Pareto distribution for firms' productivity, both sales sizes and employment sizes of firms are also distributed in a Pareto distribution. The proof can be provided upon request.

disadvantage industries. This pattern of overestimation implies that the predicted relationship between the exporter fraction and comparative advantage is weaker in the sample than in the population. In other words, for a more skill-abundant country (e.g., the United States) the correlation between the exporter fractions and industry skill intensities in the sample is estimated to be *less positive* than how it should be in the population, and for a less skill-abundant country (e.g., India) the correlation in the sample is estimated to be *less negative* than how it should be in the population. This suggests that the quasi-Heckscher-Ohlin pattern of the exporter fraction, which is found to be significant in the empirical analysis in Section 2.4, should be even stronger in the population of firms that includes small firms.

From Data

Recall Equation (2.23) for individual country analysis:

$$ex_share_i = \gamma + \theta \cdot skill_i + \varepsilon_i$$

Since ex_share_i on the left-hand side is the ratio of the number of exporters in a sample to the number of domestic producers in the sample, this equation can also be expressed as follows:

$$\left(\frac{EX_i}{DOM_i} \right) \cdot \left(\frac{1 - ex_i / EX_i}{1 - dom_i / DOM_i} \right) = \gamma + \theta \cdot skill_i + \varepsilon_i \quad (2.B.1)$$

where EX_i , ex_i , DOM_i , and dom_i denote the numbers of exporters in a firm population, exporters in small firms omitted from the sample, domestic producers in the population, and domestic producers in the small firms in Industry i , respectively. Hence, if the second element of the left-hand side of Equation (2.B.1), $\frac{1 - ex_i / EX_i}{1 - dom_i / DOM_i}$, would be positively

(negatively) correlated to the industry skill intensity $skill_i$, the estimation of Equation

(2.23) would provide a positive (negative) estimate for the coefficient θ even though the exporter fraction in the population is not correlated to the skill intensity at all. Or, at least, a regression would overestimate or underestimate the correlation in the population.⁶¹

To check whether such spurious correlation or the estimation bias is crucial in the present empirical analysis, I examine the cross-industry correlation for the term

$\frac{1 - ex_i / EX_i}{1 - dom_i / DOM_i}$ and the skill intensity term $skill_i$ for Chile and Colombia, for which

data are available at the firm level.⁶² Since in the data I cannot observe omitted small firms with less than ten employees, I measure ex_i and dom_i using a group of the smallest firms in the data for each country, i.e., firms with (more than 10 and) less than 20 workers; and measure the population counterparts EX_i and DOM_i from all firms included in the data. The variables are for 25 three-digit ISIC industries and of the average over the years 1990 through 96 for Chile, and 1981-91 for Colombia.

The estimated correlation between $\frac{1 - ex_i / EX_i}{1 - dom_i / DOM_i}$ and $skill_i$ for Chile is -0.255

with the p-value of 0.219. This implies that the coefficient θ estimated in Section 2.4, which is positive and significant, might be underestimated (but not significantly) in the sample, and therefore the population coefficient could be even more positive.⁶³ For Colombia, the estimated correlation between the two terms is almost zero (the correlation coefficient is -0.085 with the p-value of 0.686). These results suggest that sample

⁶¹ The empirical result could be misled if overestimation would be the case for a positive coefficient estimate $\hat{\theta}$, or if underestimate would be the case for a negative coefficient estimate.

⁶² The data for India and the United States do not have sufficient firm-level information for the same examination.

⁶³ Note that this is consistent with what the theoretical model suggests.

truncation should not cause estimation bias and thus the empirical finding in Section 2.4 should be valid.

2.8 Appendix C

This appendix performs an exercise to examine from a pure empirical perspective whether the relative factor abundance of a country and the relative factor intensity of an industry have a significant influence on the fraction of exporters among domestic firms in that industry in that country. For this purpose, I estimate a variant of Equation (2.26) using a larger pooled dataset that is composed as follows:

- An observation is for one country, one industry, and one year. That is, data for seven years (1990-1996) are used for Chile, and data for eleven years (1981-1991) are used for Colombia. The data for India and the United States are for a single year (1998 for India, 1992 for the U.S.).
- Industry skill intensity is allowed to vary across countries and years (i.e., a standard Heckscher-Ohlin assumption is relaxed).⁶⁴
- Factor abundance in each country is assumed to be invariant over periods, and measured using the data provided by Hall and Jones (1999).
- For every country, the manufacturing industries are classified according to the two-digit 1987 U.S. SIC; i.e., there are 17 observations⁶⁵ for each country in each year.

This way I can increase the number of observations in the dataset to 340.

⁶⁴ Schott (2003) shows that in reality different countries employ different factor mixes for production in the same industry classified according to the three-digit ISIC.

⁶⁵ Tobacco, petroleum and coal, and miscellaneous industry categories are excluded.

The equation is estimated by OLS including various combinations of dummies for one or more groups (dummies for countries, industries, and/or years). Hence, the regression equation is as follows:

$$ex_share_{ict} = \theta_1 \cdot skill_{ict} + \theta_2 \cdot skill_{ict} \cdot \log(S/U)_c + (\eta_i + \mu_c + \nu_t + \varepsilon_{ic}) \quad (2.C.1)$$

where η_i , μ_c , and ν_t denote industry-specific, country-specific, and year-specific intercepts, respectively.

The results of the regressions are shown in Tables 2.C.1 through 2.C.8. With any combination of the dummies (or with no dummies), the estimate of the coefficient of interest, θ_2 , is positive and significant at the 5% level or better. Two exceptions are (i) when only country-specific dummies are included, and (ii) both country-specific dummies and year-specific dummies are simultaneously included. However, even in these cases, the coefficient estimate is positive and large (above 0.8) and the p-value of the estimate is not very far from 0.1. These results indicate that the data strongly suggest an empirical relationship between comparative advantage and exporter selection.

2.9 Appendix D

This appendix describes how the alternative measure of industry skill intensity (or the skilled-worker cost shares of an industry) that is used in Subsection 2.4.3 is computed following Morrow (2008). For the details of Morrow's calculation, see his paper.

Morrow obtains the data on wages, educational attainment, and ages of workers from the March U.S. Current Population Survey for the years 1988-92. These data are used for a Mincerian (log of) wage regression. He groups the educational attainment of workers into the following four categories: 0-11 grades of school completed, 12th grade

completed, 1-3 years of college, and 4 or more years of college. The equation for his Mincerian regression is the following:

$$\log(w_{it}) = \alpha_0 + \alpha_1 \cdot age_{it} + \alpha_2 \cdot age_{it}^2 + \beta_{12th} \cdot D_{12th} + \beta_{college1-3} \cdot D_{college1-3} + \beta_{college4+} \cdot D_{college4+} + \gamma_t + \varepsilon_{it}$$

where i indexes a worker, t indexes time. $D_{12th} = 1$ if the worker has completed 12th grade, $D_{college1-3} = 1$ if the worker has attained 1-3 years of college education, and $D_{college4+} = 1$ if the worker has attained 4 or more years of college education. γ_t is time-specific intercepts. The coefficient for each education-group dummy (β) indicates the wage for a worker in that education group relative to the wage for a worker in the benchmark (the lowest) education group in logarithm, when the two workers differ only in their educational attainment. That is; $w_{edu} / w_0 = \exp(\beta_{edu})$ where edu is some education group and 0 denotes the benchmark education group. He reports the following coefficient estimates for the three education-group dummies:

β_{12th}	$\beta_{college1-3}$	$\beta_{college4+}$
0.2939	0.4755	0.8128

He also reports the share of workers in each education category, as well as the total number of workers reported in the Current Population Survey, for the 25 manufacturing industries classified according to the three-digit ISIC. Using these numbers, I calculate the share of workers in each education category for the 17 manufacturing industries classified according to the two-digit U.S. SIC, using the concordance presented in Table 2.9. The obtained shares of workers in the three education categories in the 17 industries are shown in the table below.

Shares of Workers in Different Education Categories in Total Employment:
for 17 Two-digit 1987 U.S. SIC Manufacturing Industries

SIC	Total No. Workers	12th Grade Completion or Less	1-3 Years of College	4+ Years of College
20	6,427	0.714	0.174	0.111
22	3,059	0.798	0.127	0.075
23	3,369	0.834	0.110	0.056
24	1,891	0.757	0.162	0.081
25	2,118	0.777	0.145	0.077
26	2,358	0.652	0.212	0.136
27	6,132	0.501	0.246	0.253
28	3,773	0.447	0.230	0.323
30	3,068	0.691	0.189	0.121
31	601	0.800	0.110	0.090
32	1,944	0.703	0.176	0.121
33	2,400	0.691	0.205	0.105
34	3,911	0.688	0.201	0.110
35	3,179	0.624	0.243	0.133
36	10,699	0.501	0.243	0.256
37	7,501	0.553	0.251	0.196
38	2,225	0.493	0.246	0.261

Source: Author's calculation from Morrow (2008).

From these data, I calculate the cost share of skilled workers in each two-digit U.S. SIC manufacturing industry. I define skilled labor by workers with one or more years of college education (i.e., workers in the highest two education categories). While Morrow's benchmark education category is 0-11 grades of schooling, he does not report the share of workers in this category. Instead, he reports the share of workers with high school education or less, which combines the lowest two education categories of workers (0-11 grades and 12th grade). Hence, I use the group of workers with 12th or lower grade of education as my benchmark category, and compute the skilled-labor cost share in each industry as follows:

$$skill_i = \frac{(wage_{college4+} / wage_{0-12th}) \cdot R_{college4+} + (wage_{college1-3} / wage_{0-12th}) \cdot R_{college1-3}}{(wage_{college4+} / wage_{0-12th}) \cdot R_{college4+} + (wage_{college1-3} / wage_{0-12th}) \cdot R_{college1-3} + R_{0-12th}}$$

$$= \frac{\exp(\beta_{college4+} - \beta_{12th}) \cdot R_{college4+} + \exp(\beta_{college1-3} - \beta_{12th}) \cdot R_{college1-3}}{\exp(\beta_{college4+} - \beta_{12th}) \cdot R_{college4+} + \exp(\beta_{college1-3} - \beta_{12th}) \cdot R_{college1-3} + R_{0-12th}}$$

where R_{0-12th} is the share of workers with 12th or less grade, $R_{college1-3}$ is the share of workers with 1-3 years of college, and $R_{college4+}$ is the share of workers with 4 or more years of college. The calculated skilled-labor cost shares in the 17 manufacturing industries are reported in Table 2.12.

Table 2.1: Fractions of Exporting Firms in Manufacturing Industries (2-digit U.S. SIC): Chile, Colombia, India, and the United States

SIC	Industry	Chile (avg. 1990-96)			Colombia (avg. 1981-91)			India (FY 1997/98)			United States (1992)		
		# active firms (a)	# exporters (b)	ratio (a)/(b)	# active firms (a)	# exporters (b)	ratio (a)/(b)	# active firms (a)	# exporters (b)	ratio (a)/(b)	# active firms (a)	# exporters (b)	ratio (a)/(b)
20	Food and kindred products	1,523	301	19.8%	1,332	98	7.4%	4,690	353	7.5%	20,641	4,563	22.1%
21	Tobacco products	3	2	69.6%	15	3	23.5%	600	27	4.5%	114	67	58.8%
22	Textile mill products	358	77	21.6%	466	65	13.9%	3,489	615	17.6%	5,868	1,613	27.5%
23	Apparel and other textile products	322	44	13.5%	983	87	8.9%	644	349	54.2%	22,935	2,801	12.2%
24	Lumber and wood products	355	83	23.3%	179	12	6.8%	718	16	2.2%	35,245	3,266	9.3%
25	Furniture and fixtures	144	19	13.1%	202	9	4.5%	111	3	2.7%	11,620	2,058	17.7%
26	Paper and allied products	74	27	36.3%	142	21	14.6%	680	54	7.9%	6,401	1,840	28.7%
27	Printing and publishing	210	22	10.3%	358	31	8.7%	722	29	4.0%	65,349	4,495	6.9%
28	Chemicals and allied products	257	116	45.0%	421	102	24.3%	2,314	513	22.2%	11,982	5,502	45.9%
29	Petroleum and coal products	22	11	50.0%	27	4	14.4%	262	14	5.3%	1,961	416	21.2%
30	Rubber and miscellaneous plastics products	306	76	24.9%	390	62	16.0%	1,344	158	11.8%	15,819	5,806	36.7%
31	Leather and leather products	209	45	21.8%	356	68	19.2%	373	170	45.6%	2,032	677	33.3%
32	Stone, clay, and glass products	188	28	15.0%	393	42	10.6%	2,239	197	8.8%	16,001	2,116	13.2%
33	Primary metal industries	65	31	48.0%	93	11	11.9%	1,534	195	12.7%	6,500	2,139	32.9%
34	Fabricated metal products	405	57	14.1%	566	68	11.9%	1,312	111	8.5%	36,360	8,211	22.6%
35	Industrial machinery and equipment	172	28	16.6%	316	59	18.7%	1,933	341	17.6%	53,849	13,990	26.0%
36	Electronic and other electric equipment	59	17	28.2%	198	37	18.7%	1,425	246	17.3%	16,890	8,306	49.2%
37	Transportation equipment	105	18	17.1%	226	27	12.1%	1,068	192	18.0%	11,249	3,711	33.0%
38	Instruments and related products	19	6	33.6%	65	16	25.1%	284	69	24.3%	11,331	6,716	59.3%
39	Misc. manufacturing	63	11	16.7%	157	35	22.0%	385	90	23.4%	16,998	4,260	25.1%
Total		4,859	1,019	21.0%	6,886	858	12.5%	26,127	3,742	14.3%	369,145	82,553	22.4%

Sources: Author's calculation from the data of manufacturing census of the countries.

Table 2.2a: Human Capital to Labor Ratio (H/L)

	H/L (in logarithm)	Percentile	Rank
Chile	0.783	73	37
Colombia	0.590	36	61
India	0.409	24	89
United States	1.198	95	2
Mean	0.671		
Min	0.122		
25%	0.414		
Median	0.739		
75%	0.791		
Max	1.215		
No. of countries			127

Notes: The summary statistics, ranks and percentile ranks are among 127 countries that are covered in the source data. All the numbers are based on the statistics weighted by the amount of labor (L) reported in the source data.

Source: Hall and Jones (1999)

Table 2.2b: Tertiary Education Completion in Total Population over Age 15

	1985		1990		1995	
	[%]	(rank)	[%]	(rank)	[%]	(rank)
Chile	4.1%	[81] (24)	5.0%	[81] (29)	6.1%	[80] (29)
Colombia	2.3%	[70] (45)	3.2%	[72] (44)	3.7%	[70] (47)
India	1.2%	[45] (60)	1.7%	[45] (58)	2.0%	[44] (60)
United States	15.4%	[94] (1)	21.8%	[94] (1)	22.5%	[94] (1)
Mean	2.86%		3.79%		4.35%	
Min	0.0%		0.0%		0.1%	
25%	0.7%		1.4%		1.8%	
Median	1.2%		1.7%		2.0%	
75%	2.6%		3.5%		4.0%	
Max	15.4%		21.8%		22.5%	
No. of Countries		103		103		103

Notes: The summary statistics, ranks and percentile ranks are among 103 countries that are covered in the source data. All the numbers are based on the statistics weighted by the population over age 15.

Sources: Barro and Lee (2000) for educational attainment; World Bank (2006) for population.

Table 2.3: Individual Country Regression: Chile

Dependent variable = <i>exporter share_i</i>	
Coef. for <i>skill_i</i> (θ)	0.634** (0.299)
Intercept (γ)	0.087 (0.092)
No. of observations	25
R ²	0.16

Notes: The observations are for 25 three-digit ISIC industries excluding 314 (tobacco products), 353 (petroleum refineries), 354 (misc. petroleum and coal products), and 390 (other manufacturing products).

The standard errors are in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.4: Individual Country Regression: Colombia

Dependent variable = <i>exporter share_i</i>	
Coef. for <i>skill_i</i> (θ)	-0.091 (0.217)
Intercept (γ)	0.180** (0.068)
No. of observations	25
R ²	0.01

Notes: The observations are for 25 three-digit ISIC industries excluding 314 (tobacco products), 353 (petroleum refineries), 354 (misc. petroleum and coal products), and 390 (other manufacturing products).

The standard errors are in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.5: Individual Country Regression: India

Dependent variable = $exporter\ share_i$	
Coef. for $skill_i (\theta)$	-0.865* (0.448)
Intercept (γ)	0.388*** (0.121)
No. of observations	25
R^2	0.12

Notes: The observations are for 25 three-digit ISIC industries excluding 314 (tobacco products), 353 (petroleum refineries), 354 (misc. petroleum and coal products), and 390 (other manufacturing products).

The standard errors are in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.6: Comparison of Estimated Coefficients for Industry Skill Intensity among Three Countries (with 25 three-digit ISIC manufacturing industries)

Dependent variable = $exporter\ share_i$	Chile	Colombia	India
Coef. for $skill_i (\theta)$	0.634** (0.299)	-0.091 (0.217)	-0.865* (0.448)
$\log(H/L)$	0.783	0.590	0.409

Notes: The coefficient for each country is estimated with 25 three-digit ISIC industries excluding 314 (tobacco products), 353 (petroleum refineries), 354 (misc. petroleum and coal products), and 390 (other manufacturing products). The standard errors are in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

The log of human capital to labor ratio is from Hall and Jones (1999).

Table 2.7: Individual Country Regression: The United States

Dependent variable = $exporter\ share_i$	
Coef. for $skill_i (\theta)$	0.587* (0.303)
Intercept (γ)	0.113 (0.092)
No. of observations	17
R^2	0.20

Notes: The observations are for 17 two-digit 1987 U.S. SIC industries excluding 21 (tobacco products), 29 (petroleum and coal products), and 39 (misc. manufacturing industries).
The standard errors are in parentheses.
* indicates that the coefficient estimate is significant at the 10% level;
** indicates that the coefficient estimate is significant at the 5% level; and
*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.8: Comparison of Estimated Coefficients for Industry Skill Intensity among Four Countries (with 17 two-digit U.S. SIC manufacturing industries)

Dependent variable = $exporter\ share_i$	USA	Chile	Colombia	India
Coef. for $skill_i (\theta)$	0.587* (0.303)	0.534 (0.329)	0.248 (0.188)	-0.728 (0.579)
$\log(H/L)$	1.198	0.783	0.590	0.409

Notes: The coefficient for each country is estimated with 17 two-digit U.S. SIC industries excluding 21 (tobacco products), 29 (petroleum and coal products), and 39 (misc. manufacturing industries).
The standard errors are in parentheses.
* indicates that the coefficient estimate is significant at the 10% level;
** indicates that the coefficient estimate is significant at the 5% level; and
*** indicates that the coefficient estimate is significant at the 1% level.
The log of human capital to labor ratio is from Hall and Jones (1999).

Table 2.9: Concordance from Three-digit ISIC (revision 2) to Two-digit 1987 U.S. SIC

3-digit ISIC	Industry Description	2-digit usSIC	Industry Description
311	Food products	20	Food and kindred products
312	Animal feeds, etc		
313	Beverages		
(314)	(Tobacco products)	(21)	(Tobacco products)
321	Textiles	22	Textile mill products
322	Wearing apparel, except footwear	23	Apparel and other textile products
323	Leather products	31	Leather and leather products
324	Footwear, except rubber or plastic		
331	Wood products, except furniture		
332	Manufacture of furniture and fixtures, except primarily of metal	24	Lumber and wood products
		25	Furniture and fixtures
341	Paper and products	26	Paper and allied products
342	Printing and publishing	27	Printing and publishing
351	Industrial chemicals	28	Chemicals and allied products
352	Other chemicals		
(353)	(Petroleum refineries)		
(354)	(Misc. petroleum and coal products)	(29)	(Petroleum and coal products)
355	Rubber products	30	Rubber and misc. plastic products
356	Plastic products		
361	Manufacture of pottery, china and earthenware	32	Stone, clay, and glass products
362	Glass and products		
369	Other non-metallic mineral products		
371	Iron and steel	33	Primary metal industries
372	Non-ferrous metals		
381	Fabricated metal products	34	Fabricated metal products
382	Machinery, except electrical	35	Industrial machinery and equipment
383	Machinery electric	36	Electronic and other electric equipment
384	Transport equipment	37	Transportation equipment
385	Professional and scientific equipment	38	Instruments and related products
(390)	(Other manufactured products)	(39)	(Misc. manufacturing industries)

Note: Industries in parentheses are excluded for the estimation of the regression equations.

Source: Author's mapping.

Table 2.10: Cost Share of Skilled Labor ($skill_i$) in 17 Two-digit U.S. SIC Manufacturing Industries

SIC	Industry	$skill_i$	SIC	Industry	$skill_i$
20	Food and kindred products	0.365	31	Leather and leather products	0.320
22	Textile mill products	0.246	32	Stone, clay, and glass products	0.311
23	Apparel and otr. textile products	0.297	33	Primary metal industries	0.291
24	Lumber and wood products	0.265	34	Fabricated metal products	0.369
25	Furniture and fixtures	0.337	35	Ind. machinery and equipment	0.490
26	Paper and allied products	0.314	36	Electronic and otr. elec. equip.	0.524
27	Printing and publishing	0.551	37	Transportation equipment	0.417
28	Chemicals and allied products	0.526	38	Instruments and related products	0.630
30	Rubber and misc. plastic prod.	0.352			

Notes: The cost share of skilled labor is measured as the non-production workers wages as the share in the annual payroll in each industry.

Manufacturing industries are classified according to the twp-digit 1987 U.S. SIC. The following categories are excluded: 21 (tobacco products), 29 (petroleum and coal products), and 39 (misc. manufacturing industries).

Source: 1992 U.S. Census of Manufactures.

Table 2.11: Regression for Four Countries and 17 Two-digit U.S. SIC Manufacturing Industries

Dependent variable = $exporter\ share_{ic}$	
Coef. for $skill_i * \log(S/U)_c (\theta_2)$	0.762** (0.360)
No. of observations	68
R^2	0.64

Notes: The four countries are Chile, Colombia, India, and the United States.

The following categories are excluded from the estimation: 21 (tobacco products), 29 (petroleum and coal products), and 39 (misc. manufacturing industries).

Country-specific dummies and industry-specific dummies are both included. Robust standard errors are shown in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.12: Alternative Skilled-labor Cost Share in 17 Two-digit U.S. SIC Manufacturing Industries: Wage Share of Workers with One or More Years of College Education

SIC	Industry	$skill_i$	SIC	Industry	$skill_i$
20	Food and kindred products	0.357	31	Leather and leather products	0.261
22	Textile mill products	0.259	32	Stone, clay, and glass products	0.371
23	Apparel and otr. textile products	0.213	33	Primary metal industries	0.379
24	Lumber and wood products	0.304	34	Fabricated metal products	0.383
25	Furniture and fixtures	0.281	35	Ind. machinery and equipment	0.452
26	Paper and allied products	0.426	36	Electronic and otr. elec. equip.	0.590
27	Printing and publishing	0.590	37	Transportation equipment	0.533
28	Chemicals and allied products	0.647	38	Instruments and related products	0.598
30	Rubber and misc. plastic prod.	0.383			

Note: See Appendix D for the details of calculation.

Source: Author's calculation from Morrow (2008).

Table 2.13: Alternative Relative Skilled-labor Abundance $(S/U)_c$ in Four Countries: Percentage of Population with Tertiary Education Attainment

	Tertiary Education Attainment (%)	Period of measurement	(Ref) Period for Exporter Fraction
Chile	8.9	Average of 1990 & 95	Average over 1990-96
Colombia	3.8	Average of 1980, 85, and 90	Average over 1981-91
India	3.2	Average of 1995 & 2000	Fiscal year 1997/98
United States	27.3	1990	1992

Notes: Educational attainment data for each country are selected for periods that correspond to the data periods of the exporter fractions.

Educational attainment data are available for every five years.

Source: Barro and Lee (2000).

Table 2.14: Regression for Four Countries and 17 Two-digit U.S. SIC Manufacturing Industries with Alternative Measures of Industry Skill Intensity and Country Skill Abundance

Dependent variable = $exporter_share_{ic}$	
Coef. for $skill_i * \log(S/U)_c$ (θ_2)	0.277** (0.113)
No. of observations	68
R ²	0.66

Notes: $skill_i$ is measured as the cost share of workers with one or more years of college education.

$(S/U)_c$ is measured as the percentage of population with any tertiary schooling.

Country-specific dummies and industry-specific dummies are both included.

Robust standard errors are shown in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.15: Relative Wage of Skilled to Unskilled Labor (s/w)

	(i) Max–Min Wage Ratio	(ii) Top10%– Bottom10% Mean Wage Ratio	(iii) Top 10 th –Bottom 10 th Percentile Wage Ratio	Data Periods
Chile	11.164	7.213	5.289	Average 1984-86
Colombia	9.694	5.670	3.379	Average 1988-90
India	30.280	13.218	5.662	Average 1997-98
United States	4.577	3.321	2.462	1992

Source: OWW Database (Oostendorp, 2005)

Table 2.16: Regression with Relative Factor Price (Skilled-to-Unskilled Wage Ratio)

Dependent variable = $exporter_share_{ic}$

Wage Ratio Measure	(i) Max–Min Wage Ratio	(ii) Top10%–Bottom10% Mean Wage Ratio	(iii) Top 10 th –Bottom 10 th Percentile Wage Ratio
Coef. for $skill_i * \log(s/w)_c (\psi)$	-0.351** (0.166)	-0.492** (0.222)	-0.732** (0.302)
No. of observations	68	68	68
R ²	0.64	0.65	0.65

Notes: Pooled data for the 17 two-digit U.S. SIC manufacturing industries for Chile, Colombia, India, and the United States are used for estimation.

$skill_i$ is measured as the non-production workers wages as the share in the annual payroll in each industry.

Country-specific dummies and industry-specific dummies are both included.

Robust standard errors are shown in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.17: Relative Wage of Skilled to Unskilled Labor (s/w), Measured Based on Same Occupations

	(i) Max–Min Wage Ratio	(ii) Top10%– Bottom10% Mean Wage Ratio	(iii) Top 10 th –Bottom 10 th Percentile Wage Ratio	Data Periods
Chile	7.434	5.555	4.894	Average 1985-86
Colombia	4.665	3.494	3.001	1988
India	8.987	6.824	5.795	Average 1997-98
United States	2.974	2.560	2.359	1992

Note: The relative wage is measured based on the set of the following 25 occupations in the ILO October Inquiry codes: 30, 36, 59, 65, 67, 70, 82, 84, 85, 88, 90, 91, 95, 96, 97, 99, 100, 129, 130, 131, 133, 134, 135, 140, and 141.

Source: OWW Database (Oostendorp, 2005)

Table 2.18: Regression with Relative Factor Price (Skilled-to-Unskilled Wage Ratio) Measured Based on Same Occupations

Dependent variable = $exporter_share_{ic}$

Wage Ratio Measure	(i) Max–Min Wage Ratio	(ii) Top10%–Bottom10% Mean Wage Ratio	(iii) Top 10 th –Bottom 10 th Percentile Wage Ratio
Coef. for $skill_i * \log(s/w)_c$ (ψ)	-0.587** (0.243)	-0.645** (0.273)	-0.674** (0.288)
No. of observations	68	68	68
R^2	0.65	0.65	0.65

Notes: Pooled data for the 17 two-digit U.S. SIC manufacturing industries for Chile, Colombia, India, and the United States are used for estimation.

$skill_i$ is measured as the non-production workers wages as the share in the annual payroll in each industry.

Country-specific dummies and industry-specific dummies are both included.

Robust standard errors are shown in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.C.1: Country-Industry-Year Pooled Regression (1): Without Dummies

Dependent variable = <i>exporter share_{ict}</i>	
Coef. for <i>skill_{ict}</i> (θ_1)	-0.390*** (0.105)
Coef. for <i>skill_{ict}</i> *log(<i>S/U</i>) _c (θ_2)	0.977*** (0.165)
No. of observations	340
R ²	0.25

Notes: Data are for 17 manufacturing industries according to the two-digit 1987 U.S. SIC for the following countries and years: Chile (1990-96), Colombia (1981-1991), India (1998), and the U.S. (1992).

Robust standard errors are shown in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.C.2: Country-Industry-Year Pooled Regression (2): With Industry Dummies

Dependent variable = <i>exporter share_{ict}</i>	
Coef. for <i>skill_{ict}</i> (θ_1)	-0.758*** (0.124)
Coef. for <i>skill_{ict}</i> *log(<i>S/U</i>) _c (θ_2)	0.993*** (0.116)
No. of observations	340
R ²	0.61

Notes: Data are for 17 manufacturing industries according to the two-digit 1987 U.S. SIC for the following countries and years: Chile (1990-96), Colombia (1981-1991), India (1998), and the U.S. (1992).

Industry-specific dummies are included.

Robust standard errors are shown in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.C.3: Country-Industry-Year Pooled Regression (3): With Country Dummies

Dependent variable = <i>exporter share_{ict}</i>	
Coef. for <i>skill_{ict}</i> (θ_1)	-0.252 (0.400)
Coef. for <i>skill_{ict}</i> *log(<i>S/U</i>) _c (θ_2)	0.819 (0.593)
No. of observations	340
R ²	0.31

Notes: Data are for 17 manufacturing industries according to the two-digit 1987 U.S. SIC for the following countries and years: Chile (1990-96), Colombia (1981-1991), India (1998), and the U.S. (1992).

Country-specific dummies are included.

Robust standard errors are shown in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.C.4: Country-Industry-Year Pooled Regression (4): With Year Dummies

Dependent variable = <i>exporter share_{ict}</i>	
Coef. for <i>skill_{ict}</i> (θ_1)	-0.137 (0.193)
Coef. for <i>skill_{ict}</i> *log(<i>S/U</i>) _c (θ_2)	0.649** (0.300)
No. of observations	340
R ²	0.34

Notes: Data are for 17 manufacturing industries according to the two-digit 1987 U.S. SIC for the following countries and years: Chile (1990-96), Colombia (1981-1991), India (1998), and the U.S. (1992).

Year-specific dummies are included.

Robust standard errors are shown in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.C.5: Country-Industry-Year Pooled Regression (5): With Industry and Country Dummies

Dependent variable = $exporter\ share_{ict}$	
Coef. for $skill_{ict} (\theta_1)$	-0.663** (0.315)
Coef. for $skill_{ict} * \log(S/U)_c (\theta_2)$	0.968** (0.411)
No. of observations	340
R^2	0.66

Notes: Data are for 17 manufacturing industries according to the two-digit 1987 U.S. SIC for the following countries and years: Chile (1990-96), Colombia (1981-1991), India (1998), and the U.S. (1992).

Industry- and Country-specific dummies are both included.

Robust standard errors are shown in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.C.6: Country-Industry-Year Pooled Regression (6): With Industry and Year Dummies

Dependent variable = $exporter\ share_{ict}$	
Coef. for $skill_{ict} (\theta_1)$	-0.568*** (0.144)
Coef. for $skill_{ict} * \log(S/U)_c (\theta_2)$	0.650*** (0.192)
No. of observations	340
R^2	0.70

Notes: Data are for 17 manufacturing industries according to the two-digit 1987 U.S. SIC for the following countries and years: Chile (1990-96), Colombia (1981-1991), India (1998), and the U.S. (1992).

Industry- and Year-specific dummies are both included.

Robust standard errors are shown in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.C.7: Country-Industry-Year Pooled Regression (7): With Country and Year Dummies

Dependent variable = $exporter\ share_{ict}$	
Coef. for $skill_{ict} (\theta_1)$	-0.309 (0.418)
Coef. for $skill_{ict} * \log(S/U)_c (\theta_2)$	0.896 (0.622)
No. of observations	340
R^2	0.34

Notes: Data are for 17 manufacturing industries according to the two-digit 1987 U.S. SIC for the following countries and years: Chile (1990-96), Colombia (1981-1991), India (1998), and the U.S. (1992).

Country- and Year-specific dummies are both included.

Robust standard errors are shown in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Table 2.C.8: Country-Industry-Year Pooled Regression (8): With Industry, Country, and Year Dummies

Dependent variable = $exporter\ share_{ict}$	
Coef. for $skill_{ict} (\theta_1)$	-0.975 ^{***} (0.345)
Coef. for $skill_{ict} * \log(S/U)_c (\theta_2)$	1.140 ^{***} (0.442)
No. of observations	340
R^2	0.70

Notes: Data are for 17 manufacturing industries according to the two-digit 1987 U.S. SIC for the following countries and years: Chile (1990-96), Colombia (1981-1991), India (1998), and the U.S. (1992).

Industry-, Country- and Year-specific dummies are all included.

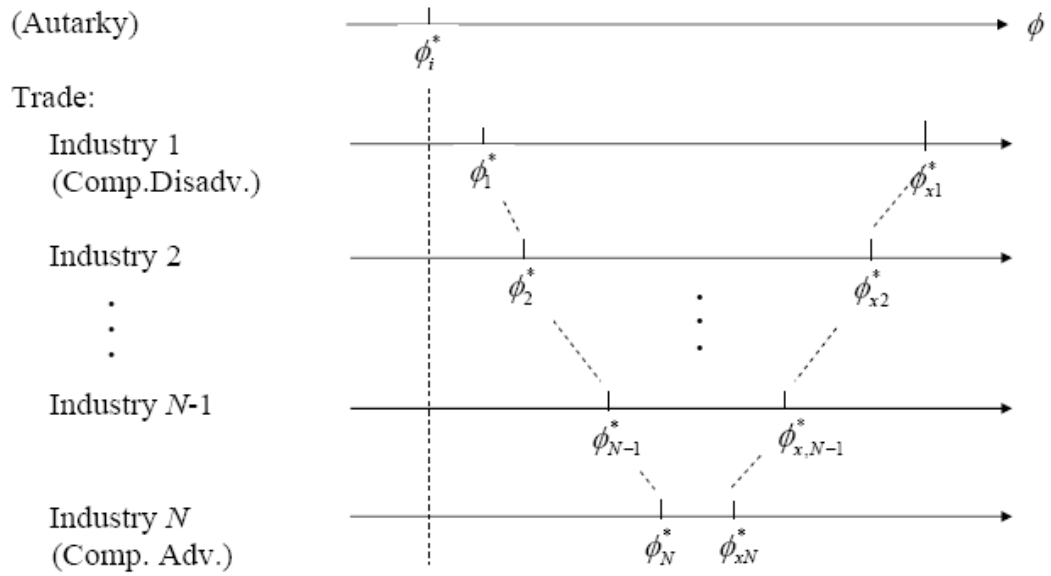
Robust standard errors are shown in parentheses.

* indicates that the coefficient estimate is significant at the 10% level;

** indicates that the coefficient estimate is significant at the 5% level; and

*** indicates that the coefficient estimate is significant at the 1% level.

Figure 2.1: “Gap” between Productivity Cutoffs in Costly-Trade Equilibrium (for Skilled-labor Abundant Country)



Note: The rank of industry skill intensities (the Cobb-Douglas production cost shares) are as follows: $\beta_1 < \beta_2 < \dots < \beta_N$.

Figure 2.2: Fraction of Exporters among All Active Firms vs Industry Skill Intensity: Chile

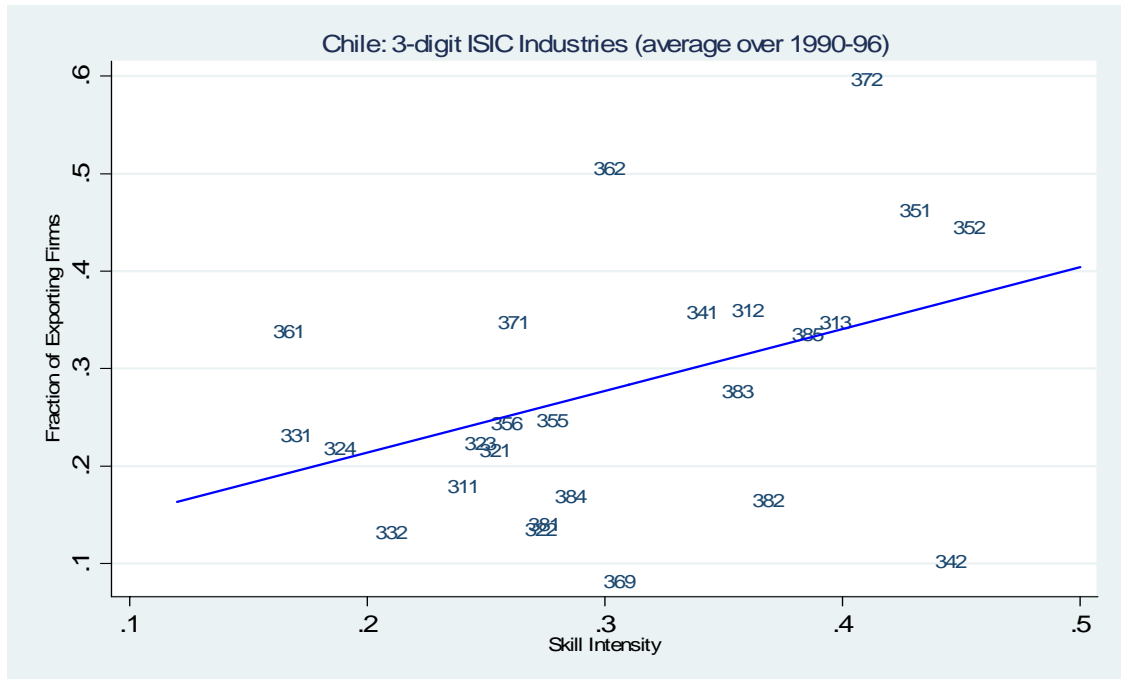


Figure 2.3: Fraction of Exporters among All Active Firms vs Industry Skill Intensity: Colombia

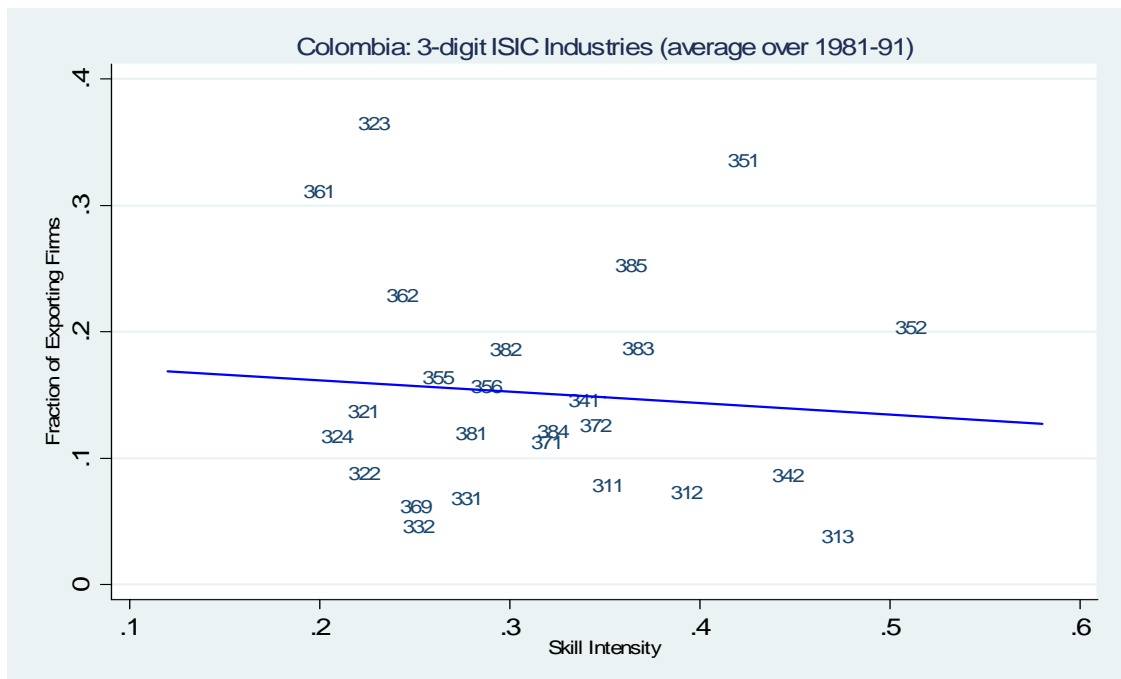


Figure 2.4: Fraction of Exporters among All Active Firms vs Industry Skill Intensity: India

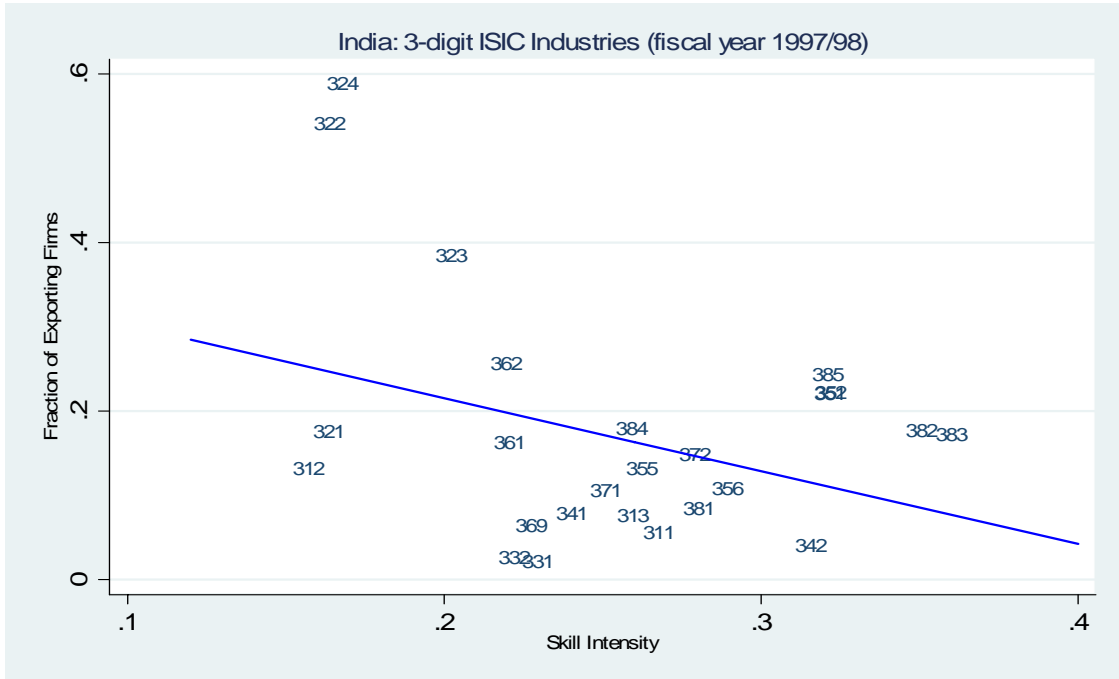
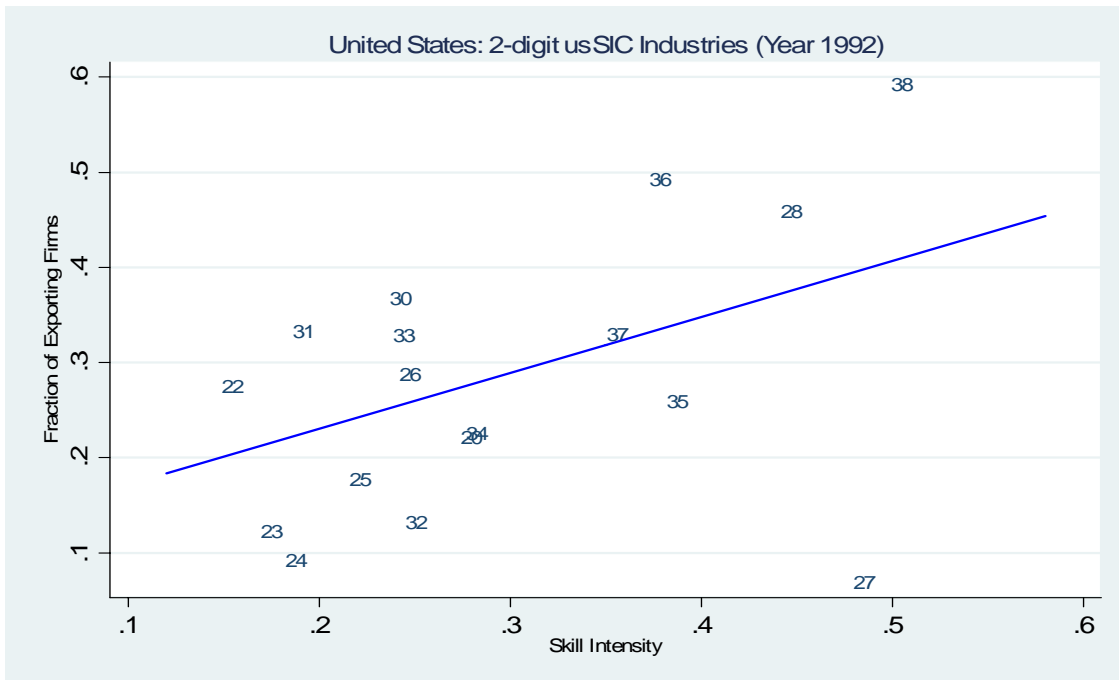


Figure 2.5: Fraction of Exporters among All Active Firms vs Industry Skill Intensity: the United States



CHAPTER III

Explaining Export Varieties: The Unexplored Role of Comparative Advantage

3.1 Introduction

The recent trade literature on export or import variety has grown rapidly. Although the increases in product variety have long been known as an important source of gains from trade, empirical studies on the significance of the growth of product varieties, or “extensive margin,” in international trade are relatively new. For example, Kehoe and Ruhl (2003) show that the trade of new goods (extensive margin) explains a larger proportion of the growth of trade following trade liberalization than the increase in the volume of previously-traded goods (intensive margin) does. Hummels and Klenow (2005) demonstrate that more than a half of greater exports of larger countries are explained by a larger variety or export margin in their exports. A series of empirical studies by Funke and Ruhwedel (2001a, 2001b, 2005) indicates that the growth of product variety in exports has a significant effect on the economic growth in various countries and regions. Feenstra and Kee (2004b, 2008) also provide evidence supporting the positive impact of export variety on productivity growth for a sample of both

developed and developing countries. Broda and Weinstein (2004) empirically show how much the increase in imported variety mattered for the welfare of United States. Their results suggest that the U.S. welfare has increased by 3% due to the increase in the extensive margin of its import.⁶⁶

Literature has investigated product varieties in international trade, or the extensive margin, as an influential factor on various aspects of the economy such as productivity, growth, and welfare. However, influential factors *on the extensive margin*, or what determine the patterns of varieties in trade, have not been much explored, except for a very few pieces such as Hummels and Kleknow (2005) that has shown the effect of the size of the economy on the extensive margin and Kehoe and Ruhl (2003) demonstrating that the reduction of trade friction is associated with an increase in variety in exports. In addition, although the preceding research has examined the cross-country patterns of product varieties in international trade, few studies have investigated the patterns of traded varieties *across industries*. To fill this gap in literature, in the current paper we examine a determinant of the cross-industry patterns of product varieties in the exports of countries, as well as how the patterns differ across countries. Specifically, we examine whether the traditional theory of comparative advantage based on factor proportions explains the observed cross-industry patterns of varieties in countries' exports.

⁶⁶ Another important branch of this recent literature focuses on the quality differentiation of exported goods. Hummels and Klenow (2005) investigate the "quality margin" in exports in addition to the extensive and intensive margins. Hallak (2006a) attempts to identify the effect of product quality on the direction of international trade. The paper empirically investigates whether importers at a higher income level tend to buy more varieties of products from exporters with higher income as well because they tend to produce higher quality products. In a related paper Hallak applies his framework of product quality and uses sectoral level data to provide evidence for the Linder hypothesis according to which international trade is more intensive between countries with similar income levels than those that differ (Hallak, 2006b). Choi, Hummels and Xiang (2006) explore the effect of income distribution on varieties in trade, whose key insight is that consumers with higher income will buy goods with higher quality rather than buy greater quantities of goods that vary in the quality dimension.

Moreover, most of the existing studies on the variety or extensive margin in trade have built on the framework of the monopolistic competition model by Krugman (1979), which first brought product variety in international trade into focus. However, empirical research on the topic has not been well connected to heterogeneous firm models that have been recently developed and widespread, while Feenstra and Kee (2008) is an attempt to this new direction.⁶⁷ Our study contributes to this new literature by considering the modern framework of heterogeneous firms together with the traditional framework of factor proportion theory to explore, both theoretically and empirically, the role for the comparative advantage in export variety.

This study first follows the previous chapter of this dissertation, which extends Melitz (2003) and Bernard, Redding, and Schott (2007) to a broader framework, to develop a theoretical model in which countries vary in factor endowment, industries differ in factor intensity, and firms are heterogeneous in productivity within industries. The paper next derives a prediction that relates product varieties in a country's exports to the degree of relative factor intensity of industries. This prediction is empirically tested using data on the U.S. imports from Feenstra, Romalis, and Schott (2002) that finely classify traded goods according to the ten-digit Harmonization System (HS). The study also employs the data on factor use in various industries from the U.S. Census of Manufactures, as well as the data on factor abundance of a number of countries from Hall and Jones (1999). The empirical analysis supports our semi-Heckscher-Ohlin prediction on product varieties in exports; i.e, countries export more varieties in industries in which they have a comparative advantage in terms of factor proportions.

⁶⁷ Theoretical work by Chaney (2008) also investigates the extensive (and intensive) margin in trade under the framework of a heterogeneous firm model.

The current paper also adds to the literature on firm-level heterogeneity by providing empirical evidence for an unexplored aspect of the recent models. In particular, in contrast to the existing studies, this paper performs an empirical test for a heterogeneous firm model without relying on firm-level data for a particular country but using industry-level data that are more publicly accessible and available for a broader range of countries.⁶⁸

The paper proceeds as follows. Section 3.2 develops the theoretical model in order to provide an implication for the relationship between factor proportions and export variety. Section 3.3 describes the data. A proposed empirical approach to test the theoretical prediction, as well as the results of the empirical tests, is presented in Section 3.4. Section 3.5 concludes.

3.2 The Model

We build this study on the model presented in the previous chapter II, which extends the model by Bernard, Redding, and Schott (2007) to the framework of two countries, two factors, and multiple industries. Therefore, in what follows we present only the key elements of the model to derive our prediction on the relationship between product varieties in exports and the comparative advantages of countries. Other details of the model are left to Chapter II.

3.2.1 Basic Framework

The modeled economy comprises two countries, Home (H) and Foreign (F); two factors, skilled labor (S) and unskilled labor (U); and $N (>2)$ industries. Within each industry there is a continuum of firms that are heterogeneous in productivity. Countries

⁶⁸ Feenstra and Kee (2008) also utilize country- and industry-level data for the test of a heterogeneous firm model.

differ in factor endowments: Home is relatively abundant in skilled labor, and Foreign is relatively abundant in unskilled labor; i.e., $\frac{\bar{S}^H}{\bar{U}^H} > \frac{\bar{S}^F}{\bar{U}^F}$ where \bar{S}^H (\bar{U}^H) and \bar{S}^F (\bar{U}^F) denote the total inelastic supply of (un)skilled labor in Home and Foreign, respectively.

3.2.2 Consumption

The preference of the representative consumer is described by the following utility function:

$$U = C_1^{\alpha_1} C_2^{\alpha_2} \dots C_N^{\alpha_N}, \quad \sum_{i=1}^N \alpha_i = 1 \quad (3.2.1)$$

The consumption index C_i for each industry $i=1, \dots, N$ takes the following CES (or Dixit-Stiglitz) form:

$$C_i = \left[\int_{\omega \in \Omega_i} q_{\omega}^{\rho} d\omega \right]^{\frac{1}{\rho}} \quad (3.2.2)$$

where ω indexes product varieties within an industry, Ω_i denotes a set of available varieties in Industry i , and q_{ω} represents the quantity of each variety consumed. The ideal price index for Industry i is defined as follows:

$$P_i = \left[\int_{\omega \in \Omega_i} p_{i,\omega}^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}} \quad (3.2.3)$$

where $\sigma = \frac{1}{1-\rho} > 1$ is the constant elasticity of substitution across varieties.

3.2.3 Production and Export

Each firm produces a unique variety of products. A firm's production technology, which exhibits economies of scale, is described by the following cost function:

$$\begin{aligned}\Gamma_{i,\omega}^H &= \left[f_i + \frac{q_{i,\omega}}{\phi_{i,\omega}} \right] \cdot (s^H)^{\beta_i} (w^H)^{1-\beta_i} \\ \Gamma_{i,\omega}^F &= \left[f_i + \frac{q_{i,\omega}}{\phi_{i,\omega}} \right] \cdot (s^F)^{\beta_i} (w^F)^{1-\beta_i}\end{aligned}\tag{3.2.4}$$

where s is the wage for skilled labor, w is the wage for unskilled labor, and the superscripts H and F denote Home and Foreign, respectively. The intensities of the two factors in each industry (β_i and $1-\beta_i$) are common across countries, but the firm-specific productivity level $\phi_{i,\omega}$ varies the marginal cost across firms. The industries are ranked according to the skilled-labor intensity (β_i) such that $0 < \beta_1 < \beta_2 \dots < \beta_{N-1} < \beta_N < 1$.⁶⁹

The optimal pricing of each firm equals a constant markup ($1/\rho$) over the marginal cost of production. Therefore, for domestic sales, each firm charges the following price for its product:

$$\begin{aligned}p_{i,\omega}^H(\phi_{i,\omega}) &= \frac{(s^H)^{\beta_i} (w^H)^{1-\beta_i}}{\rho \phi_{i,\omega}} \\ p_{i,\omega}^F(\phi_{i,\omega}) &= \frac{(s^F)^{\beta_i} (w^F)^{1-\beta_i}}{\rho \phi_{i,\omega}}\end{aligned}\tag{3.2.5}$$

Firms can also export their products by incurring the (amortized per-period) fixed costs $f_{xi}(s)^{\beta_i} (w)^{1-\beta_i}$ ($f_{xi} > 0$), as well as the variable “iceberg” shipping costs such that only $1/\tau_i$ ($\tau_i > 1$) of the shipped quantity reaches to the other country. The optimal price of a firm’s product for exporting is thus as follows:

⁶⁹ To be accurate, β_i indicates the Cobb-Douglas cost share of skilled labor. However, since the equilibrium relative factor intensity in each industry is $\frac{S_i}{S_i + U_i} = \frac{\beta_i}{\beta_i + (1-\beta_i) \cdot (s/w)}$ for any relative wage $s/w > 0$, $S_i/(S_i+U_i)$ is larger for a larger β_i .

$$\begin{aligned}
P_{xi,\omega}^H(\phi) &\equiv \tau_i \cdot P_{i,\omega}^H(\phi) = \frac{\tau_i (s^H)^{\beta_i} (w^H)^{1-\beta_i}}{\rho \phi_{i,\omega}} \\
P_{xi,\omega}^F(\phi) &\equiv \tau_i \cdot P_{i,\omega}^F(\phi) = \frac{\tau_i (s^F)^{\beta_i} (w^F)^{1-\beta_i}}{\rho \phi_{i,\omega}}
\end{aligned} \tag{3.2.6}$$

Accordingly, a firm's revenue from the domestic sales is:

$$\begin{aligned}
r_{i,\omega}^H(\phi_{i,\omega}) &= \alpha_i Y^H \left(\frac{(s^H)^{\beta_i} (w^H)^{1-\beta_i}}{\rho \phi_{i,\omega} P_i^H} \right)^{1-\sigma} \\
r_{i,\omega}^F(\phi_{i,\omega}) &= \alpha_i Y^F \left(\frac{(s^F)^{\beta_i} (w^F)^{1-\beta_i}}{\rho \phi_{i,\omega} P_i^F} \right)^{1-\sigma}
\end{aligned} \tag{3.2.7}$$

and the revenue from the overseas sales is:

$$\begin{aligned}
r_{xi,\omega}^H(\phi) &= \alpha_i Y^F \left(\frac{\tau_i (s^H)^{\beta_i} (w^H)^{1-\beta_i}}{\rho \phi_{i,\omega} P_i^F} \right)^{1-\sigma} \\
r_{xi,\omega}^F(\phi) &= \alpha_i Y^H \left(\frac{\tau_i (s^F)^{\beta_i} (w^F)^{1-\beta_i}}{\rho \phi_{i,\omega} P_i^H} \right)^{1-\sigma}
\end{aligned} \tag{3.2.8}$$

where Y^H and Y^F are the total national incomes of Home and Foreign, respectively.

3.2.4 Zero Profit

Firms need to maintain at least zero profit in each of the domestic and export markets. Firms do not export if they are not profitable enough to satisfy the zero-profit condition for the export market. Firms do not even serve the domestic market if they are not profitable enough to fulfill the zero-profit condition for the domestic market. The zero-profit condition for a firm in each market is described such that the firm's revenue net of the variable costs equals the fixed costs:

(For the domestic market)

$$\begin{aligned}\frac{r_{i,\omega}^H(\phi_{i,\omega})}{\sigma} &= f_i(s^H)^{\beta_i} (w^H)^{1-\beta_i} \Leftrightarrow r_i^H(\phi_i^{*H}) = \sigma f_i(s^H)^{\beta_i} (w^H)^{1-\beta_i} \\ \frac{r_{i,\omega}^F(\phi_{i,\omega})}{\sigma} &= f_i(s^F)^{\beta_i} (w^F)^{1-\beta_i} \Leftrightarrow r_i^F(\phi_i^{*F}) = \sigma f_i(s^F)^{\beta_i} (w^F)^{1-\beta_i}\end{aligned}\quad (3.2.10)$$

(For the export market)

$$\begin{aligned}\frac{r_{xi,\omega}^H(\phi_{i,\omega})}{\sigma} &= f_{xi}(s^H)^{\beta_i} (w^H)^{1-\beta_i} \Leftrightarrow r_{xi}^H(\phi_{xi}^{*H}) = \sigma f_{xi}(s^H)^{\beta_i} (w^H)^{1-\beta_i} \\ \frac{r_{xi,\omega}^F(\phi_{i,\omega})}{\sigma} &= f_{xi}(s^F)^{\beta_i} (w^F)^{1-\beta_i} \Leftrightarrow r_{xi}^F(\phi_{xi}^{*F}) = \sigma f_{xi}(s^F)^{\beta_i} (w^F)^{1-\beta_i}\end{aligned}\quad (3.2.11)$$

ϕ_i^* and ϕ_{xi}^* in the above equations denote the productivity “cutoffs” for firms serving the domestic market (or “domestic producers”) and exporters, respectively. The first cutoff divides domestic producers from firms exiting from the domestic market, and the second divide exporting firms from domestic producers.⁷⁰

3.2.5 Entry and Equilibrium under Costly Trade

To enter the domestic market, firms must incur a sunk entry cost, which takes the following form:

$$f_{ei}(s)^{\beta_i} (w)^{1-\beta_i}, \quad f_{ei} > 0 \quad (3.2.12)$$

Firms discover their productivity after the entry. The productivity parameter ϕ is randomly drawn from a distribution $G(\phi)$, which is common across countries. Each firm, or a potential entrant, decides to enter (to realize its own productivity by paying the sunk entry cost) if its pre-entry or *ex ante* expected future profit stream is at least as large as

⁷⁰ We focus only on the case in which exporters are more productive than domestic producers; i.e., $\phi_{xi}^* > \phi_i^*$. The reasons and conditions to be satisfied for this are described in Chapter II.

the sunk entry cost. In stationary equilibrium, the *ex ante* expected future profit exactly equals the entry cost, which determines the free-entry condition described as follows:

$$\begin{aligned} V_i^H &= \frac{1 - G(\phi_i^{*H})}{\delta} [\pi_i^H(\bar{\phi}_i^H) + \chi_i^H \cdot \pi_{xi}^H(\bar{\phi}_{xi}^H)] = f_{ei}(s^H)^{\beta_i} (w^H)^{1-\beta_i} \\ V_i^F &= \frac{1 - G(\phi_i^{*F})}{\delta} [\pi_i^F(\bar{\phi}_i^F) + \chi_i^F \cdot \pi_{xi}^F(\bar{\phi}_{xi}^F)] = f_{ei}(s^F)^{\beta_i} (w^F)^{1-\beta_i} \end{aligned} \quad (3.2.13)$$

$\delta < 1$ is an exogenous probability of a firm's "death" in each period. $1 - G(\phi_i^*)$ is the (*ex*

ante) probability of successful entry or survival in the domestic market. $\chi_i \equiv \frac{1 - G(\phi_{xi}^*)}{1 - G(\phi_i^*)}$

is the probability for a successful entrant or domestic producer to be an exporter, given

that $\phi_{xi}^* > \phi_i^*$. $\pi_i(\bar{\phi}_i)$ is the per-period domestic profit of the averagely productive

domestic producer, and $\pi_{xi}(\bar{\phi}_{xi})$ is the per-period export profit of the averagely

productive exporter. The average productivity levels of domestic producers (or survivors)

$\bar{\phi}_i$ and of exporters $\bar{\phi}_{xi}$ are defined, respectively, as follows:

$$\begin{aligned} \bar{\phi}_i(\phi_i^*) &= \left[\frac{1}{1 - G(\phi_i^*)} \int_{\phi_i^*}^{\infty} \phi^{\sigma-1} g(\phi) d\phi \right]^{\frac{1}{\sigma-1}} \\ \bar{\phi}_{xi}(\phi_{xi}^*) &= \left[\frac{1}{1 - G(\phi_{xi}^*)} \int_{\phi_{xi}^*}^{\infty} \phi^{\sigma-1} g(\phi) d\phi \right]^{\frac{1}{\sigma-1}} \end{aligned} \quad (3.2.14)$$

The zero-profit conditions (3.2.10) and (3.2.11) and the free-entry condition (3.2.13)

jointly determine the two productivity cutoffs, ϕ_i^* and ϕ_{xi}^* , for the respective two

countries *H* and *F*.

3.2.6 Mass of Firms and Export Varieties

Now we examine how many firms in each country will export to the overseas market in each industry. In our model, the number of firms is measured by the size of the

“mass” of the continuum of firms. M_i denotes the mass of domestic producers, and M_{ix} denotes the mass of the exporting firms. Only a portion of the domestic producers will be exporters, and that fraction is determined by the two cutoff productivity levels. That is, in equilibrium, the *ex ante* probability for a domestic producer to be an exporter is equal to the *ex post* fraction of exporters among domestic producers, such that:

$$\begin{aligned}\chi_i^H &= M_{xi}^H / M_i^H \Leftrightarrow M_{xi}^H = \chi_i^H \cdot M_i^H \\ \chi_i^F &= M_{xi}^F / M_i^F \Leftrightarrow M_{xi}^F = \chi_i^F \cdot M_i^F\end{aligned}\tag{3.2.15}$$

Our concern is with the relative size of the exporter mass between the two countries in each industry, M_{xi}^H / M_{xi}^F , and how it will differ across industries in relation to the relative factor intensities of the industries and the relative factor abundance of each country. To derive and examine M_{xi}^H / M_{xi}^F , we consider the equilibrium price indexes of Industry i in the two countries, which are composed of the number and average price of domestically produced products, as well as those of products imported from the other country:

$$P_i^H = [M_i^H (p_i^H (\bar{\phi}_i^H))^{1-\sigma} + \chi_i^F \cdot M_i^F (\tau_i \cdot p_i^F (\bar{\phi}_{xi}^F))^{1-\sigma}]^{\frac{1}{1-\sigma}}\tag{3.2.16}$$

$$P_i^F = [M_i^F (p_i^F (\bar{\phi}_i^F))^{1-\sigma} + \chi_i^H \cdot M_i^H (\tau_i \cdot p_i^H (\bar{\phi}_{xi}^H))^{1-\sigma}]^{\frac{1}{1-\sigma}}\tag{3.2.17}$$

Dividing Equation (3.2.16) by (3.2.17) in both sides yields the following equation:

$$\left(\frac{P_i^H}{P_i^F}\right)^{1-\sigma} = \frac{M_i^H (p_i^H (\bar{\phi}_i^H))^{1-\sigma} + \chi_i^F \cdot M_i^F \cdot \tau_i^{1-\sigma} (p_i^F (\bar{\phi}_{xi}^F))^{1-\sigma}}{M_i^F (p_i^F (\bar{\phi}_i^F))^{1-\sigma} + \chi_i^H \cdot M_i^H \cdot \tau_i^{1-\sigma} (p_i^H (\bar{\phi}_{xi}^H))^{1-\sigma}}\tag{3.2.18}$$

By rearranging this equation, we can derive the following expression for the ratio of the masses of domestic producers in the two countries:

$$\frac{M_i^H}{M_i^F} = \frac{\left(\frac{P_i^H}{P_i^F}\right)^{1-\sigma} (p_i^F (\bar{\phi}_i^F))^{1-\sigma} - \chi_i^F \cdot \tau_i^{1-\sigma} (p_i^F (\bar{\phi}_{xi}^F))^{1-\sigma}}{(p_i^H (\bar{\phi}_i^H))^{1-\sigma} - \left(\frac{P_i^H}{P_i^F}\right)^{1-\sigma} \cdot \chi_i^H \cdot \tau_i^{1-\sigma} (p_i^H (\bar{\phi}_{xi}^H))^{1-\sigma}} \quad (3.2.19)$$

By combining Equations (3.2.15) and (3.2.19) and rearranging further, we obtain the following expression for the ratio of the exporter masses in the two countries:⁷¹

$$\frac{M_{xi}^H}{M_{xi}^F} = \frac{\chi_i^H}{\chi_i^F} \cdot \frac{\left[\left(\frac{Y^H}{Y^F}\right) \left(\frac{f_i}{f_{xi}}\right) \left(\frac{\phi_{xi}^{*F}}{\phi_i^{*F}}\right)^{\sigma-1} - \chi_i^F \left(\frac{\bar{\phi}_{xi}^F}{\bar{\phi}_i^F}\right)^{\sigma-1} \right]}{\tau_i^{\sigma-1} \left[1 - \chi_i^H \left(\frac{Y^H}{Y^F}\right) \left(\frac{f_{xi}}{f_i}\right) \left(\frac{\phi_i^{*H}}{\phi_{xi}^{*H}}\right)^{\sigma-1} \left(\frac{\bar{\phi}_{xi}^H}{\bar{\phi}_i^H}\right)^{\sigma-1} \right]} \cdot \left(\frac{p_i^H (\bar{\phi}_i^H)}{p_i^F (\bar{\phi}_i^F)} \right)^{\sigma-1} \quad (3.2.20)$$

That is, the relative size of the exporter mass in each industry depends on the ratio of (or the “gap” between) the two productivity cutoffs, ϕ_{xi}^* / ϕ_i^* , and the ratio of the average productivity of exporters to that of domestic producers, $\bar{\phi}_{xi} / \bar{\phi}_i$, as well as the ratio of the average domestic price of products between the two countries, $p_i^H (\bar{\phi}_i^H) / p_i^F (\bar{\phi}_i^F)$.

For the purpose of the cross-industry comparison of this relative exporter mass, we impose the following two assumptions:

Assumption 1: $f_i = f_j, f_{ix} = f_{jx}$, and $\tau_i = \tau_j$ for $i \neq j$

Assumption 2: $G(\phi_i) = 1 - \left(\frac{\phi_i}{\phi_i^*}\right)^k$ for $i = 1, 2, \dots, N; k > 2\sigma$

The first assumption implies that (i) both fixed costs for production and fixed costs for export, *adjusted for the difference due to factor intensity difference*, are identical across industries; and also that (ii) the “iceberg” shipping cost for export is the same for all industries. The second assumption means that (i) the *ex ante* distribution of firm productivity is common (not only across countries but also) across industries, and that (ii)

⁷¹ See Appendix for the derivation of Equation (3.2.20).

the distribution is a Pareto distribution with $\underline{\phi}_i$ as the minimum value for productivity drawn in Industry i ($\phi_i \in [\underline{\phi}_i, +\infty)$) and k as a shape parameter indicating the dispersion of productivity distribution.⁷² We assume $k > 2\sigma$, as assumed in the previous chapter, for the variances of both drawn productivities and sizes of firms (measured as domestic sales) to be finite.

By examining Equation (3.2.20) across industries under Assumptions 1 and 2, we derive the following proposition regarding the relative size of the masses of exporters between the two countries.

$$\text{Proposition: If } \frac{\bar{S}^H}{\bar{U}^H} > \frac{\bar{S}^F}{\bar{U}^F} \text{ and } \beta_i > \beta_j, \text{ then } \frac{M_{ix}^H}{M_{ix}^F} > \frac{M_{jx}^H}{M_{jx}^F}.$$

Proof: See Appendix.

This proposition implies that the mass of exporters in a country relative to the mass in the other country will be larger in industries in which the country has a comparative advantage. That is, the relatively skill-abundant country has a larger exporter mass than the other country in a more skill-intensive industry, and vice versa.

Can we predict the relative size of the mass of exporters under free trade with FPE? It is well-known that with FPE the cross-industry patterns of production and trade are indeterminate when the number of industries (sectors) is greater than the number of input factors (e.g., Melvin (1968)). This indeterminacy will also apply to our model,⁷³ and under free trade with FPE there exist multiple equilibrium allocations of the two

⁷² Chaney (2008) brings some rationale of the use of a Pareto distribution for this type of the model.

⁷³ We can see this indeterminacy in the relative size of the mass of exporters in Equation (3.2.19). Under free trade with FPE, $\tau_i = 1$, $\chi_i = 1$ (since all active firms will be exporters), the price of a product variety will be the same in the two market, and the industry price index will be equal in the two countries. Hence, both numerator and denominator of the right-hand side of the equation is zero, which implies the indeterminacy of M_i^H/M_i^F .

factors across industries. As an overall tendency, however, the production resources will on average be allocated more to industries in which the country has its comparative advantage (for both factors in the country to be fully employed), so that the mass of firms will *on average* be larger in the comparative advantage industries.

Finally, we present the key prediction for the product varieties in exports. Since each firm is considered to produce a unique variety of differentiated product, the mass of exporting firms in a country, which is examined above, represents the number of product varieties exported from the country in each industry. Therefore, the above Proposition has the following implication on export varieties, which is expressed as the following prediction:

Prediction: For a certain pair of countries, international trade will exhibit the following cross-industry pattern: The relatively skilled-labor abundant country will export more product varieties in more skill-intensive industries (industries with greater β). In contrast, the relatively unskilled-labor abundant country will export more varieties in more unskilled-labor intensive industries (industries with smaller β).

3.3 The Data

An empirical test of the prediction of our model requires data for three variables: the number of product varieties exported from each country in each industry, factor endowment in each exporting country, and factor intensity in each industry.

For the product varieties in exports, we use the data on the U.S. imports in the years of 1990, 1995, and 2000 that are from Feenstra, Romalis, and Schott (2002). The data contain information on the U.S. imports of each good classified according to the

disaggregated ten-digit Harmonized System (HS) exported from each country. The data also map each ten-digit HS code onto different and more aggregated industry classifications such as the four-digit U.S. Standard Industrial Classification (SIC, the 1987 version) and the six-digit North American Industry Classification System (NAICS, the 1997 version). These different levels of classification in the data enable us to count the number of product varieties in each industry by defining “products” or “varieties” according to the ten-digit HS and “industries” according to the four-digit SIC.⁷⁴ Due to the limitation of the availability of the data on industry factor intensity, our empirical analysis focuses on trade in manufacturing industries (the codes 2011 through 3999 in the four-digit SIC, and 311111 through 339999 in the six-digit NAICS). Table 3.1 provides the numbers of exporters, numbers of product varieties, and total import values in the U.S. total imports and manufacturing imports in each of the three years. In these three years, manufacturing industries represent 94% of the total U.S. imports in terms of the number of product varieties, and 83% through 86% in terms of value.

The data for the factor endowment of each country are from Hall and Jones (1999). Since our theoretical model is embedded in a two-factor framework with skilled labor (S) and unskilled labor (U), we use the data on human capital per worker as the measure of the abundance of skilled labor relative to unskilled labor (S/U) in each country. The data on human capital per worker are estimated as of 1988 and available for 127 countries in their study.

Our theoretical model assumes a common factor intensity for each industry across countries. To measure this world common factor intensity of each industry, we use the data from the U.S. Census of Manufactures for the years of 1992, 1997, and 2002. The

⁷⁴ See the following section for further details.

1992 census applies the U.S. SIC (1987 version), while the 1997 and 2002 censuses use NAICS to classify manufacturing industries.⁷⁵ For each classified industry, the censuses report the number of production workers (average per worker) separately from the total employment. Therefore, we measure industry *unskilled-labor* intensity as the share of production workers in the total employment, and accordingly *skilled-labor* intensity as the share of non-production workers (i.e., one minus unskilled-labor intensity). We thus obtain the skill intensities for 458 four-digit SIC industries from the 1992 census that are combined with the U.S. import data for 1990, and the skill intensities for 473 six-digit NAICS industries from the 1997 and 2002 censuses that are combined with the 1995 and 2000 import data, respectively.

The data for our empirical analysis includes 115 countries whose factor endowment measure is available in Hall and Jones (1999) and from which the U.S. imported in any one or more manufacturing industry in the years 1990, 1995, and 2000.⁷⁶ Table 3.2 lists these 115 countries, and Table 3.3 provides the summary statistics of the relative factor endowment (the skilled-labor to unskilled-labor ratio: S/U) of these countries with the lists of the ten most and least skilled labor-abundant countries. The data also include 394 (four-digit SIC) manufacturing industries for 1990, and 383 and 384 (six-digit NAICS) industries for 1995 and 2000, respectively, in which the U.S. imported from one or more countries in each year. Tables 3.4.1 through 3.4.3 present the summary statistics of the intensities of the two factors (S and U) of these manufacturing

⁷⁵ NAICS has been modified for the 2002 census (2002 version) from the previous 1997 version. However, for manufacturing industries, the two versions are identical.

⁷⁶ Of the 115 countries, the following three countries are included only in the data for 1990: Czechoslovakia, the U.S.S.R., and Yugoslavia.

industries, as well as the ten most and least skilled-labor intensive industries, for the three respective years.

Figures 3.1.1 through 3.1.3 display the number of countries from which the U.S. imported in each manufacturing industry for each year. In each table, the industries are sorted (from left to right) in the order of skilled-labor intensity. Figures 3.2.1 through 3.2.3 and 3.3.1 through 3.3.3 plot the number of exporting countries and the total number of product varieties in the U.S. imports in each industry, respectively, against the industry skilled-labor intensity. These figures indicate that the U.S., one of the world's most skilled-labor abundant countries, tended to import more varieties from more countries in relatively unskilled-labor intensive industries, while the U.S. has increased imports in relatively skill intensive industries and thus the trend has become unclear in recent years.

3.4 Empirical Tests

As stated in Section 3.2, our model provides one key prediction: A country will export more varieties of products in industries in which the country has a comparative advantage, in terms of factor proportions, than it will in other industries. In this section we empirically test this implication using the data described in the previous section.

3.4.1 Measuring Exported Varieties

Our model explains the number of product varieties in each industry that are exported from each country to a common importer—in this case, the U.S.— in terms of two elements: the relative factor abundance of the exporting country and the relative factor intensity of the industry. As described in the previous section, we define a variety as each ten-digit HS good and an industry as each four-digit SIC (for 1990) or six-digit

NAICS (for 1997 and 2000). We thus measure the number of product varieties in Industry i exported from Country c , or n_{ic} , as follows:

$$n_{ic} \equiv \text{No. of ten-digit HS goods exported from Country } c \text{ in a four-digit SIC or six-digit NAICS Industry } i$$

Some four-digit SIC or six-digit NAICS industries may contain by nature more ten-digit HS goods in their catalogue than other industries, and thus in the U.S. imports we may observe more varieties in those industries than in other industries, regardless of the role of the comparative advantage. Therefore, for a proper cross-industry comparison, we use the following normalized measure for the number of varieties:⁷⁷

$$n_share_{ic} = \frac{n_{ic}}{N_i}$$

where N_i is the total number of varieties that the U.S. imports from the world in Industry i : $N_i = \sum_c n_{ic}$.⁷⁸ It should be noted that the imports of the same ten-digit commodities from different countries are considered as different product varieties, following the theoretical assumption that products are differentiated across firms and thus across countries.

3.4.2 Regressions for Aggregate North and South

We first test our two-country, two-factor, and multi-industry model with the data for country aggregates. We divide the 115 countries into two groups to construct two country aggregates, one of which consists of countries that are relatively skilled-labor

⁷⁷ This variable is consistent with the idea of the “relative size of firm mass” described in Proposition in Section 3.2. Here, due to the limitation of the employed data, the number of exported varieties from one country in one industry is expressed as the relative value to the number of varieties exported from the rest of the world in that industry, instead of the ratio to the number of varieties exported from the trading partner (i.e., the U.S.).

⁷⁸ Accordingly, the total number of varieties in each industry, N_i , includes the number of varieties exported to the U.S. from countries other than 115 countries in the sample.

abundant (or with high S/U). We refer to this group as the “North.” The other consists of countries that are relatively unskilled-labor abundant (or with low S/U), which we call the “South.” The North consists of 51 countries whose S/U is above the average of all the 115 countries, and the South comprises other 64 countries.⁷⁹ Table 3.5 lists the names of the countries constituting each of the aggregates North and South. Table 3.6 compares the within-group averages of relative factor abundance S/U .

The following equation is estimated using the OLS for the aggregate North and South:⁸⁰

$$\log(n_share_{i,A}) = \gamma + \theta \cdot skill_i + \varepsilon_i \quad (3.4.1)$$

$$\text{where } n_share_{i,A} = \sum_{c \in A} n_share_{i,c}, A = \{\text{North, South}\}$$

$skill_i$ = skill intensity of Industry i .

Equation (3.4.1) is estimated for the three respective years. The result of the estimation is shown in Table 3.7. For all the three years, the result is consistent with the prediction of the model. That is, the estimated coefficient for the industry skill intensity is positive for the North, indicating that the relatively skilled-labor abundant North exports more varieties in more skill-intensive industries; and the coefficient estimate is negative for the South, which implies that in the relatively unskilled-labor abundant South the number of varieties in exports is higher as the industry is less skill intensive (or more unskilled-labor intensive). The result of the analysis for the country aggregates thus

⁷⁹ We also attempted the following two other “cutoffs” for S/U to divide the countries into the aggregates North and South: above or below the 75 percentile (29 countries in the North, 86 in the South), and above or below 0.7 relative to S/U of the U.S. (25 countries in the North, 90 in the South). These alternative groupings are also indicated in Table 3.5. The qualitative results of the estimation, however, are the same regardless of the cutoffs.

⁸⁰ $n_share_{i,c}$ is skewed in distribution, and therefore scaled to logarithm for the regressions to adjust for potential heteroskedasticity. We do not scale the factor intensity measure ($skill_i$) to logarithm, but the results do not change even though the log-scaled intensity is used.

supports the semi-Heckscher-Ohlin prediction of the model about the product varieties in exports.⁸¹

3.4.3 Pooled Regression for Dependent Parameter Specification

We next use the pooled data for all the individual exporting countries to estimate cross-industry patterns of exports in terms of product varieties. We consider the following regression model:

$$\log(n_share_{ic}) = \gamma + \Pi_c \cdot skill_i + \varepsilon_{ic} \quad (3.4.2)$$

The slope coefficient for skilled-labor intensity, Π_c , would differ across exporter countries. The theory predicts that the value of the slope coefficient will be higher for countries with greater relative endowment of skilled labor, and lower for countries with smaller relative skilled labor endowment (or greater relative endowment of unskilled labor). This pattern is indeed observed in the result of the estimation of Equation (3.4.2) for each individual exporting country. Figures 3.4.1 through 3.4.3 plot the slope coefficient $\hat{\Pi}_c$ estimated from each individual country regression against the relative skilled-labor abundance of the country (in logarithmic scale, $\log(S/U)$). The figures exhibit the tendency that the coefficient Π_c is greater for a more skill-abundant country, which is consistent across years.⁸² To confirm this pattern in the pooled regression, we impose the following structure on the slope coefficient Π_c :

$$\Pi_c = \Pi((S/U)_c) = \theta_1 + \theta_2 \cdot \log(S/U)_c \quad (3.4.3)$$

⁸¹ The level of significance is not very high for the estimate for the North in the year 2000. This should be because, as shown in Figures 3.1.3, 3.2.3, and 3.3.3, in recent years the U.S. imports from more countries in relatively skill-intensive industries. However, the estimate is more significant (at the 1% level) when the alternative cutoffs are applied to group the North and South.

⁸² To draw the fitted line in Figures 3.4.1 through 3.4.3, the cases (i.e., the results of individual country regressions) are weighted by the number of observations (i.e., the number of industries for each country in the data) in each individual country regression.

where $(S/U)_c$ is the skilled- to unskilled-labor ratio of Country c .⁸³ The theoretical prediction is that θ_1 will be negative (since Π_c will be negative for countries with low skilled-labor abundance) and θ_2 will be positive (since Π_c will be larger and to be positive for countries with higher skilled-labor abundance). By substituting Equation (3.4.3) into (3.4.2), we derive the following equation for our pooled regression:

$$\log(n_share_{ic}) = \theta_1 \cdot skill_i + \theta_2 \cdot skill_i \cdot \log(S/U)_c + \mu_c + \varepsilon_{ic} \quad (3.4.4)$$

We include country dummies, μ_c , to capture the effects of all country-specific factors other than the relative factor abundance, such as fixed and variable trade costs (of importing to the U.S.) and the size of the country.

Table 3.8 presents the result of the estimation of Equation (3.4.4) for each of the years 1990, 1995, and 2000 using the fixed-effect OLS. The estimates of both coefficients θ_1 and θ_2 have the signs that are expected from the theory, and they are highly significant (at the 1% level).⁸⁴ This result is consistent across years. Hence, the semi-Heckscher-Ohlin prediction of our economic model on the exported varieties is also supported by the pooled analysis using the U.S. import data.

Finally, using these estimates we compute the “threshold” factor abundant at which the country-specific slope coefficient for skill intensity Π_c turns from negative to positive (i.e., S/U^* such that $\Pi_c(S/U^*) = 0$). The value of the “threshold” S/U^* is 2.11 for the year 1990,⁸⁵ which is the closest to the relative factor abundance in China ($S/U =$

⁸³ We use the logarithm of the relative skill abundance to have the size of the coefficient estimate for θ_2 invariant to which of S or U is on the denominator.

⁸⁴ This result does not change when the natural-scaled measure of S/U is used in the regression instead of $\log(S/U)$; i.e., $\hat{\theta}_1$ is negative and $\hat{\theta}_2$ is positive, both significant at the 1% level.

⁸⁵ $S/U^* = \exp(-\frac{\hat{\theta}_1}{\hat{\theta}_2}) = \exp(-\frac{-1.80}{2.42}) \approx 2.10$. The value is also computed for the years 1995 and 2000 in the same way.

2.09, the 39th most skilled-abundant) among the 115 countries. The threshold S/U^* is 2.25 for 1995, which is the closest to S/U in Greece (=2.25, the 29th out of 115); and is 2.32 for 2000 that is the closest to S/U in Taiwan (=2.31, the 26th). These values for the skill abundance can be interpreted as the cutoff to divide countries into the North and South for the respective years, which is more accurate than the cutoff value used in the previous subsection to divide the countries into the two groups.⁸⁶

3.4.4 Alternative Measure of Export Varieties

For checking the robustness of the results of our empirical tests, we also employ an alternative measure of product varieties in countries' exports that are frequently used in literature. Following Feenstra and Kee (2004a) and Hummels and Klenow (2005),⁸⁷ as an alternative to our original measure of export varieties n_share_{ic} , we use the following measure of “relative product variety” (Hummels & Klenow use the term of the “extensive margin”) in a country's export:

$$RV_{ic} \equiv \frac{\sum_{\omega \in \Omega_i^c} p_{\omega}^* x_{\omega}^*}{\sum_{\omega \in \Omega_i^*} p_{\omega}^* x_{\omega}^*}$$

The asterisk * denotes the “benchmark country” for comparison, which is the aggregate of all countries in the world.⁸⁸ ω denotes a ten-digit HS good; Ω_i^c is a subset of ten-digit HS goods belonging to Industry i (defined by the four-digit SIC or six-digit NAICS) that are exported from a particular country c to the U.S.; and Ω_i^* is a whole set of all the ten-digit HS goods in Industry i that are exported to the U.S. from all countries (other than the U.S. itself) in the world. p_{ω}^* and x_{ω}^* are the price and quantity of Product ω exported

⁸⁶ The average (S/U) of the 115 countries that is used as the cutoff in the previous subsection is 1.88, which is a little lower than these values.

⁸⁷ Broda and Weinstein (2006) also employ this measure of “relative variety.”

⁸⁸ This “benchmark” world aggregate includes countries other than the 115 countries in our data.

by the “benchmark country” (i.e., $p_{\omega}^* x_{\omega}^*$ is the value of the exports of Product ω from the world country aggregate to the U.S.).⁸⁹

We replace the dependent variable in Equation (3.4.4) for the pooled regression with this alternative measure of “relative variety” RV_{ic} , both in the natural scale and logarithm, and estimate the following resulted equations for each of the years 1990, 1995, and 2000 using the fixed-effect OLS with country dummies (μ_c):

$$RV_{ic} = \theta_1 \cdot skill_i + \theta_2 \cdot skill_i \cdot \log(S/U)_c + \mu_c + \varepsilon_{ic} \quad (3.4.5)$$

$$\log(RV_{ic}) = \theta_1 \cdot skill_i + \theta_2 \cdot skill_i \cdot \log(S/U)_c + \mu_c + \varepsilon_{ic} \quad (3.4.6)$$

The results are shown in Table 3.9. In both Equations (3.4.5) and (3.4.6), the estimate of the coefficient θ_1 is negative and the estimate of θ_2 is positive, both of which are significant at the 1% level, throughout the years. The results are consistent with the prediction from our model as the result in the previous subsection is, and thus confirm that the result of the empirical test is robust across measures of export varieties.

3.5 Conclusion

In this paper, we have investigated the relationship between export varieties and the exporting country’s comparative advantage in terms of factor proportions. We have generalized the heterogeneous-firm models by Melitz (2003) and Bernard, Redding, and Schott (2007) to the framework with multiple industries, and have derived a prediction that relates product varieties in a country’s exports to the relative factor intensity of exported industries. To test the prediction we have employed the disaggregated data on the U.S. imports, as well as the data on skill abundance in countries and the skill

⁸⁹ Note that this RV_{ic} is a value-based measure while our original measure n_share_{ic} is based on number counting. However, the two measures are similar in the sense that both define industries by the four-digit SIC or six-digit NAICS and product varieties by the ten-digit HS.

intensities of manufacturing industries. The results of a variety of empirical tests provide strong evidence for our semi-Heckscher-Ohlin prediction: countries tend to export more varieties of products in industries in which they have their respective comparative advantages.

3.6 Appendix

3.6.1 Derivation of Equations (3.2.20)

By combining the revenue equations (3.2.7) and (3.2.8) and the zero-profit conditions (3.2.10) and (3.2.11), we can derive the following equations for the ratio of the two productivity cutoffs for each of Home and Foreign:

$$\frac{\phi_{xi}^{*H}}{\phi_i^{*H}} = \tau_i \left(\frac{P_i^H}{P_i^F} \right) \left(\frac{Y^H}{Y^F} \cdot \frac{f_{xi}}{f_i} \right)^{\frac{1}{\sigma-1}} \quad (3.A.1)$$

$$\frac{\phi_{xi}^{*F}}{\phi_i^{*F}} = \tau_i \left(\frac{P_i^F}{P_i^H} \right) \left(\frac{Y^F}{Y^H} \cdot \frac{f_{xi}}{f_i} \right)^{\frac{1}{\sigma-1}} \quad (3.A.2)$$

The ratio of the industry price indexes in the two countries can be derived, by rearranging the Equations (3.A.1) and (3.A.2), as follows:

$$\frac{P_i^H}{P_i^F} = \tau_i^{-1} \left(\frac{\phi_{xi}^{*H}}{\phi_i^{*H}} \right) \left(\frac{Y^F}{Y^H} \right)^{\frac{1}{\sigma-1}} \left(\frac{f_i}{f_{xi}} \right)^{\frac{1}{\sigma-1}} = \tau_i \left(\frac{\phi_i^{*F}}{\phi_{xi}^{*F}} \right) \left(\frac{Y^F}{Y^H} \right)^{\frac{1}{\sigma-1}} \left(\frac{f_{xi}}{f_i} \right)^{\frac{1}{\sigma-1}} \quad (3.A.3)$$

Substituting this Equation (3.A.3) to Equation (3.2.19) and re-arranging yields the following:

$$\frac{M_i^H}{M_i^F} = \frac{\tau_i^{1-\sigma} \left[\left(\frac{Y^H}{Y^F} \right) \left(\frac{f_i}{f_{xi}} \right) \left(\frac{\phi_i^{*F}}{\phi_{xi}^{*F}} \right)^{1-\sigma} (p_i^F (\bar{\phi}_i^F))^{1-\sigma} - \chi_i^F (p_i^F (\bar{\phi}_{xi}^F))^{1-\sigma} \right]}{(p_i^H (\bar{\phi}_i^H))^{1-\sigma} - \chi_i^H \left(\frac{Y^H}{Y^F} \right) \left(\frac{f_{xi}}{f_i} \right) \left(\frac{\phi_{xi}^{*H}}{\phi_i^{*H}} \right)^{1-\sigma} (p_i^H (\bar{\phi}_{xi}^H))^{1-\sigma}} \quad (3.A.4)$$

The optimal pricing equation (3.2.5) implies that the ratio of the prices charged by two firms with different productivity *in the same market* can be expressed as the ratio of the two productivities, i.e.:

$$p_i(\phi') = \left(\frac{\phi}{\phi'}\right) \cdot p_i(\phi) \quad (3.A.5)$$

Using this, the price charged by a firm with the average exporter productivity *in the domestic market* is expressed, using the average price of domestic producers, as follows:

$$p_i^H(\bar{\phi}_{xi}^H) = \left(\frac{\bar{\phi}_i^H}{\bar{\phi}_{xi}^H}\right) \cdot p_i^H(\bar{\phi}_i^H) \quad (3.A.6)$$

$$p_i^F(\bar{\phi}_{xi}^F) = \left(\frac{\bar{\phi}_i^F}{\bar{\phi}_{xi}^F}\right) \cdot p_i^F(\bar{\phi}_i^F) \quad (3.A.7)$$

Substituting these equations (3.A.6) and (3.A.7) into Equation (3.A.4) and re-arranging the terms yields the following expression for the relative size of the masses of domestic producers:

$$\frac{M_i^H}{M_i^F} = \frac{\left[\left(\frac{Y^H}{Y^F}\right) \left(\frac{f_i}{f_{xi}}\right) \left(\frac{\phi_{xi}^{*F}}{\phi_i^{*F}}\right)^{\sigma-1} - \chi_i^F \left(\frac{\bar{\phi}_{xi}^F}{\bar{\phi}_i^F}\right)^{\sigma-1} \right]}{\tau_i^{\sigma-1} \left[1 - \chi_i^H \left(\frac{Y^H}{Y^F}\right) \left(\frac{f_{xi}}{f_i}\right) \left(\frac{\phi_i^{*H}}{\phi_{xi}^{*H}}\right)^{\sigma-1} \left(\frac{\bar{\phi}_{xi}^H}{\bar{\phi}_i^H}\right)^{\sigma-1} \right]} \cdot \left(\frac{p_i^H(\bar{\phi}_i^H)}{p_i^F(\bar{\phi}_i^F)} \right)^{\sigma-1} \quad (3.A.8)$$

Equation (3.2.20) is derived from this (3.A.8) and Equation (3.2.15).

3.6.2 Proof of Proposition

Without the loss of generality, Industry i is assumed to be more skill intensive than Industry j ($\beta_i > \beta_j$). Then, from the previous chapter II, the relationship of the probability for a domestic producer to be an exporter between the two industries is:

$\chi_i^H > \chi_j^H$ for the relatively skill abundant Home; and $\chi_i^F < \chi_j^F$ for the relatively

unskilled-labor abundant Foreign. These two inequalities is equivalent to the following inequality:

$$\frac{\chi_i^H}{\chi_i^F} > \frac{\chi_j^H}{\chi_j^F} \quad (3.A.9)$$

Recall now Equation (3.2.20) for the relative exporter mass:

$$\frac{M_{xi}^H}{M_{xi}^F} = \frac{\chi_i^H}{\chi_i^F} \cdot \frac{[(\frac{Y^H}{Y^F})(\frac{f_i}{f_{xi}})(\frac{\phi_{xi}^{*F}}{\phi_i^{*F}})^{\sigma-1} - \chi_i^F (\frac{\bar{\phi}_{xi}^F}{\phi_i^F})^{\sigma-1}]}{\tau_i^{\sigma-1} [1 - \chi_i^H (\frac{Y^H}{Y^F})(\frac{f_{xi}}{f_i})(\frac{\phi_i^{*H}}{\phi_{xi}^{*H}})^{\sigma-1} (\frac{\bar{\phi}_{xi}^H}{\phi_i^H})^{\sigma-1}]} \cdot \left(\frac{p_i^H (\bar{\phi}_i^H)}{p_i^F (\bar{\phi}_i^F)} \right)^{\sigma-1}$$

Let us rewrite this as follows:

$$\frac{M_{ix}^H}{M_{ix}^F} = \left(\frac{\chi_i^H}{\chi_i^F} \right) \cdot \frac{A_i}{\tau_i^{\sigma-1} \cdot B_i} \cdot \left(\frac{p_i^H (\bar{\phi}_i^H)}{p_i^F (\bar{\phi}_i^F)} \right)^{\sigma-1} \quad (3.A.10)$$

where $A_i \equiv (\frac{Y^H}{Y^F})(\frac{f_i}{f_{xi}})(\frac{\phi_{xi}^{*F}}{\phi_i^{*F}})^{\sigma-1} - \chi_i^F (\frac{\bar{\phi}_{xi}^F}{\phi_i^F})^{\sigma-1}$ and

$B_i \equiv 1 - \chi_i^H (\frac{Y^H}{Y^F})(\frac{f_{xi}}{f_i})(\frac{\phi_i^{*H}}{\phi_{xi}^{*H}})^{\sigma-1} (\frac{\bar{\phi}_{xi}^H}{\phi_i^H})^{\sigma-1}$. The relative exporter mass thus depends on the

ratio of the fractions of exporters among active firms in the two countries (χ_i^H / χ_i^F), the

terms A_i and B_i , and the relative average price of domestic products in the two countries

($p_i^H (\bar{\phi}_i^H) / p_i^F (\bar{\phi}_i^F)$). Let us first examine these four factors separately.

- $\frac{\chi_i^H}{\chi_i^F}$ vs $\frac{\chi_j^H}{\chi_j^F}$: As shown in the inequality (3.A.9), $\frac{\chi_i^H}{\chi_i^F} > \frac{\chi_j^H}{\chi_j^F}$.
- A_i vs A_j : As shown in Chapter II, the ratio of (or the ‘‘gap’’ between) the two productivity cutoffs is larger in the country’s comparative *disadvantage* industry.

Therefore, for the relatively skill-scarce Foreign, $\frac{\phi_{xi}^{*F}}{\phi_i^{*F}} < \frac{\phi_{xj}^{*F}}{\phi_j^{*F}}$. Hence, with

Assumption 1, the first term is larger in A_i than A_j .

To examine the second term, we first consider the ratio of the two productivity

averages, $\frac{\bar{\phi}_{xi}}{\bar{\phi}_i}$. From Assumption 2, the productivity distribution is the same across

industries and has a Pareto form. Therefore, by substituting the Pareto density function in Assumption 2 into the definition of these productivity averages (3.2.14), and with some algebra, we can show that the ratio of the two productivity averages

is indeed equal to the ratio of the two productivity cutoffs; i.e., $\frac{\bar{\phi}_{xi}}{\bar{\phi}_i} = \frac{\phi_{xi}^*}{\phi_i^*}$. In

addition, the Pareto assumption implies that $\chi_i = \left(\frac{\phi_{xi}^*}{\phi_i^*}\right)^{-k}$. Thus, the second term of

A_i equals $\left(\frac{\phi_{xi}^*}{\phi_i^*}\right)^{-(1+k-\sigma)}$ while the second term of A_j equals $\left(\frac{\phi_{xj}^*}{\phi_j^*}\right)^{-(1+k-\sigma)}$. Hence, for

the skill-scarce Foreign, the second term is smaller in A_i than in A_j (i.e.,

$$\frac{\phi_{xi}^{*F}}{\phi_i^{*F}} > \frac{\phi_{xj}^{*F}}{\phi_j^{*F}} \Leftrightarrow \left(\frac{\phi_{xi}^{*F}}{\phi_i^{*F}}\right)^{-(1+k-\sigma)} < \left(\frac{\phi_{xj}^{*F}}{\phi_j^{*F}}\right)^{-(1+k-\sigma)}.$$

These relationship of the two terms implies $A_i > A_j$.

- B_i vs B_j : Using the equalities $\frac{\bar{\phi}_{xi}}{\bar{\phi}_i} = \frac{\phi_{xj}^*}{\phi_j^*}$ and $\chi_i = \left(\frac{\phi_{xi}^*}{\phi_i^*}\right)^{-k}$ that are implied by

Assumption 2, we obtain $B_i = 1 - \chi_i^H \left(\frac{Y^H}{Y^F}\right) \left(\frac{f_{xi}}{f_i}\right)$ and $B_j = 1 - \chi_j^H \left(\frac{Y^H}{Y^F}\right) \left(\frac{f_{xj}}{f_j}\right)$. Since

Assumption 1 implies $\frac{f_{xi}}{f_i} = \frac{f_{xj}}{f_j}$, $\chi_i^H > \chi_j^H$ indicates $B_i < B_j$.

- $\frac{\bar{p}_i^H}{\bar{p}_i^F}$ vs $\frac{\bar{p}_j^H}{\bar{p}_j^F}$: From the optimal pricing Equation (3.2.5), the relative average price

depend on two factors: the ratio of the average productivity of active firms in the two countries, and the relative factor prices. The relative average price takes the following form:

$$\frac{\bar{p}_i^H}{\bar{p}_i^F} = \left(\frac{\bar{\phi}_i^F}{\bar{\phi}_i^H}\right) \left\{ \left(\frac{s^H / w^H}{s^F / w^F}\right)^{\beta_i} \left(\frac{w^H}{w^F}\right) \right\} \quad (3.A.11)$$

The second term of the right-hand side of (3.A.11) is smaller for Industry i since

$\frac{s^H}{w^H} < \frac{s^F}{w^F}$ (see Chapter II for the proof) and $\beta_i > \beta_j$. The first term equals one since

we assume a common productivity distribution across countries, which implies that

$\bar{\phi}_i^H = \bar{\phi}_i^F$. Hence, the relative price follows the comparative advantages of the

countries; i.e., $\frac{\bar{p}_i^H}{\bar{p}_i^F} < \frac{\bar{p}_j^H}{\bar{p}_j^F}$.

From these results of the four elements in Equation (3.A.10), it is shown that

$\frac{M_{ix}^H}{M_{ix}^F} > \frac{M_{jx}^H}{M_{jx}^F}$, which implies that each country has a larger mass of exporters in its

comparative advantage industries relative to the other country. ■

Table 3.1: U.S. Imports and Varieties

		Total Import	Manufacturing Import
Number of Exporting Countries	1990	153	153
	1995	169	169
	2000	174	173
Number of Varieties	1990	182,375	171,322
	1995	219,329	206,334
	2000	259,181	243,598
Total Import Value (in million \$)	1990	495,260	409,953
	1995	743,505	643,128
	2000	1,216,888	1,024,664

Notes:

1. The data are from Feenstra, Romalis, and Schott (2002).
2. Manufacturing imports are the imports in the industries classified as the 4-digit 1987 U.S. SIC 2011 through 3999 (for 1990) or the 6-digit NAICS 311111 through 339999 (for 1995 and 2000).
3. Exporting countries in this table include overseas territories of countries.
4. The number of varieties is defined as the number of goods classified by the 10-digit Harmonization System (HS) that the U.S. imports *from each exporter*. (I.e., the same 10-digit HS goods imported from different exporters are counted as different varieties.)
5. Import value is the customs value of general imports. “General Imports measure the total physical arrivals of merchandise from foreign countries, whether such merchandise enters consumption channels immediately or is entered into bonded warehouses or Foreign Trade Zones under Customs custody” (U.S. International Trade Administration).

Table 3.2: Country List (115 countries)

Algeria	Guinea	Peru
Angola	Guinea-Bissau	Philippines
Argentina	Guyana	Poland
Australia	Haiti	Portugal
Austria	Honduras	Reunion
Bangladesh	Hong Kong	Rwanda
Barbados	Hungary	Saudi Arabia
Belgium	Iceland	Senegal
Benin	India	Seychelles
Bolivia	Indonesia	Sierra Leone
Brazil	Iran	Singapore
Burkina Faso	Ireland	Somalia
Burundi	Israel	South Africa
Cameroon	Italy	South Korea
Canada	Jamaica	Spain
Central African Republic	Japan	Sri Lanka
Chad	Jordan	Sudan
Chile	Kenya	Suriname
China	Madagascar	Sweden
Colombia	Malawi	Switzerland
Congo	Malaysia	Syria
Costa Rica	Mali	Taiwan
Cote d'Ivoire	Malta	Tanzania
Cyprus	Mauritania	Thailand
Czechoslovakia*	Mauritius	Togo
Denmark	Mexico	Trinidad and Tobago
Dominican Republic	Morocco	Tunisia
Ecuador	Mozambique	Turkey
Egypt	Netherlands	U.S.S.R.*
El Salvador	New Zealand	Uganda
Fiji	Nicaragua	United Kingdom
Finland	Niger	Uruguay
France	Nigeria	Venezuela
Gabon	Norway	Yugoslavia*
Gambia	Oman	Zaire
Germany	Pakistan	Zambia
Ghana	Panama	Zimbabwe
Greece	Papua New Guinea	
Guatemala	Paraguay	

Note: The data for Years 1995 and 2000 do not include three countries marked with an asterisk (*).

Table 3.3: Factor Abundance of Countries: Skilled Labor (S) to Unskilled Labor (U) Ratio

Summary Statistics:

Variables	Mean	Std. Dev.	Min.	Max.
S/U	1.879	0.553	1.075	3.369
$\log(S/U)$	0.589	0.290	0.072	1.215

Number of countries: 115

10 most skilled-labor abundant countries:

Country	S/U	$\log(S/U)$
New Zealand	3.369	1.215
Hungary	3.086	1.127
Norway	3.010	1.102
Canada	3.008	1.101
Denmark	2.999	1.098
Australia	2.981	1.092
Finland	2.833	1.041
Sweden	2.825	1.039
Israel	2.818	1.036
Belgium	2.768	1.018

10 most unskilled-labor abundant countries:

Country Name	S/U	$\log(S/U)$
Niger	1.075	0.072
Guinea-Bissau	1.078	0.075
Benin	1.098	0.094
Mali	1.116	0.110
Rwanda	1.119	0.113
Gambia	1.119	0.113
Sudan	1.130	0.122
Mozambique	1.156	0.145
Central African Republic	1.184	0.169
Nigeria	1.217	0.196

Note: The relative abundance of skilled labor to unskilled labor (S/U) is measured as the human capital to labor ratio provided by Hall and Jones (1999).

Table 3.4.1: Relative Skilled-labor (*S*) and Unskilled-labor (*U*) Intensity of Manufacturing Industries: for 4-digit U.S. SIC Industries, Year 1992

Summary Statistics:

Variables	Mean	Std. Dev.	Min.	Max.
<i>S</i> -intensity	0.296	0.124	0.078	0.827
<i>U</i> -intensity	0.704	0.124	0.173	0.922

Number of manufacturing industries: 394

10 Most Skilled-labor Intensive Industries

SIC	Industry Description	<i>S</i> -intensity	<i>U</i> -intensity
2721	Periodicals	0.827	0.173
2731	Book Publishing	0.766	0.234
3571	Electronic Computers	0.718	0.282
3761	Guided Missiles & Space Vehicles	0.685	0.315
2711	Newspapers	0.676	0.324
2741	Miscellaneous Publishing	0.638	0.362
2835	Diagnostic Substances	0.633	0.367
3572	Computer Storage Devices	0.627	0.373
3826	Analytical Instruments	0.617	0.383
2086	Bottled and Canned Soft Drinks	0.604	0.396

10 Most Unskilled-labor Intensive Industries

SIC	Industry Description	<i>S</i> -intensity	<i>U</i> -intensity
2322	Men's & Boys' Underwear & Nightwear	0.078	0.922
2281	Yarn Spinning Mills	0.089	0.911
2284	Thread Mills	0.097	0.903
2211	Weaving Mills, Cotton	0.102	0.898
2436	Softwood Veneer and Plywood	0.105	0.895
2015	Poultry and Egg Processing	0.108	0.892
3263	Fine Earthenware Food Utensils	0.111	0.889
2325	Men's & Boys' Trousers & Slacks	0.116	0.884
2321	Shirts, Men's and Boys'	0.120	0.880
3144	Women's Footwear, Except Athletic	0.120	0.880

Notes:

1. The data for factor intensity is from the 1992 U.S. Census of Manufactures.
2. Industries are classified according to the 4-digit U.S. Standard Industrial Classification (SIC; 1987 version).
3. Skilled-labor (*S*) intensity is defined as the share of non-production workers in the total employment; and unskilled-worker (*U*) intensity is defined as the share of production workers. The sum of *S*-intensity and *U*-intensity is thus one for each industry.

Table 3.4.2: Relative Skilled-labor (*S*) and Unskilled-labor (*U*) Intensity of Manufacturing Industries: for 6-digit NAICS Industries, Year 1997

Summary Statistics:

Variables	Mean	Std. Dev.	Min.	Max.
<i>S</i> -intensity	0.285	0.111	0.095	0.654
<i>U</i> -intensity	0.715	0.111	0.346	0.905

Number of manufacturing industries: 383

10 Most Skilled-labor Intensive Industries

NAICS	Industry Description	<i>S</i> -intensity	<i>U</i> -intensity
334511	Search, detection, navigation, & guidance instrument	0.654	0.346
336414	Guided missile & space vehicle	0.640	0.360
334111	Electronic computer	0.639	0.361
334516	Analytical laboratory instrument	0.629	0.371
334210	Telephone apparatus	0.596	0.404
332995	Other ordnance & accessories	0.594	0.406
334517	Irradiation apparatus	0.582	0.418
312112	Bottled water	0.579	0.421
312111	Soft drink	0.568	0.432
334119	Other computer peripheral equipment	0.562	0.438

10 Most Unskilled-labor Intensive Industries

NAICS	Industry Description	<i>S</i> -intensity	<i>U</i> -intensity
321212	Softwood veneer & plywood	0.095	0.905
313111	Yarn spinning mills	0.098	0.902
315221	Men's & boys' cut & sew underwear & nightwear	0.098	0.902
315224	Men's & boys' cut & sew trouser, slack, & jean	0.104	0.896
313113	Thread mills	0.107	0.893
311615	Poultry processing	0.109	0.891
327213	Glass container	0.118	0.882
335222	Household refrigerator & home freezer	0.125	0.875
335224	Household laundry equipment	0.127	0.873
321211	Hardwood veneer & plywood	0.128	0.872

Notes:

1. The data for factor intensity is from the 1997 U.S. Census of Manufactures.
2. Industries are classified according to the 6-digit 1997 North American Industry Classification System (NAICS).
3. Skilled-labor (*S*) intensity is defined as the share of non-production workers in the total employment; and unskilled-worker (*U*) intensity is defined as the share of production workers. The sum of *S*-intensity and *U*-intensity is thus one for each industry.

Table 3.4.3: Relative Skilled-labor (*S*) and Unskilled-labor (*U*) Intensity of Manufacturing Industries: for 6-digit NAICS Industries, Year 2002

Summary Statistics:

Variables	Mean	Std. Dev.	Min.	Max.
<i>S</i> -intensity	0.301	0.120	0.087	0.711
<i>U</i> -intensity	0.699	0.120	0.289	0.913

Number of manufacturing industries: 384

10 Most Skilled-labor Intensive Industries

NAICS	Industry Description	<i>S</i> -intensity	<i>U</i> -intensity
334210	Telephone apparatus	0.711	0.289
334111	Electronic computer	0.704	0.296
334119	Other computer peripheral equipment	0.670	0.330
334511	Search, detection, navigation, & guidance instrument	0.666	0.334
334517	Irradiation apparatus	0.664	0.336
334516	Analytical laboratory instrument	0.662	0.338
334515	Electricity measuring & testing instrument	0.660	0.340
333295	Semiconductor machinery	0.639	0.361
336414	Guided missile & space vehicle	0.628	0.372
336415	Guided missile & space vehicle propulsion unit & parts	0.619	0.381

10 Most Unskilled-labor Intensive Industries

NAICS	Industry Description	<i>S</i> -intensity	<i>U</i> -intensity
321212	Softwood veneer & plywood	0.087	0.913
313111	Yarn spinning mills	0.101	0.899
311615	Poultry processing	0.108	0.892
335222	Household refrigerator & home freezer	0.126	0.874
336111	Automobile	0.130	0.870
327213	Glass container	0.131	0.869
313210	Broadwoven fabric mills	0.131	0.869
321113	Sawmills	0.135	0.865
311611	Animal (except poultry) slaughtering	0.135	0.865
311411	Frozen fruit, juice, & vegetable	0.139	0.861

Notes:

1. The data for factor intensity is from the 2002 U.S. Census of Manufactures.
2. Industries are classified according to the 6-digit 2002 North American Industry Classification System (NAICS).
3. Skilled-labor (*S*) intensity is defined as the share of non-production workers in the total employment; and unskilled-worker (*U*) intensity is defined as the share of production workers. The sum of *S*-intensity and *U*-intensity is thus one for each industry.

Table 3.5: List of Countries in Aggregate North and South

North (51 countries)		South (64 Countries)	
Argentina ^{#,##}	Norway	Algeria	Mali
Australia	Panama ^{#,##}	Angola	Mauritania
Austria ^{#,##}	Peru ^{#,##}	Bangladesh	Mauritius
Barbados	Philippines ^{#,##}	Benin	Mexico
Belgium	Poland	Bolivia	Mozambique
Canada	South Korea	Brazil	Nicaragua
Chile ^{#,##}	South Africa ^{#,##}	Burkina Faso	Niger
China ^{#,##}	Spain ^{#,##}	Burundi	Nigeria
Costa Rica ^{#,##}	Sri Lanka ^{#,##}	Cote d'Ivoire	Oman
Cyprus	Sweden	Cameroon	Pakistan
Czechoslovakia [*]	Switzerland	Central African Republic	Papua New Guinea
Denmark	Taiwan ^{##}	Chad	Paraguay
Ecuador ^{#,##}	Thailand ^{#,##}	Colombia	Portugal
Egypt ^{#,##}	Trinidad and Tobago ^{#,##}	Congo	Reunion
Fiji ^{##}	United Kingdom	Dominican Republic	Rwanda
Finland	Uruguay ^{#,##}	El Salvador	Saudi Arabia
France ^{#,##}	U.S.S.R. [*]	Gabon	Senegal
Germany	Venezuela ^{#,##}	Gambia	Seychelles
Greece ^{##}	Yugoslavia [*]	Ghana	Sierra Leone
Guyana ^{#,##}		Guatemala	Singapore
Hong Kong		Guinea	Somalia
Hungary		Guinea-Bissau	Sudan
Iceland		Haiti	Suriname
Ireland		Honduras	Syria
Israel		India	Togo
Italy ^{#,##}		Indonesia	Tunisia
Japan		Iran	Turkey
Malaysia ^{#,##}		Jamaica	Uganda
Malta ^{##}		Jordan	Tanzania
Morocco ^{#,##}		Kenya	Zaire
Netherlands		Madagascar	Zambia
New Zealand		Malawi	Zimbabwe

Notes:

1. The aggregate North consists of countries whose skilled-to-unskilled labor ratio (S/U) is above the average of the 115 countries (1.879); and the aggregate South consists of countries with S/U below the average.
2. Countries marked with # are grouped into the South if the 75 percentile value of S/U is applied to the North-South cutoff (22 countries in the North and 93 in the South); and countries with ## are grouped into the South if the 0.7 of the U.S. relative factor endowment (S/U) is applied to the cutoff (26 in the North and 89 in the South).
3. Countries marked with * are not included in the data for Years 1995 and 2000.

Table 3.6: Skilled-to-Unskilled Labor Ratios (S/U) of North and South

	S/U (average within group)	$\log(S/U)$ (average within group)
North	2.40	0.862
South	1.47	0.371

Notes:

1. Human capital to labor ratio in Hall and Jones (1999) is used as the measure of the relative factor abundance, or the ratio of skilled- to unskilled-labor (S/U), for each country.
2. The North comprises 51 countries that have the highest S/U , and the South comprises 64 countries with the lowest S/U . See Table 3.5 for the list of the countries in each group.

Table 3.7: Regressions for Aggregate North and South

Dependent Variable: Log of aggregate no. of varieties as the share in the total no. of varieties imported by the U.S. ($\log(n_share_{i,A})$)

Year 1990:	North	South
Industry skill intensity ($skill_i$)	0.260 ^{***} (0.049)	-1.21 ^{***} (0.201)
Constant	-0.275 ^{***} (0.016)	-1.56 ^{***} (0.061)
Observations	394	385
R ²	0.08	0.12
Year 1995:	North	South
Industry skill intensity ($skill_i$)	0.260 ^{***} (0.048)	-1.25 ^{***} (0.186)
Constant	-0.326 ^{***} (0.015)	-1.48 ^{***} (0.057)
Observations	383	378
R ²	0.08	0.12
Year 2000:	North	South
Industry skill intensity ($skill_i$)	0.131 [*] (0.070)	-0.866 ^{***} (0.143)
Constant	-0.325 ^{***} (0.020)	-1.52 ^{***} (0.046)
Observations	384	379
R ²	0.02	0.10

Notes:

1. Regressions estimate Equation (3.4.1) in the text for each year.
2. Robust standard errors are in parentheses.
3. ***, **, and * indicate that the coefficient estimate is significant at the 1% level, 5% level, and 10% level, respectively.

Table 3.8: Pooled Regressions for Individual Exporters

Dependent Variable: Log of no. of exported varieties in each industry as the share in the total no. of varieties imported by the U.S. ($\log(n_share_{ic})$)

Year 1990:

$skill_i$	-1.80 ^{***} (0.433)
$skill_i * \log(S/U)_c$	2.42 ^{***} (0.522)
Observations	17,050
R ²	0.15

Year 1995:

$skill_i$	-2.38 ^{***} (0.456)
$skill_i * \log(S/U)_c$	2.93 ^{***} (0.537)
Observations	17,469
R ²	0.17

Year 2000:

$skill_i$	-1.98 ^{***} (0.363)
$skill_i * \log(S/U)_c$	2.35 ^{***} (0.450)
Observations	19,037
R ²	0.18

Notes:

1. Regressions estimate Equation (3.4.4) in the text for each year. Country-specific dummies are included.
2. $skill_i$ is skilled-labor intensity of each industry, and $(S/U)_c$ is skilled-to-unskilled labor endowment ratio in each country.
3. Standard errors in parentheses are clustered by country.
4. ***, **, and * indicate that the coefficient estimate is significant at the 1% level, 5% level, and 10% level, respectively.

Table 3.9: Pooled Regressions using Alternative Measure of Export Varieties

Dependent Variable: Measure of “Relative Product Variety” in exports (RV_{ic}), in natural scale or logarithm

Year 1990:	RV_{ic}	$\log(RV_{ic})$
$skill_i$	-0.030*** (0.007)	-1.63*** (0.501)
$skill_i * \log(S/U)_c$	0.044*** (0.011)	2.37*** (0.629)
Observations	17,050	17,048
R ²	0.06	0.08
Year 1995:		
$skill_i$	-0.027*** (0.005)	-2.13*** (0.572)
$skill_i * \log(S/U)_c$	0.037*** (0.007)	2.90*** (0.671)
Observations	17,469	17,469
R ²	0.07	0.10
Year 2000:		
$skill_i$	-0.015*** (0.004)	-1.26*** (0.454)
$skill_i * \log(S/U)_c$	0.019*** (0.007)	1.89*** (0.573)
Observations	19,037	19,036
R ²	0.07	0.10

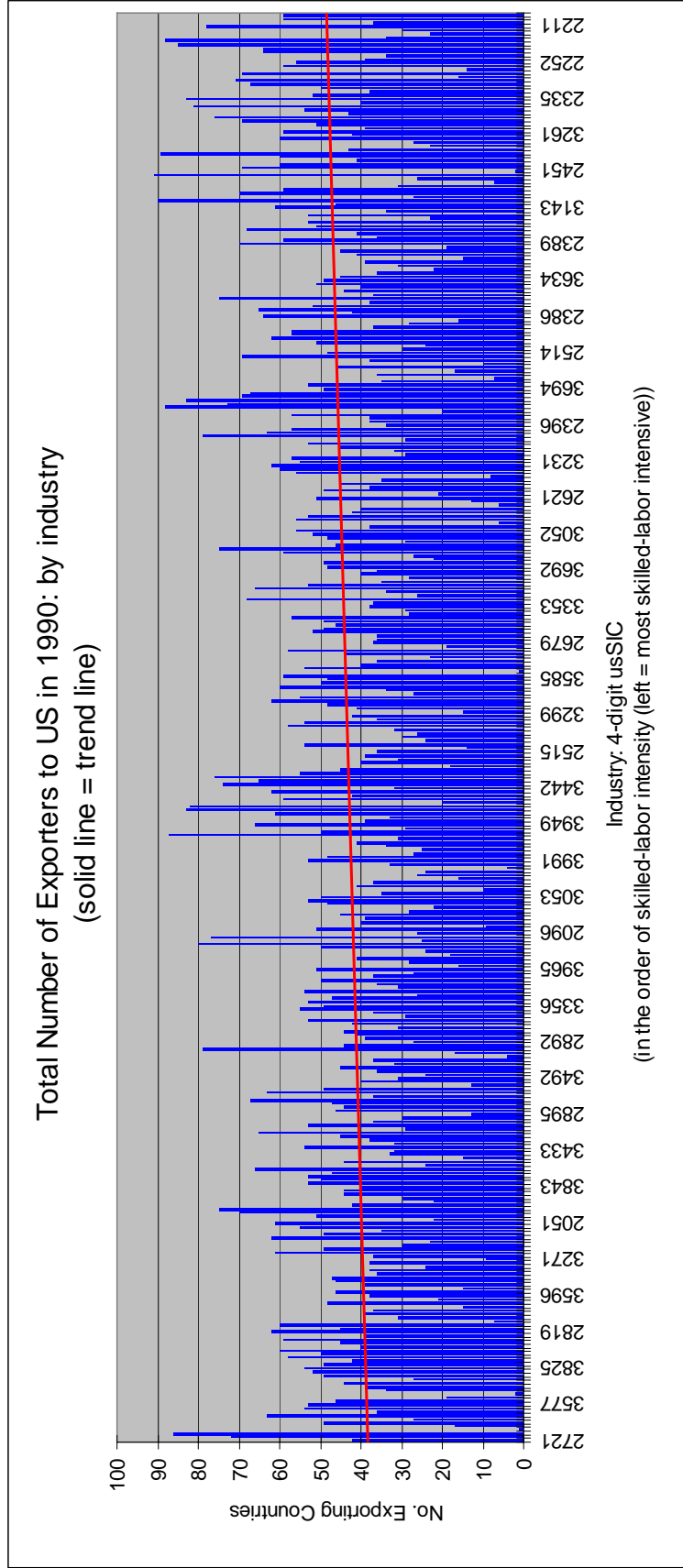
Notes:

1. The measure of relative product variety is defined as follows:

$$RV_{ic} \equiv \frac{\sum_{\omega \in \Omega_i^c} p_{\omega}^* x_{\omega}^*}{\sum_{\omega \in \Omega_i^*} p_{\omega}^* x_{\omega}^*}$$

2. Regressions estimate Equations (3.4.5) and (3.4.6) in the text for each year. Country-specific dummies are included.
3. $skill_i$ is skilled-labor intensity of each industry, and $(S/U)_c$ is skilled-to-unskilled labor endowment ratio in each country.
4. Standard errors in parentheses are clustered by country.
5. ***, **, and * indicate that the coefficient estimate is significant at the 1% level, 5% level, and 10% level, respectively.

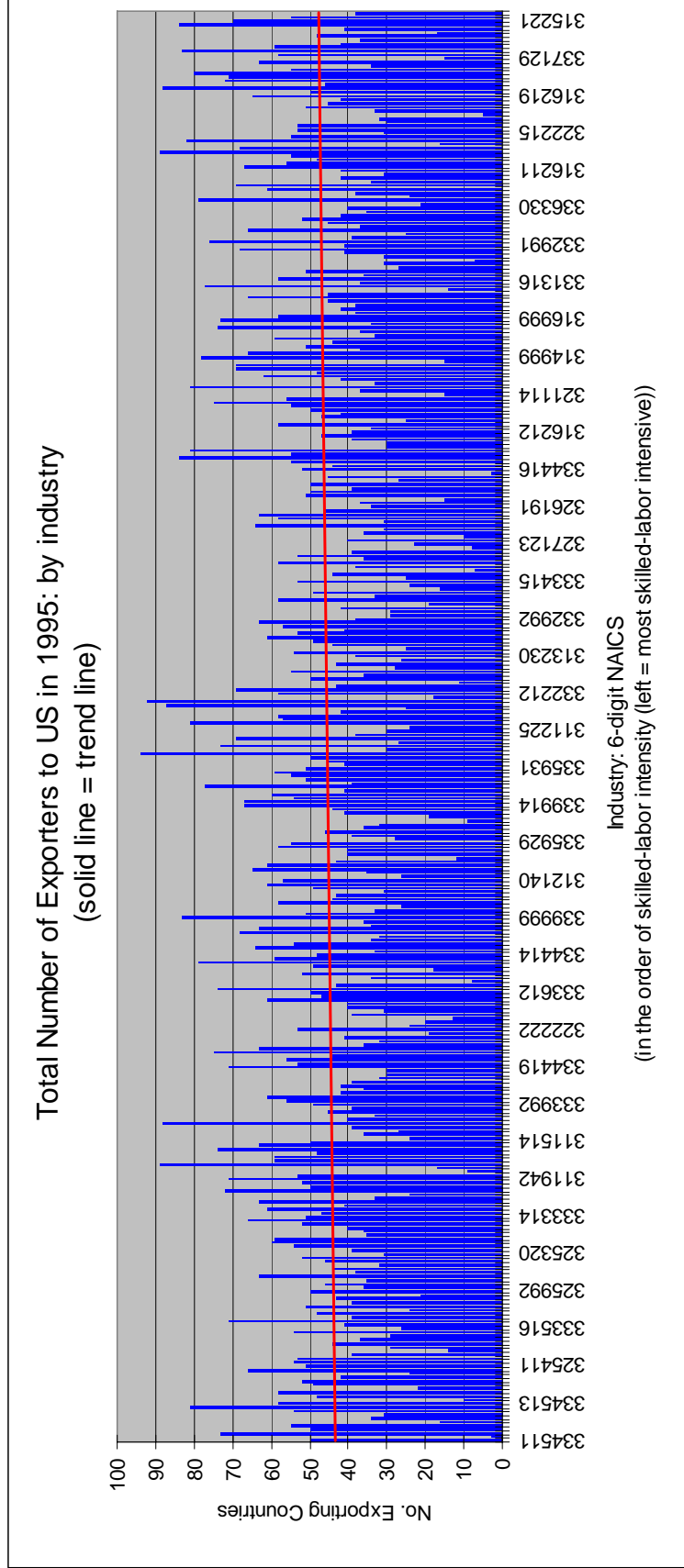
Figure 3.1.1: Number of Exporters to the U.S. in Each Manufacturing Industry; Year 1990



Notes:

1. Industries are classified according to the 4-digit 1987 U.S. Standard Industrial Classification (SIC).
2. 394 manufacturing industries are listed in the order of skilled-labor intensity; the left is the most skilled-labor intensive, and the right is the least.
3. Skilled-labor intensity is defined as the share of non-production workers in the total number of employees.

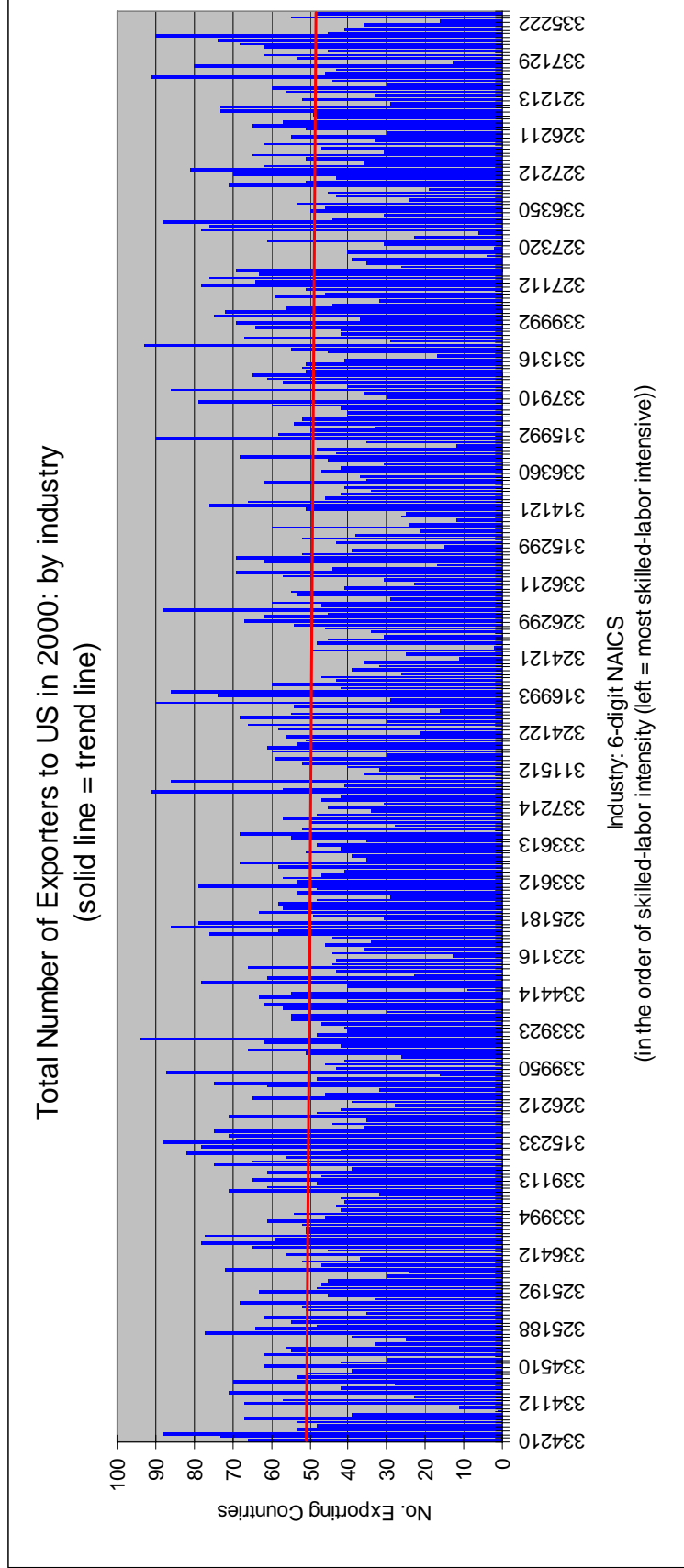
Figure 3.1.2: Number of Exporters to the U.S. in Each Manufacturing Industry; Year 1995



Notes:

1. Industries are classified according to the 6-digit 1997 North American Industry Classification System (NAICS).
2. 383 manufacturing industries are listed in the order of skilled-labor intensity; the left is the most skilled-labor intensive, and the right is the least.
3. Skilled-labor intensity is defined as the share of non-production workers in the total number of employees.

Figure 3.1.3: Number of Exporters to the U.S. in Each Manufacturing Industry; Year 2000



Notes:

1. Industries are classified according to the 6-digit 2002 North American Industry Classification System (NAICS).
2. 384 manufacturing industries are listed in the order of skilled-labor intensity; the left is the most skilled-labor intensive, and the right is the least.
3. Skilled-labor intensity is defined as the share of non-production workers in the total number of employees.

Figure 3.2.1: Number of Exporters vs Industry Skilled-labor Intensity in the U.S. Manufacturing Imports; Year 1990

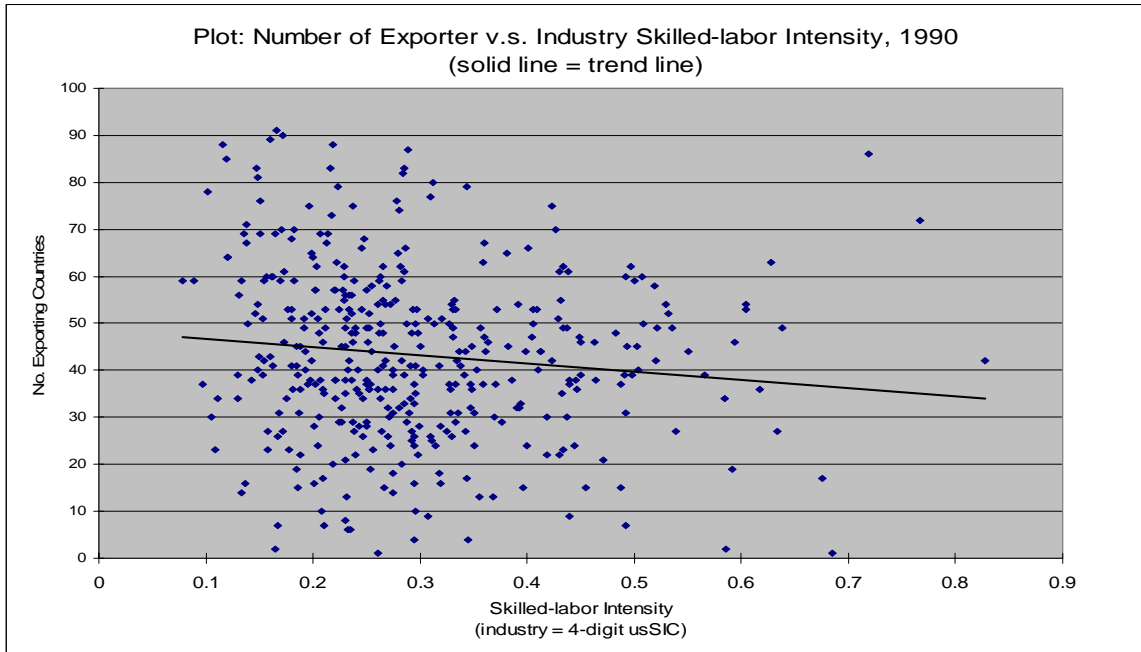


Figure 3.2.2: Number of Exporters vs Industry Skilled-labor Intensity in the U.S. Manufacturing Imports; Year 1995

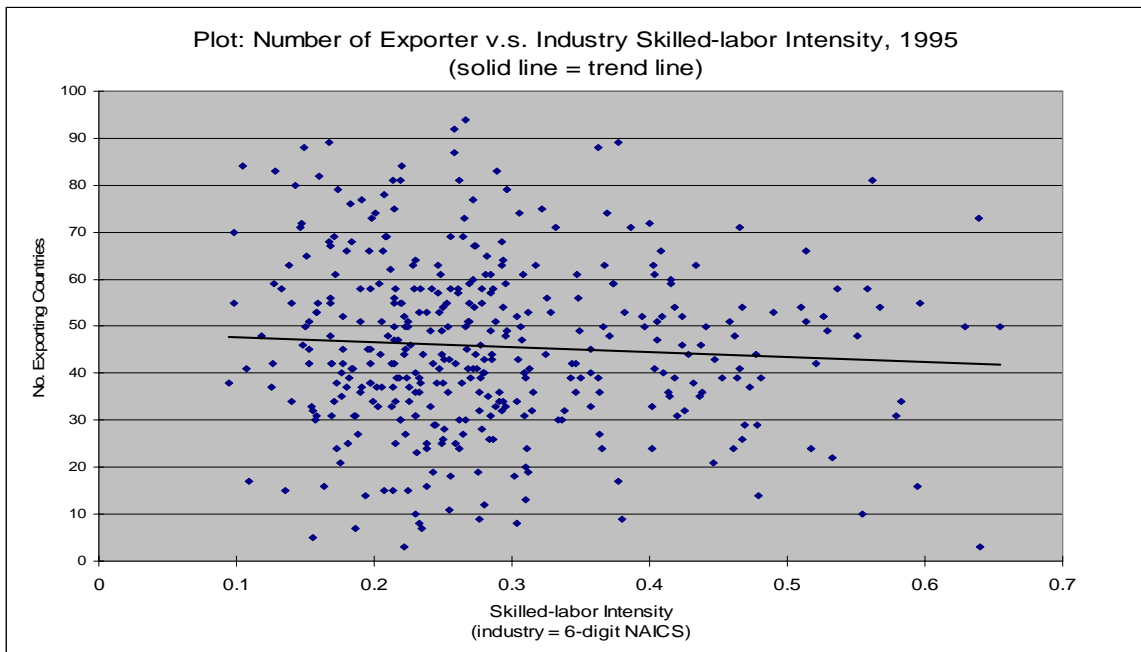
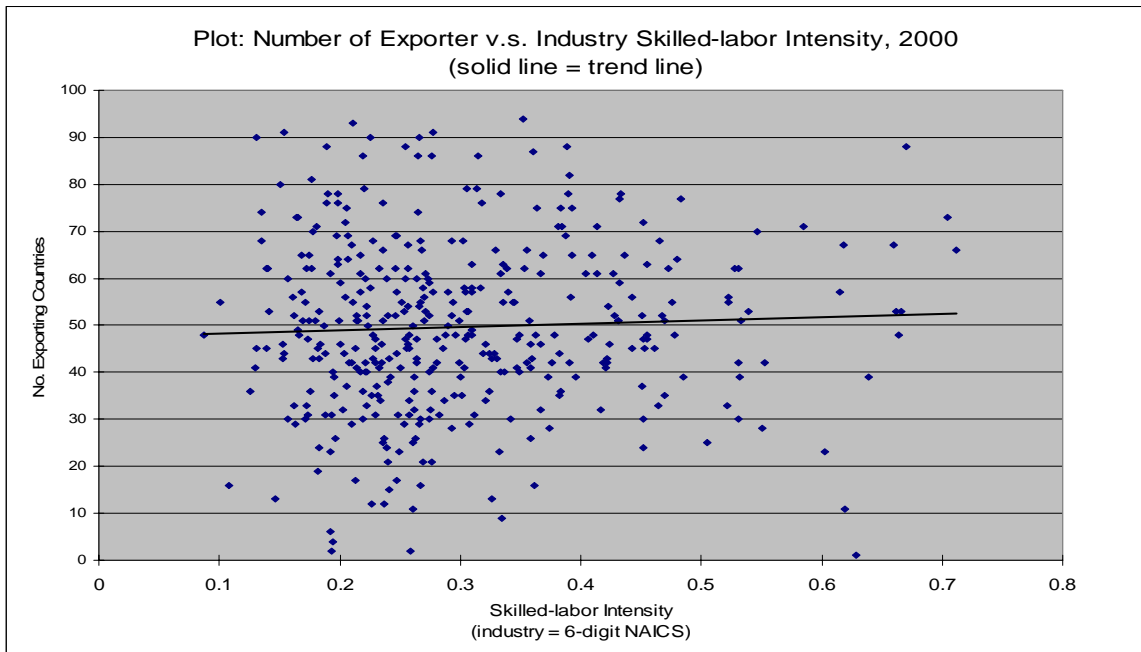


Figure 3.2.3: Number of Exporters vs Industry Skilled-labor Intensity in the U.S. Manufacturing Imports; Year 2000



Notes on Figures 3.2.1 through 3.2.3:

1. Manufacturing industries are classified according to the 4-digit 1987 U.S. SIC for the year 1990, and according to the 6-digit 1997 NAICS for the years 1995 and 2000.
2. The number of exporters is the number of countries from which the United States imports in each manufacturing industry.
3. Skilled-labor intensity is defined as the share of non-production workers in the total number of employees in each industry.

Figure 3.3.1: Number of Varieties vs Industry Skilled-labor Intensity in the U.S. Manufacturing Imports; Year 1990

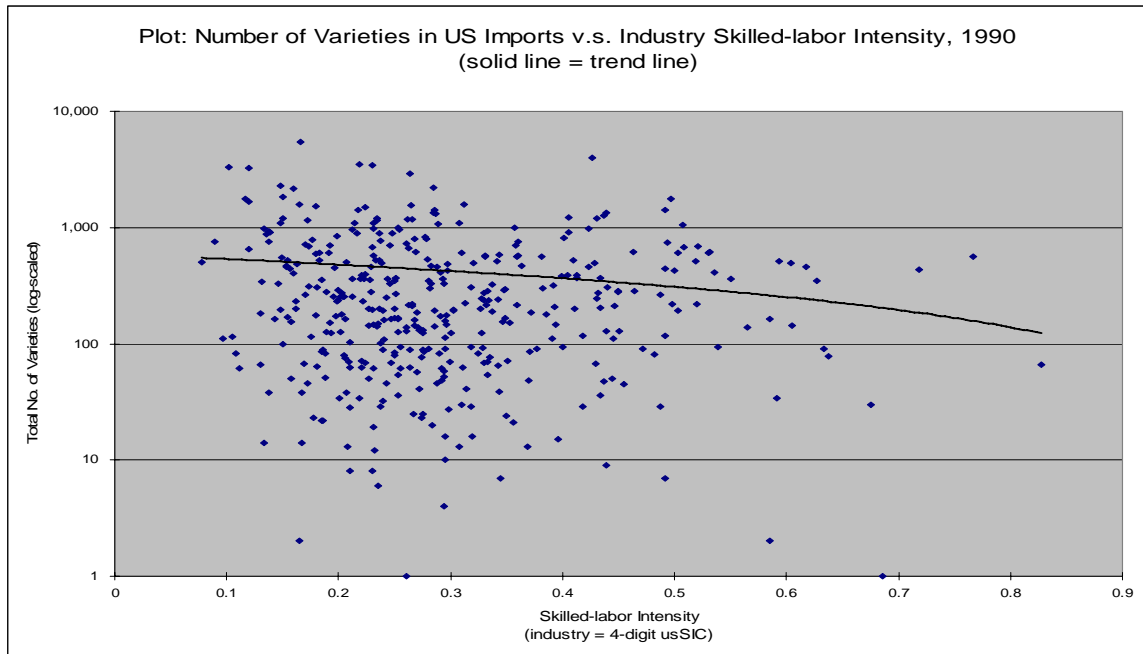


Figure 3.3.2: Number of Varieties vs Industry Skilled-labor Intensity in the U.S. Manufacturing Imports; Year 1995

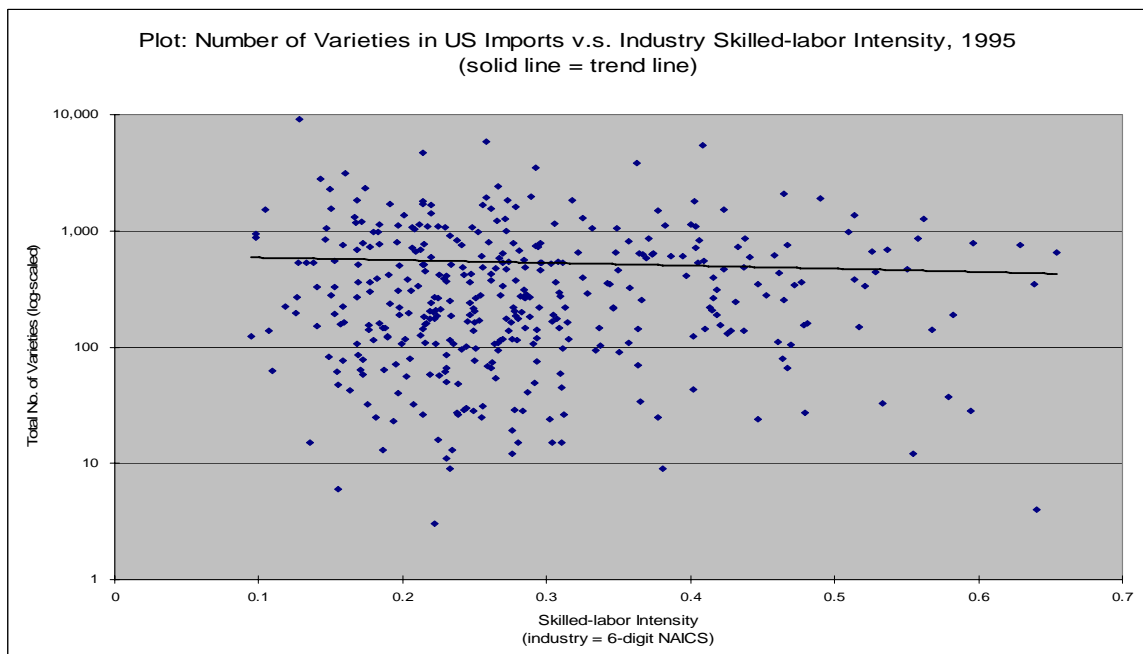
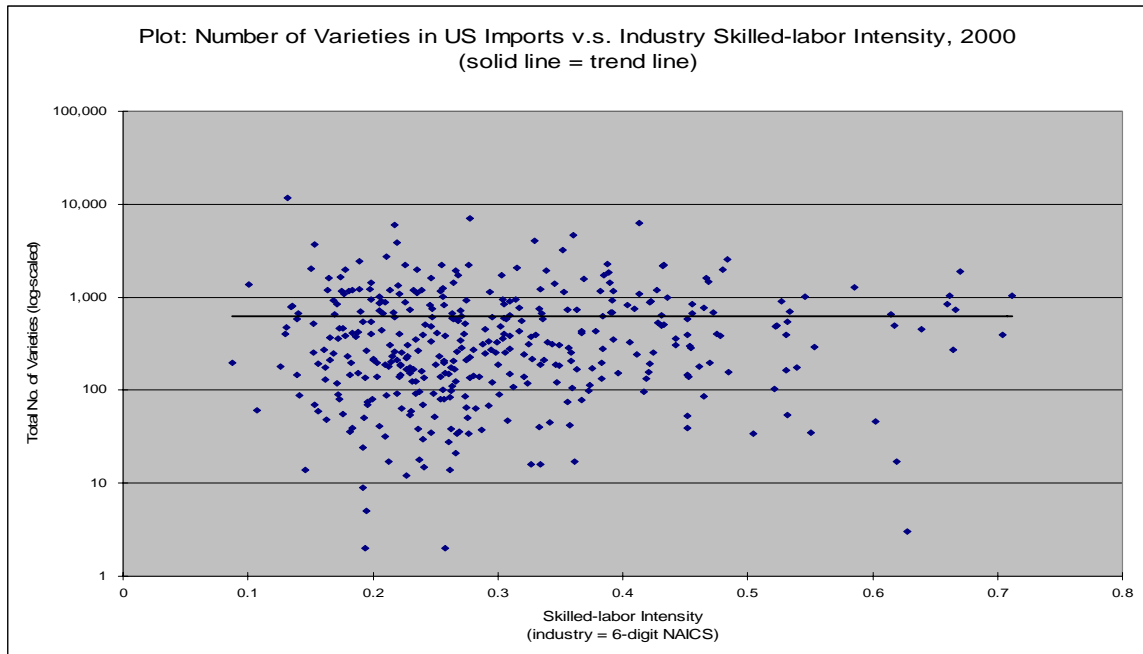


Figure 3.3.3: Number of Varieties vs Industry Skilled-labor Intensity in the U.S. Manufacturing Imports; Year 2000



Notes on Figures 3.3.1 through 3.3.3:

1. Manufacturing industries are classified according to the 4-digit 1987 U.S. SIC for the year 1990, and according to the 6-digit 1997 NAICS for the years 1995 and 2000.
2. The number of varieties in each industry is defined as the number of 10-digit HS goods that the U.S. imports *from each country* in each 4-digit SIC industry (i.e., the same 10-digit HS products imported from different countries are counted as different varieties).
3. Skilled-labor intensity is defined as the share of non-production workers in the total number of employees in each industry.

Figure 3.4.1: Individual Exporter Country Regression for 1990: Slope Coefficient vs Skill Abundance of the Country

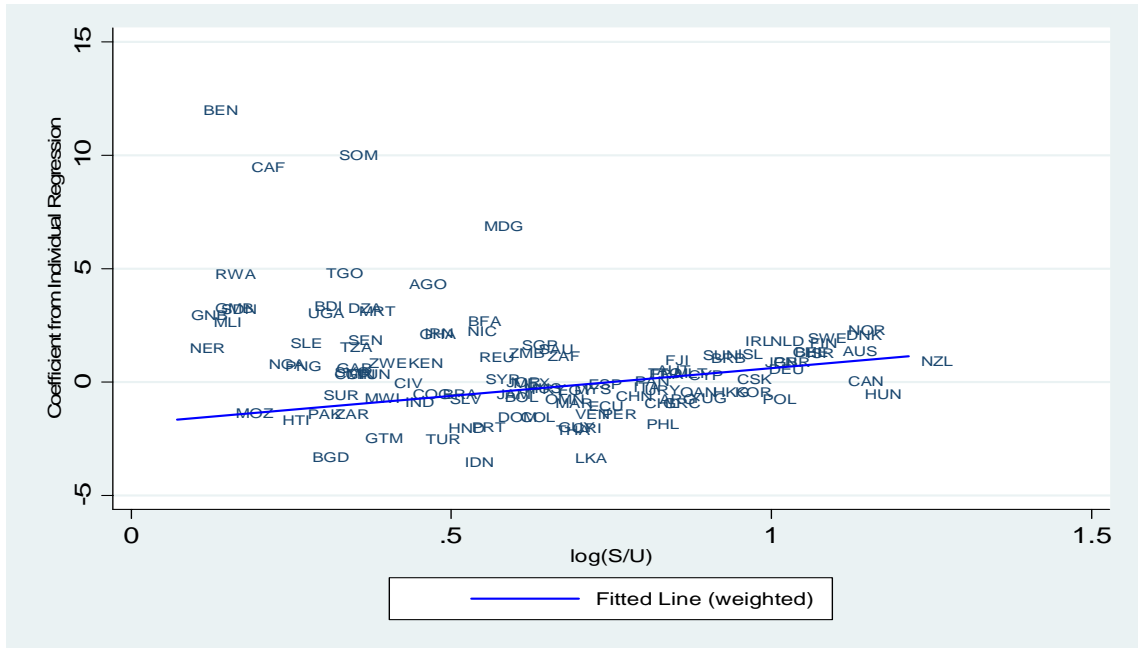


Figure 3.4.2: Individual Exporter Country Regression for 1995: Slope Coefficient vs Skill Abundance of the Country

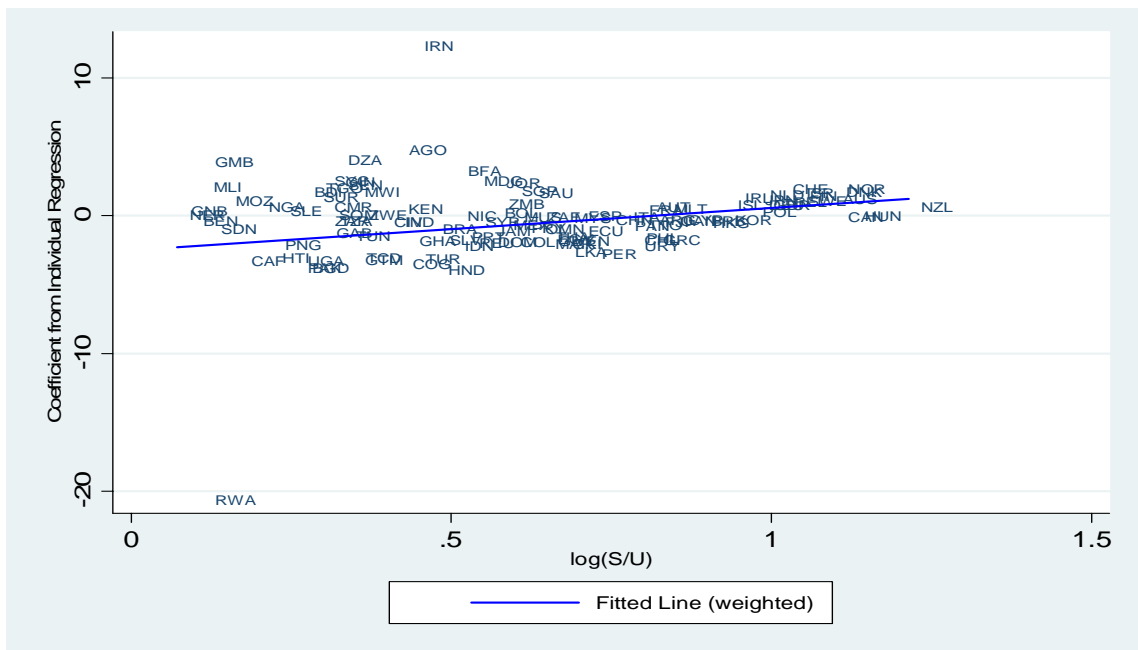
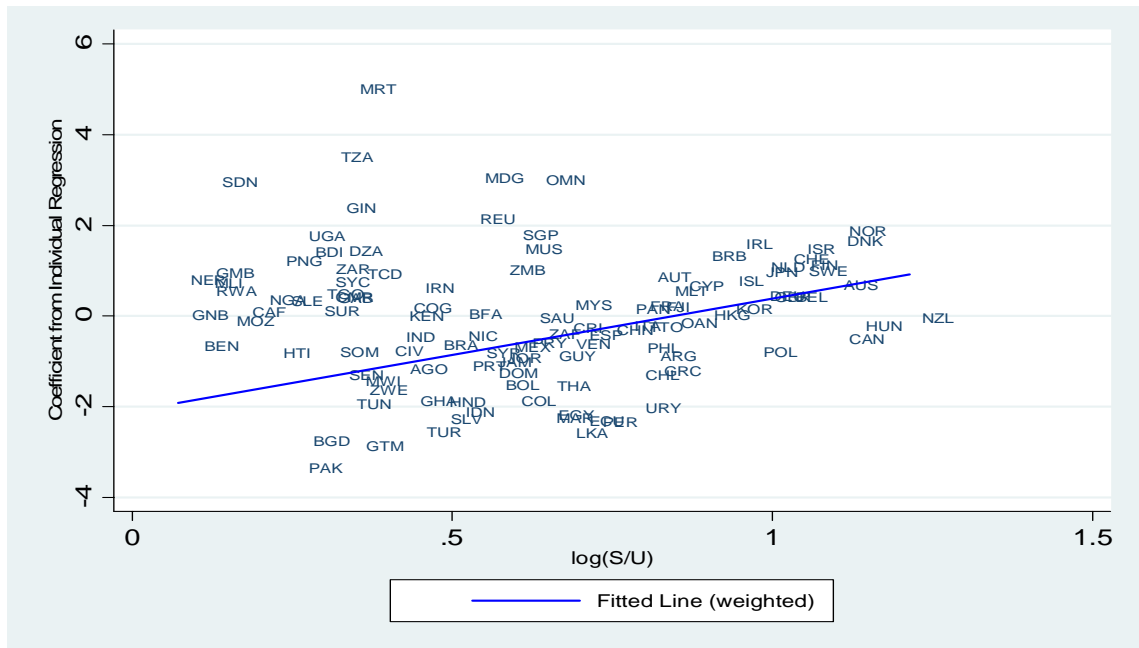


Figure 3.4.3: Individual Exporter Country Regression for 2000: Slope Coefficient vs Skill Abundance of the Country



Notes on Figures 3.4.1 through 3.4.3:

1. The individual regressions estimate the equation $\log(n_share_{i,c}) = \gamma_c + \Pi_c skill_i + \varepsilon_{i,c}$, where i indexes 4-digit SIC industries (for the year 1990) or 6-digit NAICS industries (for the years 1995 and 2000), and c indexes exporter countries. The regression is performed for each country to estimate the country-specific slope coefficient $\hat{\Pi}_c$ for each year.
2. The figures plot $\hat{\Pi}_c$ for each country (marked by the ISO country code) against the skilled-labor to unskilled-labor ratio of the country ($(S/U)_c$) in logarithm.
3. The fitted line in each figure is based on the weighted regression of $\hat{\Pi}_c$ on $\log(S/U)_c$ with the observations weighted by the number of 4-digit SIC industries for each country in the sample. (That is, the weight is the number of observation used for each individual country regression.)

CHAPTER IV

Revisiting the Revisited: An Alternative Test of the Monopolistic Competition Model of International Trade

4.1 Introduction

New Trade Theory is characterized by a model of international trade with monopolistic competition among the varieties of differentiated products in an industry. This theory was originally motivated by the fact that a large part of international trade is intra-industry rather than inter-industry,^{90,91} a characteristic that neo-classical trade theory such as the Heckscher-Ohlin (H-O) Model or the Ricardian Model cannot explain. The monopolistic competition models of international trade, first presented in the works of Krugman (1979, 1980) and Helpman (1981), have been widely employed and applied in numerous studies of international trade.

This type of model has implications for the volume of trade; in particular, as Helpman and Krugman (1985, Chapter 8) have demonstrated, the volume of trade among a group of countries, as a share in the total income of the country group, will be larger as the sizes the economies of individual countries in the group are more similar to each

⁹⁰ The significance of intra-industry trade has been reported by, for example, Grubel and Lloyd (1975).

⁹¹ On the other hand, it is debated in literature whether such intra-industry trade, or “trade overlap,” observed in the data is a matter of the aggregation of sectors or commodities. See Finger (1975).

other. In other words, if two regions have the same total sizes of their economies and consist of the same number of countries, the region in which countries are more equal in GDP will trade more within that region.

Although this theoretical implication is clear-cut and has an empirically testable form, only a few studies have directly examined this implication empirically. Helpman (1987) employed time-series data on 14 OECD countries and graphically showed the positive relationship between the volume of trade among the countries as a fraction of their total GDP and the similarity in their respective GDPs. Hummels and Levinsohn (1995) performed more formal empirical tests using panel data on bilateral trade flows between pairs of the same 14 OECD countries, as well as those of another 14 non-OECD countries. They expected that the data on trade between the OECD countries would fit the monopolistic competition model while it would not be the case for trade between the non-OECD countries, because the former was likely to be more intra-industry trade of horizontally differentiated products⁹² that the theoretical model considers, while the latter did not seem to be characterized as such. Their results, however, showed that GDP similarity between two trading countries well explained the volume of bilateral trade between them, both for the OECD and non-OECD countries, which left a puzzle. Debaere (2005) re-examined the study by Hummels and Levinsohn, and claimed that their empirical approach may not have been able to properly assess the impact of the income similarity on bilateral trade, and this was why their results were puzzling. He thus presented a modified equation explaining the relationship between the volume of trade

⁹² In literature two types of product differentiation are distinguished: horizontal product differentiation and vertical product differentiation. The former arises when products of a similar quality vary in certain characteristics, while the latter arises when products differ in quality. The product differentiation discussed in the current paper is horizontal differentiation, which the monopolistic competition model considers.

and GDP similarity between countries, and estimated it using updated data for the same set of OECD and non-OECD countries. From the estimation results he concluded that positive correlation between the volume of trade and size similarity among trading countries was significant only for the OECD countries but not for the non-OECD countries, and thus the puzzle was not present any more.⁹³

These studies attempted to test the monopolistic competition model in the context of *aggregate* trade, which includes all types of traded goods. However, not all goods that are internationally traded are differentiated products, and the trade of those non-differentiated products may be driven by other mechanisms than the one that is described by the monopolistic competition model. In fact, to expand the tested implication—that the volume of trade will increase as trading economies become more equal in size—to the level of aggregate trade, they assumed that all industries were internally differentiated in terms of product varieties, or alternatively that perfect specialization of production took place in every sector. These assumptions are very restrictive and thus may not be realistic.

In this paper, I propose an alternative empirical approach to testing the implication of the monopolistic competition model for the volume of trade among countries. The key is to focus on the trade of differentiated products. I review the model and derive the equation for the volume of bilateral trade of differentiated products without imposing such restrictive assumptions as those mentioned above. The derived alternative equation suggests that the simple GDP similarity between trading economies does not predict the volume of bilateral trade of differentiated products. The equation, however, implies that the volume of bilateral trade of differentiated products, as a share

⁹³ The appendix to this chapter reviews the work by Hummels & Levinsohn (1995) and by Debaere (2005).

in the domestic production of these products in the two trading countries, will be proportional to the two countries' GDP similarity *adjusted for how symmetric the countries are in their production structure*. In other words, the volume of trade of differentiated products between two countries will be larger as the countries are more similar in GDP, as well as in the share of the differentiated sectors in GDP.

This implication must be tested with data on trade and production in the sectors of differentiated products. Therefore, in addition to data on aggregate trade and GDP such as those used in the previous studies, I employ disaggregated data on trade and production in manufacturing industries for a range of countries. I also use the information on product characteristics classified by Rauch (1999) to define the “differentiated sectors.” Furthermore, to handle zero-trade observations in the data, I apply non-linear estimation methods in addition to the benchmark OLS estimation of log-linear forms of the volume-of-trade equations.

The empirical analysis, especially the result of the estimation with a non-linear method that handles zero-trade observations, shows that the tested implication of the monopolistic competition model—that the volume of bilateral trade per production will be larger as two trading countries are more similar in GDP and more symmetric in production structure—is supported by the data for both OECD and non-OECD countries, not only for the differentiated-sector trade but also for aggregate trade. Therefore, in terms of the relationship between the volume of trade and the size similarity, we go back to Hummels and Levinsohn's puzzle, contrary to Debaere's conclusion. However, using a unique approach that separates trade of differentiated products from aggregate trade, this paper also demonstrates two other things: (i) bilateral trade flows among OECD

countries, especially in the sectors of differentiated products, are well explained by the monopolistic competition model; but (ii) trade flows among non-OECD countries are not equally well-explained by the model. This finding suggests that there should be some other mechanism that makes trade patterns among lower-income countries different from those among rich countries.

This study offers some insight for a series of empirical studies on the gravity equation, to which the monopolistic competition model provides a theoretical basis. Most studies have estimated the gravity equation for aggregate trade. For example, Feenstra, Markusen and Rose (2001), Evenett and Keller (2002), and Haveman and Hummels (2004) use the gravity equation for aggregate trade to test which theory of international trade is the most likely to explain the actual trade flows, following Deardorff (1998) pointing out that multiple trade theories can derive the gravity equation. The point of Feenstra et al. is the existence of a home-market effect that may distinguish the monopolistic competition model from others, while Evenett and Keller, as well as Haveman and Hummels, focus on the elasticity of national income with respect to the volume of trade, which will be smaller than unity if specialization in production is incomplete. However, aggregate trade involves the trade of various products, some of which the monopolistic competition model fits well, but others may be characterized by product homogeneity and incomplete specialization; thus all trade should not be explained by a single model in a unified manner.⁹⁴ In contrast, Harrigan (1994) and Jensen (2000) have estimated the gravity equation at the sectoral level using data on trade

⁹⁴ Feenstra, Markusen and Rose also divide trade into three categories according to Rauch (1999) to estimate their gravity equation, but the explanatory variables are for the aggregate; i.e., GDPs of exporter and importer countries.

and production in manufacturing industries.⁹⁵ They, however, do not explicitly consider differences in product characteristics (differentiated versus homogeneous) across manufacturing industries, to which this paper pays careful attention.⁹⁶

The remainder of this paper is organized as follows. The next section derives the equation explaining the volume of trade in the differentiated sectors, and discusses its implication in comparison with the equation for aggregate trade that has been used in the existing literature. The section presenting the empirical approaches follows. The data employed for the empirical analysis are described in the fourth section. The results of the analysis are presented and discussed in the fifth section, which is followed by the concluding section.

4.2 Monopolistic Competition Model and Volume of Trade

In this section, to account for the volume of trade I derive two formulas from the monopolistic competition model of international trade introduced by Helpman and Krugman (1985, Chapters 6-8). This model is characterized as follows: (i) some sectors have a number of product varieties (I hereinafter call these sectors “differentiated sectors”); (ii) each of the product varieties in a differentiated sector is produced monopolistically competitively by a single firm; and (iii) consumers throughout the world have identical preferences that are characterized by a two-tier utility function: the upper-

⁹⁵ Harrigan introduces a variety of proxies for scale economies in his equation to see whether the home-market effect would be significant, which would indicate a monopolistic competition rather than Armington preference for national varieties. Jensen’s interest is in the size of the estimated elasticity of volume of imports to the importer’s income.

⁹⁶ Other empirical work such as Anderson and van Wincoop (2003) carefully derives a structural gravity-type equation from a generalized monopolistic competition model, but due to the unobservability of variables, their attention is limited to a certain factor such as distance or trade cost. Lai and Zhu (2004), on the other hand, have made an extended effort to measure as many variables as possible to estimate their structural and generalized volume-of-trade equation with data.

tier utility is homothetic, and the sub-utility over product varieties within a sector takes a CES functional form.

Here I consider an equilibrium of frictionless trade so that the price of each good or horizontally differentiated product is equal throughout the world. In this free-trade equilibrium, every product *in the differentiated sectors* produced in each country will be divided among all consumers worldwide, according to their share of world income. The volume of exports from one country to another is thus expressed as follows:

$$EX_i^j = \sum_{s \in D} y_j p_s Q_{s,i} + \sum_{s \in H} EX_{s,i}^j, \quad (4.1)$$

where D : group of the differentiated sectors;
 H : group of homogeneous sectors;
 i, j : scripts for countries ($i \neq j$);
 $EX_{s,i}^j$: exports from Country i to Country j in Sector s ;
 $Q_{s,i}$: Country i 's production in Sector s ;
 p_s : equilibrium price of (differentiated) products in Sector s
 y_j : Country j 's GDP share in the world ($= Y_j/Y_w$)

Note that the volume of trade between a specific pair of countries in the sectors of homogeneous products (or “homogeneous sectors”), $EX_{s,i}^j$ for $s \in H$, is indeterminate. That is, although a country will export a homogeneous product when the amount of the product that the country domestically produces is greater than the amount it consumes, how much of the country's product will be exported to which country(ies) cannot be determined because, in the free-trade equilibrium, importing countries will be indifferent about from which country(ies) they import the homogeneous product to supply their domestic demand.

4.2.1 Aggregate volume of trade

The version of the formula for the *aggregate* volume of trade, which has been employed in studies such as Helpman (1987), Hummels and Levinsohn (1995), and Debaere (2005), further assumes the following:

- (A1) Each country in the world is also completely specialized in production in the homogeneous sectors. That is, every homogeneous product is produced by no more than one country. Under this assumption, any product produced by a sole producer country (i.e., a sole exporter) will be imported by all other countries, and how much each country imports will be determined according to the country's share of world income. Therefore, no indeterminacy will be left for the quantities of bilateral trade, and the volume of exports in both homogeneous and differentiated sectors from Country i to Country j is expressed as follows:

$$EX_i^j = \sum_{s \in D, H} y_j p_s Q_{s,i} .$$

- (A2) Products in any sector are tradable, i.e., there exist no non-traded sectors.⁹⁷ Under this assumption, the aggregate value of a country's production over the sectors equals its income, or GDP. That is;

$$\sum_{s \in D, H} p_s Q_{s,i} = Y_i ;$$

$$EX_i^j = y_j Y_i$$

where Y_i is GDP of Country i .

Therefore, following Helpman (1987), the *aggregate* bilateral trade volume between Countries i and j is expressed as follows:

⁹⁷ This assumption (A2) can be replaced with the following weaker assumption to derive Equation (4.2A) below.

(A2'): Every country has an equal share of non-traded sectors in its GDP.

$$VT_{ij} \equiv EX_i^j + EX_j^i = y_j Y_i + y_i Y_j = 2Y_i Y_j / Y_w = y_{ij} \cdot 2(Y_i Y_j / Y_{ij})$$

$$\Leftrightarrow VT_{ij} / Y_{ij} = y_{ij} [1 - (Y_i / Y_{ij})^2 - (Y_j / Y_{ij})^2]$$

where $Y_{ij} = Y_i + Y_j$: Country i - j pair's total GDP

$y_{ij} = Y_{ij} / Y_w$: Country i - j pair's share of world GDP

The term in the square brackets on the right-hand side of the second equation indicates the similarity of GDPs, or the similarity of the sizes of the economy, of two trading countries. This term takes a greater value as the size of the two countries become more equal, and takes the maximum value of 0.5 when the two countries are exactly equal in GDP; i.e., $Y_i / Y_{ij} = Y_j / Y_{ij} = 1/2$.⁹⁸ Using this index of size similarity,⁹⁹ the equation is expressed as follows:

$$VT_{ij} / Y_{ij} = y_{ij} \cdot sim_{ij}, \quad (4.2A)$$

$$\text{where } sim_{ij} = [1 - (Y_i / Y_{ij})^2 - (Y_j / Y_{ij})^2].$$

This Equation (4.2A) implies that the volume of *aggregate* bilateral trade, as a share in the total income (GDP) of the two trading countries, will be greater as their respective national incomes are more similar.

4.2.2 Volume of Trade in the Differentiated Sectors

The two assumptions A1 and A2 are very restrictive. Since Equation (4.2A) can be derived only with these restrictive assumptions, its validity should be limited accordingly. However, by focusing our attention on the differentiated sectors, it is possible to derive an alternative formula that can explain the volume of trade in such sectors in a similar way but without imposing these assumptions. Since countries are

⁹⁸ Note that $Y_j / Y_{ij} = 1 - Y_i / Y_{ij}$. In theory, this index takes the minimum value of zero when two countries are completely dissimilar in GDP; i.e., $Y_i / Y_{ij} = 0$ and $Y_j / Y_{ij} = 1$, or vice versa.

⁹⁹ Helpman (1987), as well as Hummels & Levinsohn (1995), calls this term the “dispersion” index, while Debaere (2005) names it the “similarity” index. I follow the latter since this index being larger means two countries being more similar in income.

considered to be completely specialized in production of unique varieties in the differentiated sectors, by taking the first term of Equation (4.1), export from Country i to Country j in the differentiated sectors is described as follows:

$$EX_i^{j,D} = \sum_{s \in D} y_j p_s Q_{s,i} = y_j X_i^D,$$

where $EX_i^{j,D}$: export in the differentiated sectors from Country i to Country j

X_i^D : value of Country i 's domestic production in the differentiated

$$\text{sectors: } X_i^D \equiv \sum_{s \in D} p_s Q_{s,i}.$$

Therefore, the volume of trade *in the differentiated sectors* between Countries i and j is expressed as follows:

$$VT_{ij}^D = EX_i^{j,D} + EX_j^{i,D} = y_j X_i^D + y_i X_j^D = y_{ij} \{ (Y_j/Y_{ij}) \cdot X_i^D + (Y_i/Y_{ij}) \cdot X_j^D \}$$

$$\Leftrightarrow \frac{VT_{ij}^D}{X_{ij}^D} = y_{ij} \left(1 - \frac{X_i^D}{X_{ij}^D} \cdot \frac{Y_i}{Y_{ij}} - \frac{X_j^D}{X_{ij}^D} \cdot \frac{Y_j}{Y_{ij}} \right) = y_{ij} \left[1 - \left(\frac{X_i^D/Y_i}{X_{ij}^D/Y_{ij}} \right) \cdot \left(\frac{Y_i}{Y_{ij}} \right)^2 - \left(\frac{X_j^D/Y_j}{X_{ij}^D/Y_{ij}} \right) \cdot \left(\frac{Y_j}{Y_{ij}} \right)^2 \right],$$

where VT_{ij}^D : volume of trade in differentiated sectors between Countries i and j

X_{ij}^D : Countries i and j 's total domestic production in the differentiated

$$\text{sectors } (X_{ij}^D \equiv X_i^D + X_j^D).$$

The term in the square brackets in this equation is similar to the size similarity index in Equation (4.2A) for aggregate trade, but this term depends not only on two countries' relative income sizes but also on the sizes of production in the differentiated sectors of the countries (X_i^D, X_j^D). The GDP share term for each country ($(Y_i/Y_{ij})^2$ or $(Y_j/Y_{ij})^2$) is “weighted” by the term $(X_i^D/Y_i)/(X_{ij}^D/Y_{ij}^D)$, and this “weight” term indicates how large the share of the differentiated sectors in GDP is in each country, relative to the overall GDP share of the differentiated sectors in the two countries. In other words, this term indicates GDP similarity between two countries *adjusted for how symmetric the two countries are*

in their production structure. This term takes a larger value as two countries are more similar in the size of their economies *and* more symmetric in production structure. I thus call this term the *production structure-adjusted size* (or *GDP*) *similarity*, and re-write the equation as follows:

$$VT_{ij}^D / X_{ij}^D = y_{ij} \cdot sim_{ij}^* \quad (4.2D)$$

$$\text{where } sim_{ij}^* = \left[1 - \left(\frac{X_i^D / Y_i}{X_{ij}^D / Y_{ij}} \right) \cdot \left(\frac{Y_i}{Y_{ij}} \right)^2 - \left(\frac{X_j^D / Y_j}{X_{ij}^D / Y_{ij}} \right) \cdot \left(\frac{Y_j}{Y_{ij}} \right)^2 \right].$$

Equation (4.2D) implies that the volume of bilateral trade *in the differentiated sectors*, as a share in the two countries' *total production* in those sectors, is predicted by the size similarity between the two trading countries *adjusted for how symmetric their production structures are*. That is, two countries will trade more in the differentiated sectors as the two countries are more similar in GDP and more symmetric in production.

4.2.3 Discussion on Production Structure-adjusted Size Similarity

As mentioned above, the volume of bilateral trade in the differentiated sectors, as a share in the two countries' domestic production in those sectors, is proportional to the similarity in size between the countries that is adjusted for the symmetry of the country pair's production structure. This adjusted index of GDP similarity takes a larger value as two trading countries are more similar in GDP and more symmetric in production structure. This is true in general, i.e., for more common cases in which a country with larger GDP is a larger producer in the differentiated sectors than the other country.¹⁰⁰ However, this index is in fact even greater for less common cases in which a country with

¹⁰⁰ For instance, one country has 70% of two countries' total GDP and 60% of their differentiated-sector production.

smaller GDP is a larger producer in the differentiated sectors;¹⁰¹ i.e., the two countries are dissimilar or asymmetric in an extreme manner.^{102, 103} This is because, according to the monopolistic competition model of trade, a trade flow between countries will be larger when the exporter has larger production and the importer has larger income. Therefore, having the sizes of GDP and sectoral production adjusted (or normalized), the trade flow in the sector will be larger when one country imports the whole domestic production of the other country (for a hypothetical case in which one country has 100% of a country pair's GDP but no production in the considered sector, while the other country has zero income but 100% of the country pair's production in that sector), rather than when two countries exchange a half of their respective production (for another hypothetical case in which two countries are exactly equal in both GDP and sectoral production).

4.3 Empirical Approaches to Estimate Volume-of-Trade Equations

In this section, I describe empirical specifications to estimate the volume-of-trade equations derived in the preceding section, to test how well bilateral trade is explained by the size similarity of two trading economies. Each approach is taken to estimate both Equation (4.2A) for aggregate trade and Equation (4.2D) for trade in the differentiated sectors. The results of the estimation from each approach, which is presented in the fifth section, are compared to examine how the proposed alternative model for the differentiated-sector trade differs from the conventional model for aggregate trade.

¹⁰¹ For example, one country has 30% of two countries' total GDP and 80% of their differentiated-sector production.

¹⁰² In fact, in such a case the adjusted similarity index takes a value over 0.5 and up to 1, compared to the case in which two countries are perfectly similar and symmetric ($sim^* = 0.5$).

¹⁰³ In the data used in the current study, the number of such uncommon cases for the OECD countries is 228 out of the total 3,630 observations; and 2,144 out of 14,565 for the non-OECD countries. See Section 4.4 for the detailed description of the data.

4.3.1 OLS Estimation of Log-linearized Form

As a benchmark, I first estimate the volume-of-trade equations in a log-linearized form by the OLS. Recalling Equations (4.2A) and (4.2D), but also considering other potential factors that may affect bilateral trade flows:¹⁰⁴

$$VT_{ijt} / Y_{ijt} = sim_{ijt}^{\beta_1} \cdot y_{ijt}^{\beta_2} \cdot \mu_{ij} \cdot \varepsilon_{ijt} \quad (4.2A')$$

$$VT_{ijt}^D / X_{ijt}^D = sim_{ijt}^{*\beta_1} \cdot y_{ijt}^{\beta_2} \cdot \mu_{ij} \cdot \varepsilon_{ijt} \quad (4.2D')$$

Although the underlying monopolistic competition model explains a core mechanism determining the volume of trade as Equations (4.2A) and (4.2D) suggest (with both β_1 and β_2 equaling one), real trade flows may be affected by other factors. For example, the literature on the gravity equation suggests that bilateral trade flows will be affected by geographic factors such as distance, border sharing, and commonness of language. The term μ_{ij} is included in the equations to capture these factors that are specific to country pairs, as well as other unobserved potential country pair-specific (but time-invariant) factors affecting bilateral trade flows. The last term ε_{ijt} captures idiosyncratic disturbances to recorded trade flows or measurement errors in data, which are assumed to be log-normally distributed. Taking the logarithm of both sides of the two equations (4.2A') and (4.2D') yields the following linearized equations:

$$\log(VT_{ijt} / Y_{ijt}) = \beta_1 \cdot \log(sim_{ijt}) + \beta_2 \cdot \log(y_{ijt}) + \mu_{ij} + \varepsilon_{ijt} \quad (4.3A)$$

$$\log(VT_{ijt}^D / X_{ijt}^D) = \beta_1 \cdot \log(sim_{ijt}^*) + \beta_2 \cdot \log(y_{ijt}) + \mu_{ij} + \varepsilon_{ijt} \quad (4.3D)$$

Equation (4.3A) for the volume of aggregate bilateral trade is the same as the main empirical specification that is employed by Debaere (2005).¹⁰⁵ Equation (4.3D), which is

¹⁰⁴ Since panel data are used for the estimation, here and in the rest of this paper, variables in the equations are expressed with script t to denote a time period.

designed to account for the volume of bilateral trade *in the differentiated sectors*, is an alternative empirical approach that this paper proposes. Both equations are estimated by OLS regression with country pair-specific dummies (μ_{ij}). Year-specific dummies are also included for the estimation in order to capture any trend in or shocks to trade flows that are common for all countries in the world.

Equations (4.3A) and (4.3D) are estimated separately for the samples of OECD and non-OECD countries.¹⁰⁶ This is to examine whether trade among OECD countries and trade among non-OECD countries are equally well explained by the volume-of-trade equations, following the studies by Hummels and Levinsohn (1995) and Debaere (2005). These studies separated a group of OECD countries from that of non-OECD countries for estimation, based on the understanding that intra-industry trade of differentiated products, which the monopolistic competition model primarily aims to explain, is dominant in trade among OECD countries, while trade among non-OECD countries should not be mainly characterized by horizontal product differentiation. Their expectation was thus that the aggregate version of the volume-of-trade equation (4.3A) would describe bilateral trade well for OECD countries but not for non-OECD countries. Although Hummels and Levinsohn found a result that was counter to this expectation (i.e., the data support the model for both country groups), Debaere's re-examination found empirical support for the model only for OECD countries, as initially expected. In contrast, the current study focuses on trade of differentiated products, which the monopolistic competition model aims to explain for any country group. Therefore, it is expected that the proposed equation (4.3D) for the differentiated-sector trade should explain both trade among

¹⁰⁵ See the appendix for more details of the empirical approach of Debaere (2005), as well as Hummels and Levinsohn (1995).

¹⁰⁶ See the next section for the list of the countries included in each sample.

OECD countries and trade among non-OECD countries equally well, while the conventional equation (4.3A) for aggregate trade would not.

An empirical issue here in estimating Equations (4.3A) and (4.3D) is the treatment of zero-trade observations. A considerable number of country pairs in both OECD and non-OECD groups have no bilateral trade in the differentiated sectors in certain years. In the data used in this study, observations with no differentiated-sector trade are less than one percent of all the observations in the OECD sample, while such zero-trade observations comprise more than 60% in the non-OECD sample.¹⁰⁷ For the estimation of the log-linear equations, these zero-valued observations bring the problem of undefined logarithmic values in the left-hand side. To handle this problem, for the benchmark estimation I (i) omit such zero-trade observations and use only observations with positive differentiated-sector trade; but also (ii) include these zero-trade observations for estimation by replacing zero with a very small positive number.^{108, 109}

4.3.2 Non-linear Model for Zero-trade Observations: Poisson Quasi-maximum Likelihood Estimation

Although replacing zero with a small positive number has been a convention in estimating a logarithmic form, it is not ideal. It is more desirable if there exists an other appropriate alternative estimation method that can treat zero in the value of trade as it is. Hummels and Levinsohn (1995) estimated (by the OLS) their volume-of-trade equation in a level form, instead of a logarithmic form, for their non-OECD sample to avoid

¹⁰⁷ The details of the data are described in the next section.

¹⁰⁸ Debaere (2005) also applies a similar procedure to handle zero-trade observations in estimating his log-linear model.

¹⁰⁹ This number must be at least smaller than the minimum non-zero value of trade in the used data. The minimum value of the bilateral trade per production (VT_{ijt}^D/X_{ijt}^D) in the data is 9.4e-9, and I thus chose 10^{-9} (1.0e-9) for the positive small number replacing zero.

omitting zero-trade observations. Debaere (2005) also employed similar level specifications¹¹⁰ and estimated the equations by the Tobit method to keep zero-trade observations in his non-OECD data. The cost of using such level forms of the equation was that (i) they had to give up estimating separately the impact of the two variables of interest, the country pair's size similarity and the country pair's share of the world GDP; or (ii) as in one of Debaere's two level specifications, for separate estimation of the effects of the two variables they had to abandon the strict consistency of a regression equation with the theoretical monopolistic competition model. (See Appendix for further details of the empirical approaches of Hummels and Levinsohn (1995) and Debaere (2005).)

In the current paper, I employ an alternative method to handle zero-trade observations, which can both maintain the structural consistency of the regression equation with the theoretical model and separately estimate the impacts of the two variables of interest. The alternative is the (fixed-effect) Poisson quasi-maximum likelihood (PQML) estimation. The Poisson regression is usually applied for count data, but it is also applicable to non-negative continuous variables. Hausman, Hall and Griliches (1984) developed the conditional fixed-effect PQML method in the panel data context, which has been shown by Wooldridge (1999) to be consistent and robust across distributional assumptions when the conditional mean of the dependent variable is an

¹¹⁰ The level forms of the volume-of-trade equation in the two studies are not the same. Hummels and Levinsohn (1995) used the value of (aggregate) trade (VT_{ij}) as the dependent variable, while Debaere (2005) employed the volume of aggregate trade as the share in GDP (VT_{ij}/Y_{ij}). Hummels and Levinsohn's approach thus left the term of the country pair's GDP (Y_{ij}) in the right-hand side of the equation, about which Debaere argued in terms of its relevance for assessing the impact of the size similarity between trading countries.

exponential-class function of the linear combination of regressors.¹¹¹ The PQML method has also been applied to the estimation of the gravity equation by Silva and Tenreiro (2006) for cross-sectional data and by Westerlund and Wilhelmsson (2006) for panel data. These studies have shown by simulation that with zero-trade observations the PQML method has the advantage of smaller potential estimation bias compared to the OLS estimation of a logarithmic form of the equation. I thus employ the PQML method and estimate the following form of the volume-of-trade equations:

$$VT_{ijt} / Y_{ijt} = sim_{ijt}^{\beta_1} \cdot y_{ijt}^{\beta_2} \cdot \mu_{ij} + \varepsilon_{ijt}$$

$$\Leftrightarrow VT_{ijt} / Y_{ijt} = \exp[\beta_1 \cdot \log(sim_{ijt}) + \beta_2 \cdot \log(y_{ijt}) + \mu_{ij}] + \varepsilon_{ijt} \quad (4.4A)$$

$$VT_{ijt}^D / X_{ijt}^D = sim_{ijt}^{*\beta_1} \cdot y_{ijt}^{\beta_2} \cdot \mu_{ij} + \varepsilon_{ijt}$$

$$\Leftrightarrow VT_{ijt}^D / X_{ijt}^D = \exp[\beta_1 \cdot \log(sim_{ijt}^*) + \beta_2 \cdot \log(y_{ijt}) + \mu_{ij}] + \varepsilon_{ijt} \quad (4.4D)$$

The main difference from the benchmark log-linear form (4.3A) or (4.3D) is that in the above form the stochastic error term ε_{ijt} is additive, instead of multiplicative as in Equations (4.2A') and (4.2D').

4.3.3 Tobit Estimation of Log-linearized Form

For the purpose of robustness check of the OLS estimation of the log-linear form, I also apply the Tobit regression to estimate the volume-of-trade equations. Even for the Tobit estimation, zero-trade observations in the data bring the issue of the undefined logarithm of zero in principle. However, in the specific data used in the current study,¹¹² bilateral trade is recorded in thousands of U.S. dollars, and thus no (or zero) value is

¹¹¹ That is, $E[y|\mathbf{x}] = \alpha \cdot \exp(\mathbf{x}\boldsymbol{\beta})$ where y is the dependent variable, \mathbf{x} is the vector of regressors, $\boldsymbol{\beta}$ is the vector of coefficients, and α is a scalar.

¹¹² The details of the employed trade data are described in the next section.

recorded when the value of bilateral trade is less than \$500 (rounded to zero thousands).

Using this feature of the employed data, I apply the Tobit estimation to the following log-linear specification, which is slightly different from Equations (4.3A) and (4.3D):

$$\log(VT_{ijt}) = \log(Y_{ijt}) + \beta_1 \cdot \log(sim_{ijt}) + \beta_2 \cdot \log(y_{ijt}) + \mu_{ij} + \varepsilon_{ijt} \quad (4.5A)$$

$$\log(VT_{ijt}) = \log(VT_{ijt}^*) \quad \text{if } VT_{ijt}^* > 0.5 \text{ (\$500)}$$

$$\log(VT_{ijt}) = \log(0.5) \quad \text{if } VT_{ijt}^* \leq 0.5 \text{ (\$500)}$$

$$\log(VT_{ijt}^D) = \log(X_{ijt}^D) + \beta_1 \cdot \log(sim_{ijt}^*) + \beta_2 \cdot \log(y_{ijt}) + \mu_{ij} + \varepsilon_{ijt} \quad (4.5D)$$

$$\log(VT_{ijt}^D) = \log(VT_{ijt}^{D*}) \quad \text{if } VT_{ijt}^{D*} > 0.5 \text{ (\$500)}$$

$$\log(VT_{ijt}^D) = \log(0.5) \quad \text{if } VT_{ijt}^{D*} \leq 0.5 \text{ (\$500)}$$

where VT_{ijt} or VT_{ijt}^D is the observed or recorded value of bilateral trade in the data, while VT_{ijt}^* or VT_{ijt}^{D*} is the underlying actual trade value.¹¹³ The following two things should be noted for this estimation approach. First, a country pair's total production (X_{ijt}^D in the differentiated-sector equation or Y_{ijt} in the aggregate equation) is now moved from the denominator of the left-hand side to the right-hand side of the equation. The variable is thus included as one of the regressors, but the coefficient for this variable is restricted to be one for estimation. Secondly, all the zero values for bilateral trade in the data are replaced with \$500 or 0.5 in thousands of dollars.

4.4 The Data

To estimate Equations (4.3A) and (4.3D) through (4.5A) and (4.5D) presented in the previous section, data on trade, GDP, and industrial production have been collected for various countries.

¹¹³ It should be noted that the unconditional fixed-effect Tobit model will generally be biased due to the problem of incidental parameters (Hsiao, 2003; pp. 48-9, 243).

The data on bilateral trade are from the NBER-Statistics Canada Trade Data compiled by Feenstra, Lipsey, and Bowen (1997) for the period 1970-1992, and the UCD-Statistics Canada Trade Data that is compiled by Feenstra (2000) to supplement for the period up to 1997. The dataset contains trade flows between each pair of countries. Goods in the trade flows are classified according to the four-digit Standard International Trade Classification (SITC, Revision 2). The value of each trade flow is recorded in thousands of nominal U.S. dollars.

The data on GDP measured in current U.S. dollars are from the World Development Indicators (World Bank, 2005). Both GDP of each country and the world total GDP have been collected to compute the world income (GDP) share of each country pair (y_{ij}).¹¹⁴

The data on industrial production are from the United Nation's Industrial Statistics Database (INDSTAT3; UNIDO, 2003), which contains the annual data on manufacturing production in countries for the years of 1960-2000. Manufacturing industries are classified according to the three-digit International Standard Industrial Classification (ISIC, Revision 2). The data on gross output in nominal U.S. dollars are used.

The data for the current study cover 89 countries for the years 1970 through 1997. These countries all have population above one million as of the year 1997. The countries are divided into two groups, OECD countries and non-OECD countries, according to the actual OECD membership as of the year 1973.¹¹⁵ As a result, the data include 20

¹¹⁴ Note that the world GDP (Y_w) in this study also counts GDP of countries that are not included in the sample, and thus is greater than the sum of GDP of the 89 sample countries.

¹¹⁵ 1973 is the year in which New Zealand joined the OECD. New Zealand was the newest member until Mexico joined in 1994.

countries (190 bilateral pairs) in the OECD group and 69 countries (1,808 pairs¹¹⁶) in the non-OECD group. Table 1 lists the countries and years included in the data for each group. The bilateral trade flows between the OECD countries represent 33.8% of the world total flows on average over the period 1970-1997 (with an annual share ranging 0.3% through 62.0%); and the flows between the non-OECD countries represent 1.0% on average over the period (with an annual share ranging 0.5% through 1.5%). The panel data are kept unbalanced to retain as many observations in the data as possible.¹¹⁷

4.4.1 Industry/commodity classifications for the production data and trade data

Since the trade data and the production data are based on different classification schemes, mapping one classification onto the other is required to merge the two datasets using a common classification.¹¹⁸ In the production data 28 manufacturing industries are classified according to the three-digit ISIC, while in the trade data goods are classified into over a thousand categories according to the four-digit SITC. The mapping thus requires condensing the four-digit SITC (Revision 2) into the three-digit ISIC (Revision 2). I have mapped the trade data onto the three-digit ISIC using the concordance information sourced from the OECD, which is available on Jon Haveman's Industry Concordances web page

¹¹⁶ The number of country pairs in the data is less than ${}_{69}C_2 = 2,346$. This is because the 69 countries include countries that appear in the data as one of a country pair in any year(s), while some country pairs have no years for which production or GDP data are available for both countries. For instance, the data for Mexico are available only for 1994-97 while the data for Hong Kong are available only for 1973-90. As a result, bilateral trade between these two countries is not included in the data for any year.

¹¹⁷ I cannot make the panel balanced for the entire 190 + 1,808 country pairs for the 28 years due to the lack of data for one or more variables for some countries in some years.

¹¹⁸ While the ISIC for the production data is based on industrial activities, the SITC for the trade data is based on commodity characteristics. Since the two classifications are based on different principles, the mapping cannot necessarily be one-to-one.

(<http://www.macalester.edu/research/economics/PAGE/HAVEMAN/Trade.Resources/TradeConcordances.html>).¹¹⁹

Next, to separate the differentiated sectors from other (non-differentiated) sectors, I follow Rauch (1999), which classifies the four-digit SITC commodities into three categories based on the degree of product differentiation: goods traded on an organized exchange (homogeneous goods), reference priced goods, and differentiated goods. Although the production data, which are classified according to ISIC, cannot be simply mapped onto Rauch's three categories, there are ten three-digit ISIC manufacturing industries whose corresponding four-digit SITC categories are all classified as Rauch's differentiated goods. These industries are: 322 (wearing apparel), 324 (footwear), 332 (furniture), 355 (rubber products), 356 (plastic products), 361 (pottery, china, and earthenware), 362 (glass and products), 382 (non-electric machinery), 384 (transport equipment), and 385 (professional and scientific equipment). I therefore group these 10 three-digit industries as representative of the differentiated sectors, and accordingly compute bilateral trade and production in these differentiated sectors for each country pair for each year. These 10 differentiated manufacturing industries comprise 31.2% of the world aggregate trade on average, with the share in each year ranging from 24.3 to 37.0% during the period of 1970-1997.¹²⁰ These shares in the total trade flows among the 89 sample countries are: 41% on average with annual shares ranging 33 through 49% for the OECD countries; and 13% on average with annual shares ranging 9 through 21% for the non-OECD countries.

¹¹⁹ The original mapping is from the five-digit SITC to the three-digit ISIC. However, since the trade data have only the detail of the four-digit classification, I disregarded the details of the five-digit SITC in the original concordance.

¹²⁰ Note that the differentiated-sector industries are selected only from manufacturing industries.

4.4.2 Zero-trade Observations

In the OECD group, while all country pairs have positive bilateral trade flows in all the 28 years, 28 out of 3,630 observations (for 190 country pairs for 28 years) have zero trade in the differentiated sectors. In the non-OECD group, 4,551 out of 14,565 observations (for 1,808 country pairs for 28 years) have no trade flows, and additional 2,798 observations have zero flows in the differentiated sectors.

Figures 4.1A through 4.2D plot bilateral trade per production vs the size similarity index with a trend line fitted by locally weighted regression (Lowess¹²¹). Figures 4.1A and 4.1D are for the OECD countries, and 4.2A and 4.2D are for the non-OECD countries. The left panels (Figures 4.1A and 4.2A) plot the value of aggregate trade per GDP against the index of GDP similarity between two countries (sim_{ijt}). The right panels (Figures 4.1D and 4.2D) plot the value of trade per production in the differentiated sectors against the index of production structure-adjusted GDP similarity (sim^*_{ijt}). All the variables are in logarithms and mean-differenced, which correspond to the benchmark OLS estimation with dummies. The vertical and horizontal lines indicate zeros, which are the means of the mean-differenced variables. While the trend line exhibits some positive slope on all the figures, the positive relationship between the two variables does not seem to be very clear except for Figure 4.1A for aggregate trade between the OECD countries.

4.5 Empirical Results

4.5.1 OLS Estimation of Log-linear Form

The results of the benchmark OLS estimation of the log-linear form of the volume-of-trade equations are presented in Tables 4.2 and 4.3. In each table, the second

¹²¹ Locally weighted scatterplot smoothing. The smoothing parameter (or bandwidth) is 0.8 for the trend line in these figures.

through fourth columns show the results for the OECD countries, and the fifth through seventh columns show the results for the non-OECD countries. For each country group, one column shows the result of the estimation of Equation (4.3A) for aggregate trade, and one column shows the result of the estimation of Equation (4.3D) for the differentiated-sector trade. For the purpose of comparison, the sectoral equation (4.3D) is also estimated for a group of three-digit ISIC manufacturing industries that are not included in the differentiated sector.¹²² The estimation result for these “non-differentiated” sectors (indicated as “ND”) is shown in another column for each country group.

The lower part of the tables shows the results of the tests, in the p-values, of the hypotheses that (i) the coefficient for the index of size similarity equals one; (ii) the coefficient for a country pair’s world GDP share equals one; and (iii) these two coefficients are jointly equal to one. These hypotheses are what the monopolistic competition model suggests when international trade is frictionless. It should be noted, however, that in reality various kinds of trade friction exist, and not all of them may be controlled for by country-pair specific dummies in the estimation. Having such trade friction, the coefficient estimates may be different from (smaller than) one even though the estimation suggests a positive and significant relationship between the volume of trade and the respective determinants.

Table 4.2 shows the result of the OLS estimation using observations with positive trade values but excluding zero-trade cases. In the following, to focus on the tested

¹²² The “non-differentiated” sector group consists of the following 17 three-digit ISIC industries: 311 (food products), 313 (beverages), 314 (tobacco), 321 (textiles), 323 (leather products), 331 (wood products), 341 (paper and products), 342 (printing and publishing), 351 (industrial chemicals), 352 (other chemicals), 353 (petroleum refineries), 354 (miscellaneous petroleum and coal products), 369 (other non-metallic mineral products), 371 (iron and steel), 372 (non-ferrous metals), 381 (fabricated metal products), and 383 (electric machinery). The miscellaneous category 390 is excluded from both differentiated and non-differentiated groups.

prediction on the relationship between the volume of bilateral trade per production and the size similarity between trading countries, I put my main focus on the estimate of the coefficient for the similarity index (β_1).¹²³ The result indicates that among the OECD countries the positive relationship between the volume of trade per production and the size similarity index is significant for both aggregate and differentiated-sector trade. This relationship is also positive for trade in non-differentiated sectors but less significant. In addition, the size of the coefficient estimate is the largest for the differentiated sectors ($\hat{\beta}_1 = .858$), it is smallest for the non-differentiated sectors ($\hat{\beta}_1 = .312$), and the case for aggregate trade falls in between ($\hat{\beta}_1 = .422$). The difference between the estimate for the differentiated-sector case and those for the other two cases is significant.¹²⁴ On the other hand, for the non-OECD countries, the coefficient is estimated to be positive and significant (at the 1% level) for all the three cases; but the difference in the value of the estimate is not significant across the cases.¹²⁵

The same equations (4.3A) and (4.3D) (, as well as (4.3ND)) are also estimated by OLS using all the observations with zero-trade values being replaced with a small positive number (10^{-9}). The result is shown in Table 4.3.¹²⁶ For the OECD countries, the overall result is the same as the previous case, except that now the estimate for the non-

¹²³ The estimates of the coefficient for the countries' world GDP share (β_2) are discussed in a later subsection.

¹²⁴ The hypothesis that $\hat{\beta}_1$ is the same between the aggregate case and the differentiated-sector case is rejected at the 5% level of significance.

¹²⁵ The p-value of the test of $\hat{\beta}_1$ being equal between the differentiated-sector case (with the largest value) and the non-differentiated-sector case (with the smallest value) is 0.30.

¹²⁶ It should be noted that the result is somewhat sensitive to the choice of the small positive number for zero-trade values, except for the case of aggregate trade between the OECD countries. In particular, when a much smaller number (such as 10^{-18} or smaller) is applied, the estimate of coefficient for the similarity index (β_1) is insignificant (or its p-value exceeds 10%) for the differentiated-sector equation even for the OECD countries. On the other hand, for the non-OECD countries the result for the differentiated sectors does not qualitatively change in terms of the signs and significance of the estimates of two coefficients (β_1 and β_2).

differentiated sector is not significant even at the 10% level. However, for the non-OECD countries, the coefficient estimate is insignificant for all the three cases.¹²⁷ The point estimate for differentiated-sector trade is larger than that in the other two cases, but the difference is not significant.¹²⁸ In other words, for the non-OECD countries, the OLS estimation of the log-linear form of the volume-of-trade equation gives a different picture depending on whether zero-trade observations are excluded or included.

4.5.2 Alternative Estimation of the Log-linear Form: Tobit

The Tobit estimation of the log-linear equations is also performed to see the robustness of the result when both zero- and nonzero-trade observations are included. Equations (4.5A) and (4.5D) are estimated for aggregate and differentiated-sector trade, respectively. As in the OLS estimation, Equation (4.5D) is also estimated for non-differentiated sectors (ND). The result is shown in Table 4.4. The overall picture is similar to Table 3 for the OLS estimation having zero-trade observations included, but the coefficient estimate $\hat{\beta}_1$ increases its significance in the differentiated-sector equation (4.5D) for both country groups. In particular, for the non-OECD countries the estimate is weakly significant (at the 10% level) in (4.5D) while it is insignificant in other two equations (4.5A) and (4.5ND).¹²⁹ This result indicates that the separation of the differentiated sectors in estimating the volume-of-trade equation, which the current paper proposes, gives evidence of the prediction of the monopolistic competition model more clearly than the conventional aggregate trade approach does.

¹²⁷ Note that the result for aggregate trade is consistent with Debaere's (2005).

¹²⁸ The p-value of the test of $\hat{\beta}_1$ being equal between the differentiated-sector case (with the largest value) and the non-differentiated-sector case (with the smallest value) is 0.22.

¹²⁹ However, the difference in the estimate across the three cases is not significant for the non-OECD countries. On the other hand, for the OECD countries, the estimate in the differentiated-sector equation is significantly larger than that in the other two cases at the 1% significance level.

4.5.3 Poisson Quasi-maximum Likelihood (PQML) Estimation

The above three estimation methods do not treat the zero value in the trade data as it is. On the other hand, the proposed estimation of Equations (4.4A) and (4.4D) by the Poisson quasi-maximum likelihood (PQML) procedure can treat zeros in observations as they are. Table 4.5 presents the result of the PQML estimation. Equation (4.4D) is also estimated for the non-differentiated sectors (ND).

The result for the OECD countries is consistent with the estimation results by the previous three methods, while the estimated coefficient for the similarity index is significant at the 1% level not only in the aggregate and differentiated-sector equations but also in the non-differentiated-sector equation. In other words, the estimation shows that among the OECD countries the positive correlation between the volume of trade per production and the adjusted size similarity is indicated even in the non-differentiated sectors. However, this may be because these non-differentiated sectors comprise manufacturing industries. These industries are excluded from the “pure” differentiated sectors, but that does not mean that products in these industries are all homogeneous. A more important thing in the estimation result is that the size of the estimated coefficient is the largest for the differentiated-sector trade, the median for the aggregate trade, and the smallest for the non-differentiated sectors. The coefficient estimate in the differentiated-sector equation is significantly larger than the estimate in the other two equations.¹³⁰

For the non-OECD countries, the coefficient estimate is also significant in all the three equations (4.4A), (4.4D) and (4.4ND), at least at the 5% level. However, the

¹³⁰ The difference is significant at the 10% level between (4.4D) and (4.4A), and at the 5% level between (4.4D) and (4.4ND).

estimate for the differentiated sector is the smallest and least significant,¹³¹ which is counter to the expectation from the theory. In other words, the result of the PQML estimation implies that, among the non-OECD countries, the positive correlation between the volume of trade per production and the size similarity between countries is more striking as international trade contains more *non*-differentiated products.

4.5.4 Comparison of Four Approaches to Estimation

The above four estimation approaches give consistent results for the OECD countries, but for the non-OECD countries they provide different results from each other. To see which method describes the data, especially for the non-OECD countries, better than the others, I use the Akaike Information Criteria (AIC)¹³² for the four estimation specifications. The AIC measures the goodness of fit of an empirical model, and a model with a lower AIC value is preferred to that with a higher AIC value. Table 4.6 compares the value of the AIC of each estimated model for the two country groups and the three versions (A, D, and ND). For any country group and any version, the estimated model by the PQML has the lowest AIC value, the OLS with only positive-trade observations gives the next lowest, the Tobit gives the third, and the estimated model by the OLS including (value-replaced) zero-trade observations has the highest AIC value. This comparison indicates that, for any case, the equation estimated by the PQML describes the data the best.

4.5.5 Summary and Discussion

As described above, the result for the OECD countries is consistent across the four estimation approaches. The estimated coefficient for the size similarity index is

¹³¹ However, the difference of the estimate between (4.4D) and the other two equations is not significant (the p-value is 0.15).

¹³² Akaike (1974).

positive and significant not only in the differentiated-sector-trade equation but also in the aggregate-trade equation. The estimate for the differentiated sectors, however, is significantly larger than that in the other cases, and is also close to one.¹³³ On the other hand, the estimation for the non-differentiated sectors gives a smaller and less significant coefficient estimate than the other two cases, implying that the monopolistic competition model does not describe trade in the non-differentiated sectors as well as it does trade in the differentiated sectors. Therefore, this study, by separating differentiated (and non-differentiated) sectors from aggregate trade in estimation, clearly demonstrates that the positive correlation between the volume of trade among OECD countries and size similarity among the countries, which has been found in the previous studies, is driven by such correlation in trade of the differentiated products, as the monopolistic competition model suggests.

On the other hand, for the non-OECD countries, the results are mixed in the four approaches. Some methods estimate the coefficient for the similarity index being insignificant even in the differentiated-sector equation, but other methods estimate the coefficient being significant even for the non-differentiated sectors. However, the estimation by the PQML, which has econometric advantages (small potential estimation bias with zero-valued data) and better describes the data with a lower AIC value than the other three approaches, shows that the coefficient for the size similarity index is significant regardless of whether the traded sectors are differentiated or not. This result brings us back to Hummels and Levinsohn's puzzle; and also implies that Debaere's finding may be due to his way of handling zero-trade observations in estimation.

¹³³ The p-value of the test of the hypothesis that the coefficient equals zero ranges from 0.38 through 0.85 across the four estimation procedures.

Moreover, the current study deepens the puzzle. That is, the estimation indicates that for the non-OECD countries the correlation between the volume of trade and the size similarity between trading economies is weaker in the differentiated sectors than in the less differentiated sectors, while the correlation should be driven by product differentiation if the monopolistic competition model applies. The current study thus implies that some different mechanism from horizontal product differentiation may underlie the observed relationship between the volume of trade and the size similarity among these lower-income countries.

4.5.6 World GDP Share of Trading Countries

So far the analysis has been focused on the significance of the size similarity of two trading economies, which is one of the two determinants of the volume of trade per production in the model. In this subsection, I briefly discuss the estimation results for the other determinant: the GDP of two trading countries as a share in the world GDP (or, more simply, the country pair's world GDP share, y_{ij}). According to the monopolistic competition model, two countries' world GDP share should also be positively correlated with the volume of bilateral trade as a share in the countries' total production.

The results of the estimation from the four different approaches are as shown in Tables 4.2 through 4.5. For the OECD countries, the coefficient for the world GDP share (β_2) is insignificant in any estimation for any country group and trading sector. This result suggests that among rich countries how large trading countries are in the world may not be very important for the volume of trade per production. Exceptions, however are the estimates in the differentiated- and non-differentiated-sector equations, (4.4D) and (4.4ND), estimated by the PQML. In these cases the coefficient is estimated to be

positive and significant. In particular, for the differentiated sectors the estimate is fairly large (but smaller than one) and very significant (at the 1% level). This should be additional evidence that the monopolistic competition model explains the flows of trade in differentiated sectors among rich countries. On the other hand, for the non-OECD countries, the result varies across estimations. However, in the estimation by the PQML and the log-linear OLS without zero-trade observations that give the two lowest AIC values, the coefficient is positive and significant for all sectors. This result implies that trading countries' world income share plays an important role in determining the volume of trade among non-rich countries. This finding is consistent with the study by Jensen (2000) that estimates equations for bilateral one-way trade (import or export) derived from the monopolistic competition model. He has also found that the importer's income (GDP) is not significant for trade between rich countries but significant for trade between middle-income or poor countries.¹³⁴

4.5.7 Robustness Check: Alternative Groupings of Differentiated Sectors

Finally, for the purpose of checking the robustness of the estimation results, I re-estimate the volume-of-trade equations by varying criteria for selection of the group of the differentiated (and non-differentiated) sectors. The first alternative is to include in the differentiated sectors the three-digit ISIC industries in which corresponding Rauch's "differentiated" four-digit SITC goods share more than 90% of the world trade value throughout the period of 1970-1997. This grouping adds the following three industries as differentiated sectors to the 10 industries in the benchmark grouping: 323 (leather

¹³⁴ However, Jensen re-estimated the coefficient by replacing country pair-specific dummies with direct measures of barriers to trade such as bilateral distance and the importing country's tariff. As a result, he found that the importer's income is rather insignificant when the importing country is poor than when the importer is a rich country, which was counter to his initial finding.

products), 342 (printing and publishing), 383 (electric machinery). The second alternative is to include the three-digit ISIC industries that include none of Rauch's four-digit SITC goods "traded in an organized market," or homogeneous goods. The grouping further adds to the first alternative the following five industries: 313 (beverages), 352 (other chemicals), 354 (miscellaneous petroleum and coal products), 369 (other non-metallic mineral products), and 371 (iron and steel). Table 7 compares the benchmark and these two alternative groupings of the differentiated sectors by showing which three-digit ISIC manufacturing industries are included. Note that these two alternative groupings of the differentiated sectors cover broader sets of industries than the benchmark, and the second grouping includes more industries than the first.

The results of estimation by the respective four methods are presented in Tables 4.8.1 through 4.8.4 for the first alternative differentiated-sector grouping, and in Tables 4.9.1 through 4.9.4 for the second alternative grouping. The estimation results for both alternative groupings do not differ from the results of the estimation for the benchmark differentiated-sector grouping that are shown in Tables 4.2 through 4.5; and they thus confirm that the estimation results are robust across groupings of (non-)differentiated sectors.

It should also be noted that, for the OECD countries, the estimated coefficient for the size similarity is smaller in the differentiated-sector equation (D), and so is it in the non-differentiated-sector equation (ND), for the grouping with a broader range of industries (i.e., the first alternative compared to the benchmark; and the second alternative compared to the first). This finding for the OECD countries is consistent with what the model suggests, since the correlation between the volume of trade and the

(adjusted) size similarity is less clear as the sectors consist of less differentiated or more homogeneous industries, in which the monopolistic competition model does not primarily aim to describe the trade. However, for the non-OECD countries, the coefficient estimate in the non-differentiated-sector equation (ND) is *larger* for a non-differentiated-sector grouping that covers *less* differentiated industries. This implies that for lower-income countries the correlation between the trade volume and the size similarity among trading economies is greater as the traded sectors are more homogeneous, which is counter to the theoretical expectation. Varying the grouping of sectors in estimation thus underlines the puzzle in the results for non-rich countries.

4.6 Conclusion

This paper proposes an alternative approach to testing the monopolistic competition model of international trade. The monopolistic competition model, in which the main driving force of international trade is horizontal product differentiation, suggests that the volume of trade will be larger as trading countries are more similar in the size of the economy. In the preceding studies such as Hummels and Levinsohn (1995) and Debaere (2005), this implication of the model has been tested for the relationship between aggregate trade and GDP similarity among countries, while aggregate trade includes sectors that are not characterized by product differentiation.

In contrast to the existing literature, this paper focuses on trade of differentiated products that the monopolistic competition model directly aims to describe. The paper derives the equation for the volume of trade of differentiated products under less restrictive assumptions than those required to derive the aggregate-trade equation. The derived equation predicts that the volume of trade *in the differentiated sectors* will be

larger as the trading countries are more similar in GDP *and more symmetric in production structure*. This prediction is tested using the disaggregated data on trade and manufacturing production for various countries, in which industries are classified into the differentiated and non-differentiated sectors using the information on the degree of product differentiation provided by Rauch (1999). The test employs not only the conventional OLS regression for the log-linearized form of the equation but also the non-linear estimation methods such as PQML to handle zero-trade cases in the data.

The result shows that the predicted positive correlation between the volume of trade and the size similarity among countries is significant for both aggregate and differentiated sectors, regardless of whether the trade is among the OECD or non-OECD countries. This result, contrary to Debaere's conclusion, brings us back to the puzzle presented by Hummels and Levinsohn. Moreover, the proposed alternative approach in this paper reveals the following. First, for OECD countries the relationship between trade and the size similarity is shown more evidently by separating the differentiated sectors from aggregate trade, indicating that the monopolistic competition model explains very well trade in the differentiated sectors among OECD countries. Secondly, however, for non-OECD countries the predicted relationship between the volume of trade and the size similarity among countries is more pronounced in the *non*-differentiated sectors than in the differentiated sectors, which is counter to what is suggested by the model. The second point implies that trade flows among non-rich countries may be driven or crucially influenced by some other mechanism than what is described by the monopolistic competition model.

4.7 Appendix A

This appendix is to review empirical approaches of the two preceding studies; Hummels and Levinsohn (1995) and Debaere (2005). Both studies estimated some versions of the equation for the volume of aggregate bilateral trade, which are derived from the monopolistic competition model based on the two assumptions A1 and A2 described in the second section of this paper. The derivation of the equation is left to the section.

4.7.1 Hummels and Levinsohn (1995)

Hummels and Levinsohn estimated the following forms:¹³⁵

$$\text{for OECD countries: } \log(VT_{ijt}) = \beta \cdot \log(Y_{ijt} \cdot sim_{ijt}) + \mu_{ij} + \varepsilon_{ijt} \quad (*)$$

$$\text{for non-OECD countries: } VT_{ijt} = \beta \cdot (Y_{ijt} \cdot sim_{ijt}) + \eta_{ij} + \varepsilon_{ijt} \quad (**)$$

$$\text{where } sim_{ij} = [1 - (Y_i / Y_{ij})^2 - (Y_j / Y_{ij})^2].$$

Some points should be noted, in terms of differences from the equation applied in the current paper. First, they used the (logarithm of) the volume of aggregate trade itself as the dependent variable, rather than the volume of trade per GDP as in Equation (4.3A) in this paper. A country pair's GDP, which appears as the denominator on the left-hand side in Equation (4.3A), was put on the right-hand side as the product term with the size similarity index in their forms. Secondly, they accordingly estimated only one coefficient for the product term of GDP and the similarity index¹³⁶, but did not estimate the impacts of the two factors separately. Thirdly, they assumed, as Helpman (1987) did, that the world income share of a pair of two countries would not change (at least much) across

¹³⁵ Notations are not the same as those used in the original paper.

¹³⁶ Imposing the restriction that the coefficients for the two elements are the same is not a problem by itself, since the model suggests that the both elements are strictly proportional to the volume of trade. However, Debaere claims an econometric problem in this approach, as described later.

years, so that the term for the world income share (y_{ij}) was considered to be time-invariant and thus merged into the country pair-specific dummies η_{ij} in their equation for the OECD countries. (In the equation for the non-OECD, the time-invariant income share term was absorbed into the slope coefficient β .) They estimated the equation in the log-linear form (*) for the OECD countries but in the *level* form (**) for the non-OECD countries to keep observations with zero trade ($VT_{ijt} = 0$) in their estimation. They used balanced panel data on bilateral aggregate trade among 14 OECD countries in 1962-1983 to estimate Equation (*), and data for 14 non-OECD countries in 1962-1977 to estimate Equation (**). They applied the pooled OLS, random-effect OLS, and fixed-effect OLS regressions to both equations. In any case, they obtained an estimate for the coefficient β that was positive and significant for both country groups.

4.7.2 Debaere (2005)

Debaere started with a claim that the result of Hummels and Levinsohn, which was counter to the expectation for non-OECD countries, may have been driven by a high correlation between the volume of trade and GDP of country pairs rather than a correlation between trade and the size similarity of trading economies. He argued that, although the size similarity would not at all relate to, and thus be totally independent of, the volume of bilateral trade, the coefficient estimate for the product term of GDP and the similarity index ($Y_{ijt} \cdot sim_{ijt}$) would be significant if GDP (Y_{ijt}) is highly correlated to the volume of trade. This is in fact highly likely since in general the absolute volume of trade of large countries is greater than that of small countries.¹³⁷ Therefore, he used regression

¹³⁷ However, it should be noted that Hummels and Levinsohn seem to have noticed this issue by themselves. In fact, as they mentioned in their paper (Hummels and Levinsohn, 1995; pp. 808, footnote 14), they also estimated an equation separating the term for income size or GDP (Y_{ijt}) from the similarity index, from which they concluded that the impact of the similarity index was still significant.

equations whose dependent variable was the volume of bilateral aggregate trade as the share in GDP of the country pair. His benchmark is the estimation of the log-linear equation, which was the same as Equation (4.3A) in this paper, by the OLS with country pair-specific and year-specific dummies. For zero-trade observations in his non-OECD data, he applies a similar “replacement method” to the one that is used in the current paper.¹³⁸

In addition to his benchmark log-linear form, he estimated the following two level forms of the equation for the volume of aggregate trade per GDP:

$$VT_{ijt} / Y_{ijt} = \beta \cdot (y_{ijt} \cdot sim_{ijt}) + \mu_{ij} + \varepsilon_{ijt}$$

$$VT_{ijt} / Y_{ijt} = \beta_1 \cdot sim_{ijt} + \beta_2 \cdot y_{ijt} + \mu_{ij} + \varepsilon_{ijt}$$

These equations were estimated by the OLS for OECD countries, and by the Tobit regression for non-OECD countries. (The regressions also included year-specific dummies.) For the estimation, he constructed balanced panel data on bilateral (aggregate) trade and GDP for 14 OECD countries and 12 non-OECD countries for the period of 1970 through 1989. The results of the OLS estimation of his benchmark log-linear equation led him to conclude that the monopolistic competition model was supported for OECD countries but not for non-OECD countries, as he expected (and Hummels and Levinsohn also expected initially).¹³⁹

4.7.3 Countries in the Data

The table below lists countries that Hummels and Levinsohn selected for each of their OECD and non-OECD groups. The 14 countries in their OECD data are the same as

¹³⁸ See the third section of this paper.

¹³⁹ Although Debaere claimed that the results of his other estimations showed support for this conclusion, the evidence does not seem to be very clear but mixed.

those originally chosen by Helpman (1987). Debaere selected exactly the same sets of OECD and non-OECD countries as those in Hummels and Levinsohn's study, except that he excluded Congo and Cote d'Ivoire from the non-OECD group due to the unavailability of the data for these countries.¹⁴⁰ Note that the data in the current study cover a broader range of countries for both OECD and non-OECD groups (see Table 4.1).

OECD countries (14)		Non-OECD countries (14 [*])	
Austria	Italy	Brazil	Nigeria
Belgium	Japan	Cameroon	Norway
Canada	Netherlands	Columbia	Pakistan
Germany	Sweden	Congo [*]	Peru
Denmark	Switzerland	Cote d'Ivoire [*]	Philippines
France	United Kingdom	Greece	Paraguay
Ireland	United States	South Korea	Thailand

Note: Countries marked with asterisk (*) are not included in the data used by Debaere (2005).

¹⁴⁰ Hummels and Levinsohn, as well as Debaere, included Greece and Norway in their *non*-OECD group, while these two countries have been the original OECD members since 1961. In contrast, both countries are included in the OECD group for the current study.

Table 4.1: List of Countries and Years in Data

OECD (20 Countries)*		Non-OECD (69 Countries)			
Country	Years	Country	Years	Country	Years
Australia	1970-92	Albania	1993, 96	Morocco	1976
Austria	1970-97	United Arab Emirates	1977-78, 81	Moldova	1990-92
Belgium	1970-84	Argentina	1984-90, 93-96	Madagascar	1970-77
Canada	1970-94	Armenia	1994-97	Mexico	1994-97
Germany (West)	1971-84	Azerbaijan	1990-94	Macedonia	1990-96
Denmark	1970-91	Benin	1974-81	Mongolia	1993
Spain	1970-92	Bangladesh	1970-92, 95	Mozambique	1986-87, 91
Finland	1970-94	Bolivia	1981, 96, 97	Malawi	1970-75, 79-85
France	1970-79	Chile	1970-97	Malaysia	1970-97
United Kingdom	1970-92, 94, 95	Colombia	1970-97	Nigeria	1981-85, 91-96
Greece	1970-97	Costa Rica	1970-83, 91-97	Nicaragua	1970-85
Italy	1970-91	Dominican Republic	1970-84	Nepal	1997
Japan	1970-97	Algeria	1970-80	Oman	1994-97
Netherlands	1970-80	Ecuador	1970-97	Pakistan	1970-91
Norway	1970-91	Egypt	1970-96	Panama	1970-79, 92-95, 97
New Zealand	1970-89	Ethiopia	1981-96	Peru	1982-92, 94-96
Portugal	1970-89, 93-95	Gabon	1980-82, 91-95	Philippines	1970-97
Sweden	1970-97	Ghana	1970-87	Poland	1989-97
Turkey	1970-95	Gambia	1975-82	Russia	1993-97
United States	1970-95	Guatemala	1971-88, 91-95, 97	Saudi Arabia	1989
		Hong Kong	1973-90	Sudan	1972, 76
		Honduras	1971-75, 81-96	Senegal	1974-84, 89-90, 95, 97
		Croatia	1990-92	El Salvador	1970-85, 95-97
		Hungary	1970-97	Syria	1971-1979
		Indonesia	1994-96	Thailand	19774, 75, 77, 79, 82, 84, 88, 89, 91, 93, 94
		India	1970-97	Tunisia	1970-81
		Iran	1974-77, 79-90, 93	Tanzania	90-91
		Iraq	1970-77	Uganda	1971, 89
		Israel	1970-89	Uruguay	1971-86, 91-97
		Jordan	1971, 74-97	Venezuela	1970-97
		Kyrgyz Republic	1994	Serbia & Montenegro	1994-97
		Korea (South)	1970-96	South Africa	1970, 72-86, 96
		Kuwait	1970-97	Zambia	1970-75, 80-82
		Liberia	1984	Zimbabwe	1970-86, 96
		Sri Lanka	1970-74, 79-85, 96, 97		

Note: The OECD countries are grouped according to the OECD membership as of Year 1973.

Table 4.2: Result of OLS Estimation, with Positive-Trade Observations

Dependent Variable:	OECD Countries			Non-OECD Countries		
	log(Volume of Trade Per Production)			log(Volume of Trade Per Production)		
	Aggregate Eq. (4.3A)	Differentiated Eq. (4.3D)	Non-Diff'ed Eq. (4.3ND)	Aggregate Eq. (4.3A)	Differentiated Eq. (4.3D)	Non-Diff'ed Eq. (4.3ND)
log(similarity)	0.422***	0.858***	0.312*	0.577***	0.675***	0.562***
(s.e.)	(0.133)	(0.201)	(0.175)	(0.150)	(0.148)	(0.150)
[p-value]	[0.002]	[0.000]	[0.076]	[0.000]	[0.000]	[0.000]
log(world GDP share)	-0.163	0.284	-0.069	0.586***	0.514***	0.736***
(s.e.)	(0.163)	(0.270)	(0.195)	(0.147)	(0.159)	(0.137)
[p-value]	[0.318]	[0.293]	[0.724]	[0.000]	[0.001]	[0.000]
R-square	0.01	0.12	0.05	0.02	0.03	0.03
# observations	3,617	3,617	3,628	7,216	7,216	9,040
<i>(Tests for Coefficient = 1: P-values)</i>						
coef. for similarity = 1	0.000	0.481	0.000	0.005	0.028	0.004
coef. for income share = 1	0.000	0.009	0.000	0.005	0.002	0.054
coef. for similarity = coef. for i-share = 1	0.000	0.011	0.000	0.006	0.007	0.011

Notes: All variables are in logarithm. The OECD group includes 20 countries and the non-OECD group includes 69 countries, both for years 1970-97. Observations with zero trade in differentiated sectors are excluded from the regression. Country pair-specific and year-specific dummies are included in the regression. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Differentiated Sectors: 3-digit ISIC = 322, 324, 332, 355, 356, 361, 362, 382, 384, and 385.

Non-differentiated Sectors: 3-digit ISIC = 311, 313, 314, 321, 323, 331, 341, 342, 351, 352, 353, 354, 369, 371, 372, 381, and 383.

Table 4.3: Result of OLS Estimation, with All Observations

Dependent Variable:	OECD Countries			Non-OECD Countries		
	log(Volume of Trade Per Production)			log(Volume of Trade Per Production)		
	Aggregate Eq. (4.3A)	Differentiated Eq. (4.3D)	Non-Diff'ed Eq. (4.3ND)	Aggregate Eq. (4.3A)	Differentiated Eq. (4.3D)	Non-Diff'ed Eq. (4.3ND)
log(similarity)	0.397***	0.793***	0.195	0.062	0.246	0.012
(s.e.)	(0.138)	(0.235)	(0.226)	(0.242)	(0.202)	(0.230)
[p-value]	[0.004]	[0.001]	[0.389]	[0.796]	[0.224]	[0.960]
log(world GDP share)	-0.155	0.281	-0.105	0.000	0.733***	0.314
(s.e.)	(0.169)	(0.305)	(0.219)	(0.254)	(0.225)	(0.240)
[p-value]	[0.361]	[0.358]	[0.632]	[0.999]	[0.001]	[0.192]
R-square	0.01	0.09	0.02	0.00	0.07	0.03
# observations	3,630	3,630	3,630	14,565	14,565	14,565
<i>(Tests for Coefficient = 1: P-values)</i>						
coef. for similarity = 1	0.000	0.379	0.001	0.000	0.000	0.000
coef. for income share = 1	0.000	0.020	0.000	0.000	0.236	0.004
coef. for similarity = coef. for i-share = 1	0.000	0.033	0.000	0.000	0.001	0.000

Notes: All variables are in logarithm. The OECD group includes 20 countries and the non-OECD group includes 69 countries, both for years 1970-97. Observations with zero trade in differentiated sectors are included in the regression. Country pair-specific and year-specific dummies are included in the regression. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Differentiated Sectors: 3-digit ISIC = 322, 324, 332, 355, 356, 361, 362, 382, 384, and 385.

Non-differentiated Sectors: 3-digit ISIC = 311, 313, 314, 321, 323, 331, 341, 342, 351, 352, 353, 354, 369, 371, 372, 381, and 383.

Table 4.4: Result of Tobit Estimation

Dependent Variable:	OECD Countries			Non-OECD Countries		
	log(Volume of Trade Per Production)			log(Volume of Trade Per Production)		
	Aggregate Eq. (4.5A)	Differentiated Eq. (4.5D)	Non-Diff'ed Eq. (4.5ND)	Aggregate Eq. (4.5A)	Differentiated Eq. (4.5D)	Non-Diff'ed Eq. (4.5ND)
log(similarity)	0.397***	1.03***	0.287**	0.070	0.484*	0.265
(s.e.)	(0.133)	(0.171)	(0.142)	(0.269)	(0.261)	(0.274)
[p-value]	[0.003]	[0.000]	[0.044]	[0.794]	[0.063]	[0.334]
log(world GDP share)	-0.353	-0.517	-0.290	-0.784**	-0.068	-0.199
(s.e.)	(0.226)	(0.474)	(0.419)	(0.317)	(0.313)	(0.326)
[p-value]	[0.118]	[0.276]	[0.489]	[0.013]	[0.828]	[0.541]
R-square	0.81	0.63	0.77	0.31	0.37	0.31
# observations	3,630	3,630	3,630	14,565	14,565	14,565
<i>(Tests for Coefficient = 1: P-values)</i>						
coef. for similarity = 1	0.000	0.851	0.000	0.001	0.048	0.007
coef. for income share = 1	0.000	0.001	0.002	0.000	0.001	0.000
coef. for similarity = coef. for i-share = 1	0.000	0.002	0.000	0.000	0.001	0.000

Notes: All variables are in logarithm. Log of GDP (for the aggregate specification) or log of sectoral production (for the differentiated-sector specification) is included as a regressor, but the coefficient for the term is constrained to be 1. The OECD group includes 20 countries and the non-OECD sample includes 69 countries, both for years 1970-97. All observations are included, and left-censored at the value of ln(\$500). Country pair-specific and year-specific dummies are included in the regressions. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Differentiated Sectors: 3-digit ISIC = 322, 324, 332, 355, 356, 361, 362, 382, 384, and 385.

Non-differentiated Sectors: 3-digit ISIC = 311, 313, 314, 321, 323, 331, 341, 342, 351, 352, 353, 354, 369, 371, 372, 381, and 383.

Table 4.5: Results of Poisson Quasi-maximum Likelihood (PQML) Estimation

Dependent Variable:	OECD Countries			Non-OECD Countries		
	log(Volume of Trade Per Production)			log(Volume of Trade Per Production)		
	Aggregate Eq. (4.4A)	Differentiated Eq. (4.4D)	Non-Diff'ed Eq. (4.4ND)	Aggregate Eq. (4.4A)	Differentiated Eq. (4.4D)	Non-Diff'ed Eq. (4.4ND)
log(similarity)	0.628***	0.875***	0.497***	0.862***	0.434**	0.710***
(s.e.)	(0.120)	(0.149)	(0.140)	(0.311)	(0.194)	(0.182)
[p-value]	[0.000]	[0.000]	[0.000]	[0.006]	[0.025]	[0.000]
log(world GDP share)	0.102	0.583***	0.297**	0.652***	0.664***	0.527***
(s.e.)	(0.125)	(0.167)	(0.150)	(0.134)	(0.192)	(0.146)
[p-value]	[0.412]	[0.000]	[0.048]	[0.000]	[0.001]	[0.000]
# observations	3,630	3,630	3,630	12,329	10,218	11,869
<i>(Tests for Coefficient = 1: P-values)</i>						
coef. for similarity = 1	0.002	0.399	0.000	0.657	0.004	0.111
coef. for income share = 1	0.000	0.013	0.000	0.009	0.080	0.001
coef. for similarity = coef. for i-share = 1	0.000	0.034	0.000	0.026	0.008	0.005

Notes: The dependent variable is in level, while all the regressors are in logarithm. The OECD group includes 20 countries and the non-OECD group includes 69 countries, both for years 1970-97. All observations are included. The conditional fixed-effect PQML estimation follows Hausman et al. (1984), including time-specific dummies. Observations for country pairs that have data for only one year or whose volume of trade is zero for the entire period (1970-97) are omitted for the estimation. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Differentiated Sectors: 3-digit ISIC = 322, 324, 332, 355, 356, 361, 362, 382, 384, and 385.

Non-differentiated Sectors: 3-digit ISIC = 311, 313, 314, 321, 323, 331, 341, 342, 351, 352, 353, 354, 369, 371, 372, 381, and 383.

Table 4.6: Akaike Information Criteria (AIC) of Estimated Models

Estimated Models	OECD Countries			Non-OECD Countries		
	Volume-of-trade Equation for:					
	Aggregate (A)	Differentiated Sectors (D)	Non-diffed Sectors (ND)	Aggregate (A)	Differentiated Sectors (D)	Non-diffed Sectors (ND)
OLS (3): excluding zero-trade observations	806.60	4,434.27	1,620.14	19,985.79	22,575.58	29,521.20
OLS (3): including zero-trade observations	2,595.51	6,921.75	3,722.83	68,980.41	70,014.69	72,312.18
Tobit (5)	2,523.50	5,419.98	3,048.18	50,750.46	37,745.48	47,753.59
PQML (4)	134.10	198.73	148.90	99.50	126.52	123.00

Note: The number in the parentheses () following the name of estimation method indicates the equation number in the text.

Table 4.7: Alternative Groupings of Differentiated Sectors

ISIC	Industry	Benchmark Grouping	Alternative Grouping (1)	Alternative Grouping (2)
311	Food products			
313	Beverages			X
314	Tobacco			
321	Textiles			
322	Wearing apparel, except footwear	X	X	X
323	Leather products		X	X
324	Footwear, except rubber or plastic	X	X	X
331	Wood products, except furniture			
332	Furniture, except metal	X	X	X
341	Paper and products			
342	Printing and publishing		X	X
351	Industrial chemicals			
352	Other chemicals			X
353	Petroleum refineries			
354	Miscellaneous petroleum and coal products			X
355	Rubber products	X	X	X
356	Plastic products	X	X	X
361	Pottery, china, earthenware	X	X	X
362	Glass and products	X	X	X
369	Other non-metallic mineral products			X
371	Iron and steel			X
372	Non-ferrous metals			
381	Fabricated metal products			
382	Machinery, except electrical	X	X	X
383	Electric machinery		X	X
384	Transport equipment	X	X	X
385	Professional and scientific equipment	X	X	X
Number of manufacturing industries included in the differentiated sector group		10	13	18

Notes:

1. Manufacturing industries are classified according to the three-digit ISIC (Revision 2).
2. "X" indicates an industry included in the differentiated sector group for each grouping. The corresponding non-differentiated sector group comprises manufacturing industries that are not marked with "X."
3. The miscellaneous category ISIC 390 (other manufactured products) is excluded from the list.

Table 4.8.1: Result of OLS Estimation for Alternative Sector Grouping (1), with Positive-Trade Observations

Dependent Variable:	OECD Countries			Non-OECD Countries		
	log(Volume of Trade Per Production)			log(Volume of Trade Per Production)		
	Aggregate Eq. (4.3A)	Differentiated Eq. (4.3D)	Non-Diff'ed Eq. (4.3ND)	Aggregate Eq. (4.3A)	Differentiated Eq. (4.3D)	Non-Diff'ed Eq. (4.3ND)
log(similarity)	0.423***	0.741***	0.571***	0.510***	0.592***	0.571***
(s.e.)	(0.133)	(0.206)	(0.157)	(0.149)	(0.147)	(0.157)
[p-value]	[0.002]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]
log(world GDP share)	-0.166	0.217	0.663***	0.569***	0.578***	0.663***
(s.e.)	(0.164)	(0.274)	(0.141)	(0.147)	(0.149)	(0.141)
[p-value]	[0.312]	[0.429]	[0.000]	[0.000]	[0.000]	[0.000]
R-square	0.01	0.11	0.02	0.02	0.03	0.02
# observations	3,619	3,619	8,905	7,562	7,562	8,905
<i>(Tests for Coefficient = 1: P-values)</i>						
coef. for similarity = 1	0.000	0.210	0.006	0.001	0.006	0.006
coef. for income share = 1	0.000	0.005	0.017	0.003	0.005	0.017
coef. for similarity = coef. for i-share = 1	0.000	0.012	0.010	0.001	0.004	0.010

Notes: All variables are in logarithm. The OECD group includes 20 countries and the non-OECD group includes 69 countries, both for years 1970-97. Observations with zero trade in differentiated sectors are excluded from the regression. Country pair-specific and year-specific dummies are included in the regression. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Differentiated Sectors: 3-digit ISIC = 322, 323, 324, 332, 342, 355, 356, 361, 362, 382, 383, 384, and 385.

Non-differentiated Sectors: 3-digit ISIC = 311, 313, 314, 321, 331, 341, 351, 352, 353, 354, 369, 371, 372, and 381.

Table 4.8.2: Result of OLS Estimation for Alternative Sector Grouping (1), with All Observations

Dependent Variable:	OECD Countries			Non-OECD Countries		
	log(Volume of Trade Per Production)			log(Volume of Trade Per Production)		
	Aggregate Eq. (4.3A)	Differentiated Eq. (4.3D)	Non-Diff'ed Eq. (4.3ND)	Aggregate Eq. (4.3A)	Differentiated Eq. (4.3D)	Non-Diff'ed Eq. (4.3ND)
log(similarity)	0.397***	0.680***	0.150	0.062	0.131	0.017
(s.e.)	(0.138)	(0.239)	(0.235)	(0.242)	(0.208)	(0.233)
[p-value]	[0.004]	[0.005]	[0.523]	[0.796]	[0.528]	[0.940]
log(world GDP share)	-0.155	0.223	-0.242	0.000	0.792***	0.237
(s.e.)	(0.169)	(0.309)	(0.225)	(0.254)	(0.223)	(0.241)
[p-value]	[0.361]	[0.471]	[0.284]	[0.999]	[0.000]	[0.326]
R-square	0.01	0.09	0.00	0.00	0.08	0.02
# observations	3,630	3,630	3,630	14,565	14,565	14,565
<i>(Tests for Coefficient = 1: P-values)</i>						
coef. for similarity = 1	0.000	0.183	0.000	0.000	0.000	0.000
coef. for income share = 1	0.000	0.013	0.000	0.000	0.351	0.002
coef. for similarity = coef. for i-share = 1	0.000	0.034	0.000	0.000	0.000	0.000

Notes: All variables are in logarithm. The OECD group includes 20 countries and the non-OECD group includes 69 countries, both for years 1970-97. Observations with zero trade in differentiated sectors are included in the regression. Country pair-specific and year-specific dummies are included in the regression. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Differentiated Sectors: 3-digit ISIC = 322, 323, 324, 332, 342, 355, 356, 361, 362, 382, 383, 384, and 385.

Non-differentiated Sectors: 3-digit ISIC = 311, 313, 314, 321, 331, 341, 351, 352, 353, 354, 369, 371, 372, and 381.

Table 4.8.3: Result of Tobit Estimation for Alternative Sector Grouping (1)

Dependent Variable:	OECD Countries			Non-OECD Countries		
	log(Volume of Trade Per Production)			log(Volume of Trade Per Production)		
	Aggregate Eq. (4.5A)	Differentiated Eq. (4.5D)	Non-Diff'ed Eq. (4.5ND)	Aggregate Eq. (4.5A)	Differentiated Eq. (4.5D)	Non-Diff'ed Eq. (4.5ND)
log(similarity)	0.397***	0.960***	0.238*	0.070	0.421	0.239
(s.e.)	(0.133)	(0.176)	(0.138)	(0.269)	(0.258)	(0.282)
[p-value]	[0.003]	[0.000]	[0.085]	[0.794]	[0.103]	[0.396]
log(world GDP share)	-0.353	-0.830*	-0.370	-0.784**	-0.089	-0.213
(s.e.)	(0.226)	(0.486)	(0.374)	(0.317)	(0.298)	(0.325)
[p-value]	[0.118]	[0.088]	[0.323]	[0.013]	[0.766]	[0.513]
R-square	0.81	0.65	0.76	0.31	0.37	0.30
# observations	3,630	3,630	3,630	14,565	14,565	14,565
<i>(Tests for Coefficient = 1: P-values)</i>						
coef. for similarity = 1	0.000	0.820	0.000	0.001	0.025	0.007
coef. for income share = 1	0.000	0.000	0.000	0.000	0.000	0.000
coef. for similarity = coef. for i-share = 1	0.000	0.001	0.000	0.000	0.000	0.000

Notes: All variables are in logarithm. Log of GDP (for the aggregate specification) or log of sectoral production (for the differentiated-sector specification) is included as a regressor, but the coefficient for the term is constrained to be 1. The OECD group includes 20 countries and the non-OECD sample includes 69 countries, both for years 1970-97. All observations are included, and left-censored at the value of ln(\$500). Country pair-specific and year-specific dummies are included in the regressions. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Differentiated Sectors: 3-digit ISIC = 322, 323, 324, 332, 342, 355, 356, 361, 362, 382, 383, 384, and 385.

Non-differentiated Sectors: 3-digit ISIC = 311, 313, 314, 321, 331, 341, 351, 352, 353, 354, 369, 371, 372, and 381.

Table 4.8.4: Results of Poisson Quasi-maximum Likelihood (PQML) Estimation for Alternative Sector Grouping (1)

Dependent Variable:	OECD Countries			Non-OECD Countries		
	log(Volume of Trade Per Production)			log(Volume of Trade Per Production)		
	Aggregate Eq. (4.4A)	Differentiated Eq. (4.4D)	Non-Diff'ed Eq. (4.4ND)	Aggregate Eq. (4.4A)	Differentiated Eq. (4.4D)	Non-Diff'ed Eq. (4.4ND)
log(similarity)	0.628***	0.843***	0.418***	0.862***	0.362*	0.722***
(s.e.)	(0.120)	(0.143)	(0.139)	(0.311)	(0.190)	(0.193)
[p-value]	[0.000]	[0.000]	[0.003]	[0.006]	[0.057]	[0.000]
log(world GDP share)	0.102	0.589***	0.201	0.652***	0.681***	0.466***
(s.e.)	(0.125)	(0.178)	(0.131)	(0.134)	(0.167)	(0.147)
[p-value]	[0.412]	[0.001]	[0.125]	[0.000]	[0.000]	[0.002]
# observations	3,630	3,630	3,630	12,329	10,478	11,824
<i>(Tests for Coefficient = 1: P-values)</i>						
coef. for similarity = 1	0.002	0.271	0.000	0.657	0.001	0.149
coef. for income share = 1	0.000	0.021	0.000	0.009	0.057	0.000
coef. for similarity = coef. for i-share = 1	0.000	0.064	0.000	0.026	0.001	0.001

Notes: The dependent variable is in level, while all the regressors are in logarithm. The OECD group includes 20 countries and the non-OECD group includes 69 countries, both for years 1970-97. All observations are included. The conditional fixed-effect PQML estimation follows Hausman et al. (1984), including time-specific dummies. Observations for country pairs that have data for only one year or whose volume of trade is zero for the entire period (1970-97) are omitted for the estimation. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Differentiated Sectors: 3-digit ISIC = 322, 323, 324, 332, 342, 355, 356, 361, 362, 382, 383, 384, and 385.

Non-differentiated Sectors: 3-digit ISIC = 311, 313, 314, 321, 331, 341, 351, 352, 353, 354, 369, 371, 372, and 381.

Table 4.9.1: Result of OLS Estimation for Alternative Sector Grouping (2), with Positive-Trade Observations

Dependent Variable:	OECD Countries			Non-OECD Countries		
	log(Volume of Trade Per Production)			log(Volume of Trade Per Production)		
	Aggregate Eq. (4.3A)	Differentiated Eq. (4.3D)	Non-Diff'ed Eq. (4.3ND)	Aggregate Eq. (4.3A)	Differentiated Eq. (4.3D)	Non-Diff'ed Eq. (4.3ND)
log(similarity)	0.424***	0.660***	0.263*	0.516***	0.576***	0.569***
(s.e.)	(0.133)	(0.204)	(0.153)	(0.147)	(0.160)	(0.157)
[p-value]	[0.002]	[0.001]	[0.087]	[0.000]	[0.000]	[0.000]
log(world GDP share)	-0.166	0.108	-0.034	0.548***	0.395**	0.785***
(s.e.)	(0.164)	(0.266)	(0.185)	(0.145)	(0.156)	(0.142)
[p-value]	[0.315]	[0.685]	[0.853]	[0.000]	[0.011]	[0.000]
R-square	0.01	0.07	0.05	0.03	0.05	0.02
# observations	3,622	3,622	3,628	7,960	7,960	8,642
<i>(Tests for Coefficient = 1: P-values)</i>						
coef. for similarity = 1	0.000	0.097	0.000	0.001	0.008	0.006
coef. for income share = 1	0.000	0.001	0.000	0.002	0.000	0.131
coef. for similarity = coef. for i-share = 1	0.000	0.003	0.000	0.001	0.000	0.022

Notes: All variables are in logarithm. The OECD group includes 20 countries and the non-OECD group includes 69 countries, both for years 1970-97. Observations with zero trade in differentiated sectors are excluded from the regression. Country pair-specific and year-specific dummies are included in the regression. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Differentiated Sectors: 3-digit ISIC = 313, 322, 323, 324, 332, 342, 352, 354, 355, 356, 361, 362, 369, 371, 382, 383, 384, and 385.

Non-differentiated Sectors: 3-digit ISIC = 311, 314, 321, 331, 341, 351, 353, 372, and 381.

Table 4.9.2: Result of OLS Estimation for Alternative Sector Grouping (2), with All Observations

Dependent Variable:	OECD Countries			Non-OECD Countries		
	log(Volume of Trade Per Production)			log(Volume of Trade Per Production)		
	Aggregate Eq. (4.3A)	Differentiated Eq. (4.3D)	Non-Diff'ed Eq. (4.3ND)	Aggregate Eq. (4.3A)	Differentiated Eq. (4.3D)	Non-Diff'ed Eq. (4.3ND)
log(similarity)	0.397***	0.572**	0.142	0.062	0.052	0.049
(s.e.)	(0.138)	(0.236)	(0.212)	(0.242)	(0.218)	(0.234)
[p-value]	[0.004]	[0.016]	[0.501]	[0.796]	[0.810]	[0.833]
log(world GDP share)	-0.155	0.081	-0.065	0.000	0.456**	0.492**
(s.e.)	(0.169)	(0.300)	(0.209)	(0.254)	(0.224)	(0.240)
[p-value]	[0.361]	[0.787]	[0.755]	[0.999]	[0.042]	[0.041]
R-square	0.01	0.05	0.02	0.00	0.05	0.04
# observations	3,630	3,630	3,630	14,565	14,565	14,565
<i>(Tests for Coefficient = 1: P-values)</i>						
coef. for similarity = 1	0.000	0.071	0.000	0.000	0.000	0.000
coef. for income share = 1	0.000	0.003	0.000	0.000	0.015	0.035
coef. for similarity = coef. for i-share = 1	0.000	0.008	0.000	0.000	0.000	0.000

Notes: All variables are in logarithm. The OECD group includes 20 countries and the non-OECD group includes 69 countries, both for years 1970-97. Observations with zero trade in differentiated sectors are included in the regression. Country pair-specific and year-specific dummies are included in the regression. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Differentiated Sectors: 3-digit ISIC = 313, 322, 323, 324, 332, 342, 352, 354, 355, 356, 361, 362, 369, 371, 382, 383, 384, and 385.

Non-differentiated Sectors: 3-digit ISIC = 311, 314, 321, 331, 341, 351, 353, 372, and 381.

Table 4.9.3: Result of Tobit Estimation for Alternative Sector Grouping (2)

Dependent Variable:	OECD Countries			Non-OECD Countries		
	log(Volume of Trade Per Production)			log(Volume of Trade Per Production)		
	Aggregate Eq. (4.5A)	Differentiated Eq. (4.5D)	Non-Diff'ed Eq. (4.5ND)	Aggregate Eq. (4.5A)	Differentiated Eq. (4.5D)	Non-Diff'ed Eq. (4.5ND)
log(similarity)	0.397***	0.82]***	0.214*	0.070	0.203	0.301
(s.e.)	(0.133)	(0.161)	(0.126)	(0.269)	(0.278)	(0.279)
[p-value]	[0.003]	[0.000]	[0.088]	[0.794]	[0.464]	[0.281]
log(world GDP share)	-0.353	-0.712	-0.142	-0.784**	-0.102	-0.064
(s.e.)	(0.226)	(0.488)	(0.317)	(0.317)	(0.314)	(0.324)
[p-value]	[0.118]	[0.144]	[0.655]	[0.013]	[0.745]	[0.843]
R-square	0.81	0.69	0.75	0.31	0.36	0.30
# observations	3,630	3,630	3,630	14,565	14,565	14,565
<i>(Tests for Coefficient = 1: P-values)</i>						
coef. for similarity = 1	0.000	0.266	0.000	0.001	0.004	0.012
coef. for income share = 1	0.000	0.001	0.000	0.000	0.000	0.001
coef. for similarity = coef. for i-share = 1	0.000	0.002	0.000	0.000	0.000	0.001

Notes: All variables are in logarithm. Log of GDP (for the aggregate specification) or log of sectoral production (for the differentiated-sector specification) is included as a regressor, but the coefficient for the term is constrained to be 1. The OECD group includes 20 countries and the non-OECD sample includes 69 countries, both for years 1970-97. All observations are included, and left-censored at the value of ln(\$500). Country pair-specific and year-specific dummies are included in the regressions. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Differentiated Sectors: 3-digit ISIC = 313, 322, 323, 324, 332, 342, 352, 354, 355, 356, 361, 362, 369, 371, 382, 383, 384, and 385.

Non-differentiated Sectors: 3-digit ISIC = 311, 314, 321, 331, 341, 351, 353, 372, and 381.

Table 4.9.4: Results of Poisson Quasi-maximum Likelihood (PQML) Estimation for Alternative Sector Grouping (2)

Dependent Variable:	OECD Countries			Non-OECD Countries		
	log(Volume of Trade Per Production)			log(Volume of Trade Per Production)		
	Aggregate Eq. (4.4A)	Differentiated Eq. (4.4D)	Non-Diff'ed Eq. (4.4ND)	Aggregate Eq. (4.4A)	Differentiated Eq. (4.4D)	Non-Diff'ed Eq. (4.4ND)
log(similarity)	0.628***	0.763***	0.403***	0.862***	0.469**	0.742***
(s.e.)	(0.120)	(0.138)	(0.134)	(0.311)	(0.194)	(0.191)
[p-value]	[0.000]	[0.000]	[0.003]	[0.006]	[0.015]	[0.000]
log(world GDP share)	0.102	0.536***	0.225*	0.652***	0.514**	0.585***
(s.e.)	(0.125)	(0.164)	(0.136)	(0.134)	(0.201)	(0.139)
[p-value]	[0.412]	[0.001]	[0.098]	[0.000]	[0.011]	[0.000]
# observations	3,630	3,630	3,630	12,329	10,831	11,659
<i>(Tests for Coefficient = 1: P-values)</i>						
coef. for similarity = 1	0.002	0.086	0.000	0.657	0.006	0.176
coef. for income share = 1	0.000	0.005	0.000	0.009	0.016	0.003
coef. for similarity = coef. for i-share = 1	0.000	0.018	0.000	0.026	0.004	0.012

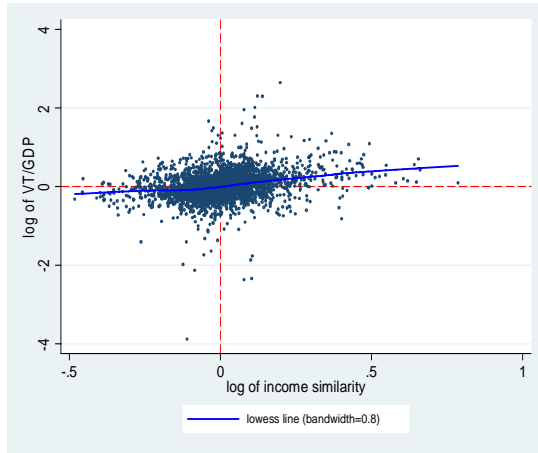
Notes: The dependent variable is in level, while all the regressors are in logarithm. The OECD group includes 20 countries and the non-OECD group includes 69 countries, both for years 1970-97. All observations are included. The conditional fixed-effect PQML estimation follows Hausman et al. (1984), including time-specific dummies. Observations for country pairs that have data for only one year or whose volume of trade is zero for the entire period (1970-97) are omitted for the estimation. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Differentiated Sectors: 3-digit ISIC = 313, 322, 323, 324, 332, 342, 352, 354, 355, 356, 361, 362, 369, 371, 382, 383, 384, and 385.

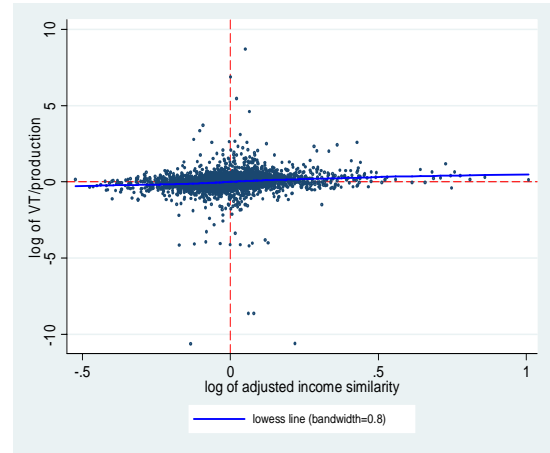
Non-differentiated Sectors: 3-digit ISIC = 311, 314, 321, 331, 341, 351, 353, 372, and 381.

Figures 4.1: Volume of Bilateral Trade per Production vs Size Similarity Index; for OECD Countries (in logarithm; mean-differenced)

4.1A: Aggregate Trade

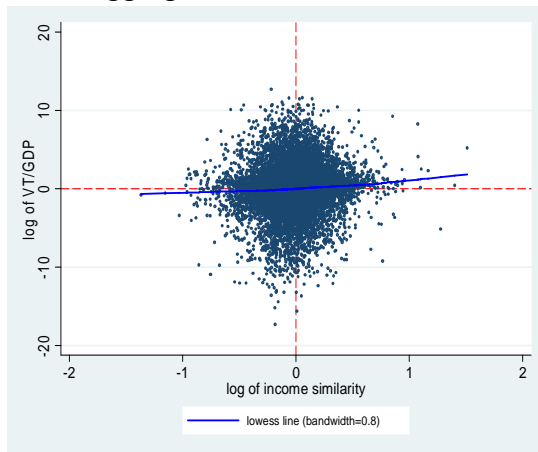


4.1D: Differentiated-sector Trade

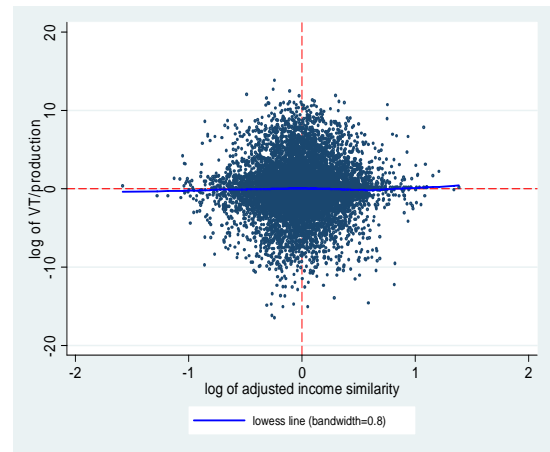


Figures 4.2: Volume of Bilateral Trade per Production vs Size Similarity Index; for Non-OECD Countries (in logarithm; mean-differenced)

4.2A: Aggregate Trade



4.2D: Differentiated-sector Trade



Notes: The GDP similarity index (for 4.1A and 4.2A) or the production structure-adjusted size similarity index (for 4.1D and 4.2D) is on the horizontal axis, and the volume of bilateral trade as the share in GDP (for 4.1A and 4.2A) or production (for 4.1D and 4.2D) on the vertical. All the variables are in logarithm and mean-differenced (for the fixed-effect OLS). The vertical and horizontal lines indicate zero. The solid line in each figures is the trend lines fitted by locally weighted regression (Lowess) with the bandwidth = 0.8.

CHAPTER V

Conclusion

This dissertation has investigated the patterns of international trade in various dimensions such as firms (Chapter II), products (Chapter III), and the volume of flows (Chapter IV), in terms of the roles of three forces: comparative advantage, monopolistic competition, and firm-level heterogeneity. Although these concepts have risen at different stages of the evolution of trade theory, these three are key elements of today's research in international trade.

The second and third chapters rest on an economic model that combines the traditional framework with the recent theoretical development. This integrated model has illustrated how comparative advantage operates when all the three forces function and interact. Having empirical analyses added to the theoretical investigation, the two chapters have demonstrated the significance of the traditional comparative advantage even in the firm-level and product-level phenomena of international trade, which the factor proportion theory did not consider in its original form.

In contrast, Chapter IV has put the research focus on monopolistic competition and its key feature—horizontal product differentiation—with the other two elements left aside. Through the empirical examination using the information on product

characteristics, the chapter has shown that monopolistic competition well explains trade flows among a group of relatively homogeneous countries (such as OECD countries) but does not very well when trading countries are more diverse (such as non-OECD countries). This finding also suggests the potential importance of country-level factors, which include comparative advantage highlighted in the preceding chapters, as determinants of trade patterns.

BIBLIOGRAPHY

- [1] Aitken, Brian, Gordon H. Hanson, and Ann E. Harrison (1997), “Spillovers, Foreign Investment, and Export Behavior,” *Journal of International Economics*, Vol. 43, No. 1-2, pp. 103–132
- [2] Akaike, Hirotugu (1974), “A New Look at the Statistical Model Identification,” *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, pp. 716–723
- [3] Alvarez, Roberto and Ricardo A. Lopez (2005), “Exporting and Performance: Evidence from Chilean Plants,” *Canadian Journal of Economics*, Vol. 38, No. 4, pp. 1384–1400
- [4] Alvarez, Roberto and Ricardo A. Lopez (2006), “Entry and Exit in International Markets: Evidence from Chilean Data,” CAEPR Working Paper No. 2006–014, Center for Applied Economics and Policy Research, Indiana University
- [5] Anderson, James E. (1979), “A Theoretical Foundation for the Gravity Equation,” *American Economic Review*, Vol. 69, No. 1, pp. 106–116
- [6] Anderson, James E. and Eric van Wincoop (2003), “Gravity with Gravitas: A Solution to the Border Puzzle,” *American Economic Review*, Vol. 93, No. 1, pp. 170–192
- [7] Anderson, James E. and Eric van Wincoop (2004), “Trade Costs,” *Journal of Economic Literature*, Vol. 42, No. 3, pp. 691–751
- [8] Aw, Bee Y., Mark J. Roberts, and Tor Winston (2007), “Export Market Participation, Investment in R&D and Worker Training, and the Evolution of Firm Productivity,” *The World Economy*, Vol. 30, No. 1, pp. 83–104
- [9] Baldwin, Robert E. (1971), “Determinants of the Commodity Structure of U.S. Trade,” *American Economic Review*, Vol. 61, No. 1, pp. 126–146
- [10] Baldwin, John R. and Wulong Gu (2003), “Export-market Participation and Productivity Performance in Canadian Manufacturing,” *Canadian Journal of Economics*, Vol. 36, No. 3, pp. 634–657

- [11] Banerjee, Abhijit V. and Esther Duflo (2005), "Growth Theory through the Lens of Development Economics," in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 1, Part 1, pp. 473–552
- [12] Barro, Robert J. and Jong-Wha Lee (1993), "International Comparisons of Educational Attainment," *Journal of Monetary Economics*, Vol. 32, No. 3, pp. 363–394
- [13] Barro, Robert J. and Jong-Wha Lee (2000), "International Data on Educational Attainment: Updates and Implications," CID Working Paper No. 42, Harvard University
- [14] Bergoeing, Raphael and Andrea Repetto (2006), "Micro Efficiency and Aggregate Growth in Chile," *Cuadernos de Economía*, Vol. 43, No. 127, pp. 169–191
- [15] Bergstrand, Jeffrey H. (1985), "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence," *Review of Economics and Statistics*, Vol. 67, No. 3, pp. 474–481
- [16] Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum (2003), "Plants and Productivity in International Trade," *American Economic Review*, Vol. 93, No. 4, pp. 1268–1290
- [17] Bernard, Andrew B. and J. Bradford Jensen (1995), "Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987," *Brookings Papers on Economic Activity, Microeconomics*, Vol. 1995, pp. 67–119
- [18] Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott (2007), "Comparative Advantage and Heterogeneous Firms," *Review of Economic Studies*, Vol. 74, No. 1, pp. 31–66
- [19] Bilal, Mark and Peter J. Klenow (2000), "Does Schooling Cause Growth?," *American Economic Review*, Vol. 90, No. 5, pp. 1160–1183
- [20] Blalock, Garrick and Paul J. Gertler (2004), "Learning from Exporting Revisited in A Less Developed Setting," *Journal of Development Economics*, Vol. 75, No. 2, pp. 397–416
- [21] Broda, Christian and David E. Weinstein (2004), "Variety Growth and World Welfare," *American Economic Review*, Vol. 94, No. 2, pp. 139–144
- [22] Castellani, Davide and Antonello Zanfei (2007), "Internationalisation, Innovation and Productivity: How Do Firms Differ in Italy?," *The World Economy*, Vol. 30, No. 1, pp. 156–176

- [23] Carree, Martin A., Luuk Klomp, and A. Roy Thurik (2000), "Productivity Convergence in OECD Manufacturing Industries," *Economic Letters*, Vol. 66, No. 3, pp. 337–345
- [24] Chaney, Thomas (2008), "Distorted Gravity: The Intensive and Extensive Margins of International Trade," Mimeo, University of Chicago
- [25] Choi, Yo C., David Hummels, and Chong Xiang (2006), "Explaining Import Variety and Quality: the Role of the Income Distribution," NBER Working Paper No. 12531
- [26] Clerides, Sofronis K., Saul Lach, and James R. Tybout (1998), "Is Learning by Exporting Important? Micro-dynamic Evidence from Colombia, Mexico, and Morocco," *Quarterly Journal of Economics*, Vol. 113, No. 3, pp. 903–947
- [27] Cogan, John F. (1981), "Fixed Costs and Labor Supply," *Econometrica*, Vol. 49, No. 4, pp. 945–963
- [28] Davis, Donald R. and David E. Weinstein (2003), "Market Access, Economic Geography, and Comparative Advantage: An Empirical Assessment," *Journal of International Economics*, Vol. 59, No. 1, pp. 1–23
- [29] Davis, Donald R. and David E. Weinstein (2001), "An Account of Global Factor Trade," *American Economic Review*, Vol. 91, No. 5, pp. 1423–1453
- [30] Deardorff, Alan V. (1980), "The General Validity of Law of Comparative Advantage," *Journal of Political Economy*, Vol. 88, No. 5, pp. 941–957
- [31] Deardorff, Alan V. (1982), "The General Validity of the Heckscher-Ohlin Theorem," *American Economic Review*, Vol. 72, No. 4, pp. 683–694
- [32] Deardorff, Alan V. (1984), "Testing Trade Theories and Predicting Trade Flows," in Ronald W. Jones and Peter B. Kenen, eds., *Handbook of International Economics*, Vol. 1, pp. 467–517
- [33] Deardorff, Alan V. (1998), "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?," in Jeffrey A. Frankel, eds., *The Regionalization of the World Economy*, University of Chicago Press, Chicago, I.L.
- [34] Debaere, Peter (2005), "Monopolistic Competition and Trade, Revisited: Testing the Model without Testing for Gravity," *Journal of International Economics*, Vol. 66, No. 1, pp. 249–266
- [35] Dixit, Avinash K. and Victor Norman (1980), *Theory of International Trade*, Cambridge University Press, Cambridge, U.K.

- [36] Dornbusch, Rudiger, Stanley Fischer, and Paul A. Samuelson (1980), “Heckscher-Ohlin Trade Theory with a Continuum of Goods,” *Quarterly Journal of Economics*, Vol. 95, No. 2, pp. 823–839
- [37] Eaton, Jonathan and Samuel Kortum (2002), “Technology, Geography, and Trade,” *Econometrica*, Vol. 70, No. 5, pp. 1741–1779
- [38] Eaton, Jonathan and Akiko Tamura (1994), “Bilateralism and Regionalism in Japanese and U.S. Trade and Direct Foreign Investment Patterns,” *Journal of the Japanese and International Economies*, Vol. 8, No. 4, pp. 478–510
- [39] Ederington, Josh and Phillip McCalman (2008), “Endogenous Firm Heterogeneity and the Dynamics of Trade Liberalization,” *Journal of International Economics*, Vol. 74, No. 2, pp. 422–440
- [40] Evenett, Simon J. and Wolfgang Keller (2002), “On Theories Explaining the Success of the Gravity Equation,” *Journal of Political Economy*, Vol. 110, No. 2, pp. 281–316
- [41] Evenett, Simon J. and Anthony J. Venables (2002), “Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows,” Mimeo, Center for Economic Policy Research
- [42] Falvey, Rod, David Greenaway, and Zhihong Yu (2004), “Intra-industry Trade Between Asymmetric Countries with Heterogeneous Firms,” GEP Research Paper 2004/05, Leverhulme Centre for Research on Globalisation and Economic Policy, University of Nottingham
- [43] Falvey, Rodney E. and Henryk Kierzkowski (1987), “Product Quality, Intra-industry Trade and Imperfect Competition,” in Henryk Kierzkowski, eds., *Protection and Competition in International Trade: Essays in Honor of W. M. Corden*, Basil Blackwell, Oxford and London, U.K.
- [44] Feenstra, Robert C. (1994), “New Product Varieties and the Measurement of International Prices,” *American Economic Review*, Vol. 84, No. 1, pp. 157–177
- [45] Feenstra, Robert C. (2000), “World Trade Flows, 1980-1997,” Center for International Data, Institute of Governmental Affairs, University of California, Davis
- [46] Feenstra, Robert C. and Hiau L. Kee (2004a), “On the Measurement of Product Variety in Trade,” *American Economic Review*, Vol. 94, No. 2, pp. 145–149
- [47] Feenstra, Robert C. and Hiau L. Kee (2004b), “Export Variety and Country Productivity,” World Bank Policy Research Working Paper 3412

- [48] Feenstra, Robert C. and Hiau L. Kee (2008), "Export Variety and Country Productivity: Estimating the Monopolistic Competition Model with Endogenous Productivity," *Journal of International Economics*, Vol. 74, No. 2, pp. 500–518
- [49] Feenstra, Robert C., Robert E. Lipsey, and Harry P. Bowen (1997), "World Trade Flows, 1970-1992, with Production and Tariff Data," NBER Working Paper No. 5910
- [50] Feenstra, Robert C., James R. Markusen, and Andrew K. Rose (2001), "Using the Gravity Equation to Differentiate among Alternative Theories of Trade," *Canadian Journal of Economics*, Vol. 34, No. 2, pp. 430–447
- [51] Feenstra, Robert C., John Romalis and Peter K. Schott (2002), "U.S. Imports, Exports, and Tariff Data, 1989–2001," NBER Working Paper No. 9387
- [52] Feenstra, Robert C., Tzu-Han Yang, and Gary G. Hamilton (1999), "Business groups and product variety in trade: evidence from South Korea, Taiwan and Japan," *Journal of International Economics*, Vol. 48, No. 1, pp. 71–100
- [53] Finger, J. Michael (1975), "Trade Overlap and Intra-Industry Trade," *Economic Inquiry*, Vol. 13, No. 4, pp. 581–589
- [54] Freeman, Richard B. and Remco H. Oostendorp (2000), "Wages Across the World: Pay Across Occupations and Countries," NBER Working Paper No. 8058
- [55] Funke, Michael and Ralf Ruhwedel (2001a), "Product Variety and Economic Growth: Empirical Evidence for the OECD Countries," IMF Staff Papers, Vol. 48, No. 2
- [56] Funke, Michael and Ralf Ruhwedel (2001b), "Export Variety and Export Performance: Empirical Evidence for East Asia," *Journal of Asian Economics*, Vol. 12, No. 4, pp. 493–505
- [57] Funke, Michael and Ralf Ruhwedel (2005), "Export Variety and Economic Growth in East European Transition Economies," *Economics of Transition*, Vol. 13, No. 1, pp. 25–50
- [58] Ghironi, Fabio and Marc J. Melitz (2005), "International Trade and Macroeconomic Dynamics with Heterogeneous Firms," *Quarterly Journal of Economics*, Vol. 120, No. 3, pp. 865–915
- [59] Girma, Sourafel., Richard Kneller, and Mauro Pisu (2005), "Exports versus FDI: An Empirical Test," *Weltwirtschaftliches Archiv*, Vol. 141, No. 2, pp. 193–218

- [60] Greenaway, David and Richard Kneller (2007), "Firm Heterogeneity, Exporting and Foreign Direct Investment," *Economic Journal*, Vol. 117, No. 517, pp. F134–F161
- [61] Greenaway, David and Zhihong Yu (2004), "Firm Level Interactions Between Exporting and Productivity: Industry-specific Evidence," *Weltwirtschaftliches Archiv*, Vol. 140, No. 3, pp. 376–392
- [62] Grubel, Herbert G. and Peter J. Lloyd (1975), *Intra-Industry Trade: The Theory and Measurement of International Trade in Differentiated Products*, Macmillan, London, U.K.
- [63] Hall, Robert E. and Charles I. Jones (1999), "Why Do Some Countries Produce So Much More Output Per Worker Than Others?," *Quarterly Journal of Economics*, Vol. 114, No. 1, pp. 83–116
- [64] Hallak, Juan C. (2006a), "Product Quality and the Direction of Trade," *Journal of International Economics*, Vol. 68, No. 1, pp. 238–265
- [65] Hallak, Juan C. (2006b), "A Product-Quality View of the Linder Hypothesis," NBER Working Paper No. 12712
- [66] Hallak, Juan C. and Jagadeesh Sivadasan (2006), "Productivity, Quality and Export Behavior under Minimum Quality Requirements," Mimeo, University of Michigan
- [67] Harkness, Jon (1978), "Factor Abundance and Comparative Advantage," *American Economic Review*, Vol. 68, No. 5, pp. 784–800
- [68] Harrigan, James (1994), "Scale Economies and the Volume of Trade," *Review of Economics and Statistics*, Vol. 76, No. 2, pp. 321–328
- [69] Harrigan, James (1996), "Openness to Trade in Manufactures in the OECD," *Journal of International Economics*, Vol. 40, No. 1-2, pp. 23–39
- [70] Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches (1984), "Econometric Models for Count Data with an Application to Patents-R&D Relationship," *Econometrica*, Vol. 52, No. 4, pp. 909–938
- [71] Haveman, Jon and David Hummels (2004), "Alternative Hypotheses and the Volume of Trade: the Gravity Equation and the Extent of Specialization," *Canadian Journal of Economics*, Vol. 37, No. 1, pp. 199–218
- [72] Helpman, Elhanan (1981), "International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition: A Chamberlin-

- Heckscher-Ohlin Approach,” *Journal of International Economics*, Vol. 11, No. 3, pp. 305–340
- [73] Helpman, Elhanan (1987), “Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries,” *Journal of the Japanese and International Economies*, Vol. 1, No. 1, pp. 62–81
- [74] Helpman, Elhanan (2006), “Trade, FDI, and the Organization of Firms,” *Journal of Economic Literature*, Vol. 33, No. 3, pp. 589–630
- [75] Helpman, Elhanan and Paul R. Krugman (1985), *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, MIT Press, Cambridge, M.A.
- [76] Helpman, Elhanan, Marc J. Melitz, and Yona Rubinstein (2007), “Estimating Trade Flows: Trading Partners and Trading Volumes,” NBER Working Paper No. 12927
- [77] Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple (2004), “Export versus FDI with Heterogeneous Firms,” *American Economic Review*, Vol. 94, No. 1, pp. 300–316
- [78] Hendricks, Lutz (2002), “How Important Is Human Capital for Development? Evidence from Immigrant Earnings,” *American Economic Review*, Vol. 92, No. 1, pp. 198–219
- [79] Hsiao, Cheng (2003), *Analysis of Panel Data*, Cambridge University Press, New York, N.Y.
- [80] Hummels, David and Peter J. Klenow (2005), “The Variety and Quality of A Nation’s Trade,” *American Economics Review*, Vol. 95, No. 3, pp. 704–723
- [81] Hummels, David and James Levinsohn (1995), “Monopolistic Competition and International Trade: Reconsidering the Evidence,” *Quarterly Journal of Economics*, Vol. 110, No. 3, pp. 799–836
- [82] Hummels, David and Volodymyr Lugovskyy (2005), “Trade in Ideal Varieties: Theory and Evidence,” NBER Working Paper No. 11828
- [83] Jensen, Paul E. (2000), “Analysis of Bilateral Trade Patterns With Panel Data,” *Review of International Economics*, Vol. 8, No. 1, pp. 86–99
- [84] Kehoe, Timothy J. and Kim J. Ruhl (2003), “How Important is the New Goods Margin in International Trade?”, Federal Reserve Bank of Minneapolis Research Department Staff Report 324

- [85] Klenow, Peter J. and Andres Rodriguez-Clare (1997), “The Neoclassical Revival in Growth Economics: Has It Gone Too Far?,” *NBER Macroeconomics Annual*, Vol. 12, pp. 73–103
- [86] Krugman, Paul (1979), “Increasing Returns, Monopolistic Competition, and International Trade,” *Journal of International Economics*, Vol. 9, No. 4, pp. 469–479
- [87] Krugman, Paul (1980), “Scale Economies, Product Differentiation, and the Pattern of Trade,” *American Economic Review*, Vol. 70, No. 5, pp. 950–959
- [88] Lai, Huiwen and Susan C. Zhu (2004), “The Determinants of Bilateral Trade,” *Canadian Journal of Economics*, Vol. 37, No. 2, pp. 459–483
- [89] Lancaster, Kelvin (1980), “Intra-industry Trade under Perfect Monopolistic Competition,” *Journal of International Economics*, Vol. 10, No. 2, pp. 151–175
- [90] Lopez, Ricardo A. (2005), “Trade and Growth: Reconciling the Macroeconomic and Microeconomic Evidence,” *Journal of Economic Surveys*, Vol. 19, No. 4, pp. 623–648
- [91] McCallum, John (1995), “National Borders Matter: Canada-U.S. Regional Trade Patterns,” *American Economic Review*, Vol. 85, No. 3, pp. 615–623
- [92] Melitz, Marc J. (2003), “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, Vol. 71, No. 6, pp. 1695–1725
- [93] Melitz, Marc J. and Gianmarco I. P. Ottaviano (2008), “Market Size, Trade, and Productivity,” *Review of Economic Studies*, Vol. 75, No. 1, pp. 295–316
- [94] Melvin, James R. (1968), “Production and Trade with Two Factors and Three Goods,” *American Economic Review*, Vol. 58, No. 5, pp. 1249–1268
- [95] Morrow, Peter M. (2008), “East is East and West is West: A Ricardian-Heckscher-Ohlin Model of Comparative Advantage,” Mimeo, University of Toronto
- [96] Oostendorp, Remco H. (2005), “The Standardized ILO October Inquiry 1983-2003,” Mimeo, Free University Amsterdam
- [97] Psacharopoulos, George (1994), “Returns to Investment in Education: A Global Update,” *World Development*, Vol. 22, No. 9, pp. 1325–1343
- [98] Psacharopoulos, George and Ying C. Ng (1992), “Earnings and Education in Latin America: Assessing Priorities for Schooling Investments,” World Bank Policy Research Working Paper No. 1056

- [99] Psacharopoulos, George and Harry A. Patrinos (2002), “Returns to Investment in Education: A Further Update,” World Bank Policy Research Working Paper No. 2881
- [100] Rauch, James E. (1999), “Networks Versus Markets in International Trade,” *Journal of International Economics*, Vol. 48, No. 1, pp. 7–35
- [101] Roberts, Mark J. and James R. Tybout, eds. (1996), *Industrial Evolution in Developing Countries*, Oxford University Press, New York, N.Y.
- [102] Roberts, Mark J. and James R. Tybout (1997), “The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs,” *American Economic Review*, Vol. 87, No. 4, pp. 545–564
- [103] Romalis, John (2004), “Factor Proportions and the Structure of Commodity Trade,” *American Economic Review*, Vol. 94, No. 1, pp. 67–97
- [104] Samuelson, Paul A. (1949), “International Factor-Price Equalization Once Again,” *Economic Journal*, Vol. 59, No. 234, pp. 181–197
- [105] Saxonhouse, Gary R. (1989), “Differentiated Products, Economies of Scale, and Access to the Japanese Market,” in Robert C. Feenstra, eds., *Trade Policies for International Competitiveness*, University of Chicago Press, Chicago, I.L.
- [106] Schott, Peter K. (2003), “One Size Fits All? Heckscher-Olin Specialization in Global Production,” *American Economic Review*, Vol. 93, No.3, pp. 686–708
- [107] Schott, Peter K. (2004), “Across-Product versus Within-Product Specialization in International Trade,” *Quarterly Journal of Economics*, Vol. 119, No. 2, pp. 647–678
- [108] Silva, J. M. C. Santos and Silvana Tenreyro (2006), “The Log of Gravity,” *Review of Economics and Statistics*, Vol. 88, No. 4, pp. 641–658
- [109] Sivadasan, Jagadeesh (2007), “Product Market Liberalization and Productivity: Evidence from India,” Mimeo, University of Michigan
- [110] Stern, Robert M. and Keith E. Maskus (1981), “Determinants of the Structure of U.S. Foreign Trade, 1958-76,” *Journal of International Economics*, Vol. 11, No. 2, pp. 207–224
- [111] Stolper, Wolfgang F. and Paul A. Samuelson (1941), “Protection and Real Wage,” *Review of Economic Studies*, Vol. 9, No. 1, pp. 58–73

- [112] United Nations Industrial Development Organization (UNIDO) (2003), *Industrial Statistics Database at the 3-digit level of ISIC Code (Rev.2) (INDSTAT3 2003 ISIC Rev.2)*, New York, N.Y.
- [113] U. S. Bureau of the Census (1992), *Selected Characteristics of Manufacturing and Wholesale Establishments That Export: 1992*, Manufacturing Analytical Report Series AR92-2, U. S. Department of Commerce, Washington D.C.
- [114] Wagner, Joachim (2007), "Exports and Productivity: A Survey of the Evidence from Firm-level Data," *The World Economy*, Vol. 30, No. 1, pp. 60–82
- [115] Wooldridge, Jeffrey M. (1999), "Distribution-Free Estimation of Some Nonlinear Panel Data Models," *Journal of Econometrics*, Vol. 90, No. 1, pp. 77–97
- [116] Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, M.A.
- [117] World Bank (2006), *World Development Indicators 2006*, Washington D.C.
- [118] Westerlund, Joakim and Fredrik Wilhelmsson (2006), "Estimating the Gravity Model without Gravity Using Panel Data," Mimeo, Lund University
- [119] Yeaple, Stephen R. (2005), "A Simple Model of Firm Heterogeneity, International Trade, and Wages," *Journal of International Economics*, Vol. 65, No. 1, pp. 1–20