

DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 10-10

First submitted November 9, 2010, Revised November 24, 2010 as The distribution of firms and countries' size with heterogeneous firms, Revised April 14, 2011

The Home-market Effect across Industries with Heterogeneous Firms

Ha Nguyen
University of Colorado at Boulder

April 14, 2011

Department of Economics



University of Colorado at Boulder
Boulder, Colorado 80309

© April 2011 Ha Nguyen

The home-market effect across industries with heterogeneous firms

Ha Nguyen*

Department of Economics, University of Colorado at Boulder

April 14, 2011

Abstract

According to Krugman's home-market effect hypothesis a large country has more firms or products in an increasing returns to scale sector than does a small country. However, the large country's share of firms (or products) across industries, which are subject to increasing returns to scale, may vary with the characteristics of an industry. This study builds a model of monopolistic competition with heterogeneous firms to investigate several characteristics of an industry that affect the size of the home market effect. Our model predicts that industries with low trade costs, high fixed domestic costs, low fixed export costs, and high productivity dispersion will concentrate more in the large country. Using 3-digit SIC industries from 28 high income countries, the model's predictions are empirically tested. The empirical results are consistent with the predictions of the theoretical model.

Keywords: Home market effect, countries' size, industry characteristics, heterogeneous firms, distribution of firms, Market Structure

JEL Classification: F1, L1

*Address: The Department of Economics, e-mail: ha.nguyen@colorado.edu. I am very grateful to Professors: Robert McNown, Stephen Yeaple, Keith Maskus, Ardeshtir Sepehri, Thibault Fally, and Mattias Nilsson for advice and suggestions. All remaining errors are mine.

1 Introduction

The hypothesis of the "home market effect", which was first introduced by Krugman (1980), suggests two predictions: a large country has more products (or firms) in its increasing-returns to scale sector than does a small country and the large country's share of products (firms) in the increasing-returns sector exceeds its share of size. The second prediction implies that the large country is a net exporter in its increasing returns sector.

Although a large country can produce more products than does a small country, the large country's share of products may not be uniform across industries. Or we can say that the distribution of firms across industries between the large country and the small country is not similar. This difference can depend on industry characteristics. This study will investigate which industry characteristics affect that difference. This paper does not examine Krugman's second prediction (net exporter) of the hypothesis of the home market effect, so we prefer using the term "the distribution of firms across industries" or "difference in the number of products across industries" to using "home market effect" in this study.

Hanson and Xiang (2004)¹ was the first to examine how the strength of home-market effects varies with industry characteristics. They found that industries with high trade costs and low elasticity of substitution concentrate more in large countries. However, we think that some other industry characteristics such as fixed costs or productivity dispersion may affect the distribution of firms between large and small countries across industries.

We build a model based on the mechanism of heterogeneous firms (Melitz (2003)) to examine whether other industry characteristics affect the distribution of firms between large and small countries across industries (or the home market effect). Our model includes two countries; each country has many differentiated product industries in the increasing returns sector and one homogeneous product industry in the constant return sector. Labor is the only production factor in the model. As a result, our model shows that industries with low trade costs, high fixed domestic costs, low fixed export costs, high productivity

¹Some studies (i.e. Helpman and Krugman (1987), Amiti (1998), Hanson and Xiang (2004), Holmes and Stevens (2005)) have examined which country characteristics or industry characteristics influence the home market effect.

dispersion, and high elasticity of substitution will concentrate more in the large country, or the difference in the number of products between the large country and the small country is larger in these industries. Among these industry characteristics, the model shows that the impact of trade costs, fixed costs, and productivity dispersion on the distribution of firms (products) across industries between the large country and the small country is uniform. The impact of substitution elasticity depends on the relationship between fixed domestic costs and fixed export costs of the industry. In this study, we assume that fixed domestic costs are smaller than fixed export costs, so industries with high substitution elasticity will concentrate more in the large country.

Economics of scale can be a key factor to explain why industries with low trade costs, low fixed export costs, and high fixed domestic costs locate more in large countries. Because of economies of scale, the production costs of firms in the increasing returns to scale sector of the large country are lower than those in the small country. As a result, firms in the large country will produce products with lower prices. When trade costs and fixed export costs of industries are low, products with low prices from the large country will easily penetrate the small country and defeat the high price products of the small country. So, industries with low trade costs or low fixed export costs will tend to concentrate in the large country. Industries with high fixed domestic costs have high economics of scale, so the large country will attract more firms in these industries.

This study also shows that industries with high productivity dispersion and high elasticity of substitution concentrate more in the large country. Firms with low productivity can not operate in the small country due to high competitive pressures, but these firms still can still operate in the large country because of the diversity of consumer demand in the large country. So, industries with high productivity dispersion prefer concentrating in the large country to concentrating in the small country. Industries with high substitution elasticity have less differentiated goods or few substitutes, and when trade liberalization occurs, consumers choose and buy cheaper goods from the large country. Firms of the small country which cannot compete with firms of the large country may exit the market. This explains why industries with high elasticity of substitution tend to concentrate in the large country.

Our empirical results from examining the distribution of 3-digit SIC manufacturing industries in 28 high income countries support for the predictions from the theoretical model. To build the empirical model, we use the method of Hummels and Klenow (2002) (or Hummels and Klenow (2005)) to measure the distribution of firms across industries, and we use the industrial data of the US to represent the characteristics of industries.

As a result, our model finds that, in addition to the industry characteristics found in Hanson and Xiang (2004), other characteristics also affect the distribution of industries, such as fixed costs and productivity dispersion. In addition, the impact of the similar characteristics in our model also has some differences from Hanson and Xiang (2004). Our model finds that industries with low trade costs tend to concentrate in the large country, while Hanson and Xiang (2004) predict the opposite. However, the effect of this variable in their theoretical model is not uniform: this proposition fails for industries with very high trade costs. If we assume that fixed domestic costs are smaller than fixed export costs, our model suggests that industries with a high elasticity of substitution will locate more in the large country likewise contrasting with Hanson and Xiang (2004). These differences originate from the differences in the models: Our model is based on the mechanism of heterogeneous firms and has the appearance of a homogeneous product sector. While Hanson and Xiang (2004) use the mechanism of homogeneous firms and don't use the homogeneous product sector in their model.

Our empirical method is also different from the one of Hanson and Xiang (2004). They use the method of difference-in-difference to study the impact of industry characteristics (trade costs and substitution elasticity) on home market effects. One disadvantage of the difference-in-difference model is that we are not able to study the combinative effect of many industry characteristics (like our study) on the distribution of firms across industries. Besides, the difference-in-difference method can't incorporate industry variables in the regression model², so we are not able to observe the impact level of the industry characteristics on the distribution of firms across industries. We use an alternative empirical method to overcome these limitations.

²This method uses industry characteristics to choose treatment and control groups

To sum up, our paper contributes to the existing literature in two ways: First, this paper formulates a model of monopolistic competition with heterogeneous firms to study the distribution of firms across industries between large and small countries. In comparison with previous studies, our model incorporates three additional industry characteristics: fixed export costs, fixed domestic costs, and productivity dispersion, which are found to influence the distribution of firms across industries (or home market effect of industries). Second, this paper uses an alternative approach to empirically test the distribution of firms across industries between the large country and the small country. The results would be of interest to policy makers in both developed and developing countries, in terms of potential identifying industries these countries should invest and develop to compete in globalized trade.

The rest of the paper is organized as follows: section 2 introduces a model with heterogeneous firms and discusses its predictions. Section 3 describes the empirical methods used to examine the predictions from the theoretical model. Section 4 presents some data analysis and discusses the results of the empirical model. Section 5 concludes with discussion of some implications.

2 The Model

2.1 Set up

Assume that there are two countries (i, j) , and each country has $H+1$ industries. One industry produces a homogeneous product z with constant return to scale, while the remaining H industries produce a continuum of differentiated products with increasing returns to scale. Each firm is a monopolist for the variety which it produces. Let β_h denote the share of income spent on differentiated goods for sector h . The share of income spent on the homogeneous sector is then $1 - \sum_{h=1}^H \beta_h$. The homogeneous good z is considered as the numeraire and it can be freely traded. The price of good z is set to 1, so that if every country producing this good will have identical wage rate ($=1$). On the demand side, assume that all individuals in country i have the same utility function:

$$\max U = (1 - \sum_{h=1}^H \beta_h) \ln z + \sum_{h=1}^H \frac{\beta_h}{\alpha_h} \ln \left(\int_0^{n_h^i} x_h^i(v)^{\alpha_h} dv \right)$$

where $x_h^i(v)$ is the consumption of country i on a variety v produced by industry h . Let n_h^i denote the number of varieties produced by industry h . The parameter $\sigma_h = \frac{1}{1-\alpha_h} > 1$ is the constant elasticity of substitution across varieties in industry h with $\alpha_h > 0$. The budget constraint of country i is then

$$z + \sum_{h=1}^H \int_0^{n_h^i} p_h(v) x_h^i(v) dv = Y_i$$

where Y_i denotes total expenditure on all goods in country i . Combining the utility function with the budget constraint yields the following demand for each variety produced by an industry h in country i :

$$x_h^i(v) = \frac{\beta_h Y_i p_h(v)^{-\sigma_h}}{P_h^i^{1-\sigma_h}}$$

Where $P_h^i = \left(\int_0^{n_h^i} p_h(v)^{1-\sigma_h} dv \right)^{\frac{1}{1-\sigma_h}}$ is country i 's ideal price index for industry h and $p_h(v)$ is the price of variety v in country i .

2.2 Firms

Labor is the only input and the number of units of labors (a) needed to produce one unit of product varies across firms. In addition, a firm must pay a overhead production cost of f_d^h units of labor to produce a positive amount in each period. The overhead production costs refer to an ongoing expense of operating a firm such as accounting fees, advertising, rent, and utilities costs. This overhead fixed cost is assumed to be identical across firms operating in each industry. So the production cost of a firm is $ax_{ii}^h(v) + f_d^h$. If the firm sells its products to the foreign market, it must pay a fixed cost of f_x^h units of labor per foreign market in each period. The fixed export costs include costs of establishing the distribution network, advertising, or administrative costs in the foreign market.

In addition, an exporting firm in industry h must face an iceberg transportation cost of $\tau_{ij}^h \geq 1$. The production cost of an exporting firm is then given by $\tau_{ij}^h ax_{ij}^h(v) + f_x^h$. Assume

that the fixed cost and the distribution function of a in each industry are identical in two countries. In addition, transport costs are assumed to be identical between two countries, that is, $\tau_{ji}^h = \tau_{ij}^h = \tau^h$.

Each firm chooses the price of its variety to maximize its profit, taking as given the price charged by other firms. Since a is the number of units of labor required to produce one unit of the product in industry h in country i , $\frac{1}{a}$ is considered the productivity of a firm in industry h . Firms having a productivity larger than $\frac{1}{a_D^h}$ produce and sell their products in the domestic market and firms with the productivity $\frac{1}{a_D^h}$ earn zero profits. The set of firms with $\frac{1}{a} > \frac{1}{a_X^h}$ produce products for the domestic market and for the exporting market. The set of firms with $\frac{1}{a_D^h} \leq \frac{1}{a} \leq \frac{1}{a_X^h}$ produce for the domestic market only. The set of firms with $\frac{1}{a} \leq \frac{1}{a_D^h}$ earn a negative profit and do not produce.

The profit of a firm in industry h in country i selling its product in the domestic market is

$$\pi_d^{ih} = p_{ii}^h(v)x_{ii}^h(v) - (ax_{ii}^h(v) + f_d^h)$$

The profit of an exporting firm is

$$\pi_x^{ih} = p_{ij}^h(v)x_{ij}^h(v) - (a\tau x_{ij}^h(v) + f_x^h)$$

The price which a firm will set for the domestic market is $p_{ii}^h(v) = (\frac{\sigma_h}{\sigma_h - 1})a = \frac{a}{\alpha_h}$ and for the foreign market $p_{ij}^h(v) = \frac{\tau^h a}{\alpha_h}$. Substituting domestic value, exporting value, and the prices into the profit equations, the profits of firms in industry h in the domestic market (i) and the exporting market(j) are:

$$\pi_d^{ih} = a^{1-\sigma_h} B_h^i - f_d^h$$

$$\pi_x^{ih} = a^{1-\sigma_h} \tau^{1-\sigma_h} B_h^j - f_x^h$$

$$\text{with } B_h^i = A_h^i \alpha_h^{\sigma_h - 1} (1 - \alpha_h) \text{ and } A_h^i = \frac{\beta_h Y_i}{\int_0^{\frac{1}{a_X^h}} p(v)^{(1-\sigma_h)} dv}.$$

Since firms with the productivity level $\frac{1}{a_D^h}$ earn zero profit in the domestic market, and the firms with productivity $\frac{1}{a_X^h}$ earn the zero profit in the exporting market (the profit of these firms in the domestic market is positive), we can determine the cutoff levels of productivity

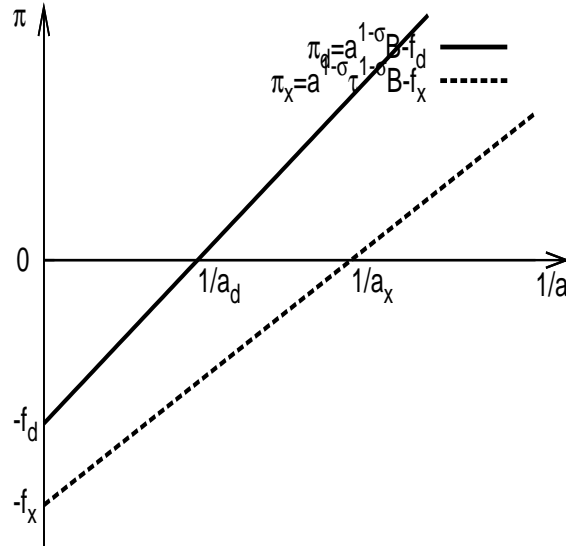


Figure 1: Profit from domestic sales and exports

through the equations of profit equal to zero:

$$\begin{aligned} (a_D^{ih})^{1-\sigma_h} B_h^i &= f_d^h \Rightarrow a_D^{ih} = \left(\frac{f_d^h}{B_h^i} \right)^{\frac{1}{1-\sigma_h}} \\ ((\tau^h a_x^{ih})^{1-\sigma_h}) B_h^j &= f_x^h \Rightarrow \tau^h a_x^{ih} = \left(\frac{f_x^h}{B_h^j} \right)^{\frac{1}{1-\sigma_h}} \end{aligned}$$

Since fixed costs are assumed to be the same in both countries, the distribution function $G(\cdot)$ is also the same in both countries. In addition, since the trade costs are also the same between two countries. The cutoff levels of productivity are also equal in both countries. This means that $a_D^{ih} = a_D^{jh} = a_D^h$ and $a_X^{ih} = a_X^{jh} = a_X^h$. These results imply $B_h^i = B_h^j = B_h$ (see Appendix C). These results hold for each of H industries in country i and country j . In the following sections we focus on industry h in country i and j and drop the h subscript.

2.3 Entry firms and market size

The price index of industry h in country i includes the product prices of domestic firms and the one of exporting firms from country j in industry h .

$$\begin{aligned}\int_0^{n_i^e} p(v)^{1-\sigma} dv &= n_i \int_0^{a_D} \left(\frac{a}{\alpha}\right)^{1-\sigma} dG(a) + n_j \int_0^{a_X} \left(\tau \frac{a}{\alpha}\right)^{1-\sigma} dG(a) \\ &= n_i \left(\frac{1}{\alpha}\right)^{1-\sigma} V(a_D) + n_j \tau^{1-\sigma} \left(\frac{1}{\alpha}\right)^{1-\sigma} V(a_X)\end{aligned}\quad (1)$$

Parameters n_i, n_j are considered the entry firms in country i and j in industry h . Substituting the above results into (1) yields:

$$n_i V(a_D) + n_j \tau^{1-\sigma} V(a_X) = \frac{(1-\alpha)\beta Y_i}{B} \quad (2)$$

Similarly for country j

$$n_j V(a_D) + n_i \tau^{1-\sigma} V(a_X) = \frac{(1-\alpha)\beta Y_j}{B} \quad (3)$$

Using equations (2) and (3) and solving for $\frac{n_i}{n_j}$:

$$\frac{n_i}{n_j} = \frac{\frac{Y_i}{Y_j} - \frac{\tau^{1-\sigma} V(a_X)}{V(a_D)}}{1 - \frac{\tau^{1-\sigma} V(a_X)}{V(a_D)} \frac{Y_i}{Y_j}} = \frac{\lambda - \rho}{1 - \rho \lambda} \quad (4)$$

Where $\lambda = \frac{Y_i}{Y_j}$ and $\rho = \frac{\tau^{1-\sigma} V(a_X)}{V(a_D)}$. If we assume that the productivity of firms in the industry ($x = 1/a$) has a Pareto distribution in $x \geq \theta$ with the cumulative distribution function of x : $F(x) = 1 - (\frac{\theta}{x})^k$. Here, k denotes the dispersion parameter of productivity. Industries with low value of k have high productivity dispersion and industries with high value of k have low productivity dispersion. From that, the cumulative distribution function of a will be: $G(a) = P(\frac{1}{x} \leq a) = P(x \geq \frac{1}{a}) = 1 - F(\frac{1}{a}) = 1 - (1 - (\theta a)^k) = (\theta a)^k$ for $a \leq \frac{1}{\theta}$. The population density function is

$$dG(a) = k\theta(\theta a)^{k-1} da$$

$V(a_D)$ and $V(a_X)$ are

$$\begin{aligned}V(a_D) &= \int_0^{a_D} a^{1-\sigma} dG(a) = c a_D^{k-(\sigma-1)} \\ V(a_X) &= \int_0^{a_D} a^{1-\sigma} dG(a) = c a_X^{k-(\sigma-1)}\end{aligned}$$

From here, we can find that

$$\frac{V(a_D)}{V(a_X)} = \left(\frac{f_X}{f_D} \tau^{\sigma-1} \right)^{\frac{k-\sigma-1}{\sigma-1}}$$

As a result,

$$\rho = \frac{1}{\tau^{\sigma-1}} \left(\frac{f_D}{f_X \tau^{\sigma-1}} \right)^{\frac{k-\sigma+1}{\sigma-1}} < 1$$

Unlike Helpman and Krugman (1987)'s model in which ρ depends only on trade costs and the elasticity of substitution, here ρ depends on two additional characteristics of the industry, namely, fixed costs and productivity dispersion.

From equation (4), we have:

$$\frac{\partial (n_i/n_j)}{\partial \lambda} = \frac{1 - \rho^2}{(1 - \lambda \rho)^2} > 0 \quad (5)$$

Equation (5) states that the difference in the number of firms (or products) of industry (h) between two countries has a positive relationship with the difference in size of two countries. If λ is larger than 1 ($\lambda > 1$), it can be shown that $\frac{1-\rho^2}{(1-\lambda\rho)^2} > 1$, indicating that the larger market attracts a disproportionate share of firms in industry h (the home market effects). The coefficient $\frac{1-\rho^2}{(1-\lambda\rho)^2}$ shows the level of difference in the number of products of an industry h between the large country and the small country. Let $g(\rho) = \frac{1-\rho^2}{(1-\lambda\rho)^2}$, we have additionally the following result:

$$\frac{\partial g}{\partial \rho} = \frac{2(\lambda - \rho)(1 - \rho\lambda)}{(1 - \rho\lambda)^4} > 0 \quad (6)$$

Equation (6) indicates that the coefficient $\frac{1-\rho^2}{(1-\lambda\rho)^2}$ is not uniform across industries: this coefficient will be larger if ρ is larger. In other words, higher the value of ρ , the larger would be the difference in the number of products between two countries or the home market effect of an industry. Since ρ depends on the characteristics of industries, the difference in the number of products (or the distribution of firms across industries) depends on industry characteristics. To find the effects of an industry characteristic on (ρ) , we assume that the other characteristics are constant.

The impact of trade costs: The derivative of ρ with respect to trade costs shows that:

$$\frac{\partial \rho}{\partial \tau} = \left(\frac{f_d}{f_x} \right)^{\frac{k-\sigma+1}{\sigma-1}} \left(\frac{-k}{\tau^{k+1}} \right) < 0 \quad (7)$$

When trade costs decrease across industries, the difference in the number of products between two countries become wider. It suggests that firms of industries with low trade costs will concentrate more in the large country. Since the production costs of firms in the large country are lower than those in the small country because of economics of scale, making the prices of products of the large country cheaper. When trade costs are low, low-priced products of the large country will easily penetrate into the small country market. Consequently, high-priced products of the small country can not compete with low-priced products of the large country and firms of the small country can exit markets when trade liberalization occurs.

The impact of fixed costs: Derivatives of ρ with respect to fixed domestic costs and fixed export costs yield:

$$\begin{aligned} \frac{\partial \rho}{\partial f_d} &= \left(\frac{k-\sigma+1}{\sigma-1} \right) \frac{1}{\tau^k} (f_x)^{-\frac{k-\sigma+1}{\sigma-1}} (f_d)^{\frac{k}{\sigma-1}-2} > 0 \\ \frac{\partial \rho}{\partial f_x} &= \left(-\frac{k-\sigma+1}{\sigma-1} \right) \frac{1}{\tau^k} (f_d)^{\frac{k-\sigma+1}{\sigma-1}} (f_x)^{-\frac{k}{\sigma-1}} < 0 \end{aligned} \quad (8)$$

An increase in the fixed domestic costs leads to a higher value of ρ , while the increase of fixed export costs makes ρ decrease. This implies that high fixed domestic costs and low fixed export costs induce firms to locate more in the large country in order to take advantage of economics of scale.

The impact of the productivity dispersion and the elasticity of substitution: The derivatives of ρ with respect to the productivity dispersion and the elasticity of substitution yield:

$$\begin{aligned} \frac{\partial \rho}{\partial k} &= \left(\frac{1}{\sigma-1} \right) \left(\frac{1}{\tau^{\sigma-1}} \right) \left(\frac{f_d}{f_x \tau^{\sigma-1}} \right)^{\frac{k-\sigma+1}{\sigma-1}} \ln \left(\frac{f_d}{f_x \tau^{\sigma-1}} \right) \\ \frac{\partial \rho}{\partial \sigma} &= \left(\frac{-k}{(\sigma-1)^2} \right) \left(\frac{1}{\tau^k} \right) \left(\frac{f_d}{f_x} \right)^{\frac{k-\sigma+1}{\sigma-1}} \ln \left(\frac{f_d}{f_x} \right) \end{aligned} \quad (9)$$

Since we assume that only some firms with high productivity can export to foreign markets, this implies that $f_d < f_x \tau^{\sigma-1}$ and hence $\frac{\partial \rho}{\partial k} < 0$. The negative correlation between ρ and the productivity dispersion indicates that industries with high productivity dispersion (low k) will locate more in the large country. Although firms with low productivity can not operate in the small country due to high competitive pressures, firms with low productivity can still operate in the large country because of the diversity of consumer demand in the large country. So, industries with high productivity dispersion prefer concentrating in the large country to concentrating in the small country.

If the fixed domestic costs are smaller than the fixed export costs ($f_d < f_x$), $\frac{\partial \rho}{\partial \sigma} > 0$ implies that industries with high elasticity of substitution (high σ) will locate more in the large country. If fixed domestic costs are larger than fixed export costs ($f_d > f_x$), $\frac{\partial \rho}{\partial \sigma} < 0$ implies industries with low elasticity of substitution (low σ) will concentrate more in the large country. In this study, we assume that ($f_d < f_x$): industries with high elasticity of substitution should locate more in the large country. Industries with high substitution elasticity have less differentiated goods or few substitutes, and when trade liberalization occurs, consumers choose and buy cheaper goods from large countries. Firms of the small country which cannot compete with firms of the large country may exit market. This explains why industries with high elasticity of substitution tend to concentrate in the large country.

2.4 The model of homogenous firms for many differentiated product industries

When we assume that all domestic firms are homogeneous, all these firms can participate in export markets. In this case, $f_d = f_x \tau^{\sigma-1}$ and our model of heterogeneous firms becomes the model of homogeneous firms (like Helpman and Krugman (1987)) but for many industries. In this case, ρ of industries depends only on trade costs and elasticity of substitution of industries: $\rho = \tau^{1-\sigma}$. As τ and σ increase across industries, ρ decreases across industries, and so does $\frac{n_i}{n_j}$ (the difference in the number of firms (or products) between the large country and the small country). It should be noted that the impact of trade costs across industries on the home market effect in this model is the opposite of the one predicted by

Hanson and Xiang (2004).

This difference in the impact of trade costs on the home market effects can be explained by differences in assumptions of the models. Both models are the models with homogeneous firms. The model of homogeneous firms in our paper has the appearance of the homogeneous product sector, this leads to wages equal between the large country and the small country. In contrast, there is not the homogeneous product sector in Hanson and Xiang (2004), and the wage in the large country is higher than the one in the small country. Moreover, in Hanson and Xiang's model, the effects of trade costs on home market effect across industries are not uniform. Their proposition doesn't hold true for industries with very high trade costs.

3 Empirical model

3.1 Empirical method

Equation (4) in the theoretical part suggests a positive relationship between $\frac{n_i}{n_j}$ and λ ($= \frac{Y_i}{Y_j}$). Expressing this relation in a log linear form is as follows::

$$\log \left(\frac{n_{ih}}{n_{jh}} \right) = \beta_h \log \left(\frac{Y_i}{Y_j} \right) + u_{ij} \quad (10)$$

From the theoretical part, we know that industries with larger home-market effects (larger ρ_h) will have larger β_h . It means that $\rho_1 > \rho_2 > \dots > \rho_h > \dots$, then $\beta_1 > \beta_2 > \dots > \beta_h > \dots$, where, $\beta_1, \beta_2, \beta_h$ denote coefficients of the above regression equation for industries 1, 2,... h .

We have already shown that industries with lower trade costs, higher domestic-fixed costs, lower-export fixed costs, and high productivity dispersion will concentrate more in the large countries. This implies that we will have $\alpha_2 < 0$, $\alpha_3 > 0$, $\alpha_4 > 0$, and $\alpha_5 < 0$ in the following relationship:

$$\beta_h = \alpha_1 + \alpha_2 \tau_h + \alpha_3 \text{disp}_h + \alpha_4 f_{dh} + \alpha_5 f_{xh} \quad (11)$$

Where τ_h denotes trade costs, disp_h the productivity dispersion, f_{dh} and f_{xh} the fixed costs in domestic and export markets. Since the productivity dispersion effect includes

the elasticity of substitution effect, we do not study the separate effect of the substitution elasticity on the distribution of industries. We will explain this issue in more detail later.

Substituting equation (11) into the regression equation (10) yields:

$$\log\left(\frac{n_{ih}}{n_{jh}}\right) = \alpha_0 + \alpha_1 \log\left(\frac{Y_i}{Y_j}\right) + \alpha_2(\tau_h) \log\left(\frac{Y_i}{Y_j}\right) + \alpha_3(\text{disp}_h) \log\left(\frac{Y_i}{Y_j}\right) + \alpha_4(f_{dh}) \log\left(\frac{Y_i}{Y_j}\right) + \alpha_5(f_{xh}) \log\left(\frac{Y_i}{Y_j}\right) + u_{ijh}$$

We predict that $\alpha_1 > 0$, $\alpha_2 < 0$, $\alpha_3 > 0$, $\alpha_4 > 0$, and $\alpha_5 < 0$.

Data for the number of firms or establishments is not available for many countries. According to this model, the ratio of the number of exporting firms (or products) for industry (h) in two countries is equivalent to the ratio of firms (or products) in two countries. So, we use the ratio of the number of exporting firms (products) $\left(\frac{EM_{ih}}{EM_{jh}}\right)$ to represent the ratio of firms (products) $\left(\frac{n_{ih}}{n_{jh}}\right)$ of two countries ($\frac{n_{ih}}{n_{jh}} \equiv \frac{EM_{ih}}{EM_{jh}}$).

EM_{ih} , the extensive margin of export of country i in industry h , is measured by the method used by Hummels and Klenow (2002) (or Hummels and Klenow (2005)). As a result, we have the following regression model to empirically assess the impact of industry characteristics on the distribution of industries between the large country and the small country:

$$\log\left(\frac{EM_{ih}}{EM_{jh}}\right) = \alpha_0 + \alpha_1 \log\left(\frac{Y_i}{Y_j}\right) + \alpha_2(\tau_h) \log\left(\frac{Y_i}{Y_j}\right) + \alpha_3(\text{disp}_h) \log\left(\frac{Y_i}{Y_j}\right) + \alpha_4(f_{dh}) \log\left(\frac{Y_i}{Y_j}\right) + \alpha_5(f_{xh}) \log\left(\frac{Y_i}{Y_j}\right) + u_{ijh} \quad (12)$$

It is predicted that $\alpha_1 > 0$, $\alpha_2 < 0$, $\alpha_3 > 0$, $\alpha_4 > 0$, and $\alpha_5 < 0$.

3.2 The extensive margins of export

In studying the role of new varieties in price indexes, Feenstra (1994) showed how to use the data of expenditure to measure the product-variety changes of each country across time. Many studies have adopted this method to compare product varieties or export varieties across countries.³ Hummels and Klenow (2002) (or Hummels and Klenow (2005)) used this method to define the extensive and intensive margins of countries' exports and im-

³.i.e Feenstra et al. (1997); Feenstra and Kee (2004); Hummels and Klenow (2002); Hummels and Klenow (2005); and Feenstra and Kee (2008)

ports.⁴ In this study, we use their methods to measure the relative number of export products of two countries.

Using the method of Feenstra (1994), Hummels and Klenow (2002) define the extensive margins of exports of country i as follows:

$$EM_t^{i,exp} = \frac{\sum_j \sum_{s \in I_t^{ij}} X_t^{Wjs}}{X_t^W} \quad (13)$$

$EM_t^{i,exp}$ is the extensive margin of an exporter i in year t . I_t^{ij} is the set of products s exported from country i to country j . X_t^{Wjs} is the value of export of product s from the world to country j . $\sum_{s \in I_t^{ij}} X_t^{Wjs}$ is the total value of export of the world to country j in products that country i exports to country j ($s \in I_t^{ij}$). X_t^W is the total export of all countries. The extensive margin of exports employs a weighted count of the number of categories to measure the extensive margins of countries in year t with the weights to be the world trade in each category.

Hummels and Klenow (2005) use a similar approach but they calculate the extensive margin of exporter at each destination. They then determined an average value of all destinations to calculate the extensive margin of exports for each country. In this case, the extensive margin of export of country i at destination d is:

$$EM_t^{id,exp} = \frac{\sum_{s \in I_t^{id}} X_t^{Wds}}{\sum_{s \in I_t^{Wd}} X_t^{Wds}}$$

To measure the extensive margins of an export country to all countries, Hummels and Klenow (2005) use the geometric mean of the extensive margin over all destinations to represent the extensive margin of each export country. In particular, the extensive margin of country i is calculated at each destination ($d \in M_{-i}$), where M_{-i} is the set of countries for which import data from country i is available. We then take the geometric average of country i 's extensive margin across the M_{-i} markets to calculate the extensive margin of

⁴Hummels and Klenow (2002) is a working paper, while Hummels and Klenow (2005) is a version of Hummels and Klenow (2002) published in the AER. Hummels and Klenow (2002) measures the extensive and intensive margins of export of a country at all destinations, while Hummels and Klenow (2005) measure them at each destination, then get the average value to represent the extensive margin of exports of countries

export for country i :

$$EM_t^{i,exp} = \prod_{d \in M_{-i}} \left(EM_t^{id,exp} \right)^{w_{id}} \quad (14)$$

w_{id} is weights which are measured as follows:

$$w_{id} = \frac{\frac{s_{id} - s_{Wd}}{\log(s_{id}) - \log(s_{Wd})}}{\sum_{d \in M_{-i}} \frac{s_{id} - s_{Wd}}{\log(s_{id}) - \log(s_{Wd})}}$$

Here w_{id} is the logarithmic mean of s_{id} and s_{Wd} and $\sum_{d \in M_{-i}} w_{id} = 1$. s_{id} is the share of export of country i to country d relative to the total export of country i $\left(s_{id} = \frac{X_i^d}{\sum_{d \in M_{-i}} X_i^d} \right)$, and s_{Wd} is the share of export of the other countries (except to country i) to country d relative to the total export of these countries $s_{Wd} = \frac{\sum_{l \in M_{-i-d}} X^{ld}}{\sum_{l \in M_{-i-d}} X^{lW}}$.

4 Data and empirical results

4.1 Data for variables of regression models

Since this paper is on how characteristics of industries affect the distribution of firms across industries between a large country and a small country, the characteristics of an industry are assumed to be homogeneous across countries. We choose a sample of 28 industrial countries (Table (8) in Appendix) with the assumption that industry characteristics of these countries are similar. In addition, 4-digit ISIC classification with 125 manufacturing industries is used to classify the manufacturing industries in these countries. If data on an industrial characteristic is available for all countries, we use the average value across countries to represent the industrial characteristic (i.e., import tariff barriers). However, we cannot approach most of data on industrial characteristics of countries except for the U.S. So, we use data on U.S. industrial characteristics to represent the industrial characteristics in our study. The U.S. is a large market, so firms (or products) in industries are diverse. In addition, technology and technique for industries in the U.S. are also typical for these in other industrial countries. Therefore, we think that the industrial characteristics of the U.S. can suitably represent those of other industrial countries.

Dependent variable: Trade flow data at HS6 level from CEPII⁵ is used to measure the extensive margin of export for a country as presented (13) (or (14)).

GDP: From the results of the theoretical part, the GDP of countries is used to represent a country's size. GDP Data (at constant prices of 2000) is from the World Development Indicator.

Variable trade costs (τ_h): The simple average tariffs (t) of high income countries are used to represent trade costs of industries and is the ratio between the sum of all the tariff rates and the number of import categories. This data is from TRAINS database. We assume that goods in an industry have equal importance, so we use simple average tariffs to represent trade costs of industries instead of using the weighted average tariffs. We know that the weighted average tariffs tend to be down-biased since the amount of low-tariff goods is higher than high-tariff goods. Therefore, the trade-weighted average tariff cannot be a good proxy for the trade costs of all goods in an industry.

Fixed domestic and export costs (f_{dh} and f_{xh}): Fixed domestic costs (f_{dh}) are considered the overhead costs that refer to ongoing expenses of a firm's operation such as management salaries, advertising, insurance, rent, and utilities. We use expense data from the Annual Survey of Manufacturers (1997) to calculate these costs. The expense categories are presented in the Appendix. Fixed domestic costs for a firm in industry h (f_d) are calculated by dividing the total of these costs by the total number of firms in the industry (h).

We are not able to measure fixed-export costs directly (f_{xh}). However, the studies of multinational firms show that industries with high firm-level economies of scale encourage FDI, not the concentration of production within a single country. This implies that industries with high firm-level economies of scale tend to produce in many countries so the number of firms (products) of a large country relative to that of a small country are lower. Based on this idea, we consider firm-level economies of scale to represent fixed-export costs. As a result, firm-level economies of scale are expected to have a negative relationship with the dependent variable. Following the approach in previous studies (i.e. Brainard (1997)), we use the average ratio of the number of nonproduction workers relative to the total employ-

⁵www.cepii.org

ment in each industry to represent firm-level economies of scale of that industry. This data is from the Annual Survey of Manufacturers (1997).

Productivity dispersion ($disp_h$): Productivity is assumed to have a Pareto distribution with shape parameter k . However, we cannot measure this parameter directly. According to Helpman et al. (2004), a Pareto distribution of productivity implies that a firms' sales also have the same distribution with shape parameter $k - \sigma + 1$. This parameter can be measured by the standard deviation of the logarithm of firm sales and is used to represent the productivity dispersion. If the standard deviation of the logarithm of firm sales ($disp$) in an industry is large, the productivity dispersion of that industry is high ($k - \sigma + 1$ low).

As mentioned in the theoretical part, we assume that $f_d < f_x$. This implies that industries with high productivity dispersion (k low) and high elasticity of substitution (σ high) (low $k - \sigma + 1$) will locate more often in a large country. Since $k - \sigma + 1$ is measured by the standard deviation of the logarithm of firm sales ($disp$), $disp$ can represent both the productivity dispersion and the elasticity of substitution.

We use the output of 10-digit NAICS U.S. products (about 7500 products) to calculate the industry-productivity dispersions. In this case, we consider each firm that produces a product; thus, the product output is also the firm's sale. The method of using product sales to calculate the productivity dispersion is similar to the method used by Nunn and Treffer (2008). They don't approach firm-level data and use the export sale of U.S. products to calculate the productivity dispersion of industries.

4.2 Data analysis

As mentioned above, industries which are disproportionately located in large countries (or have higher home market effects) will have higher β_1 in the following difference regression:

$$\log \left(\frac{EM_{ih}}{EM_{jh}} \right) = \beta_0 + \beta_1 \log \left(\frac{Y_i}{Y_j} \right) + u_{ij} \quad (15)$$

We call (β_1) the coefficient that shows the strength of the home-market effect of industry h or level of the distribution of firms of industry h . The 4-digit ISIC classification has 125

manufacturing industries. However, due to the limited availability of export data, we only estimate the coefficient (β_1) for 118 industries.⁶

The predictions of the theoretical model imply that this coefficient (β_1) should have a negative relationship with trade costs and fixed-export costs and a positive relationship with fixed domestic costs and productivity dispersion. First, we use graphs to visually summarize the relationships between the industry characteristics and this coefficient. When we combine industrial characteristics and the coefficient (β_1), only 110 industries have available data on all industrial characteristics. Figure (2) shows the relationship between the industry characteristics on the vertical axis and the home-market effect coefficients of industries on the horizontal axis. From the graphs, we can see that there are some outliers in the relationship between industry characteristics and the home market-effect coefficients; for example, one outlier in the relationship between firm scale and the coefficients (β_1), and two outliers in the relationship between the fixed domestic costs and coefficients⁷. Therefore, we drop these observations. Figure (3) shows the relationships after dropping these outliers. The results of the figures are consistent with the predictions from the theoretical model: industries with low trade costs (or low tariff barriers), high productivity dispersion, high domestic fixed costs, and high firm-level economics of scale (which represents export fixed costs) tend to concentrate in large countries. The results of the following simple relationship (Table 1) seem to affirm the results from the figure's analysis:

$$\beta_h = \alpha_1 + \alpha_2 \tau_h + \alpha_3 disp_h + \alpha_4 f_{dh} + \alpha_5 f_{xh}$$

In brief, the impact of fixed domestic costs, fixed export costs, productivity dispersion, and trade costs on the home-market effect of industries have the predicted signs.

⁶The data of the following industries are not available- 1911: Tanning and dressing of leather; 2892: Treatment & coating of metals; 3720: Recycling of non-metal waste and scrap; 1712: Finishing of textiles; 3710: Recycling of metal waste and scrap; 2731: Casting of iron and steel; 2230: Reproduction of recorded media; 2891: Metal forging/pressing/stamping/roll-forming; 2732: Casting of non-ferrous metals

⁷2109: Other articles of paper and paperboard, 2221:Printing, 2927:Weapons and ammunition,

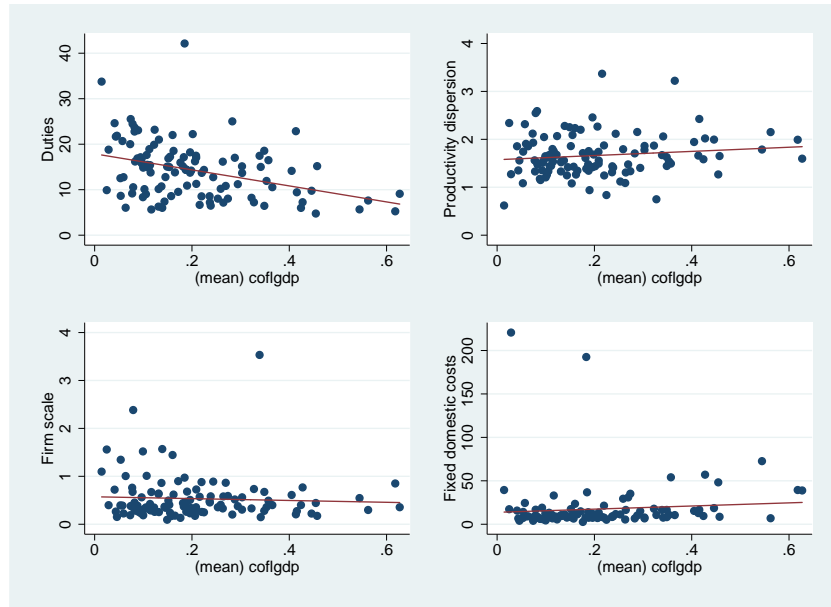


Figure 2: The relationship of industrial characteristics

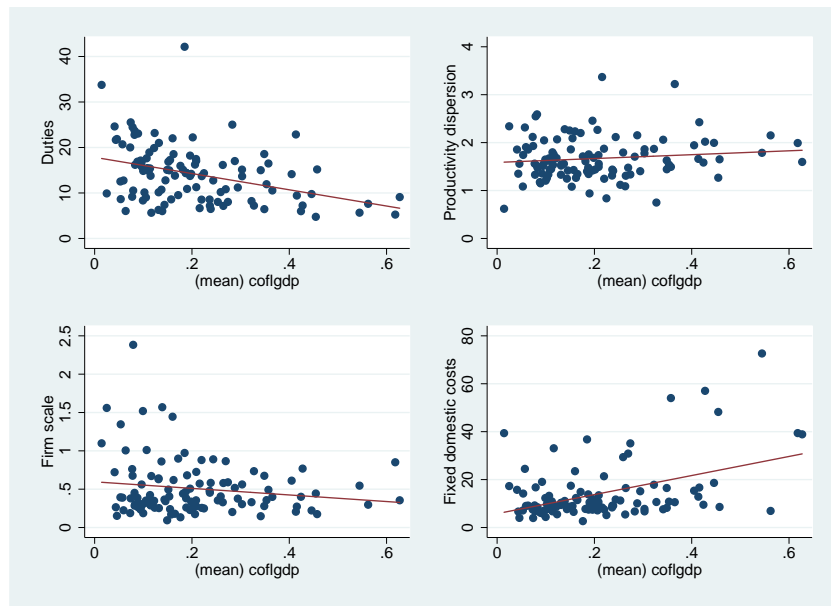


Figure 3: The relationship of industrial characteristics

Table 1: The relationship of industrial characteristics

VARIABLES	(1)	(2)	(3)	(4)	(5)
Firm level economies of scale (-)	-0.054** (0.021)				-0.117*** (0.000)
Fixed domestic costs (+)		0.005*** (0.000)			0.005*** (0.000)
Productivity dispersion (+)			0.033 (0.218)		0.036* (0.080)
Trade costs (-)				-0.008*** (0.000)	-0.008*** (0.000)
Constant	0.217*** (0.000)	0.123*** (0.000)	0.127*** (0.006)	0.298*** (0.000)	0.218*** (0.000)
Observations	112	112	107	112	107
R-squared	0.047	0.172	0.014	0.156	0.452

pval in parentheses

*** p<0.01, ** p<0.05, * p<0.1

This table is results of the relationships of industrial characteristics

4.3 Results of the main regression model

We use the export data to measure the firm (or variety) ratio across industries of a country pair. One country pair can have different characteristics from another pair. Two countries are in economic unity or have a free trade agreement or are in similar geographical locales and thus industries in these countries may have some common group effects. As a result, u_{ijh} (in model 12) can be decomposed into two parts: $u_{ijh} = v_{ij} + \varepsilon_{ijh}$, where v_{ij} is the country pair-level fixed effects or an unobserved (group) cluster effect ($v_{ij} \sim [0, \sigma_v^2]$) and ε_{ijh} is the idiosyncratic error ($\varepsilon_{ijh} \sim [0, \sigma_\varepsilon^2]$). In addition, we can consider each industry as a cluster since countries can produce these industries due to some similar reasons- for example, technology-intensive industries or high economic-value industries. So, u_{ijh} can be decomposed into three parts: $u_{ijh} = v_{ij} + e_h + \varepsilon_{ijh}$, here ($e_h \sim [0, \sigma_e^2]$) represents an unobserved cluster effect at industry level. As a result, the model (12) can be rewritten as follow:

$$\begin{aligned} \log\left(\frac{EM_{ih}}{EM_{jh}}\right) = & \alpha_0 + \alpha_1 \log\left(\frac{Y_i}{Y_j}\right) + \alpha_2(\tau_h) \log\left(\frac{Y_i}{Y_j}\right) + \alpha_3(disph) \log\left(\frac{Y_i}{Y_j}\right) + \alpha_4(f_{dh}) \log\left(\frac{Y_i}{Y_j}\right) + \\ & + \alpha_5(f_{sh}) \log\left(\frac{Y_i}{Y_j}\right) + v_{ij} + e_h + \varepsilon_{ijh} \end{aligned} \quad (16)$$

Since the regression model has common group effects in the error terms or the intracluster correlation, the usual OLS standard errors can be seriously biased (Moulton (1990)). In particular, the standard errors of the usual OLS method may be remarkably low. The bias in conventional standard errors become increasingly large in absolute value as the number of clusters decrease and the intracluster correlation increases. If other hypotheses of classical regression are still satisfied, the usual OLS estimator of coefficients remains unbiased and normally distributed. However, the usual OLS estimator is not efficient and the standard errors are incorrectly estimated. Consequently, tests based on the usual standard errors are no longer valid, which is why we need to control the presence of clusters in the regression model.

According to Cameron and Trivedi (2005), we can use the following estimation techniques to control for clustering: the OLS estimator with cluster-corrected standard errors, the GLS estimator (the random-effects model), and the within estimator (the fixed-effects model). When the unobserved cluster effects (v_{ij} and e_h) are uncorrelated with the model's explanatory variables, the OLS estimator with cluster-corrected standard errors and the random-effect estimator are consistent. In this case, the cluster-robust standard errors of the OLS estimator converges to the true standard error as the number of clusters is large and the cluster size is fixed. In addition, when the unobserved cluster effects are independent of the explanatory variables, the GLS estimator (the random-effects model) also gives an efficient estimator, which is even more efficient than the cluster-corrected OLS estimator. If v_{ij} (or e_h) are correlated with other dependent variables, the OLS estimator and the random-effect estimator are inconsistent. In this case, the fixed estimators should be used instead. However, the fixed-effects estimator drops all cluster-invariant regressors.

Table (2) presents the regression results when country pairs are randomly chosen from all countries in the sample (i.e. not based on any pre-determined criteria). In general,

coefficients of different regression methods have predicted signs and statistical significance across different regression methods. Since industry characteristics are assumed to be similar across countries in our study, the industry characteristics are also similar across country pairs. Consequently, the estimation coefficients of OLS, random-effects, or fixed-effects estimations are quite similar. As the variable $(\log(\frac{Y_i}{Y_j}))$ is invariant across industries, its effect is removed from the model in the fixed-effects estimator (Column (8)).

We find that there is a relatively minor change in standard errors from the usual OLS method to the heteroscedastic-corrected OLS (Column (1) and (2)). However, when the cluster-robust variance estimator for country pairs and industries are used, there is a substantial change in the standard errors of coefficients. All t-ratios become significantly smaller (Columns (3) and (4) compared with columns (1)). For example, the t-ratio for *lgdp* in the usual OLS is 25.05, whereas this value for OLS with standard errors corrected for country pairs is 17.44 and 4.66 for industries. When the standard errors are corrected for both country pairs and industries, the t-ratio decreases to 4.59. These results suggest that ignoring intracluster correlation causes inflation in the OLS t-ratios and the cluster effects at the industry level are stronger than those at the country-pair level.

The Breusch-Pagan tests for the country-pair and industry random-effects models reject the null hypothesis that the random variation in the intercept is zero. This indicates that the country-pair and industry-RE models are an improvement over the OLS regression (Columns (6) and (7) compared with Column (1)). However, the standard errors of the country pair-RE model don't result in a significant change when compared with those of the OLS estimator since the t-values of some variables in the country-pair random-effects estimator are larger than these in the usual OLS estimator (Column (6) compared with Column (1)). While, all t-values of variables in the industry random effects estimator are smaller than these in OLS estimator (column (7) and column (1)). This implies that the standard errors of the industry RE model are larger than those of the usual OLS estimators across all variables.

To compare the random-effects and fixed-effects estimators, the Hausman test shows that the industry fixed-effects regression does not change significantly from the industry random-effects regression (column (9) and (7)). We do not compare these estimators at the country-

pair level because the variable $(\log(\frac{Y_i}{Y_j}))$ is removed from the fixed-effects estimator.

In the above case, the country pairs are all built without any particular criteria from sample countries. However, if we choose any two countries to build a pair, it can sometimes be difficult to find common characteristics between the two countries. For example, we can observe the common features between the US and Canada, but not between Canada and Australia. This implies that the comparison between the U.S. and Canada pair and the Canada and Australia pair might not be reasonable. To eliminate these potential problems, we form pairs from a set of countries that belong to a preferential trade arrangement of some kind. In particular, we divide countries into four groups: members of European Union (19 countries), Canada and the US (the US-Canada Free Trade Agreement), New Zealand and Australia (British Commonwealth), and Japan and Korea (a group of Asian countries) (Sample 2 of Table (8) in Appendix). Country pairs are then built within each group. The regression results are shown in Table (3), which are quite similar to the ones of the previous case and the cluster effects across country pairs and industries are still significant. The random-effects model is still preferred to the OLS model although the results of this method donot change much in comparison. We can see this by comparing the t-values of variables of Columns (7) and (8) with Column (1). In this case, the Hausman test shows the fixed-effects regression at the industry level is an improvement over the random-effects regression at the industry level (p-value=0.015).

The above empirical results show that the impact of variables not only have predicted signs but also have high statistical significance. As a result, these results confirm the predictions of the theoretical model that industries with low trade costs (low tariff barriers), high domestic fixed costs, high firm-level economics of scale (which represent export fixed costs), and high productivity dispersion tend to concentrate in large countries. As discussed above, the productivity dispersion measured in this paper includes not only the productivity dispersion (k) but also the elasticity of substitution (σ). When the productivity dispersion increases ($k - \sigma + 1$ decreases), it can be equivalent with k decreased and σ increased (high elasticity of substitution). As a result, the positive relationship between the productivity dispersion as measured above and the distribution of firms across industries might imply

that industries with high-productivity dispersion and high elasticity of substitution are more likely to locate in large countries as predicted by the theoretical model.

Table 2: The impact of industry characteristics on the distribution of firms across industries

VARIABLES	OLS					Random effects		Fixed effects	
	Usual	Het	Country	Industry	Both	Country	Industries	Country	industry
	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(10)	(11)
LGDP (+)	0.21*** (22.94)	0.21*** (24.98)	0.21*** (11.89)	0.21*** (4.55)	0.21*** (4.34)	0.21*** (12.32)	0.22*** (15.62)		0.22*** (15.15)
Duties*LGDP (-)	-0.75*** (-26.26)	-0.75*** (-25.03)	-0.75*** (-16.77)	-0.75*** (-4.28)	-0.75*** (-4.23)	-0.75*** (-30.00)	-0.76*** (-16.68)	-0.75*** (-30.00)	-0.76*** (-16.10)
Fixed domestic cost*LGDP (+)	0.01*** (36.37)	0.01*** (25.51)	0.01*** (17.21)	0.01*** (6.43)	0.01*** (6.22)	0.01*** (41.54)	0.01*** (20.99)	0.01*** (41.54)	0.01*** (20.01)
Firm scale*LGDP (-)	-0.14*** (-26.75)	-0.14*** (-31.98)	-0.14*** (-17.03)	-0.14*** (-5.92)	-0.14*** (-5.70)	-0.14*** (-30.56)	-0.12*** (-14.59)	-0.14*** (-30.56)	-0.12*** (-13.80)
Productivity dispersion*LGDP (+)	0.04*** (10.59)	0.04*** (11.16)	0.04*** (13.83)	0.04*** (2.06)	0.04*** (2.08)	0.04*** (12.10)	0.04*** (5.73)	0.04*** (12.10)	0.04*** (5.42)
Constant	0.06*** (11.26)	0.06*** (10.99)	0.06*** (2.15)	0.06*** (12.08)	0.06*** (2.16)	0.06*** (2.21)	0.06*** (3.81)	0.34*** (32.24)	0.06*** (11.62)
Observations	37,557	37,557	37,557	37,557	37,557	37,557	37,557	37,557	37,557
R-squared	0.14	0.14	0.14	0.14	0.14			0.08	0.11
Number of Industries							107		107
Hausman test (p-value)									0.901
Number of Country-pairs						351		351	
Breusch-Pagan test(p-value)						0	0		

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Country pairs are chosen from all countries in the sample without based on any pre determined criteria.

The dependent variable is measured by the ratio of the extensive margin of exports (equation (13)).

Columns (1)-(5) are the OLS estimators.

Columns (6)-(7) are the random-effects estimators.

Columns (8)-(9) are the fixed-effects estimators.

Table 3: The impact of industry characteristics on the distribution of firms across industries

VARIABLES	OLS					Random effects		Fixed effects	
	Usual	Het	Country	Industry	Both	Country	Industries	Country	industry
	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(10)	(11)
LGDP (+)	0.24*** (19.44)	0.24*** (23.06)	0.24*** (11.93)	0.24*** (4.54)	0.24*** (4.34)	0.24*** (14.34)	0.25*** (19.72)		0.25*** (19.79)
Duties*LGDP (-)	-0.67*** (-16.61)	-0.67*** (-16.65)	-0.67*** (-8.52)	-0.67*** (-3.26)	-0.67*** (-3.11)	-0.67*** (-18.98)	-0.69*** (-16.80)	-0.67*** (-18.98)	-0.69*** (-16.84)
Fixed domestic cost*LGDP (+)	0.01*** (28.10)	0.01*** (22.88)	0.01*** (16.67)	0.01*** (6.70)	0.01*** (6.49)	0.01*** (32.11)	0.01*** (26.93)	0.01*** (32.11)	0.01*** (26.33)
Firm scale*LGDP (-)	-0.13*** (-17.47)	-0.13*** (-21.79)	-0.13*** (-24.50)	-0.13*** (-4.62)	-0.13*** (-4.66)	-0.13*** (-19.96)	-0.13*** (-17.16)	-0.13*** (-19.96)	-0.13*** (-16.97)
Productivity dispersion*LGDP (+)	0.04*** (7.44)	0.04*** (8.51)	0.04*** (12.28)	0.04* (1.82)	0.04* (1.85)	0.04*** (8.50)	0.04*** (7.25)	0.04*** (8.50)	0.04*** (7.15)
Constant	-0.02*** (-4.63)	-0.02*** (-4.58)	-0.02 (-0.93)	-0.02** (-2.39)	-0.02 (-0.89)	-0.02 (-0.91)	-0.02*** (-2.60)	-0.10*** (-20.06)	-0.02*** (-4.67)
Observations	18,511	18,511	18,511	18,511	18,511	18,511	18,511	18,511	18,511
R-squared	0.35	0.35	0.35	0.35	0.35			0.08	0.35
Number of industries							107		107
Hausman test (p-value)									0.0155
Number of countrypairs						173		173	
Breusch-Pagan test(p-value)						0	0		

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Country pairs are chosen from countries in the same region.

The dependent variable is measured by the ratio of the extensive margin of exports (equation (13)).

Columns (1)-(5) are the OLS estimators.

Columns (6)-(7) are the random effects estimators.

Columns (8)-(9) are the fixed-effects estimators.

4.4 Robustness check

The UNIDO industrial database provides data on the number of establishments. However, this data is only available for a limited number of countries and industries, so we are not able to use it as a proxy for the dependent variable. Therefore, we use the ratio of the extensive margin of exports between two countries to represent the dependent variable as mentioned above. In this part, we use the data on the number of establishments from UNIDO to test the robustness of some model results. From this database, we choose 14 OECD countries (Table (8) in Appendix) which have over 80 industries. However, the number of common industries across the countries in the sample is only 51. From those countries, we demonstrate two cases. In the first case, we use all available industries to estimate the model. As there are many industries that do not exist in every country, the estimated results can be biased. Therefore, in the second case, we estimate the model by using only the industries that exist in all countries (51 industries). The regression results across the different methods are presented in Table (4) for the first case and in Table (5) for the second case.

The signs of the explanatory variables for both cases are still consistent with our predictions across different estimation methods. The statistical significance of the explanatory variables in the second case (with 51 industries) are more significant. For example, the effect of firm-level economics of scale in the second case is statistically significant in most of the cases, while in the first case, this effect is not significant in any of the cases. In addition, by looking at the t-values of the explanatory variables, we can see that the cluster effects in the country-pair levels are not as important as in the above cases, while the cluster effects at industry levels remain strong. This suggests that intracluster correlations exist at the industry level.

Moreover, instead of calculating the extensive margin of exports as Hummels and Klenow (2002), we also use Hummels and Klenow (2005) (equation (14)) to calculate the extensive margin of export. The results in this case are not much different from the ones that were estimated in equation (13).

Most of the industry characteristics in our study model are not directly observable, therefore proxy variables employed. Many studies have shown that an industry characteristic can

be represented by different proxies. We use several alternative proxies to check the robustness of the above results. In addition, as we cannot obtain the industry-level or firm-level data for other developed countries, we still use these proxies from US data to represent industry characteristics.

In our main regression result, the number of non-production workers is used as a proxy for firm-level economics of scale (which represents export-fixed costs). However, firm-level economics of scale can also be measured by other proxies such as advertising intensity and research and development (R&D) intensity (?). When we use the ratio of industry-aggregate advertising and R&D expenses to industry-aggregate sales to represent firm-level economics of scale, the results remain statistically consistent with our prediction. Industries with low firm-level economics of scale tend to locate in large countries.

For the productivity dispersion, instead of using the industry aggregate-product output to proxy for the productivity dispersion as above, we use the industry average of firm product output to estimate the productivity dispersion of industries. The results of this are still consistent with the above findings.

Table (6) presents the regression results across different methods when we measure the dependent variable by equation (14) (Hummels and Klenow (2005)), the firm-level economics of scale as measured by the ratio of aggregate-industry advertising and R&D expenses to industry sales, and the productivity dispersion is calculated from the output per company. Other explanatory variables (trade costs and domestic-fixed costs) are still unchanged. The results of the regression methods still appear consistent with our model predictions.

Table 4: The impact of industry characteristics with data of the dependent variable from UNIDO

VARIABLES	OLS					Random effects		Fixed effects	
	Usual	Het	Country	Industry	Both	Country	Industries	Country	industry
	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(10)	(11)
LGDP (+)	0.64*** (13.44)	0.64*** (14.21)	0.64*** (6.58)	0.64*** (6.05)	0.64*** (4.72)	0.62*** (6.52)	0.67*** (10.18)		0.69*** (8.63)
Duties*LGDP (-)	-0.50*** (-3.47)	-0.50*** (-3.31)	-0.50** (-2.12)	-0.50 (-1.51)	-0.50 (-1.34)	-0.43*** (-3.71)	-0.37* (-1.85)	-0.43*** (-3.70)	-0.29 (-1.14)
Fixed domestic cost*LGDP (+)	0.00*** (5.31)	0.00*** (6.55)	0.00*** (6.97)	0.00*** (3.13)	0.00*** (3.19)	0.00*** (6.36)	0.00*** (3.19)	0.00*** (6.35)	0.00** (2.13)
Firm scale*LGDP (-)	-0.00 (-0.01)	-0.00 (-0.01)	-0.00 (-0.01)	-0.00 (-0.01)	-0.00 (-0.01)	-0.02 (-0.28)	-0.04 (-0.35)	-0.02 (-0.29)	-0.06 (-0.46)
Productivity dispersion*LGDP (+)	0.04** (2.08)	0.04** (2.20)	0.04** (2.27)	0.04 (0.92)	0.04 (0.93)	0.05*** (2.93)	0.02 (0.85)	0.05*** (2.94)	0.01 (0.32)
Constant	-0.13*** (-4.69)	-0.13*** (-4.63)	-0.13 (-0.80)	-0.13*** (-6.79)	-0.13 (-0.81)	-0.12 (-0.75)	-0.12*** (-3.17)	0.81*** (14.59)	-0.13*** (-4.85)
Observations	8,326	8,326	8,326	8,326	8,326	8,326	8,326	8,326	8,326
R-squared	0.20	0.20	0.20	0.20	0.20			0.01	0.20
Number of industries							108		108
Hausman test (p-value)									0.929
Number of countrypairs						91		91	
Breusch-Pagan test(p-value)						0	0		

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The results are estimated for all industries of countries

Columns (1)-(5) are the OLS estimators

Columns (6)-(7) are the random effects estimators

Columns (8)-(9) are the fixed-effects estimators

Table 5: The impact of industry characteristics with data of the dependent variable from UNIDO

VARIABLES	OLS					Random effects		Fixed effects	
	Usual	Het	Country	Industry	Both	Country	Industries	Country	industry
	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(10)	(11)
LGDP (+)	0.66*** (9.66)	0.66*** (9.92)	0.66*** (6.02)	0.66*** (3.83)	0.66*** (3.45)	0.66*** (6.44)	0.69*** (7.01)		0.71*** (6.17)
Duties*LGDP (-)	-0.92*** (-3.92)	-0.92*** (-4.01)	-0.92*** (-2.94)	-0.92 (-1.64)	-0.92 (-1.55)	-0.92*** (-5.01)	-0.83** (-2.43)	-0.92*** (-5.01)	-0.77* (-1.92)
Fixed domestic cost*LGDP (+)	0.01*** (3.73)	0.01*** (4.08)	0.01*** (3.83)	0.01 (1.36)	0.01 (1.36)	0.01*** (4.76)	0.00 (1.55)	0.01*** (4.76)	0.00 (0.74)
Firm scale*LGDP (-)	-0.29** (-2.55)	-0.29*** (-2.69)	-0.29*** (-2.69)	-0.29 (-1.26)	-0.29 (-1.27)	-0.29*** (-3.26)	-0.29* (-1.74)	-0.29*** (-3.26)	-0.29 (-1.49)
Productivity dispersion*LGDP (+)	0.07*** (2.63)	0.07*** (2.76)	0.07*** (3.89)	0.07 (1.09)	0.07 (1.15)	0.07*** (3.36)	0.06* (1.68)	0.07*** (3.36)	0.06 (1.37)
Constant	-0.10*** (-2.78)	-0.10*** (-2.73)	-0.10 (-0.61)	-0.10*** (-4.55)	-0.10 (-0.62)	-0.10 (-0.61)	-0.10* (-1.81)	0.87*** (11.36)	-0.10*** (-2.84)
Observations	4,641	4,641	4,641	4,641	4,641	4,641	4,641	4,641	4,641
R-squared	0.20	0.20	0.20	0.20	0.20			0.02	0.20
Number of industries							51		51
Hausman test (p-value)									0.844
Number of countrypairs						91		91	
Breusch-Pagan test(p-value)						0	0		

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The dependent variable uses number of establishments from UNIDO

The model is estimated for industries appearing across all countries

Columns (1)-(5) are the OLS estimators

Columns (6)-(7) are the random effects estimators

Columns (8)-(9) are the fixed-effects estimators

Table 6: The impact of industry characteristics-Robustness check

VARIABLES	OLS					Random effects		Fixed effects	
	Usual	Het	Country	Industry	Both	Country	Industries	Country	industry
	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(10)	(11)
LGDP (+)	0.17*** (27.89)	0.17*** (28.57)	0.17*** (11.87)	0.17*** (6.52)	0.17*** (5.86)	0.17*** (13.42)	0.16*** (17.83)		0.16*** (17.15)
Duties*LGDP (-)	-0.53*** (-27.53)	-0.53*** (-25.63)	-0.53*** (-17.95)	-0.53*** (-5.41)	-0.53*** (-5.32)	-0.53*** (-31.23)	-0.49*** (-15.98)	-0.53*** (-31.23)	-0.48*** (-15.24)
Fixed domestic cost*LGDP (+)	0.00*** (15.19)	0.00*** (10.20)	0.00*** (11.35)	0.00** (2.01)	0.00** (2.02)	0.00*** (17.23)	0.00*** (8.96)	0.00*** (17.23)	0.00*** (8.56)
Firm scale*LGDP (-)	-0.29*** (-10.72)	-0.29*** (-12.34)	-0.29*** (-10.60)	-0.29** (-2.37)	-0.29** (-2.37)	-0.29*** (-12.17)	-0.28*** (-6.49)	-0.29*** (-12.17)	-0.28*** (-6.23)
Productivity dispersion*LGDP (+)	0.03*** (6.81)	0.03*** (6.84)	0.03*** (10.44)	0.03 (1.41)	0.03 (1.43)	0.03*** (7.73)	0.03*** (4.47)	0.03*** (7.73)	0.03*** (4.33)
Constant	0.03*** (7.44)	0.03*** (7.23)	0.03 (1.43)	0.03*** (7.13)	0.03 (1.43)	0.03 (1.50)	0.03** (2.52)	0.26*** (36.98)	0.03*** (7.68)
Observations	37,206	37,206	37,206	37,206	37,206	37,206	37,206	37,206	37,206
R-squared	0.11	0.11	0.11	0.11	0.11			0.04	0.09
Number of industries							106		106
Hausman test (p-value)									0.991
Number of countrypairs						351		351	
Breusch-Pagan test(p-value)						0	0		

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The dependent variable is measure by (14)

The advertising and R&D intensity is used as a proxy of firm-level economics of scale

The productivity dispersion is measured from output per company

Columns (1)-(5) are the OLS estimators

Columns (6)-(7) are the random effects estimators

Columns (8)-(9) are the fixed-effects estimators

4.5 Discussion

As mentioned above, our model finds two factors similar to the ones in Hanson and Xiang (2004) that influence the distribution of firms across industries: trade costs and elasticity of substitution. However, the impacts of these variables in our model differ from the ones in Hanson and Xiang (2004). The discrepancy can be attributed to different approaches in building the models. We use the heterogeneous-firm model with the presence of the homogeneous-product sector, while they use the homogeneous-firm model with the nonexistence of the homogeneous product sector. We find that industries with low trade costs concentrate more in large countries and this impact on the home-market effects in our model is consistent, while Hanson and Xiang (2004) show that industries with high trade costs tend to concentrate in large countries. However, this impact in their model is not monotonic. They show that when trade costs of industries are very high, the home-market effects of these industries will decrease. Regarding the elasticity of substitution, Hanson and Xiang (2004) find that industries with low-substitution elasticities tend to concentrate in large countries and this impact is monotonic whereas the impact of this parameter in our model depends on the relationship between domestic-fixed costs and export-fixed costs. As our model assumes that domestic-fixed costs are smaller than export-fixed costs, industries with high substitution elasticities tend to locate in large countries. Hanson and Xiang have the opposite result.

In our empirical study, we use average duty rates of countries to represent trade costs of industries, while Hanson and Xiang (2004) use freight rates of the US imports to represent trade costs. Our empirical study does not examine directly the effect of the substitution elasticity on the distribution of industries as the productivity dispersion of industries in our study includes the substitution elasticity. From the method of measuring the productivity dispersion as mentioned above, the positive impact of the productivity dispersion on the distribution of firms across industries might imply that industries with high productivity dispersion and/or high elasticity of substitution will tend to locate in large countries.

As we explained above, due to economies of scale, production costs of large countries are usually lower than those of small countries; therefore, if an industry's trade costs are

low, firms are more likely to locate in large countries to save production costs. Similarly, industries with high substitution elasticities tend to concentrate in large countries as products in these industries are quite similar and these products produced by small countries cannot compete with those from large countries due to high production costs. As a result, firms in industries with high substitution elasticities are more likely to concentrate in large countries.

Hanson and Xiang (2004) argue that although large countries have higher production costs than small countries, firms in industries with high trade costs still want to move to large countries as the benefits from savings in trade costs are larger than the increase in production costs. However, when trade costs of an industry are very high, goods in this industry are not traded, so the industry's home market effect decreases. The authors didn't provide clear explanations as why industries with a low-substitution elasticity locate more frequently in large countries.

From the home market-effect coefficients estimated in equation 11, we select two groups of industries with the highest home-market effects (HME) and industries with the lowest home-market effects (Table 7). As shown in our results, industries with high home-market effects tend to concentrate in large countries, while industries with low home-market effects locate in both large and small countries. In Table 7, we can see some examples that industries with high home market effects such as basic chemicals, or basic iron and steel tend to locate in large countries, while industries with low home-market effects such as furniture or electronics locate in both large and small countries.

5 Conclusion

Based on the model of heterogeneous firms in Helpman et al. (2004), we build a model to study the impact of industry characteristics on the arrangement of firms across industries between large countries and small countries. Our model predicts that industries with low trade costs, high fixed-domestic costs, low fixed export costs, and high productivity dispersion tend to concentrate in large countries, or that the home-market effects of these industries will be higher than with other industries.

Table 7: Groups of industries with high and low home market effects

ISIC	HME	Low HME industries	ISIC	HME	High HME industries
2211	0.008	Publishing of books and other publications	2699	0.286	Other non-metallic mineral products n.e.c.
3000	0.017	Office, accounting and computing machinery	3591	0.286	Motorcycles
3313	0.019	Industrial process control equipment	1512	0.288	Processing/preserving of fish
3220	0.030	TV, radio transmitters; line comm. apparatus	3692	0.317	Musical instruments
1552	0.030	Wines	1531	0.322	Grain mill products
3120	0.032	Electricity distribution , control apparatus	2926	0.333	Machinery for textile, apparel and leather
3610	0.037	Furniture	1711	0.343	Textile fibre preparation; textile weaving
3311	0.039	Medical, surgical and orthopaedic equipment	1532	0.346	Starches and starch products
1730	0.045	Knitted and crocheted fabrics and articles	1514	0.355	Vegetable and animal oils and fats
1553	0.055	Malt liquors and malt	2923	0.380	Machinery for metallurgy
2912	0.056	Pumps, compressors, taps and valves	2927	0.381	Weapons and ammunition
2212	0.057	Publishing of newspapers, journals, etc.	1542	0.388	Sugar
1541	0.059	Bakery products	2411	0.393	Basic chemicals, except fertilizers
3430	0.060	Parts/accessories for automobiles	2710	0.409	Basic iron and steel
3312	0.062	Measuring/testing/navigating appliances, etc.	2412	0.409	Fertilizers and nitrogen compounds
1912	0.062	Luggage, handbags, etc.; saddlery, harness	2692	0.420	Refractory ceramic products
3130	0.063	Insulated wire and cable	3511	0.448	Building and repairing of ships
2520	0.070	Plastic products	2430	0.452	Man-made fibres
3110	0.071	Electric motors, generators and transformers	2813	0.459	Steam generators
3230	0.072	TV and radio receivers and associated goods	2320	0.514	Refined petroleum products

An empirical model is then developed to examine these theoretical predictions using the data of 4-digit manufacturing industries ISIC in 28 high income countries. Our empirical evidence supports the predictions from the theoretical model. Economies of scale can be a key factor to explain why the industries will locate more in large countries.

This study can provide useful lessons in determining which industries should be most highly prioritized in both developed and developing countries (especially small countries). From the results, we think that small countries should promote the development of industries with high trade costs, low fixed domestic costs, low economics of scale, and low productivity dispersion. For example, a small country may want to focus on developing a furniture industry. If small countries develop industries such basic steel, they will not be able to compete with large countries in terms of production costs.

References

- Amiti, M.**, “Inter-industry trade in manufactures: Does country size matter?,” *Journal of International Economics*, 1998, 44 (2), 231–255.
- Brainard, S.L.**, “An Empirical Assessment of the Proximity-Concentration Trade-off Between Multinational Sales and Trade,” *The American Economic Review*, 1997, 87 (4), 520–544.
- Cameron, A.C. and P.K. Trivedi**, *Microeconometrics: methods and applications*, Cambridge Univ Pr, 2005.
- Feenstra, R. and H.L. Kee**, “On the measurement of product variety in trade,” *American Economic Review*, 2004, pp. 145–149.
- **and** — , “Export variety and country productivity: estimating the monopolistic competition model with endogenous productivity,” *Journal of International Economics*, 2008, 74 (2), 500–518.
- Feenstra, R.C.**, “New product varieties and the measurement of international prices,” *The American Economic Review*, 1994, 84 (1), 157–177.
- , **D. Madani, T.H. Yang, and C.Y. Liang**, *Testing Endogenous Growth in South Korea and Taiwan*, National Bureau of Economic Research Cambridge, Mass., USA, 1997.
- Hanson, G.H. and C. Xiang**, “The Home-Market Effect and Bilateral Trade Patterns,” *The American Economic Review*, 2004, 94 (4), 1108–1129.
- Helpman, E. and P.R. Krugman**, *Market structure and foreign trade: Increasing returns, imperfect competition, and the international economy*, The MIT press, 1987.
- , **M.J. Melitz, and S.R. Yeaple**, “Export Versus FDI with Heterogeneous Firms,” *The American Economic Review*, 2004, 94 (1), 300–316.
- Holmes, T.J. and J.J. Stevens**, “Does home market size matter for the pattern of trade?,” *Journal of International Economics*, 2005, 65 (2), 489–505.

- Hummels, D. and P.J. Klenow**, “The variety and quality of a nation’s exports,” *American Economic Review*, 2005, pp. 704–723.
- Hummels, D.L. and P.J. Klenow**, “The variety and Quality of a Nation’s Trade,” *NBER Working Paper*, 2002.
- Krugman, P.**, “Scale Economies, Product Differentiation, and the Pattern of Trade,” *The American Economic Review*, 1980, 70 (5), 950–959.
- Melitz, M.**, “The Impact of Trade on Aggregate Industry Productivity and Intra-Industry Reallocations,” *Econometrica*, 2003, 71 (6), 1695–1725.
- Moulton, B.R.**, “An illustration of a pitfall in estimating the effects of aggregate variables on micro units,” *The Review of Economics and Statistics*, 1990, 72 (2), 334–338.

Appendix

A

Assume that the productivity of firms in industry h in country i has the Pareto distribution, then $dG(a) = ka^{k-1}da$. The price index of industry h in country i is

$$\begin{aligned} P_h^i &= \int_0^{n_i^e} p(v)^{1-\sigma} \\ &= n_i \int_0^{a_D^i} p_{ii}(v)^{1-\sigma} dG(a) + n_j \int_0^{a_X^j} p_{ji}(v)^{1-\sigma} dG(a) \\ &= \frac{k}{\alpha^{1-\sigma}(k-\sigma+1)} (n_i (a_D^i)^{k-\sigma+1} + n_j \tau^{1-\sigma} (a_X^j)^{k-\sigma+1}) \end{aligned}$$

We have $x_{ii} = \frac{\beta_h Y_i (p_{ii}(v))^{-\sigma_h}}{\int_0^{n_i^e} p(v)^{1-\sigma_h} dv}$ and $x_{ij} = \frac{\beta_h Y_j (p_{ij}(v))^{-\sigma_h}}{\int_0^{n_j^e} p(v)^{1-\sigma_h} dv}$

A firm with the productivity a_d^i has zero profit in the domestic market and a firm with the productivity a_x^i has zero profit in the exporting market.

$$\pi_d^i = p_{ii}(v) x_{ii}^h(v) - (a_d^i x_{ii}^h(v) + f_d) = 0$$

$$\pi_x^i = p_{ij}(v) x_{ij}^h(v) - (a_x^i \tau x_{ij}^h(v) + f_x) = 0$$

Substituting P_h^i , $x_{ii}(v)$, $x_{ij}(v)$ of country i into above equations:

$$\pi_d^i = \frac{\beta Y_i (k-\sigma+1)(1-\alpha)(a_d^i)^{1-\sigma}}{k(n_i (a_d^i)^{k-\sigma+1} + n_j \tau^{1-\sigma} (a_x^j)^{k-\sigma+1})} - f_d = 0$$

$$\pi_x^i = \frac{\beta Y_j (k-\sigma+1)(1-\alpha)\tau^{1-\sigma}(a_x^i)^{1-\sigma}}{k(n_j (a_d^j)^{k-\sigma+1} + n_i \tau^{1-\sigma} (a_x^i)^{k-\sigma+1})} - f_e = 0$$

It is similar for country j :

$$\pi_d^j = \frac{\beta Y_j (k-\sigma+1)(1-\alpha)(a_d^j)^{1-\sigma}}{k(n_j^h (a_d^j)^{k-\sigma+1} + n_j \tau^{1-\sigma} (a_x^j)^{k-\sigma+1})} - f_d = 0$$

$$\pi_x^j = \frac{\beta Y_i (k-\sigma+1)(1-\alpha)\tau^{1-\sigma}(a_x^j)^{1-\sigma}}{k(n_i (a_d^i)^{k-\sigma+1} + n_j \tau^{1-\sigma} (a_x^j)^{k-\sigma+1})} - f_e = 0$$

From these questions, the following relationships are withdrawn:

$$\frac{a_d^i}{a_x^j} = \tau \left(\frac{f_d}{f_x} \right)^{\frac{1}{1-\sigma_h}} \quad (17)$$

$$\frac{a_d^j}{a_x^i} = \tau \left(\frac{f_d}{f_x} \right)^{\frac{1}{1-\sigma_h}} \quad (18)$$

We also get the value of a_d^i and a_d^j

$$(a_d^i)^k = \frac{1}{f_d} \frac{(1-\alpha)\beta_h Y_i (k-\sigma+1)}{k \left(n_i + n_j \tau^{-k} \left(\frac{f_d}{f_x} \right)^{\frac{k-\sigma+1}{\sigma-1}} \right)}$$

$$(a_d^j)^k = \frac{1}{f_d} \frac{(1-\alpha)\beta_h Y_j (k-\sigma+1)}{k \left(n_j + n_i \tau^{-k} \left(\frac{f_d}{f_x} \right)^{\frac{k-\sigma+1}{\sigma-1}} \right)}$$

If we call the f_e is the entry cost in country i

$$\int_0^{a_d^i} (p_{ii}(v)x_{ii}(v) - f_d) dG(a) + \int_0^{a_x^i} (p_{ij}(v)x_{ij}(v) - f_x) dG(a) = f_e$$

$$\begin{aligned} & \frac{(1-\alpha)\beta_h Y_i (a_d^i)^{k-\sigma+1}}{n_i (a_d^i)^{k-\sigma+1} + n_j \tau^{1-\sigma} (a_x^j)^{k-\sigma+1}} - f_d (a_d^i)^k + \\ & + \frac{(1-\alpha)\beta_h Y_j (a_d^j)^{k-\sigma+1}}{n_j (a_d^j)^{k-\sigma+1} + n_i \tau^{1-\sigma} (a_x^i)^{k-\sigma+1}} - f_x (a_x^i)^k = f_e \\ & \left(\frac{\sigma-1}{k-\sigma+1} \right) (f_d (a_d^i)^k + f_x (a_x^i)^k) = f_e \end{aligned} \quad (19)$$

It is similar for country j :

$$\left(\frac{\sigma-1}{k-\sigma+1} \right) (f_d (a_d^j)^k + f_x (a_x^j)^k) = f_e$$

From (17), (18), (19), we have the following results

$$a_d^i = a_d^j = a_d$$

$$a_x^i = a_x^j = a_x$$

and

$$B_i = B_j = B$$

$$A_i = A_j = A$$

B Samples

Table 8: Groups of industries with high and low home market effects

Order	ISOC	Country	Regions	UNIDO Sample
1	CAN	Canada	1	Canada
2	USA	USA	1	
3	AUS	Australia	2	Australia
4	NZL	New Zealand	2	New Zealand
5	HKG	China, Hong Kong SAR		
6	JPN	Japan	3	Japan
7	KOR	Rep. of Korea	3	Rep. of Korea
8	SGP	Singapore		
9	AUT	Austria	4	Austria
10	BEL	Belgium	4	
11	CZE	Czech Rep.	4	
12	DNK	Denmark	4	Denmark
13	FIN	Finland	4	Finland
14	FRA	France	4	
15	DEU	Germany	4	Germany
16	GRC	Greece	4	
17	HUN	Hungary	4	
18	IRL	Ireland	4	
19	ISR	Israel		
20	ITA	Italy	4	
21	NLD	Netherlands	4	
22	NOR	Norway	4	
23	PRT	Portugal	4	Portugal
24	ESP	Spain	4	Spain
25	SWE	Sweden	4	Sweden
26	CHE	Switzerland	4	
27	TUR	Turkey	4	Turkey
28	GBR	United Kingdom	4	United Kingdom

Sample 1 is the main sample of our study (28 countries). In this case, country pairs are chosen from all countries in the sample ($C_{28}^2 = 378$ pairs)

Sample 2 is separated according regions. Region 1: US and Canada, region 2: Australia and New Zealand, region 3: Japan and Korea, and region 4: 19 countries in European Union. Country-pairs are formed from different regions ($1 + 1 + 1 + C_{19}^2 = 174$ pairs)

Unido sample is countries in the sample when UNIDO database is used ($C_{14}^2 = 91$ pairs)

C Fixed domestic costs

We use some expense costs in Annual Manufacturing of Survey to represent fixed domestic costs. These costs include:

- costs of electricity
- temporary staff and leased employee expenses
- Costs of software, computers, communication services
- Repair and maintenance services of building and machinery
- Advertising and promotional services
- Purchased professional and technical services
- Taxes and licenses fees