FACULTY OF LAW, ECONOMICS AND FINANCE

CREA
Discussion
Paper
2012-12

Center for Research in Economic Analysis
University of Luxembourg

The Home market effect with heterogeneous firms

available online : http://wwwfr.uni.lu/recherche/fdef/crea/publications2/discussion_papers/2011

Ha Nguyen, CREA, University of Luxembourg

December, 2012

For editorial correspondence, please contact: crea@uni.lu
University of Luxembourg
Faculty of Law, Economics and Finance
162A, avenue de la Faïencerie
L-1511 Luxembourg

UNIVERSITÉ DU LUXEMBOURG

The Home market effect with heterogeneous firms

Ha Nguyen[*]

CREA, University of Luxembourg

December 12, 2012

Abstract

This paper examines the home market effect in the framework of heterogeneous firms. The paper finds that not only trade costs but also fixed trade costs cause the home market effect and the reverse home market effect can occur as the fixed trade costs are very low. In addition, the magnitude of the home market effect varies with industry characteristics. Industries with low trade costs, high fixed production costs, low fixed export costs, and high productivity dispersion tend to be more concentrated in large countries. Finally, the negative impact of trade barriers on the home market effect is dampened by the elasticity of substitution which is contrary with the result of the homogeneous firm model. An empirical model is built to test these predictions for developed countries. The empirical results are consistent with the predictions of the theoretical model.

Keywords: Home market effect, country size, industry characteristics, heterogeneous firms, firm's location, market structure

JEL Classification: F1, R1

[*]Address: CREA, University of Luxembourg, e-mail:*ha.nguyen@uni.lu*. I am very grateful to Professors: Robert McNown, Stephen Yeaple, Keith Maskus, Ardeshir Sepehri, Thibault Fally, and Mattias Nilsson for their advice and suggestions. I am also very grateful for the comments of Professor Pierre Picard and the people who participated in The Asia Pacific Trade Seminars (2012) and the Internal Seminar of CREA at the University of Luxembourg-All remaining errors are mine- I gratefully acknowledge the financial support of the University of Luxembourg (F2R-CRE-PUL-10EGQH).

1 Introduction

The ”home market effect”, which was first introduced by Krugman (1980), offers two predictions: a large country in the presence of trade costs has more products (or firms) in its increasing-returns to scale sector than 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. Some studies such as Davis and Weinstein (1999), Davis and Weinstein (2003), Head and Ries (2001), Feenstra, Markusen, and Rose (2003) find empirical evidences consistent with the existence of the home market effect.

In addition, some studies also indicate that the strength of home market effect can vary with many factors. For instance, Helpman and Krugman (1985) show that when variable trade costs are low, the home market effect is stronger. Employing the model of monopolistic competition with many differentiated product industries, Hanson and Xiang (2004) find that industries with high transportation costs and low substitution elasticicities tend to concentrate in the larger country. Holmes and Stevens (2005) uses a mixture model of competition and oligopoly and finds that industries with higher scale economies have higher home market effect. Behrens and Picard (2007) document that industries with many multinational firms, equivalent with low fixed production costs, high transport cost, high elasticity of substitution have lower home market effect.

This paper investigate which factors cause the home market effect and how the strength of home market effect varies with industrial characteristics in the framework of monopolistic competition with heterogeneous firms. Studying the home market effect in the model of heterogeneous firms is important since this model assumes that firms’ productivity is different and only high productivity firm can enter the export market. So, firms’response to changes of trade barriers can be very different from the one predicted by the model of homogeneous firm. In the model of homogeneous firms, all firms can export their product as trade occurs. So, the change in trade barriers only affects how much output a firm can export (i.e., intensive margin). Meanwhile, in the model of heterogeneous firms, the changes of trade barriers affect not only how much output a firm can export but also how many firms can enter the foreign market (i.e., extensive margin). Firms’ response to the changes in trade barriers in this model results in both changes in intensive and extensive margins. As a result, I believe that the mechanism that results in the home market effect in the model of heterogeneous firm model can be different from the one predicted in the homogeneous firm model. Chaney (2008) finds remarkable differences in explaining trade pattern of countries across two models. Some studies such as Baldwin and Okubo (2006) and Okubo, Picard, and Thisse (2010) which are more in line with economic geography also mentioned the home market effect in the framework of heterogeneous firms. Both studies find that the large country attracts all the most efficient firms (highest productivity firms).

This paper develops a two-country version of Helpman, Melitz, and Yeaple (2004) model-but without multinationals and investigates under which conditions a ’home market effect’ occurs and how the magnitude of the home market effect varies with industrial characteristics. We assume each country has many differentiated product industries in the increasing returns sector and one homogeneous product industry in the constant return sector. Homogeneous goods can be freely traded. Labor is the only production factor in the model. If firms want to export their products to foreign markets, they have to pay fixed trade costs (or fixed export costs). This theoretical model predicts the following results:

First, as we know, the Krugman’s model with homogeneous firms predicts that when variable trade costs are zero, the home market effect will not exist. Also, since the variable trade costs are always present in the Krugman’s model, the home market effect is always possible in the increasing return sector. However, when using the model of heterogeneous firms, I find that not only variable trade costs but also fixed trade costs are necessary conditions for the home market effect. Therefore in my approach, when variable trade costs are zero, the home market effect is still able to occur as long as fixed trade costs exist. In addition, the model predicts that the reverse home market effect can still occur in the increasing return sector if the fixed trade costs are very low. This result implies that not all industries in the increasing return sector have the home market effect.

Second, the model finds that the strength of the home market effect across industries depends on many industrial characteristics. In addition to the industrial characteristics that are predicted by prior studies, the model finds that other characteristics such as fixed trade costs and the productivity dispersion also affect the home market effect. Industries with lower fixed trade costs allow firms to enter the market easier. Therefore the home market effect in these industries is stronger. Industries with low productivity dispersion usually have many low productivity firms, while industries with high productivity dispersion have many high productivity firms. As a result, all else constant, it can be easier for firms in industries with high productivity dispersion to enter the export market. This explains why industries with high productivity dispersion have a stronger home market effect.

Third, the model of heterogeneous firms also changes several results found in previous studies which use the model of homogeneous firms. The first difference is the impact of the substitution elasticity. All studies which employ the model of homogeneous firms find that industries with high substitution elasticity (i.e., less differentiated products) have a lower home market effect. Whereas the impact of the substitution elasticity in my approach depends on the relationship between the fixed production costs and the fixed trade costs. When fixed trade costs are higher than the fixed production cost, we find that industries with high elasticity tend to locate more in the large country.

In the model of homogeneous firm (.i.e Helpman and Krugman (1985)), the negative impact of trade barriers on the home market effect are magnified by high elasticity of substitution. Our paper find a contrary result: the impact of trade barriers on the home market effect model is magnified by lower elasticity of substitution. The industries with high substitution elasticities often include a larger portion of high productivity firms. Therefore when trade costs are high, most of firms in these industries still can export their products to foreign countries. As a result, the increase (or decrease) of trade barriers will not affect much the strength of the home market effect in these industries. Whereas, the industries with low elasticity of substitution include many low productivity firms. Therefore, the change of trade barriers can change the number of export firms remarkably. Hence, the impact of trade barriers on the home market effect is dampened by the elasticity of substitution. These findings are consistent with Chaney (2008) which shows that the impact of trade barriers on trade flows is dampened with the elasticity of substitution. His result is also contrary with the homogeneous firm's results.

In addition, my paper finds that the impact of trade barriers on the home market effect is magnified by the industries with high productivity dispersion. As mentioned above, the industries with less productivity dispersion have more low productivity firms. The change of trade barriers (both variable and fixed barriers), as a result, can result in a remarkable change of the number of export firms in these industries.

Except for the above differences, other findings of the paper regarding the impact of the industrial characteristics such as variable trade costs and fixed production costs on the home market effect are consistent with many studies in prior literature. Industries with high fixed production costs and low variable trade costs concentrate more in the large countries. The impact of fixed production costs are similar to the one in Holmes and Stevens (2005) and Behrens and Picard (2007), while the impact of variable trade costs are consistent with the findings of Helpman and Krugman (1985) and Behrens and Picard (2007) but different with the one of Hanson and Xiang (2004). Hanson and Xiang (2004) find that industries with high trade costs have higher home market effect. This difference in their finding is attributed to the difference in the model assumptions. Hanson and Xiang (2004) don't use the free-trade homogeneous sector in their model, while the other studies including mine use this sector. Without the presence of this sector in the model, their model implies that the larger a country, the higher its wage. Whereas, the other studies implies that wages are equal between two countries. This assumption results in the difference in the impacts of trade costs on the home market effect. However, this impact of variable trade costs in Hanson and Xiang (2004) is differential.

Based on the theoretical model, I develop an empirical model to present the relationship between the number of export products, the industrial characteristics, and the country's size. Using this empirical model, I examine whether the industry types predicted by my theoretical model tend to locate in a large country or not. My empirical analysis which use the 5-digit NAICS manufacturing industries in 28 high income countries supports the predictions of the theoretical model. I use Hummels and Klenow (2005)'s method to measure the number of export products across the countries'industries, and the US industrial data to represent the industrial characteristics. My empirical approach is also different with the one used by Hanson and Xiang (2004). These authors use the difference-in-difference method to study the impact of industry characteristics (trade costs and substitution elasticity) on the home market effect. They first use the industrial characteristics to divide industries into two groups: treatment and control groups and subsequently use the difference-in-difference method. The disadvantage of this method is that it is difficult to divide industries into two groups when there are many industrial characteristics involved, like the one we have in my study.

The rest of this 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 the data analysis and discusses the results of the empirical model. Section 5 concludes with a discussion of some of the implications.

2 The Model

2.1 Set up

Assume that there are two countries (i,j) and that each country has H+1 industries. Similar to Helpman, Melitz, and Yeaple (2004), we assume that 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 that 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 to be the numeraire, and it can be freely traded. The price of good z is set to 1, meaning that every country producing this good will have an identical wage rate (=1). On the demand side, assume that all individuals in country i have the same utility function:

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

where $x_h^i(v)$ is the consumption of country i of 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 the 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 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 fixed production cost of f_d^h units of labor to produce a positive amount in each period. The fixed production costs can include the overhead production costs, which are the costs of establishing distribution and the servicing networks in the domestic market. Assume that this cost is identical across all of the firms operating in each industry. Therefore, the production costs of a firm (or a plant) located in market i producing quantity x_{ii} are $ax_{ii}^h(v) + f_d^h$. 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 can include the costs of establishing the distribution network, administrative costs, advertising, or the costs of research and development 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 costs of a firm that exports products to a foreign market is $\tau_{ij}^h a x_{ij}^h(v) + f_x^h$. Assume that the fixed costs and the distribution function of a in each industry are identical in the two countries. In addition, the trade costs are assumed to be identical between the 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 the other firms. Because a is the number of units of labor required to produce one unit of product in industry h in country i, $\frac{1}{a}$ is considered to be the productivity of a firm in industry h. The firms that have a productivity larger than $\frac{1}{a_{ii}^h}$ produce and sell their products in the domestic market, and the firms with a productivity of $\frac{1}{a_{ij}^h}$ earn zero profits. The set of firms with $\frac{1}{a} > \frac{1}{a_{ij}^h}$ produce products for the domestic market and for the exporting market. The set of firms with $\frac{1}{a_{ij}^h} \leq \frac{1}{a} \leq \frac{1}{a_{ij}^h}$ produce for the domestic market only. The set of firms with $\frac{1}{a} \leq \frac{1}{a_{ij}^h}$ 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

$$\pi_{ii}^h = p_{ii}^h(v) x_{ii}^h(v) - (a x_{ii}^h(v) + f_d^h)$$

The additional profit of a firm from the export market is

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

The price that 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 is $p_{ij}^h(v) = \frac{\tau^h a}{\alpha_h}$. Substituting the domestic value, the exporting value, and the prices into the profit equations, the profits of the firms in industry h in the domestic market (i) and the exporting market(j) are

$$\pi_{ii}^h = a^{1-\sigma_h} B_h^i - f_d^h$$

$$\pi_{ij}^h = a^{1-\sigma_h} \tau^{1-\sigma_h} B_h^j - f_x^h$$

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

Because the firms with a productivity level of $\frac{1}{a_{ii}^h}$ earn zero profit in the domestic market, and the firms with productivity of $\frac{1}{a_{ij}^h}$ earn 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 through the

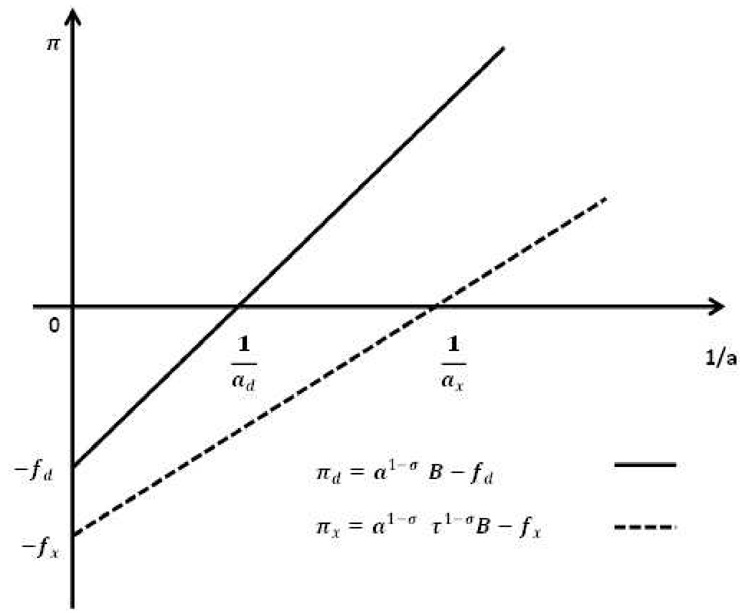

Figure 1: Profit from domestic sales and exports

equations where profit is equal to zero:

$$(a_{ii}^h)^{1-\sigma_h} B_h^i = f_d^h \Rightarrow a_{ii}^h = \left(\frac{f_d^h}{B_h^i}\right)^{\frac{1}{1-\sigma_h}}$$

$$((\tau^h a_{ij}^h)^{1-\sigma_h}) B_h^j = f_x^h \Rightarrow \tau^h a_{ij}^h = \left(\frac{f_x^h}{B_h^j}\right)^{\frac{1}{1-\sigma_h}}$$

Because fixed costs are assumed to be the same in both countries, the distribution function G(.) is also the same in both countries. In addition, because the trade costs are the same between the two countries, we can show that the cutoff levels of productivity are equal in both countries. This result means that $a_{ii}^h = a_{jj}^h = a_d$, $a_{ij}^h = a_{ji}^h = a_x^h$, which leads to $B_h^i = B_h^j = B_h$ (in the Appendix). These results hold for each of H industries in country i and country j. From the above equation, the relative cutoff level of production costs of the domestic and the foreign market is

$$\frac{a_d^h}{a_x^h} = \tau_h \left(\frac{f_d^h}{f_x^h}\right)^{\frac{1}{1-\sigma}} \tag{1}$$

In the following sections, we focus on industry h in countries i and j; therefore, we drop the h subscript to make the equations simpler.

2.3 Number of entrants and market size

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

$$\begin{aligned}\int_{v\in\Omega_i} p(v)^{1-\sigma}dv &= n_i\int_0^{a_D}(\frac{a}{\alpha})^{1-\sigma}dG(a)+n_j\int_0^{a_X}(\tau\frac{a}{\alpha})^{1-\sigma}dG(a)\\ &= n_i(\frac{1}{\alpha})^{1-\sigma}V(a_D)+n_j\tau^{1-\sigma}(\frac{1}{\alpha})^{1-\sigma}V(a_X)\end{aligned} \tag{2}$$

Parameters n_i and n_j are considered to be the number of entrants in industry h of countries i and j, Ω_i is the set of firms or products in country i . Substituting the above results into (2) yields

$$n_iV(a_D)+n_j\tau^{1-\sigma}V(a_X)=\frac{(1-\alpha)\beta Y_i}{B} \tag{3}$$

Similar to country j

$$n_jV(a_D)+n_i\tau^{1-\sigma}V(a_X)=\frac{(1-\alpha)\beta Y_j}{B} \tag{4}$$

Using equations (3) and (4) and solving for $\frac{n_i}{n_j}$:

$$\frac{n_i}{n_j}=\frac{\frac{Y_i}{Y_j}-\frac{\tau^{1-\sigma}V(a_X)}{V(a_D)}}{1-\frac{\tau^{1-\sigma}V(a_X)}{V(a_D)}\frac{Y_i}{Y_j}}=\frac{\lambda-\rho}{1-\rho\lambda} \tag{5}$$

where $\lambda=\frac{Y_i}{Y_j}$ and $\rho=\frac{\tau^{1-\sigma}V(a_X)}{V(a_D)}$. This equation is similar to the equation in Krugman (1980).

If we assume that the productivity of the firms in industry ($\varphi=1/a$) has a Pareto distribution in $\varphi\geq\varphi_o$ with the cumulative distribution function of φ, then $F(\varphi)=1-(\frac{\varphi_o}{\varphi})^k$, k denotes the dispersion parameter of productivity. The industries with a low value of k have a high productivity dispersion, and the industries with a high value of k have a low productivity dispersion. From this result, the cumulative distribution function of a will be $G(a)=P(\frac{1}{\varphi}\leq a)=P(\varphi\geq\frac{1}{a})=1-F(\frac{1}{a})=1-(1-(\varphi_oa)^k)=(\varphi_oa)^k$ for $a\leq\frac{1}{\varphi_o}$. The population density function is

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

From that, $V(a_D)$ and $V(a_X)$ become

$$V(a_D)=\int_0^{a_D}a^{1-\sigma}dG(a)=ca_D^{k-(\sigma-1)}$$

$$V(a_X)=\int_0^{a_D}a^{1-\sigma}dG(a)=ca_X^{k-(\sigma-1)}$$

From here, we can find that

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

To sum up, we have the following equations to analyze the home market effect in the model of heterogeneous firms:

$$\frac{n_i}{n_j} = \frac{\lambda-\rho}{1-\rho\lambda} \tag{6}$$

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

The equation of the home market effect is still similar to the one in Krugman (1980), however, ρ is different as we discuss bellow.

2.4 Home market effects with heterogeneous firms

In this part, I will investigate the three main points. The first point is that which conditions allow the existence of home market effect in the heterogeneous firm model. The second one is how the magnitude of the home market effect changes with industrial characteristics. The final point is how the impact of trade barriers on the home market effect changes with industrial characteristics?

Assume that firms locate in both countries, it means that n_i and n_j are positive. This requires that $\rho < \lambda < \frac{1}{\rho}$ or $\frac{1}{\rho} < \lambda < \rho$. When firms locate in both countries, the relationship between the relative number of firms and countries'size as follow:

$$\frac{\partial\,(n_i/n_j)}{\partial\lambda} = \frac{1-\rho^2}{(1-\lambda\rho)^2} \tag{7}$$

If ρ is less than one, the relation is positive. By contrast, the relation is negative. This result is different from Helpman and Krugman (1985)'s one which show that ρ is always smaller than 1 as the presence of variable trade costs ($\tau > 1$). As a result, the home market effect always exhibits in their model. ρ in this study which can be larger or smaller than 1 depends on our assumptions, especially the fixed trade costs. If we assume that only firms with high productivity can export, the cutoff level of production in the domestic market should be larger than the one in the export market, $(a_d > a_x)$. This implies that $f_x\tau^{\sigma-1} > f_d$ or ρ is smaller than 1. Therefore, there is a positive relationship between the relative number of firms and countries'size. In addition, once the fixed trade costs (f_x) are very small, ρ still can be larger than 1. In this case, the relationship between the relative number of firms and countries' relative size is negative or the reverse home market effect occurs.

The aforementioned results imply that the home market effect only occur with the case of $\rho < 1$. Together with the condition of firms locating in both countries ($\rho < \lambda < \frac{1}{\rho}$), the result of equation (7) is larger than 1 $\left(\frac{1-\rho^2}{(1-\lambda\rho)^2} > 1\right)$. This result suggests that the larger market attracts a disproportional share of the firms in the industry h or we say the industry h has the home market effect.

Proposition 1. *The home market effect occurs in the model of heterogeneous firm when* $\rho < \lambda < \frac{1}{\rho}$ *and* $\rho = \frac{1}{\tau^{\sigma-1}}\left(\frac{f_D}{f_X\tau^{\sigma-1}}\right)^{\frac{k-\sigma+1}{\sigma-1}} < 1$

When the home market effect exists, the magnitude of the home market effect seems to vary across industries. Let $g = \frac{1-\rho^2}{(1-\lambda\rho)^2}$, from the equation, we see that g presents the relation between the relative number of firms between two countries and their size or it measures the magnitude of the home market effect. In addition, g changes with respect to ρ:

$$\frac{\partial g}{\partial \rho} = \frac{2(\lambda-\rho)(1-\rho\lambda)}{(1-\rho\lambda)^4} > 0 \tag{8}$$

This result 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, the higher the value of ρ, the larger the home market effect of the industry. Because ρ depends on the characteristics of the industry, the strength of the home market effect depends on the industry characteristics. To find the effects of one industry characteristic on (ρ), we assume that the other characteristics are constant.

The impact of trade costs: The derivative of ρ with respect to the trade costs shows the following.

$$\begin{aligned} \frac{\partial \rho}{\partial \tau} &= \left(\frac{f_d}{f_x}\right)^{\frac{k-\sigma+1}{\sigma-1}}\left(\frac{-k}{\tau^{k+1}}\right) < 0 \\ \frac{\partial \rho}{\partial f_x} &= \left(-\frac{k-\sigma+1}{\sigma-1}\right)\frac{1}{\tau^k}\left(f_d\right)^{\frac{k-\sigma+1}{\sigma-1}}\left(f_x\right)^{-\frac{k}{\sigma-1}} < 0 \end{aligned} \tag{9}$$

When trade costs decrease across industries, the difference in the number of products between the two countries becomes larger. This trend suggests that firms in the industries with low trade costs will tend to concentrate in the large country. Economies of scale implies that the production costs of the firms in the large country are lower than the production costs of the firms in the small country, making the prices of the products from the large country cheaper. When the trade costs are low, the low-priced products of the large country will easily penetrate into the small country's market. Consequently, the high-priced products from the small country cannot compete with the low-priced products of the large country, and the firms in the small country will exit the markets when trade liberalization occurs.

The impact of fixed production costs:

$$\frac{\partial \rho}{\partial f_d} = \left(\frac{k-\sigma+1}{\sigma-1}\right)\frac{1}{\tau^k}(f_x)^{-\frac{k-\sigma+1}{\sigma-1}}(f_d)^{\frac{k}{\sigma-1}-2} > 0 \tag{10}$$

An increase in the fixed production costs (or fixed domestic costs) leads to a higher value of ρ. The industries with high fixed production costs tend to locate in the large country to take advantage of the economies of scale. In the large country, it will be easier for firms to attain an efficient scale.

The impact of the productivity dispersion and the elasticity of substitution:

$$\begin{aligned} \frac{\partial \rho}{\partial k} &= \left(\frac{1}{\sigma-1}\right)\left(\frac{1}{\tau^{\sigma-1}}\right)\left(\frac{f_d}{f_x\tau^{\sigma-1}}\right)^{\frac{k-\sigma+1}{\sigma-1}} ln\left(\frac{f_d}{f_x\tau^{\sigma-1}}\right) \\ \frac{\partial \rho}{\partial \sigma} &= \left(\frac{-k}{(\sigma-1)^2}\right)\left(\frac{1}{\tau^k}\right)\left(\frac{f_d}{f_x}\right)^{\frac{k-\sigma+1}{\sigma-1}} ln\left(\frac{f_d}{f_x}\right) \end{aligned} \tag{11}$$

Because we assume that only certain firms with high productivity can export to foreign markets, it 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 the industries with high productivity dispersion (low k) will locate more often in the large country, when we assume that firms' productivity has Pareto distribution. The industries with high productivity dispersion have more high productivity firms. When trade take places among countries, these firms will export their products easier, so the home market effect will be more likely to happen for these industries. Whereas, the industries with low productivity dispersion include many firms with low productivity, hence not many firms in these industries can export their products. As a result, the home market effect is lower in this case.

If the fixed domestic costs are smaller than the fixed export costs, $(f_d < f_x)$, $\frac{\partial \rho}{\partial \sigma} > 0$ implies that the industries with a high elasticity of substitution (high σ) will locate more often in the large country and vice versa, the firms in the industries with a low elasticity of substitution (low σ) will tend to concentrate in the large country. We can see the impact of substitution elasticity depends on the fixed entry costs, when the fixed entry cost is high, only firms with high productivity can enter the export market. So, industries with high elasticity of substitution include high productivity firms, firms still can enter the foreign markets when the fixed trade costs are high. As a result, the home market effect is more likely to occur in these industries. Whereas, firms in the industries with low elasticity of substitution usually have low productivity. When the fixed trade costs are high, firms cannot enter the foreign markets, the home market effect is hence lower. However, when the fixed trade costs are low, we have the opposite results.

Proposition 2. *Industries with high fixed production costs, low trade barriers (both fixed and variable), and high productivity dispersion have higher home market effect or tend to be more concentrated in the large country.*

Trade barriers are important to explain the home market effect. High trade barriers restrict trade among countries. Therefore, products from foreign countries can not have much effect on the products produced in domestic markets, so it is more difficult to observe the home market effect. However, when trade barriers decrease, products from one country can enter another country easily and the home market effect will occur. However, the change of the home market effect is not similar across industries when trade barriers change. We can see through rewriting the formula of ρ as follow:

$$log(\rho) = -klog(\tau) + \frac{k-\sigma+1}{\sigma-1}log(f_D) - \frac{k-\sigma+1}{\sigma-1}log(f_X)$$

This equation implies that the impact of trade barriers (both fixed and variable costs) on the home market effect will be stronger for industries with low productivity dispersion (high k). As mentioned above, industries with low productivity dispersion include many firms with low productivity. So, the change of trade barriers will have stronger impacts on these industries. As a result, the home market effect of these industries varies significantly with the change of trade barriers. Or the impact of trade barriers on the home market effect is dampened by industries with high productivity dispersion.

The equation also implies the impact of trade barriers on the home market effect is also dampened for industries with high elasticity of substitution (high σ). When trade barriers (fixed trade costs) change, the home market effect of industries with high elasticity of substitution changes less. Whereas, the home market effect of industries with low elasticity of substitution change more. We know industries with low elasticity of substitution include many low productivity firms. When trade barriers change, they have stronger effect on firms in these industries. As a result, the home market effect is more significant for these industries. This result is contrary to the result from the model of homogeneous firm discussed bellow. These explanations are similar to Chaney (2008) who shows that the impact of trade barriers (also fixed trade costs) on trade volume will be dampened in industries with high elasticity of substitution. In addition, the productivity dispersion has effect on the home market effect with respect to the change in both fixed and variable trade costs, while the elasticity of substitution has only effect on the home market effect with respect to on the change of fixed costs.

Proposition 3. *The impact of trade barriers on the home market effect is dampened by industries with high productivity dispersion and industries with high elasticity of substitution.*

2.5 The comparison across models

This part will summarize some important differences in the home market effect explained by the heterogeneous firm model and by the homogeneous firm model.

In the model of homogeneous firms (.i.e Helpman and Krugman (1985)), the magnitude of the home market effect usually depends on two industrial characteristics: variable trade costs and the elasticity of substitution. As in Helpman and Krugman (1985), ρ is written as follow:

$$\rho = \tau^{1-\sigma} \tag{12}$$

Although they consider only one industry in the sector of increasing return to scale. However, this equation is still held when the model is extended for many industries in the sector of increasing return. In this model, if there are not the presence of variable trade costs($\tau = 1$), the home market effect will not exist in the model. By contrast, the model implies that the home market effect always exist. While, this paper with the framework of heterogeneous firms finds that as variable trade costs are zero, the home market effect still exists if $f_d < f_x$. Furthermore, the model also implies that the reversed home market effect can occur if the fixed trade costs (f_x) are very small. When this case happens, ρ can be larger than 1. Therefore, the result of equation (12) is negative which implies the reversed home market effect.

Second, the formula of the homogeneous firm model implies that industries with low variable trade costs and the low elasticity of substitution will locate more in the large country. The impact of variable trade costs are similar to this paper's result, however the impact of the elasticity of substitution is not. The equation (12) says that industries with the low elasticity of substitution tend to locate in the large country. Meanwhile, the impact of this variable in this study depends on the relationship between fixed production costs and fixed trade costs. If fixed trade costs are larger than fixed production costs ($f_x > f_d$), industries with high elasticity of substitution tend to concentrate in the large country. While, most of studies relating to the homogeneous firm model find that industries with low elasticity of substitution locate more in the large country. The explanation of this result is also different from the one in our paper mentioned above. In the model of homogeneous firms, all firms have the same productivity and all firms can export. Industries with low elasticity of substitution usually have many firms. As a result, as trade occurs, firms in the small country have to compete with many firms from the large country. Because of disadvantage of scale, many firms from the small country can be shutdown. This explain why the home market effect is stronger for industries with low substitution elasticity in the homogeneous firm model.

Third, as in equation (12), the model of homogeneous firm implies that as trade barriers decrease (or increase), the magnitude of the home market effect will increase (or decrease) stronger for industries with high elasticity of substitution. As mentioned above, this impact is contrasted with the one in the model of heterogeneous firm. In the homogeneous firm model, the change of trade barriers doesn't change the number of export firms, it only changes output which each firm exports (intensive margin). This explains why trade volume in industries with high elasticity of substitution are stronger response to changes in trade barriers in the model of homogeneous firm. As a result, the change of the home market effect is larger when trade barriers change in the model of homogenous firm.

The impact of trade barriers in our paper is similar to many paper, but it is different from Hanson and Xiang (2004). Hanson and Xiang (2004) find that industries with high trade costs have the higher home market effect. The reason for this difference is from the assumptions of the model. Hanson and Xiang (2004) don't use the free-trade homogeneous sector in the model. This result implies that the large country has higher wages than the small country. Thus, firms in industries with high trade costs relocate to the large country since the benefits from saving trade costs are larger than the losses from paying higher production costs. However, trade costs have a differential effect on the home market effect in their model. We can see that the derivative of the home market effect to trade costs can be positive or negative in their model (page 1112). In their analysis, they assume that this derivative is positive. While, studies using homogeneous sector like our study find that industries with low trade costs will locate more in the large country. The result of this case is explained that when trade costs are low, firms can export their products easier to other countries. Firms' products from the large country with the advantage of scale economies can exclude firms in the small country. As the result, the home market effect can occur between the large country and the small country. In empirical study, Hanson and Xiang (2004) find that their results are consistent with transportation costs, while this paper's empirical results are consistent with tariff barriers or distances. A general comparison across the models are provided in table 1.

Table 1: **A comparison across models**

Industrial characteristics	Homogeneous firm model		Heterogeneous firm model with homogeneous sector
	The appearance of homogeneous sector	No homogeneous sector (Hanson and Xiang (2004))	This study
Elasticity of substitution ($\sigma \uparrow$)	HME $\downarrow$	HME $\downarrow$	HME $\uparrow\downarrow$
Trade costs ($\tau \uparrow$)	HME $\downarrow$	HME $\uparrow\downarrow$	HME $\downarrow$
Productivity dispersion ($k \uparrow$)	No	No	HME $\uparrow$
Fixed production costs ($f_d \uparrow$)	No	No	HME $\uparrow$
Fixed export costs ($f_x \uparrow$)	No	No	HME $\downarrow$

3 Empirical model

3.1 Empirical method

The number of firms (or products) that can export from country i to country d is

$$n_{id} = n_i \int_0^{a_{id}} dG(a) = n_i a_{id}^k = n_i \left(\frac{1}{\tau_{il}}\right)^k \left(\frac{f_x^d}{B_d}\right)^{\frac{k}{1-\sigma}}$$

From the theoretical model, the fixed export costs and the demand per product are similar in every country($f_x^d = f_x$ and $B_d = B$); therefore, the above equation can be rewritten as follows:

$$n_{id} = n_i \left(\frac{1}{\tau_{il}}\right)^k \left(\frac{f_x}{B}\right)^{\frac{k}{1-\sigma}}$$

We have similar equations for export from country i to other destinations d. From these equations, the average number of export products (firms) from country i is

$$\prod_{d=1}^{I} (n_{id})^{w_{id}} = n_i^{\sum_d w_{id}} \prod_{d=1}^{I} \left(\frac{1}{\tau_{id}^k}\right)^{w_{id}} \prod_{d=1}^{I} \left(\frac{f_x^d}{B_d}\right)^{\frac{k\Sigma w_{id}}{1-\sigma}}$$

Or

$$\prod_{d=1}^{I} (n_{id})^{w_{id}} = n_i \prod_{d=1}^{I} \left(\frac{1}{\tau_{id}^k}\right)^{w_{id}} \left(\frac{f_x}{B}\right)^{\frac{k}{1-\sigma}}$$

w_{id} is the export weight from country i to country d and $\sum_d w_{id} = 1$.

Similarly, the average number of export products (or firms) from country j is

$$\prod_{l=1}^{I} (n_{jd})^{w_{jd}} = n_j \prod_{d=1}^{I} \left(\frac{1}{\tau_{jd}^k}\right)^{w_{jd}} \left(\frac{f_x}{B}\right)^{\frac{k}{1-\sigma}}$$

The ratio of export products from countries i and j is

$$\frac{\prod_{d=1}^{I}(n_{id})^{w_{id}}}{\prod_{d=1}^{I}(n_{jd})^{w_{jd}}} = \frac{n_i}{n_j}\frac{\prod_{d=1}^{I}\left(\frac{1}{\tau_{id}^{k}}\right)^{w_{id}}}{\prod_{d=1}^{I}\left(\frac{1}{\tau_{jd}^{k}}\right)^{w_{jd}}}$$

Let $EM_{ih} = \prod_{d=1}^{I}(n_{id})^{w_{id}}$ and $EM_{jh} = \prod_{d=1}^{I}(n_{jd})^{w_{jd}}$. In addition, assume that the distances are used to represent the trade costs between the two countries, so that the average export distances of countries (i) and (j) are respectively defined to be $\tau_i = \prod_{d=1}^{I}(\tau_{id})^{w_{id}}$ and $\tau_j = \prod_{d=1}^{I}(\tau_{jd})^{w_{jd}}$. As a result, the above equation can be rewritten as follows:

$$\frac{EM_{ih}}{EM_{jh}} = \frac{n_i}{n_j}\left(\frac{\tau_i}{\tau_j}\right)^{-k}$$

Or

$$log\left(\frac{EM_{ih}}{EM_{jh}}\right) = log\left(\frac{n_i}{n_j}\right) - klog\left(\frac{\tau_i}{\tau_j}\right) \tag{13}$$

Equation (4) in the theoretical section suggests a positive relationship between $\frac{n_i}{n_j}$ and $\lambda\ \left(=\frac{Y_i}{Y_j}\right)$. We can express this relationship in a log linear form as follows:

$$log\left(\frac{n_{ih}}{n_{jh}}\right) = \beta_h log\left(\frac{Y_i}{Y_j}\right) + u_{ij} \tag{14}$$

From the theoretical section, we know that the industries with larger home market effects (larger ρ_h) will have larger β_h. This result means that $\rho_1 > \rho_2 > ... > \rho_h > ...$, then $\beta_1 > \beta_2 > ... > \beta_h > ...$, where, β_1, β_2, β_h denote coefficients of the above regression equation for industries 1, 2,...h. We have already shown that the industries with low trade costs, high fixed production costs, low fixed export costs, and high productivity dispersion will concentrate more in the large countries. This finding implies that we will have $\alpha_2 < 0$, $\alpha_3 > 0$, $\alpha_4 > 0$, and $\alpha_5 < 0$ in the following relationship:

$$\beta_h = \alpha_1 + \alpha_2\tau_h + \alpha_3 f_{dh} + \alpha_4 f_{hx} + \alpha_5 disp_h \tag{15}$$

Substituting equation (15) into the regression equation (14) yields:

$$log\left(\frac{n_{ih}}{n_{jh}}\right) = \alpha_0 + \alpha_1 log\left(\frac{Y_i}{Y_j}\right) + \alpha_2(\tau_h)log\left(\frac{Y_i}{Y_j}\right) + \alpha_3(f_{dh})log\left(\frac{Y_i}{Y_j}\right) + \alpha_4(f_{xh})log\left(\frac{Y_i}{Y_j}\right) + \alpha_5(disp_h)log\left(\frac{Y_i}{Y_j}\right) + u_{ijh} \tag{16}$$

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

Next, combining the equations (13) and (16), the regression model generates

$$
\begin{aligned}
log\left(\frac{EM_{ih}}{EM\,jh}\right) = \alpha_0 + \alpha_1 log\left(\frac{Y_i}{Y_j}\right) + \alpha_2(\tau_h) log\left(\frac{Y_i}{Y_j}\right) + \alpha_3(f_{dh}) log\left(\frac{Y_i}{Y_j}\right) + \alpha_4(f_{xh}) log\left(\frac{Y_i}{Y_j}\right) + \\
+ \alpha_5(disp_h) log\left(\frac{Y_i}{Y_j}\right) + \alpha_6 log\left(\frac{\tau_i}{\tau_j}\right) + u_{ijh}
\end{aligned}
\tag{17}
$$

We predict that $\alpha_1 > 0$, $\alpha_2 < 0$, $\alpha_3 > 0$, $\alpha_4 < 0$, $\alpha_5 > 0$, and $\alpha_6 < 0$. EM_{ih}, the average extensive margin of export of country i in industry h, is measured by the method used by Hummels and Klenow (2005), which is presented in the next section.

3.2 The extensive margins of exports

In studying the role of new varieties in the price indices, Feenstra (1994) showed how to use the expenditure data to measure the product-variety changes of each country across time. Many studies have adopted this method to compare the product or the export varieties across countries.[1]. Hummels and Klenow (2002) (or Hummels and Klenow (2005)) used this method to define the extensive and intensive margins of countries' exports and imports.[2]. In this study, we use their methods to measure the relative number of export products of two countries in each industry.

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

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

$EM_t^{i,exp}$ is the extensive margin of 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 the export of product s from the world to country j. $\sum_{s \in I_t^{ij}} X_t^{Wjs}$ is the total value of exports from the world into 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 the countries in year t with the weights being the total world trade in each category.

Hummels and Klenow (2005) use a similar approach but they calculate the extensive margin of the exporter at each destination. These authors subsequently determine an average value for all

[1] Feenstra et al. (1997); Feenstra and Kee (2004); Hummels and Klenow (2002); Hummels and Klenow (2005); and Feenstra and Kee (2008)

[2] Hummels and Klenow (2002) is a working paper, while Hummels and Klenow (2005) is a version of Hummels and Klenow (2002)- that was published in the AER. Hummels and Klenow (2002) measures the extensive and intensive margins of the exports of a country at all destinations, while Hummels and Klenow (2005) measures them at each destination, then takes the average value to represent the extensive margin of the exports of countries

destinations to calculate the extensive margin of exports for each country. In this case, the extensive margin of exports of country i at destination d is

$$EM_t^{id,exp} = \frac{\sum_{s \in I_t^{id}} X_t^{Wds}}{\sum_{s \in I_t^{Wd}} X_t^{Wds}} \tag{18}$$

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 export for country i:

$$EM_t^{i,exp} = \prod_{d \in M_{-i}} \left(EM_t^{id,exp} \right)^{w_{id}} \tag{19}$$

w_{id} are the weights, which are measured as follows:

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

In this equation, 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 the exports from country i to country d relative to the total exports 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 exports from the other countries (except for country i) to country d relative to the total exports of these countries. $s_{Wd} = \frac{\sum_{l \in M_{-i-d}} X^{ld}}{\sum_{l \in M_{-i-d}} X^{lW}}$.

As mentioned in Hummels and Klenow (2002), an alternative method can be used that defines the extensive margin as the number of product categories in which the country exports. However , the limitation of this simple-count method treats the products equally in calculating the extensive margin, while the method discussed above gives each category a weight equal to its share in world exports.

4 Data and empirical results

4.1 Data for the variables of the regression models

This paper examines how the characteristics of industries affect the distribution of firms across industries between a large country and a small country; therefore, the characteristics of an industry are assumed to be homogeneous across countries. We choose a sample of 28 industrial countries (Table (6) in the Appendix) with the assumption that the industry characteristics of these countries are similar. In addition, the 5-digit NAICS classification is used to classify the manufacturing industries in these countries. If the data on an industrial characteristic are available for all countries, we use the average value across these countries to represent the industrial characteristic (i.e., import tariff barriers). However, we cannot obtain most of data on the industrial characteristics of the countries, except for the U.S. Therefore, we use data on U.S. industrial characteristics to represent the industrial characteristics in the cases where we cannot obtain data from all countries. The U.S. is a large market, meaning that the firms (or products) in each industry are diverse. In addition, the technology and the techniques for industries in the U.S. are also typical of those in other industrial countries. Therefore, we believe that the industrial characteristics of the U.S. can typify those of the other industrial countries. All data used in this section are from 2002.

Dependent variable: The trade flow data at the HS6 level from CEPII[3] are used to measure the extensive margin of exports for a country according to formulas (18) and (19).

GDP: From the implication of the theoretical model, the GDP of countries is used to represent a country'size. GDP data (at 2000 constant prices) are from the World Development Indicator.

Variable trade costs (τ_h): we use the simple average tariffs (t) of high income countries to denote the trade costs of industries, while Hanson and Xiang (2004) use the freight rates of the US imports to represent the trade costs. The tariffs are chosen in our study because they can exactly represent our explanations about the impact of trade costs on the home market effect. That is, when trade barriers (equivalent to tariff barriers) between countries are decreased, the trade between two countries becomes more frequent and the products from the large country will gain an advantage over those from the small country due to economies of scale. Therefore, the industries with lower tariff barriers will have a higher home market effect. The data for tax rates are from the TRAINS database.

Fixed costs (f_{dh}, f_{xh}): f_{dh} are the fixed production costs. Previous studies use different proxies; for instance, Brainard (1997) uses the number of production workers in the median-sized plant as an industry measure of the plant-level fixed costs, while Helpman, Melitz, and Yeaple (2004) use the number of non-production workers per plant to represent a similar variable. Syverson (2004) use the average ratio of non-production workers to total employment across industry establishments as a proxy for the fixed production costs. In our paper, we follow Syverson (2004)' approach that use

[3] www.cepii.org

the average ratio of non-production worker to total employment per establishment to represent the fixed production costs. In addition, the average ratio of production worker is also used to test the robustness of the results. Data are from the Annual Survey of Manufacturers (2002).

It appears that no studies discuss the proxies for fixed export costs to date. Several studies, such as Helpman, Melitz, and Yeaple (2004), use the country-industry fixed effects as a proxy for fixed export costs. However, we believe that industries that meet high fixed barriers to enter markets need to advertise their products more. Therefore, we use the advertising intensity to represent the fixed export costs. The data for advertising costs are from the Annual Survey of Manufacturers (2002). We calculate the average advertising costs per establishment and then use the ratio of advertising costs and output per establishment to denote the advertising intensity of the industry.

Productivity dispersion ($disp_h$):

According to Helpman, Melitz, and Yeaple (2004), a Pareto distribution of productivity implies that a firms' sales also have the same distribution. This parameter can be measured by the standard deviation of the logarithm of firm sales, and these authors use this measurement 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 low). We use the firm sales obtained from the COMPUSTAT database to measure the productivity dispersion in this case.

As mentioned in the theoretical section, the effect of the elasticity of substitution on the firms' location depends on the relationship between f_D and f_X. An exact relationship between these two characteristics is hard to determine across industries. In addition, the measurement of productivity dispersion mentioned in Helpman, Melitz, and Yeaple (2004) may include the elasticity of substitution, we do not discuss the impact of the elasticity of substitution in this study.

4.2 Data analysis

The predictions of the theoretical model imply that the coefficient (β_{1h}) in the following regression for each industry h will be higher for industries that locate disproportionately in larger countries.

$$log\left(\frac{EM_{ih}}{EMjh}\right) = \beta_0 + \beta_{1h} log\left(\frac{Y_i}{Y_j}\right) + u_{ij} \tag{20}$$

This equation means that (β_1) should have a negative relationship with trade costs (tariff barriers) and fixed export costs (advertising costs) and a positive relationship with fixed production costs and productivity dispersion. First, we use graphs to visually summarize the relationships between the

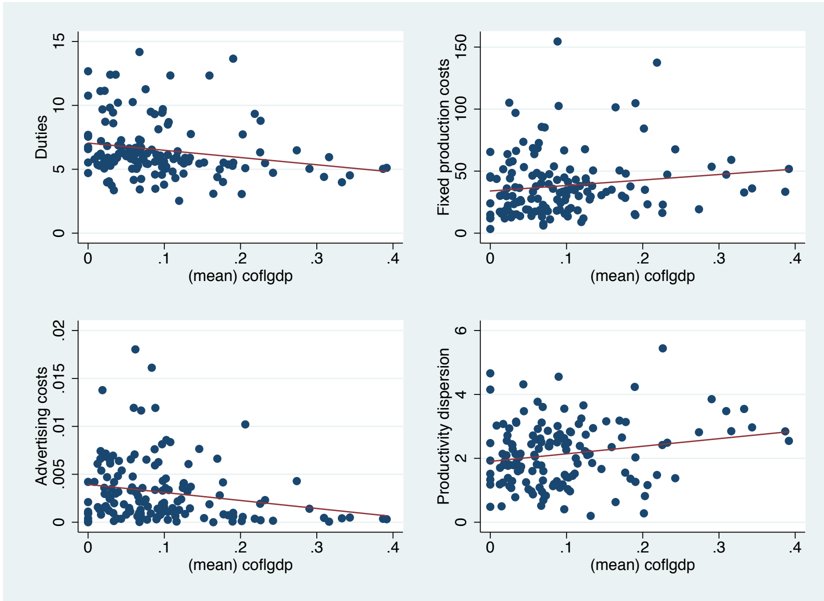

Figure 2: **The relationship of industrial characteristics and the home market effect**

industry characteristics and this coefficient. 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. The results shown in the figures appear to be consistent with the predictions from the theoretical model: the industries with low trade costs (or low tariff barriers), high production costs, low advertising costs, and high productivity dispersion tend to concentrate in the large countries. In brief, the impact of tariffs, fixed production costs, advertising costs, and productivity dispersion on the home-market effect of industries appears to have the predicted results.

4.3 Results of the regression model

There can be fixed effects at the country and the industry levels; therefore, the model (17) can be written as follows:

$$\begin{aligned} log\left(\frac{EM_{ih}}{EM\,jh}\right) &= \alpha_0 + \alpha_1 log\left(\frac{Y_i}{Y_j}\right) + \alpha_2(\tau_h) log\left(\frac{Y_i}{Y_j}\right) + \alpha_3(f_{dh}) log\left(\frac{Y_i}{Y_j}\right) + \alpha_4(f_{xh}) log\left(\frac{Y_i}{Y_j}\right) + \\ &+ \alpha_5(disp_h) log\left(\frac{Y_i}{Y_j}\right) + \alpha_6 log\left(\frac{\tau_i}{\tau_j}\right) + \mu_i + \mu_j + \eta_h + u_{ijh} \end{aligned} \tag{21}$$

where μ_i (or μ_j) are the fixed effects at the country levels, and η_h is the fixed effect at the industry level.

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

There can be two potential impacts of common groups in our model. First, when each country combines with other countries to form country pairs, the combinations of one country can be different from the combinations of another country. This situation can happen because the countries that are in economic unity, have a free agreement, or are in similar geographical locales can have similar features or characteristics that are different from other countries. Therefore, the combination of one country with other countries can be considered to be a cluster. Second, each industry is also considered to be a cluster because the countries could produce these industries for some similar reasons-for example, technology-intensive industries or high economic-value industries. As a result u_{ijh} (in model 21) can be decomposed into three parts: $u_{ijh} = v_{ij} + e_h + \varepsilon_{ijh}$, where v_{ij} is an unobserved (group) cluster effect at the country level, ($e_h \sim [0, \sigma_e^2]$) represents an unobserved cluster effect at the industry level in the estimation, and ε_{ijh} is the idiosyncratic error ($\varepsilon_{ijh} \sim [0, \sigma_\varepsilon^2]$).

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 converge to the true standard error. In addition, when the unobserved cluster effects are independent of the explanatory variables, the GLS estimator (the random-effects model) also provides an estimator, that is even more efficient than the cluster-corrected OLS estimator. If v_{ij} (or e_h) are correlated with the other dependent variables, the OLS estimator and the random-effect estimator are inconsistent. In this case, the fixed estimators should be used instead.

Table (2) presents the regression results when the country pairs are randomly chosen from all countries in the sample (i.e., not based on any pre-determined criteria). In general, the coefficients of different regression methods have predicted signs and high statistical significance across different regression methods. When the cluster-robust variance estimators for the country pairs and industries are used, there is a substantial change in the standard errors of the coefficients. All of the t-ratios become significantly smaller. For example, the t-ratio for $lgdp$ in the usual OLS is 38.92, whereas this value for the OLS with standard errors corrected for country groups and industry groups is, respectively, 15.42 and 6.17. These results suggest that ignoring the 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 random-effects models reject the null hypothesis that the random variation in the intercept is zero (the p-value is very low). This indicates that the random-effects estimations are an improvement over the OLS regression (Columns (7) and (8) compared with Columns (2) and (3)). However, we see that the t-values of the random-effects models have unremarkable change with these in the usual OLS estimator (Columns (7) and (8) compared to Columns (2) and (3)). We do not use the Hausman test to compare the random-effects and fixed-effects estimators because the fixed-effects estimators are controlled by both country and industry levels, while the random-effects estimators are controlled by either the country or the industry levels.

Instead of using the ratio of non-production workers to represent the fixed production costs, we also use the number of production workers to represent these costs. As a result, we still can get the results consistent with above discussions. The regression results are presented in Table (3).

In the above case, the country pairs are all built without any particular criteria from the sample countries. However, if we choose any two countries to build a pair, it can sometimes be difficult to find the common characteristics between the two countries. For example, we can observe the common features between the US and Canada, but not between Canada and Singapore. This difficulty implies that the comparison between the U.S. and Canada pair and the Canada and Singapore 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 regions: members of the European Union (19 countries), Canada and the US (the US-Canada Free Trade Agreement), New Zealand and Australia (the British Commonwealth), and Japan and Korea (a group of Asian countries) (Sample 2 of Table (6) in the Appendix). The country pairs are accordingly built within each region. The regression results are shown in Table (4)and are quite similar to the results from the previous case, and the cluster effects across the country pairs and the industries are still significant.

The above empirical results show that the impacts of the variables not only have the predicted signs but also have high statistical significance. As a result, these results confirm the predictions of the theoretical model that the industries with low trade costs (low tariff barriers), high fixed production costs, low advertising costs, and high productivity dispersion tend to concentrate in large countries.

Table 2: **The impact of industrial characteristics**

	OLS			Fixed effects			Random effects	
		Cluster level			Cluster level		Cluster level	
VARIABLES		Country	Industry		Country	Industry	Country	Industry
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LgGDP (+)	0.154***	0.154***	0.154***	0.134***	0.134***	0.134***	0.140***	0.146***
	(38.923)	(15.420)	(6.170)	(22.072)	(12.465)	(5.558)	(12.372)	(6.578)
Duties*LgGDP (-)	-0.006***	-0.006***	-0.006**	-0.005***	-0.005***	-0.005**	-0.006***	-0.005*
	(-12.563)	(-7.377)	(-2.071)	(-7.724)	(-3.537)	(-2.077)	(-7.498)	(-1.913)
Fixed operating costs*LgGDP (+)	2.345***	2.345***	2.345	2.318***	2.318***	2.318	2.647***	2.010
	(6.729)	(7.678)	(0.881)	(4.660)	(4.571)	(1.063)	(8.881)	(0.903)
Advertising costs*LgGDP (-)	-5.256***	-5.256***	-5.256***	-4.112***	-4.112***	-4.112**	-5.112***	-4.289***
	(-16.389)	(-12.688)	(-2.821)	(-8.986)	(-5.854)	(-2.564)	(-12.674)	(-2.632)
Productivity dispersion*LgGDP (+)	0.002***	0.002***	0.002**	0.002***	0.002***	0.002**	0.002***	0.002**
	(15.032)	(13.465)	(2.271)	(10.659)	(7.876)	(2.537)	(13.536)	(2.515)
Lg(distance) (-)	-0.135***	-0.135***	-0.135***	-0.009**	-0.009	-0.009	-0.035***	-0.137***
	(-65.122)	(-17.656)	(-12.629)	(-2.129)	(-1.280)	(-0.398)	(-6.586)	(-12.966)
Border	-0.008	-0.008	-0.008	-0.000	-0.000**	-0.000	-0.027	-0.008
	(-1.003)	(-0.324)	(-1.614)	(-0.021)	(-2.033)	(-1.001)	(-1.035)	(-1.512)
Lang	-0.025***	-0.025	-0.025***	0.000	0.000	0.000	-0.001	-0.025***
	(-4.273)	(-0.924)	(-4.505)	(0.000)	(0.030)	(0.116)	(-0.034)	(-4.465)
Constant	0.039***	0.039***	0.039***	-0.137***	-0.137***	-0.137***	0.051***	0.037***
	(12.004)	(2.629)	(10.981)	(-5.714)	(-12.607)	(-9.066)	(2.960)	(10.326)
Observations	51,976	51,976	51,976	51,976	51,976	51,976	51,976	51,976
R-squared	0.148	0.148	0.148	0.357	0.357	0.357		
Breusch-Pagan test(p-value)							0	0
Number of industries								138
Number of countrypairs							378	

t-statistics in parentheses

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

Country pairs are chosen from all countries in the sample without being based on any pre determined criteria
This case uses the ratio of non-production workers to show the fixed production costs
Columns (1)-(3) are the OLS estimators
Columns (4)-(6) are the fixed-effects estimators at country and industry levels
Columns (7)-(8) are the random-effects estimators

Table 3: **The impact of industrial characteristics**

	OLS			Fixed effects			Random effects	
		Cluster level			Cluster level		Cluster level	
VARIABLES		Country	Industry		Country	Industry	Country	Industry
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LgGDP (+)	0.157***	0.157***	0.157***	0.133***	0.133***	0.133***	0.144***	0.140***
	(34.528)	(15.087)	(5.473)	(19.471)	(11.228)	(5.085)	(12.225)	(5.590)
Duties*LgGDP (-)	-0.008***	-0.008***	-0.008**	-0.007***	-0.007***	-0.007**	-0.009***	-0.006**
	(-18.101)	(-10.037)	(-2.490)	(-10.474)	(-4.744)	(-2.250)	(-10.637)	(-2.120)
Fixed operating costs*LgGDP (+)	1.375***	1.375***	1.375	0.364	0.364	0.364	1.386***	0.457
	(6.382)	(4.405)	(0.801)	(1.203)	(0.695)	(0.255)	(4.516)	(0.325)
Advertising costs*LgGDP (-)	-6.368***	-6.368***	-6.368***	-5.316***	-5.316***	-5.316***	-6.198***	-5.470***
	(-18.799)	(-16.759)	(-2.942)	(-11.162)	(-8.526)	(-2.776)	(-16.775)	(-2.844)
Productivity dispersion*LgGDP (+)	0.014***	0.014***	0.014	0.017***	0.017***	0.017**	0.014***	0.017**
	(11.599)	(8.232)	(1.459)	(10.656)	(5.684)	(2.264)	(8.599)	(2.139)
Lg(distance)(-)	-0.134***	-0.134***	-0.134***	-0.002	-0.002	-0.002	-0.035***	-0.134***
	(-63.408)	(-18.546)	(-11.613)	(-0.410)	(-0.259)	(-0.087)	(-6.832)	(-11.854)
Border	-0.011	-0.011	-0.011**	-0.000	-0.000*	-0.000	-0.029	-0.010*
	(-1.292)	(-0.440)	(-1.995)	(-0.028)	(-1.835)	(-1.008)	(-1.141)	(-1.924)
Lang	-0.022***	-0.022	-0.022***	-0.000	-0.000	-0.000	0.002	-0.021***
	(-3.550)	(-0.864)	(-3.362)	(-0.003)	(-0.123)	(-0.558)	(0.075)	(-3.286)
Constant	0.036***	0.036**	0.036***	0.117***	0.117***	0.117***	0.049***	0.034***
	(10.691)	(2.512)	(9.429)	(4.730)	(3.896)	(3.219)	(2.864)	(8.463)
Observations	45,928	45,928	45,928	45,928	45,928	45,928	45,928	45,928
R-squared	0.161	0.161	0.161	0.384	0.384	0.384		
Breusch-Pagan test(p-value)							0	0
Number of industries								122
Number of countrypairs							378	

t-statistics in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Country pairs are chosen from all countries in the sample without being based on any pre determined criteria
This case uses the ratio of production workers to show the fixed production costs and Helpman et al. (2004)'s method to measure the productivity dispersion
Columns (1)-(3) are the OLS estimators
Columns (4)-(6) are the fixed-effects estimators at country and industry levels
Columns (7)-(8) are the random-effects estimators

Table 4: **The impact of industrial characteristics**

	OLS			Fixed effects			Random effects	
		Cluster level			Cluster level		Cluster level	
VARIABLES		Country	Industry		Country	Industry	Country	Industry
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LgGDP (+)	0.180***	0.180***	0.180***	0.126***	0.126***	0.126***	0.169***	0.180***
	(30.318)	(14.190)	(5.740)	(9.637)	(12.593)	(3.820)	(12.413)	(5.606)
Duties*LgGDP (-)	-0.005***	-0.005***	-0.005	-0.005***	-0.005***	-0.005	-0.005***	-0.005
	(-6.810)	(-3.522)	(-1.503)	(-7.480)	(-3.669)	(-1.468)	(-3.708)	(-1.505)
Fixed operating costs*LgGDP (+)	2.382***	2.382***	2.382	3.663***	3.663***	3.663	3.228***	2.599
	(4.266)	(4.732)	(0.759)	(7.121)	(7.338)	(1.094)	(6.235)	(0.801)
Advertising costs*LgGDP (-)	-3.354***	-3.354***	-3.354	-3.506***	-3.506***	-3.506	-3.356***	-3.452
	(-6.535)	(-7.101)	(-1.474)	(-7.431)	(-8.056)	(-1.445)	(-7.269)	(-1.468)
Productivity dispersion*LgGDP (+)	0.003***	0.003***	0.003**	0.003***	0.003***	0.003**	0.003***	0.003**
	(13.747)	(12.415)	(2.503)	(16.583)	(13.710)	(2.582)	(13.211)	(2.536)
Lg(distance)(-)	-0.153***	-0.153***	-0.153***	0.011	0.011	0.011	-0.003	-0.158***
	(-29.473)	(-11.032)	(-7.145)	(1.560)	(0.918)	(0.331)	(-0.263)	(-7.249)
(mean) border	0.033***	0.033	0.033***	-0.000	-0.000	-0.000	0.052	0.033***
	(3.796)	(1.046)	(5.100)	(-0.004)	(-0.335)	(-0.939)	(1.583)	(4.991)
(mean) lang	-0.031***	-0.031	-0.031***	-0.000	-0.000	-0.000	-0.023	-0.032***
	(-2.854)	(-0.997)	(-4.638)	(-0.007)	(-0.504)	(-0.673)	(-0.654)	(-4.634)
Constant	-0.045***	-0.045***	-0.045***	0.127***	0.127***	0.127***	-0.076***	-0.045***
	(-15.718)	(-3.687)	(-7.378)	(3.377)	(6.922)	(4.887)	(-5.144)	(-7.142)
Observations	23,918	23,918	23,918	23,918	23,918	23,918	23,918	23,918
R-squared	0.310	0.310	0.310	0.430	0.430	0.430		
Breusch-Pagan test(p-value)							0	0
Number of industries								138
Number of countrypairs							174	

t-statistics in parentheses

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

Country pairs are chosen from the similar regions

Columns (1)-(3) are the OLS estimators

Columns (4)-(6) are the fixed-effects estimators at country and industry levels

Columns (7)-(8) are the random-effects estimators

4.4 Robustness check

The above empirical results use the extensive margin of exports to test the predictions of the theoretical model. However, the model also implies that we can use the number of firms to test the predictions according to the regression model (16). In this section, we use the data on the number of establishments from UNIDO to test the robustness of some of the model results. The UNIDO industrial database provides data on the number of establishments. However, these data are only available for a limited number of countries and industries, unlike the trade data. From this database, we choose the 21 OECD countries (Table (6) in the Appendix) with the most available data. After combining the industrial data, only about 90 among 124 ISIC manufacturing industries are available. We use the average tariff to show trade costs, the ratio of non-production worker to show the fixed production costs, and the advertising intensity to show the fixed export costs. The regression results across the different regressions are presented in Table (5).

The signs of the explanatory variables are still consistent with our predictions across the different estimation methods. However, the statistical significance of the explanatory variables is quite similar in comparison with the case of the trade data. In addition, most of the t-values are decrease remarkably when the cluster effects are controlled. These changes in the t-values shows the evidence of the existence of the cluster impacts at the country-group levels and the industrial levels.

Table 5: **The impact of industrial characteristics with UNIDO data**

	OLS			Fixed effects			Random effects	
		Cluster level			Cluster level		Cluster level	
VARIABLES		Country	Industry		Country	Industry	Country	Industry
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LgGDP (+)	0.634***	0.634***	0.634***	0.583***	0.583***	0.583***	0.661***	0.597***
	(23.292)	(8.924)	(10.412)	(16.649)	(21.916)	(7.791)	(8.728)	(9.601)
Duties*LgGDP (-)	-0.009***	-0.009***	-0.009*	-0.001	-0.001	-0.001	-0.006**	-0.006
	(-3.434)	(-3.311)	(-1.836)	(-0.467)	(-0.325)	(-0.265)	(-2.177)	(-1.137)
Fixed operating costs*LgGDP (+)	4.810***	4.810***	4.810	7.394***	7.394***	7.394**	4.247***	7.434**
	(3.550)	(3.246)	(1.508)	(4.631)	(3.740)	(2.257)	(3.525)	(2.411)
Advertising costs*LgGDP (-)	-5.055**	-5.055***	-5.055	-3.246	-3.246	-3.246	-5.520***	-4.713
	(-2.297)	(-2.749)	(-1.080)	(-1.233)	(-1.007)	(-0.706)	(-2.974)	(-0.956)
Productivity dispersion*LgGDP (+)	0.004***	0.004***	0.004	0.003**	0.003**	0.003	0.004***	0.004
	(3.508)	(3.509)	(1.365)	(2.471)	(2.126)	(1.155)	(4.299)	(1.224)
Border	-0.471***	-0.471	-0.471***	0.012	0.012	0.012	-0.413	-0.467***
	(-6.829)	(-1.301)	(-9.022)	(0.212)	(0.705)	(0.630)	(-1.047)	(-8.981)
Lang	0.372***	0.372	0.372***	-0.000	-0.000	-0.000	0.447**	0.368***
	(6.916)	(1.599)	(8.969)	(-0.005)	(-0.012)	(-0.010)	(2.067)	(8.929)
Constant	-0.153***	-0.153	-0.153***	-0.303**	-0.303***	-0.303***	-0.243	-0.148***
	(-5.820)	(-1.078)	(-6.597)	(-2.377)	(-2.868)	(-3.258)	(-1.621)	(-6.249)
Observations	11,606	11,606	11,606	11,606	11,606	11,606	11,606	11,606
R-squared	0.156	0.156	0.156	0.607	0.607	0.607		
Breusch-Pagan test(p-value)							0	0
Number of industries								90
Number of countrypairs							210	

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The dependent variable is the number of establishments from UNIDO database
Columns (1)-(3) are the OLS estimators
Columns (4)-(6) are the fixed-effects estimators at country and industry levels
Columns (7)-(8) are the random-effects estimators

5 Conclusion

This paper studies the conditions to occur the home market effect and how the magnitude of the home market effect varies with industry characteristics in the framework of heterogeneous firms. Our model indicates that aside from trade costs, the presence of fixed costs also causes the home market effect. Once fixed trade costs of industries are very low, the model implies that the reverse home market effect can occur. These results are different from the one of the homogeneous firm model which implies that only variable trade costs are necessary condition to exist the home market effect and the home market effect are always presence in the sector of increasing return. In addition, the model predicts that industries with low trade costs, high fixed production costs, low fixed export costs, and high productivity dispersion tend to concentrate in large countries; in other words, the home-market effect of these industries will be higher than that of the other industries. In addition, the impact of trade barriers on the home market effect are dampened by industries with high elasticity of substitution. This result is also contrary with the one from the model of homogeneous firms which says that this impact is magnified with the elasticity of substitution.

An empirical model is then developed to examine how the home market effect change across industries. We use the data from 5-digit NAICS manufacturing industries in 28 high income countries. Our empirical evidence supports the predictions from the theoretical model.

According the model's data, we can see that the industries such as basic chemicals, or basic iron and steel tend to locate in large countries, while the industries such as furniture or electronics, tend to locate in both large and small countries. In general, small countries should promote industries with a low home market effect, such as the furniture industry. If small countries develop industries such basic steel, they might not be able to compete with the large countries in terms of production costs.

References

R.E. Baldwin and T. Okubo. Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting. *Journal of Economic Geography*, 6(3):323–346, 2006.

K. Behrens and P.M. Picard. Welfare, home market effects, and horizontal foreign direct investment. *Canadian Journal of Economics/Revue canadienne d'économique*, 40(4):1118–1148, 2007.

S.L. Brainard. An Empirical Assessment of the Proximity-Concentration Trade-off Between Multinational Sales and Trade. *The American Economic Review*, 87(4):520–544, 1997.

A.C. Cameron and P.K. Trivedi. *Microeconometrics: methods and applications*. Cambridge Univ Pr, 2005. ISBN 0521848059.

T. Chaney. Distorted Gravity: The Intensive and Extensive Margins of International Trade.. *American Economic Review*, 2008.

D.R. Davis and D.E. Weinstein. Economic geography and regional production structure: an empirical investigation. *European economic review*, 43(2):379–407, 1999.

D.R. Davis and D.E. Weinstein. Market access, economic geography and comparative advantage: An empirical assessment. *Journal of International Economics*, 59(1):1–24, 2003.

R. Feenstra and H.L. Kee. On the measurement of product variety in trade. *American Economic Review*, pages 145–149, 2004.

R. Feenstra and H.L. Kee. Export variety and country productivity: estimating the monopolistic competition model with endogenous productivity. *Journal of International Economics*, 74(2):500–518, 2008.

R.C. Feenstra. New product varieties and the measurement of international prices. *The American Economic Review*, 84(1):157–177, 1994.

R.C. Feenstra, 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.

R.C. Feenstra, J.R. Markusen, and A.K. Rose. Using the gravity equation to differentiate among alternative theories of trade. *Canadian Journal of Economics/Revue canadienne d'économique*, 34 (2):430–447, 2003.

G.H. Hanson and C. Xiang. The home-market effect and bilateral trade patterns. *American Economic Review*, pages 1108–1129, 2004.

K. Head and J. Ries. Increasing returns versus national product differentiation as an explanation for the pattern of us-canada trade. *American Economic Review*, pages 858–876, 2001.

E. Helpman and P.R. Krugman. *Market structure and foreign trade*. MIT press Cambridge, Mass, 1985.

E. Helpman, M.J. Melitz, and S.R. Yeaple. Export Versus FDI with Heterogeneous Firms. *American Economic Review*, 94(1):300–316, 2004.

T.J. Holmes and J.J. Stevens. Does home market size matter for the pattern of trade? *Journal of International Economics*, 65(2):489–505, 2005.

D. Hummels and P.J. Klenow. The variety and quality of a nation's exports. *American Economic Review*, pages 704–723, 2005.

D.L. Hummels and P.J. Klenow. The variety and Quality of a Nation's Trade. *NBER Working Paper*, 2002.

P. Krugman. Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review*, 70(5):950–959, 1980.

B.R. Moulton. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The Review of Economics and Statistics*, 72(2):334–338, 1990. ISSN 0034-6535.

T. Okubo, P.M. Picard, and J.F. Thisse. The spatial selection of heterogeneous firms. *Journal of international economics*, 82(2):230–237, 2010.

C. Syverson. Product substitutability and productivity dispersion. *Review of Economics and Statistics*, 86(2):534–550, 2004.

Appendix

A Samples

Table 6: **Sample of countries**

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		Singapore
9	AUT	Austria	4	Austria
10	BEL	Belgium	4	
11	CZE	Czech Rep.	4	Czech Rep.
12	DNK	Denmark	4	Denmark
13	FIN	Finland	4	Finland
14	FRA	France	4	
15	DEU	Germany	4	Germany
16	GRC	Greece	4	Greece
17	HUN	Hungary	4	Hungary
18	IRL	Ireland	4	Ireland
19	ISR	Israel		Israel
20	ITA	Italy	4	
21	NLD	Netherlands	4	Netherlands
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

B Proof of $B_i = B_j = B$

Through the zero profit equations, we can determine the cutoff level of productivity for firms of country i in the domestic market and the export market as follows:

$$\begin{aligned} \pi_{ii} = 0 \Rightarrow a_{ii}^{1-\sigma} &= \frac{f_d}{A_i} \\ \pi_{ij} = 0 \Rightarrow a_{ij}^{1-\sigma} &= \frac{f_x}{A_j} \frac{1}{\tau_{ij}^{1-\sigma}} \end{aligned} \tag{22}$$

Similarly, we have the cutoff levels for country j

$$\begin{aligned} a_{jj}^{1-\sigma} &= \frac{f_d}{A_j} \\ a_{ji}^{1-\sigma} &= \frac{f_x}{A_i}\frac{1}{\tau_{ji}^{1-\sigma}} \end{aligned} \tag{23}$$

From these equations, we have

$$\begin{aligned} \left(\frac{a_{ii}}{a_{ji}}\right)^{1-\sigma} &= \frac{f_d}{f_x}\tau_{ji}^{1-\sigma} \Rightarrow \frac{a_{ii}}{a_{ji}} = \left(\frac{f_d}{f_x}\right)^{\frac{1}{1-\sigma}} \tau_{ji} \\ \left(\frac{a_{jj}}{a_{ij}}\right)^{1-\sigma} &= \frac{f_d}{f_x}\tau_{ij}^{1-\sigma} \Rightarrow \frac{a_{jj}}{a_{ij}} = \left(\frac{f_d}{f_x}\right)^{\frac{1}{1-\sigma}} \tau_{ij} \end{aligned} \tag{24}$$

If we call the f_e is the entry cost in country i, the condition of free entry is

$$\begin{aligned} \int_0^{a_{ii}} (p_{ii}(v)x_{ii}(v) - f_d)dG(a) + \int_0^{a_{ij}} (p_{ij}(v)x_{ij}(v) - f_x)dG(a) &= f_e \\ \left(\frac{\sigma-1}{k-\sigma+1}\right)(f_d a_{ii}^k + f_x a_{ij}^k) &= f_e \end{aligned} \tag{25}$$

A similar equation is withdrawn for country j

$$\left(\frac{\sigma-1}{k-\sigma+1}\right)(f_d a_{jj}^k + f_x a_{ji}^k) = f_e \tag{26}$$

As a result, we have

$$\begin{aligned} (f_d a_{ii}^k + f_x a_{ij}^k) &= (f_d a_{jj}^k + f_x a_{ji}^k) \\ (a_{ji}^k f_d \left(\frac{f_d}{f_x}\right)^{\frac{k}{1-\sigma}} \tau_{ji}^k + f_x a_{ij}^k) &= (f_d \left(\frac{f_d}{f_x}\right)^{\frac{k}{1-\sigma}} \tau_{ij}^k a_{ij}^k + f_x a_{ji}^k) \\ a_{ji}^k \left(f_d \left(\frac{f_d}{f_x}\right)^{\frac{k}{1-\sigma}} \tau_{ji}^k - f_x\right) &= a_{ij}^k \left(f_d \left(\frac{f_d}{f_x}\right)^{\frac{k}{1-\sigma}} \tau_{ij}^k - f_x\right) \end{aligned}$$

Assume that $\tau_{ij} = \tau_{ji}$; therefore, $a_{ij} = a_{ji}$. Similarly, we have $a_{ii} = a_{jj}$