DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 10-10

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

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

[1] Some 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.

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 model[2], 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.

[2]This 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 heterogenous 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 b_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

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

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 $s_h = \frac{1}{1-a_h} > 1$ is the constant elasticity of substitution across varieties in industry h with $a_h > 0$. 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:

$$x_h^i(v) = \frac{b_h Y_i p_h(v)^{-s_h}}{P_h^{i\,1-s_h}}$$

Where $P_h^i = \left(\int_0^{n_h^i} p_h(v)^{1-s_h} dv\right)^{\frac{1}{1-s_h}}$ 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.

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, $t_{ji}^h = t_{ij}^h = t^h$.

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 $\frac{1}{a_D^{ih}}$

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(.) 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 $a_D^{ih} = a_D^{jh} = a_D$ 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*.

$$\int_0^{n_i^e} p(v)^{1\ s} dv = n_i \int_0^{a_D} (\frac{a}{a})^{1\ s} dG(a) + n_j \int_0^{a_X} (t\frac{a}{a})^{1\ s} dG(a)$$
$$= n_i(\frac{1}{a})^{1\ s} V(a_D) + n_j t^{1\ s} (\frac{1}{a})^{1\ s} V(a_X) \quad (1)$$

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:

$$n_i V(a_D) + n_j t^{1\ s} V(a_X) = \frac{(1\ \ a)bY_i}{B} \quad (2)$$

Similarly for country j

$$n_j V(a_D) + n_i t^{1\ s} V(a_X) = \frac{(1\ \ a)bY_j}{B} \quad (3)$$

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

$$\frac{n_i}{n_j} = \frac{\frac{Y_i}{Y_j}\ \ \frac{t^{1\ s}V(a_X)}{V(a_D)}}{1\ \ \frac{t^{1\ s}V(a_X)}{V(a_D)}\frac{Y_i}{Y_j}} = \frac{I\ \ r}{1\ \ rI} \quad (4)$$

Where $I = \frac{Y_i}{}$

$_j$ I= $\ s\ \ V(a_D)$

$$\frac{V(a_D)}{V(a_X)} = \quad \frac{f_X}{f_D} t^{s \quad 1} \quad ^{\frac{k \;\; s \;\; 1}{s \;\; 1}}$$

As a result,

$$r = \frac{1}{t^{s \quad 1}} \quad \frac{f_D}{f_X t^{s \quad 1}} \quad ^{\frac{k \;\; s+1}{s \;\; 1}} \quad < 1$$

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

From equation (4), we have:

$$\frac{\P \quad n_i = n_j}{\P \, l} = \frac{1 \quad r^2}{(1 \quad l \, r)^2} > 0 \qquad (5)$$

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 l is larger than 1 ($l > 1$), it can be shown that $\frac{1 \quad r^2}{(1 \quad l \, r)^2} > 1$, indicating that the larger market attracts a disproportionate share of firms in industry h (the home market effects). The coefficient $\frac{1 \quad r^2}{(1 \quad l \, r)^2}$ shows the level of difference in the number of products of an industry h between the large country and the small country. Let $g(r) = \frac{1 \quad r^2}{(1 \quad l \, r)^2}$, we have additionally the following result:

$$\frac{\P g}{\P r} = \frac{2(l \quad r)(1 \quad r \, l)}{(1 \quad r \, l)^4} > 0 \qquad (6)$$

Equation (6) indicates that the coefficient $\frac{1 \quad r^2}{(1 \quad l \, r)^2}$ is not uniform .393 Td [2.9664 Tf 11.2255 Td [(9

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

$$\frac{\partial r}{\partial \tau} = \quad \frac{f_d}{f_x}^{\frac{k \quad \sigma+1}{\sigma \quad 1}} \quad \frac{k}{\tau^{k+1}} \quad < 0 \tag{7}$$

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 r with respect to fixed domestic costs and fixed export costs yield:

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

An increase in the fixed domestic costs leads to a higher value of r, while the increase of fixed export costs makes r 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 r with respect to the productivity dispersion and the elasticity of substitution yield:

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

Since we assume that only some firms with high productivity can export to foreign markets, this implies that $f_d < f_x t^{s-1}$ and hence $\frac{\partial r}{\partial k} < 0$. The negative correlation between r 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 ($f_d < f_x$), $\frac{\partial r}{\partial s} > 0$ implies that industries with high elasticity of substitution (high s) will locate more in the large country. If fixed domestic costs are larger than fixed export costs ($f_d > f_x$), $\frac{\partial r}{\partial s} < 0$ implies industries with low elasticity of substitution (low s) 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

Hanson and Xiang (2004).

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:

$$log \frac{n_{ih}}{n_{jh}} = a_0 + a_1 log \frac{Y_i}{Y_j} + a_2(t_h) log \frac{Y_i}{Y_j} + a_3(disp_h) log \frac{Y_i}{Y_j} + a_4(f$$

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:

$$EM^{i;exp}$$

export for country *i*:

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

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 CEPII[5] 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'size. GDP Data (at constant prices of 2000) is from the World Development Indicator.

Variable trade costs (t_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 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 $k \quad s + 1$. 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 ($k \quad s + 1$ 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 (s high) (low $k \quad s + 1$) will locate more often in a large country. Since $k \quad s + 1$ 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 b_1 in the following difference regression:

log

manufacturing industries. However, due to the limited availability of export data, we only estimate the coefficient (b_1) for 118 industries.[6]

The predictions of the theoretical model imply that this coefficient (b_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 (b_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 (b_1), and two outliers in the relationship between the fixed domestic costs and coefficients[7]. 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:

$$b_h = a_1 + a_2 t_h + a_3 disp_h + a_4 f_{dh} + a_5 f_{xh}$$

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.

[6]The 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

[7]2109: 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**				-0.117***
	(0.021)				(0.000)
Fixed domestic costs (+)		0.005***			0.005***
		(0.000)			(0.000)
Productivity dispersion (+)			0.033		0.036*
			(0.218)		(0.080)
Trade costs (-)				-0.008***	-0.008***
				(0.000)	(0.000)
Constant	0.217***	0.123***	0.127***	0.298***	0.218***
	(0.000)	(0.000)	(0.006)	(0.000)	(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: $u_{ijh} = n_{ij} + e_{ijh}$, where n_{ij} is the country pair-level fixed effects or an unobserved (group) cluster effect ($n_{ij}\ [0; s_n^2]$) and e_{ijh} is the idiosyncratic error ($e_{ijh}\ [0; s_e^2]$). 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 decom-

$$log\left(\frac{EM_{ih}}{EMjh}\right) = a_0 + a_1 log\left(\frac{Y_i}{Y_j}\right) + a_2(t_h) log\left(\frac{Y_i}{Y_j}\right) + a_3(disp_h) log\left(\frac{Y_i}{Y_j}\right) + a_4(f_{dh}) log\left(\frac{Y_i}{Y_j}\right) + \\ + a_5(f_{xh}) log\left(\frac{Y_i}{Y_j}\right) + n_{ij} + e_h + e_{ijh} \quad (16)$$

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 tech-

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 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**

	OLS	Random effects	Fixed effects

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

	OLS					Random effects		Fixed effects	
VARIABLES	Usual	Het	Country	Industry	Both	Country	Industries	Country	industry
	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(10)	(11)
LGDP (+)	0.24***	0.24***	0.24***	0.24***	0.24***	0.24***	0.25***		0.25***
	(19.44)	(23.06)	(11.93)	(4.54)	(4.34)	(14.34)	(19.72)		(19.79)
Duties*LGDP (-)	-0.67***	-0.67***	-0.67***	-0.67***	-0.67***	-0.67***	-0.69***	-0.67***	-0.69***
	(-16.61)	(-16.65)	(-8.52)	(-3.26)	(-3.11)	(-18.98)	(-16.80)	(-18.98)	(-16.84)
Fixed domestic cost*LGDP (+)	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***
	(28.10)	(22.88)	(16.67)	(6.70)	(6.49)	(32.11)	(26.93)	(32.11)	(26.33)
Firm scale*LGDP (-)	-0.13***	-0.13***	-0.13***	-0.13***	-0.13***	-0.13***	-0.13***	-0.13***	-0.13***
	(-17.47)	(-21.79)	(-24.50)	(-4.62)	(-4.66)	(-19.96)	(-17.16)	(-19.96)	(-16.97)
Productivity dispersion*LGDP (+)	0.04***	0.04***	0.04***	0.04*	0.04*	0.04***	0.04***	0.04***	0.04***
	(7.44)	(8.51)	(12.28)	(1.82)	(1.85)	(8.50)	(7.25)	(8.50)	(7.15)
Constant	-0.02***	-0.02***	-0.02	-0.02**	-0.02	-0.02	-0.02***	-0.10***	-0.02***

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

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

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

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

	OLS					Random effects		Fixed effects	
VARIABLES	Usual	Het	Country	Industry	Both	Country	Industries	Country	industry
	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(10)	(11)
LGDP (+)	0.17***	0.17***	0.17***	0.17***	0.17***	0.17***	0.16***		0.16***
	(27.89)	(28.57)	(11.87)	(6.52)	(5.86)	(13.42)	(17.83)		(17.15)
Duties*LGDP (-)	-0.53***	-0.53***	-0.53***	-0.53***	-0.53***	-0.53***	-0.49***	-0.53***	-0.48***
	(-27.53)	(-25.63)	(-17.95)	(-5.41)	(-5.32)	(-31.23)	(-15.98)	(-31.23)	(-15.24)
Fixed domestic cost*lGDP (+)	0.00***	0.00***	0.00***	0.00**	0.00**	0.00***	0.00***	0.00***	0.00***
	(15.19)	(10.20)	(11.35)	(2.01)	(2.02)	(17.23)	(8.96)	(17.23)	(8.56)
Firm scale*lGDP (-)	-0.29***	-0.29***	-0.29***	-0.29**	-0.29**	-0.29***	-0.28***	-0.29***	-0.28***
	(-10.72)	(-12.34)	(-10.60)	(-2.37)	(-2.37)	(-12.17)	(-6.49)	(-12.17)	(-6.23)
Productivity dispersion*lGDP (+)	0.03***	0.03***	0.03***	0.03	0.03	0.03***	0.03***	0.03***	0.03***
	(6.81)	(6.84)	(10.44)	(1.41)	(1.43)	(7.73)	(4.47)	(7.73)	(4.33)
Constant	0.03***	0.03***	0.03	0.03***	0.03	0.03	0.03**	0.26***	0.03***
	(7.44)	(7.23)	(1.43)	(7.13)	(1.43)	(1.50)	(2.52)	(36.98)	(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 doesnot examine directly the effect of the substitution 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

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,

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		

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