Theoretical Economics Letters, 2022, 12, 1673-1692
https://www.scirp.org/journal/tel
ISSN Online: 2162-2086
ISSN Print: 2162-2078

The Role of Intra-National Trade in Trade Theory

Yong He[1], Guangzhen Sun[2]

[1]CERDI/CNRS, Universite Clermont-Auvergne, Clermont-Ferrand, France
[2]Department of Economics, University of Macau, Macao, China
Email: yong.he@uca.fr, GZSun@umac.mo

How to cite this paper: He, Y., & Sun, G. Z. (2022). The Role of Intra-National Trade in Trade Theory. *Theoretical Economics Letters, 12,* 1673-1692.
https://doi.org/10.4236/tel.2022.126092

Received: October 24, 2022
Accepted: December 24, 2022
Published: December 27, 2022

Abstract

This paper explores the implications arising from the heterogeneous firm model for a theory of trade in which intra-national trade is an intrinsic part, and international trade is its extension. Due to market competition, intra-national trade dynamically leads to an increase in productivity heterogeneity across firms, which strengthens market selection and makes allocation more efficient. It also strengthens comparative advantage, and thus stimulates international trade.

Keywords

Intra-National and International Trade, Heterogeneous Firm Model, Market Selection, Comparative Advantage

1. Introduction

The need to theorize about intra-national trade arises from certain salient facts. Chronologically, intra-national trade is the starting point for international trade. The latter is often preceded by sufficient development of intra-national trade in which a number of firms have become the champions in terms of productivity and market share. Intra-national trade differs from international trade in that it more often has transaction cost advantages due to the shorter geographic distance and similarity in language and culture. The international economy also depends on the number of countries with significant development of intra-national trade.

Ohlin (1933) has treated international trade as a special case of a more general interregional trade. To extend Ohlin's view, intra-national trade must be an intrinsic part of a comprehensive theory of trade. However, he did not analyze the nature of intra-national trade purposely. An adequate theory of intra-national trade is a necessity because it has its own driving force and allocative function, and deserves to be studied independently first, and then in the context of its relationship to international trade. If both intra-national trade and international trade contribute to increasing welfare, we still need to assess their respective contributions and, above all, to answer the following question: to what extent does taking intra-national trade into account improve our understanding of international trade?

Intra-national trade itself is still an underdeveloped area of research with only a small amount of work done. Krugman (2015) estimated that the driving force of intra-national trade, like international trade, is either comparative advantage or increasing returns and agglomeration. Based on the trade theory of monopolistic competition developed by Krugman (1979) (referred to below as the homogeneous firm model), Fujita et al. (1999: chapter 5) showed how, under the effect of increasing returns and agglomeration, the core-periphery pattern, which also applies to the intra-national market, becomes unsustainable, leading to changes in factor mobility and relative wages across regions. Combes et al. (2008) examined the adequacy of the homogeneous firm model to endogenous regional disparity for developed countries. Venables (2005) applied this model to newly industrialized countries, and analyzed how the presence of agglomeration forces leads to intra-national clustering of activities. Intra-national trade has also been addressed from the perspective of the need for improved internal market integration (Poncet, 2005), and has been shown to have a function of technology spillovers from advanced to less developed regions (He, 2017).

The main drawback of these studies is the lack of a formal framework to establish the link between intra- and international trade and, more importantly, to explain how intra-national trade affects international trade. There is no formal study focusing on the intra-national cause of international trade. Existing trade theories look for the cause of trade from an international perspective: either the increasing returns prevailing in the world market or the comparative advantage existing between countries. As will be seen below, in a conventional productivity-based Ricardian model of comparative advantage, intra-national trade is only a "black box" because of the absence of economies of scale.[1] In the model of homogeneous firms, since heterogeneity in productivity across firms is not allowed, intra-national trade cannot be adequately modeled.

Melitz (2003) constructed a trade model in which increasing returns to variety lead some low-productivity firms to survive, while high-productivity firms enjoy increasing returns to scale. Thus, welfare is also determined by the degree of heterogeneity in productivity. The new point put forward in our paper is to show that the Melitz model gives rise to the possibility of modeling intra-national trade, and to establish the link between intra- and international trade through an endogenous change in productivity heterogeneity.

[1]The Heckscher-Ohlin model based on differences in factor endowments across countries can be viewed as a productivity-based model, because intensive use of certain factors of production with abundant endowments leads to higher productivity.

This approach is complementary to existing trade theories in that it provides a new perspective on the intra-national determinants of international trade. It implies that a theory of trade that does not show how intra-national trade operates and affects international trade is, at least, incomplete. Trade relations between countries are also based on their own internal trade conditions. To put it simplistically, existing trade theories have generally shown that openness is an effective policy choice for a developing country to benefit from international trade. Intra-national trade theory, on the other hand, suggests that before opening up immediately, that country should probably first look at whether it has done enough to develop an intra-national trade network.

This paper is organized as follows: after this introduction, section II presents a framework based on a quantized Melitz model that allows us to derive reduced-form solutions for the analysis of intra-national trade. Section III first shows how intra-national trade dynamically modifies productivity heterogeneity. Second, it explores the allocative function of intra-national trade through its selection effect. Third, it analyzes why international trade is shaped by intra-national trade. Finally, before concluding, it examines the differences in the treatment of intra-national trade between the main existing trade models and our model.

2. The Framework Based on a Quantified Melitz Model

The purpose of this section is to derive the key equations for analyzing intra-national trade. The Melitz model is extended to meet our objective. The derivation processes are presented as compactly as possible, while some formal details can be found in Melitz (2003).

To begin with, following Dixit & Stiglitz (1977), demand is expressed by two equations:

$$Q \equiv U = \left[\int_0^{\Omega} q(\omega)^{\rho}\,\mathrm{d}\omega\right]^{1/\rho} \tag{1}$$

The output (or welfare) is a continuum of varieties indexed by ω. Ω is the mass of available varieties. ρ is the preference for variety with $0<\rho<1$. These varieties are substitutes, and the elasticity of substitution between any two of them is $\sigma = \dfrac{1}{1-\rho} > 1$.

Q is associated with an aggregate price:

$$P = \left[\int_0^{\Omega} p(\omega)^{1-\sigma}\,\mathrm{d}\omega\right]^{1/(1-\sigma)} = M^{1/(1-\sigma)} p(\tilde{\varphi}) \tag{2}$$

where $p(\tilde{\varphi})$ is the price associated with the average productivity level of all firms with positive profit, and M is the mass of producing firms, or available varieties.

For the purpose of analyzing intra-national trade and its connection to international trade, we focus on the derivation of reduced-form solutions of Q_a, Q_w, Q_d, and Q_x, where Q_a and Q_w are the autarkic and post-opening outputs, Q_d and

Q_x are the country's outputs for domestic consumption and exportation, and $Q_x = Q_w - Q_d$. With a unit-normalized wage, the revenue $R = L = PQ$ (L is the exogenous labor supply). Because of these relationships, the key variables to derive are then the aggregate pre- and post-opening prices P_a and P_w, which then lead to Q_a and Q_w. P_w is obtained by finding P_d and P_x, which give rise to Q_d and Q_x also. The reduced-form solutions and quantitative predictions become possible by following Helpman et al. (2004) and Chaney (2008) with the assumption that the distribution of productivities across firms follows a Pareto distribution.

To get P, from (2), we need to find M and $p(\tilde{\varphi})$. This is done in the following steps: we find firstly M. Second, $p(\tilde{\varphi})$ is found by solving for φ^*, the cutoff productivity level for which profit is zero, which is also a key variable for the analysis on intra-national trade. What follows is their execution step by step.

First, by deriving the optimal consumption and expenditure decisions, the individual demand for each variety is:

$$r(\varphi) = R\left[\frac{p(\varphi)}{P}\right]^{1-\sigma} \tag{3}$$

where φ is the productivity of the firm that produces a certain variety, $r(\varphi)$ and $p(\varphi)$ are then the revenue put on the consumption and the price of that variety.

On the production side, firm technology is represented by a cost function: $l = f_d + q/\varphi$, or the labor used by the firm is a function of the fixed cost for domestic market-destined production, and the output divided by productivity.

A conventional pricing rule: $p(\varphi) = \frac{\sigma}{(\sigma-1)\varphi}$ is yielded by the profit maximization of the firm in the face of a residual demand curve with constant elasticity.

Given $\pi = r - l$, using the cost function and pricing rule, firm's profit is:

$$\pi(\varphi) = \frac{r(\varphi)}{\sigma} - f_d \tag{4}$$

From (3), and using the pricing rule, we get $\frac{r(\tilde{\varphi})}{r(\varphi^*)} = \left(\frac{\tilde{\varphi}}{\varphi^*}\right)^{\sigma-1}$. With this result, from (4), the revenue of the firm of the cutoff productivity level, $r(\varphi^*) = \sigma f_d$ (as $\pi(\varphi^*) = 0$). Then, the revenue of the average productivity level of all firms with positive profit is:

$$r(\tilde{\varphi}) = \left(\frac{\tilde{\varphi}}{\varphi^*}\right)^{\sigma-1} \sigma f_d \tag{5}$$

Using (4) and (5), the average-productivity-level profit of all firms with positive profit is:

$$\pi(\tilde{\varphi}) = \left[\left(\frac{\tilde{\varphi}}{\varphi^*}\right)^{\sigma-1} - 1\right] f_d \tag{4.1}$$

From appendix (1),

$$\left(\frac{\tilde{\varphi}}{\varphi^*}\right)^{\sigma-1} = \frac{\theta}{\theta-(\sigma-1)} \equiv Z \tag{6}$$

where θ is the parameter of a Pareto distribution negatively reflecting the level of productivity heterogeneity.

From now on, the solutions are indexed *a* (autarchy) to distinguish from the post-opening results, indexed *w*. Using (6), (4.1) becomes:

$$\pi_a(\tilde{\varphi}) = (Z-1) f_d \tag{7}$$

(7) is called the pre-opening zero cutoff profit (ZCP) function: the average profit of all firms with productivity levels above φ^*.

Using (6) again, (5) becomes:

$$r_a(\tilde{\varphi}) = Z\sigma f_d \tag{8}$$

Then, as $R = L = Mr(\tilde{\varphi})$,

$$M_a = \frac{L}{Z\sigma f_d} \tag{9}$$

From (2), to find P_a, $p_a(\tilde{\varphi})$ is needed. It is derived in appendix (3): $p_a(\tilde{\varphi}) = \left(\frac{\sigma}{\sigma-1}\right) Z^{\frac{1}{1-\sigma}} \frac{\tau_d}{\varphi_a^*}$. With the equalization of the ZCP and the free entry (FE) functions obtained from appendix (2), in general equilibrium,

$$\varphi_a^* = \left[(z-1)\frac{f_d}{\delta f_e}\right]^{\frac{1}{\theta}} \tag{10}$$

$(z-1)\frac{f_d}{\delta f_e} > 1$, because equalizing (7) and (A2) in appendix (2), $\frac{\delta f_e}{1-G(\varphi^*)} = (Z-1) f_d$, and $0 < 1-G(\varphi^*) < 1$.

Using this result, (9), and (10), we get:

$$P_a = \frac{\sigma^{\frac{\sigma}{\sigma-1}}}{\sigma-1} \tau_d \left(\frac{f_d}{L}\right)^{\frac{1}{\sigma-1}} \left[\frac{\delta f_e}{(Z-1) f_d}\right]^{\frac{1}{\theta}} \tag{11}$$

Thus, we obtain the main reduced-form solutions of the variables that allow us to analyze intra-national trade in a closed economy: $\pi_a(\tilde{\varphi})$, $r_a(\tilde{\varphi})$, M_a, φ_a^*, and P_a. The next task is to find their corresponding values in an open economy. We will express all these values in terms of their corresponding closed-economy values.

For exporting firms, the corresponding cost function is $l = f_x + q/\varphi$, where f_x is the fixed cost per-period specific to export production. The first task is to find the ZCP function for exporting firms: the profit of their average productivity level. In the same way that we derive the ZCP function for firms before the opening, we obtain $\frac{r_x(\tilde{\varphi}_x)}{r_x(\varphi_x^*)} = \left(\frac{\tilde{\varphi}_x}{\varphi_x^*}\right)^{\sigma-1} = Z$, where $\tilde{\varphi}_x$ is the average productivity level of all exporting firms with positive profit, and φ_x^* is the productivity level for exporting firms at which profit is zero. With $\pi_x(\varphi)=\frac{r_x(\varphi)}{\sigma}-f_x$, and using $\pi_x(\varphi_x^*)=0$, following (7), $\pi_x(\tilde{\varphi}_x)=[Z-1]f_x$.

Based on (7) and the solution for $\pi_x(\tilde{\varphi}_x)$, the open-economy ZCP function is:

$$\begin{aligned}\pi_w(\tilde{\varphi})&=\pi_d(\tilde{\varphi}_d)+P_x n\pi_x(\tilde{\varphi}_x)=[Z-1](f_d+P_x nf_x)\\&=[Z-1]f_d(1+A)=(1+A)\pi_a(\tilde{\varphi})\end{aligned} \tag{7.1}$$

where $P_x=\left(\frac{f_x}{f_d}\right)^{-\frac{\theta}{\sigma-1}}\left(\frac{\tau_x}{\tau_d}\right)^{-\theta}$ is the probability of being an exporting firm, τ_d and τ_x are the variable costs of intra- and international trade, respectively (see appendix (4) for its derivation), n is the number of foreign countries, and $A=nP_x\frac{f_x}{f_d}=n\frac{f_x}{f_d}\left[\left(\frac{f_x}{f_d}\right)^{\frac{1}{\sigma-1}}\left(\frac{\tau_x}{\tau_d}\right)\right]^{-\theta}$. As we will see, A is the key factor in explaining how intra-national trade shapes international trade.

By definition, the mass of firms in an open economy: $M_w=\frac{L}{r_w(\tilde{\varphi})}$, and $r_w(\tilde{\varphi})=r_d(\tilde{\varphi}_d)+P_x nr_x(\tilde{\varphi}_x)$. Just as we derive $r_d(\tilde{\varphi}_d)=Z\sigma f_d$, from $\frac{r_x(\tilde{\varphi}_x)}{r_x(\varphi_x^*)}=Z$, and given $r_x(\varphi_x^*)=\sigma f_x$ and $r_x(\tilde{\varphi}_x)=Z\sigma f_x$, then:

$$r_w(\tilde{\varphi})=(1+A)Z\sigma f_d=(1+A)r_a(\tilde{\varphi}) \tag{8.1}$$

Therefore,

$$M_w=(1+A)^{-1}\frac{L}{Z\sigma f_d}=(1+A)^{-1}M_a \tag{9.1}$$

The FE function being the same as that of the closed economy, equalizing ZCP (7.1) and FE (A2),

$$\varphi_w^*=\varphi_d^*=\left[(z-1)\frac{f_d}{\delta f_e}\right]^{\frac{1}{\theta}}(1+A)^{\frac{1}{\theta}}=(1+A)^{\frac{1}{\theta}}\varphi_a^* \tag{10.1}$$

Price index in open economy, following Melitz (2003), is $P_w=M_t^{1/(1-\sigma)}p(\tilde{\varphi}_t)$, where $M_t=M_w+nM_x=M_w+nP_xM_w$ is the total number of domestic and exporting firms and $p(\tilde{\varphi}_t)$ is the average price of all domestic and exporting products. As $p(\tilde{\varphi}_t)=\frac{\sigma^{\frac{\sigma}{\sigma-1}}}{\sigma-1}M_t^{\frac{1}{\sigma-1}}\tau_d\left(\frac{f_d}{L}\right)^{\frac{1}{\sigma-1}}\varphi_w^{*-1}$ (see appendix (5)), using the definition of M_t, (9), (9.1), (10), (10.1), and (11),

$$P_w=\frac{\sigma^{\frac{\sigma}{\sigma-1}}}{\sigma-1}\tau_d\left(\frac{f_d}{L}\right)^{\frac{1}{\sigma-1}}\left[\frac{\delta f_e}{(Z-1)f_d}\right]^{\frac{1}{\theta}}(1+A)^{-\frac{1}{\theta}}=(1+A)^{-\frac{1}{\theta}}P_a \tag{11.1}$$

Thus, we obtain the key reduced-form solutions of the variables of an open economy: $\pi_w(\tilde{\varphi})$, $r_w(\tilde{\varphi})$, M_w, φ_w^*, and P_w.

3. Analysis of Intra-National Trade

Intra-national trade by definition covers intra- and inter-regional trade within a country. Regions differ in their technology, resource endowment, and geographic location. A region produces more goods in the most productive sectors (or, equivalently, by making intensive use of abundantly endowed resources). The productivity advantage also comes from location, because of lower transportation costs. The region then trades with other regions with higher productivity in other sectors. But despite this, there is still heterogeneity in productivity across regions. As such, productivity heterogeneity can be seen as a simplified, summary indicator of regional disparities within a country.

This section shows that intra-national trade is a mechanism for increasing productivity heterogeneity across firms. It is the key to increasing intra-national output and welfare. It is also through this increase that intra-national trade enhances international comparative advantage and, hence, shapes international trade.

3.1. Micro-Foundation and the Cause of Intra-National Trade

Given from (2), $P_a = M_a^{1/(1-\sigma)} p_a(\tilde{\varphi})$, using (9), (10), $p_a(\tilde{\varphi}) = \left(\dfrac{\sigma}{\sigma-1}\right) Z^{\frac{1}{1-\sigma}} \dfrac{\tau_d}{\varphi_a^*}$, and $Q = L/P$,

$$Q_a = \left\{ \sigma^{\frac{1}{1-\sigma}} Z^{-\frac{1}{\sigma-1}} L \left(\frac{L}{f_d}\right)^{\frac{1}{\sigma-1}} \right\} \left\{ \frac{\sigma-1}{\sigma} \frac{1}{\tau_d} Z^{\frac{1}{\sigma-1}} \left[\frac{(Z-1) f_d}{\delta f_e}\right]^{\frac{1}{\theta}} \right\} \tag{12}$$

(12) builds the micro-foundation of intra-national trade. The first brace reflects the output effects coming from consumer demand. The second brace defines the supply-side effects derived from firm behavior. σ shapes at once consumer preference to variety and firm behavior in terms of the elasticity of substitution. The most remarkable feature is that σ negatively affects the trade volume in the first brace, whereas this effect is positive in the second brace. The first brace can be rewritten as $\left\{ L M_a^{\frac{1}{\sigma-1}} \right\}$, implying that on the demand side, from $M_a = \dfrac{L}{Z\sigma f_d}$, the stronger the preference to variety, the larger the number of varieties. On the supply side, because larger elasticity of substitution allows for larger firm scale, the larger the σ, the smaller the mass of firms is.

Intra-national market makes possible to take a trade-off, and (12) allows a quantitative observation on how this is made. Intuitively, as σ exerts opposite effects, optimal intra-national trade output is achieved when marginal output gains from demand and supply sides of σ become equal.

This trade-off, however, is possible only whenever technological heterogeneity exists. To see this, without heterogeneity (θ is a positive infinity), the second brace approaches a constant equal $\frac{\sigma-1}{\sigma}\frac{1}{\tau_d}$, and in the first brace with $Z \to 1$, the number of varieties and demand side effect is maximized. This is a typical case of the homogeneous firm model in which each firm is specialized in one product, and the number of variety is maximized. With the presence of firm heterogeneity, the larger Z leads to smaller M_a in the first brace. Meanwhile the term $Z^{\frac{1}{\sigma-1}}\left[\frac{(Z-1)f_d}{\delta f_e}\right]^{\frac{1}{\theta}}$ in the second brace becomes larger. Therefore, θ is crucial for enhancing scale economies on the supply size, and for making possible the trade-off between increasing returns to variety and scale economies.

In (12), the term $Z^{\frac{1}{\sigma-1}}$ appears in both sides with opposite signs and offsets each other. Then, the only item shaped by θ is $\left[\frac{(Z-1)f_d}{\delta f_e}\right]^{\frac{1}{\theta}}$, which is negatively linked with θ. This implies that while the reduction of the mass of varieties gives rise to a loss in the gains from increasing returns to variety, this loss, however, is fully offset by the gains from productivity heterogeneity, and there is a net gain reflected by the $\left[\frac{(Z-1)f_d}{\delta f_e}\right]^{\frac{1}{\theta}}$. Thus, productivity heterogeneity has a net positive effect on trade volume.

The explanation on this positive effect is that, on the demand side, the mass of varieties (firms) is required to be large, and the cutoff productivity level to be low. On the supply side, the firms of high productivity are favored by increasing returns to scale. The trade-off between the demand- and supply-sides effects by the market leads to the coexistence of the firms heterogeneous in productivity and size, and welfare is increasing in productivity heterogeneity.

(12) also expresses the relationship between the trade volume of intra-national market and other determinants. As expected, the population (L) measuring the market size positively affects the trade volume, while the trade impacts coming from the variable and fixed trade costs (τ_d and f_d) are negative.

In summary, the micro-foundation of intra-national trade consists in the trade-off between the demand- and supply-sides effects by the market. The modelling becomes possible due to the coexistence of the two key factors: σ and θ. The driving force or the cause of intra-national trade comes from increasing returns to variety and scale economies. Their trade-off becomes possible due to the existence of the productivity heterogeneity among the firms.

3.2. Market Competition and Dynamic Transition of Productivity Heterogeneity

Melitz (2003) considered that his results can be the steady-state solutions of a dynamic model. This section extends this ideal by exploring the endogenous increase of productivity heterogeneity during transition phases in a dynamical modeling framework.

To fix the idea, let us first note that a key feature of the Melitz model is increasing profits to productivity: Using the pricing rule: $p(\varphi)=\dfrac{\sigma}{(\sigma-1)\varphi}$, (3) and (4), we can show that when $\sigma>2$, then $\dfrac{\partial\pi}{\partial\varphi}>0$, and $\dfrac{\partial^2\pi}{\partial\varphi^2}>0$. In other words, with an elasticity of substitution ensuring some minimum level of economies of scale, there are increasing profits to productivity.

Second, assuming that productivity is a positive function of investment, and that investment is a positive function of profit, the increasing profits to productivity is able to give rise to a dynamic evolution of productivity heterogeneity in the following way: on the one hand, firms are motivated to invest in improving their productivities, and on the other hand, high-productivity firms have a greater capacity to invest, and thus to increase their productivity more than others. Consequently, at the aggregate level, the growth of the average level of profit implies an even more unequal distribution of investment, and thus increases the heterogeneity of productivity. This idea allows us to model the dynamic transition of productivity heterogeneity towards its equilibrium value with an equation of motion:

$$\left|\dot{\theta}\right|=I_t(\tilde{\varphi})\left(I^*(\tilde{\varphi})-I_t(\tilde{\varphi})\right) \tag{13}$$

where $I_t(\tilde{\varphi})$ is the average investment at transition stage *t*, and $I^*(\tilde{\varphi})$ is a constant determined by the value of the average investment in steady state.

$I_t(\tilde{\varphi})$ is a function of average profit, and from (7), average profit decreases with θ. Assuming that a fixed share of average profit is used in investment, then:

$$\left|\dot{\theta}\right|=\varepsilon\pi_t(\tilde{\varphi})\left[\bar{\pi}(\tilde{\varphi})-\pi_t(\tilde{\varphi})\right]=\varepsilon\pi_t(\tilde{\varphi})\left[\frac{\sigma-1}{\gamma}fd-\pi_t(\tilde{\varphi})\right] \tag{13.1}$$

where ε is a constant, $\bar{\pi}(\tilde{\varphi})=\dfrac{\sigma-1}{\gamma}fd$ is shown in appendix (6).

(12.1) is the equation of motion that describes the transition path of productivity heterogeneity. Gradually, with the increase of average profit, θ converges to its steady-state value. At the steady state, $\left|\dot{\theta}\right|=0$ leads $\bar{\theta}=\gamma+(\sigma-1)$, and $\bar{\pi}(\tilde{\varphi})=((\sigma-1)/\gamma)f_d$. $\pi(\tilde{\varphi})=0$ is another but unstable solution. From $\partial\left|\dot{\theta}\right|/\partial\pi_t(\tilde{\varphi})=((\sigma-1)/\gamma)f_d-2\pi_t(\tilde{\varphi})=0$, to $\pi_t(\tilde{\varphi})=\frac{1}{2}((\sigma-1)/\gamma)f_d$, $\left|\dot{\theta}\right|$ is the highest, then, decreases to zero. **Figure 1** shows the phase diagram.

Even starting from a point where θ is quite high (a fairly low level of heterogeneity), the market for intra-national trade could drive θ down, or productivity heterogeneity to increase. This dynamism, however, is not automatic and depends on the development process of a country, in particular its institutional conditions that shapes its division of labor, and thus its industrial organization

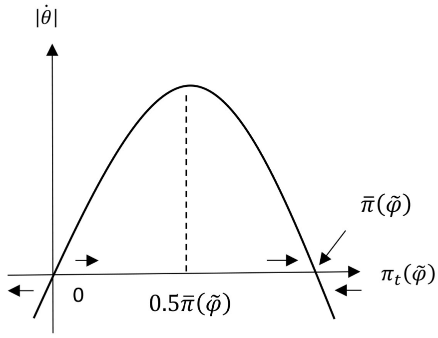

Figure 1. The evolution of θ.

(cf. (Sun, 2012)). In most underdeveloped countries, the formation of large firms faces enormous difficulties, leading to industrial organizations characterized by low productivity heterogeneity. The evolution of industrial organization can take a very long time, because σ can evolve on the one hand, and, γ, the parameter reflecting institutional constraints, defined in appendix (6), can change on the other, making stable states continuously unstable. But in general, due to intra-national trade, the higher the level of development (the higher the average profit in the model), the higher the level of productivity heterogeneity across firms.

Empirical evidence of dynamically increasing productivity heterogeneity with economic development supports this analysis. For example, Okubo & Tomiura (2010) found that in Japan, firm productivity is distributed with larger dispersions in central regions than in other regions, indicating that productivity heterogeneity is positively correlated with regional development levels. Poschke (2018) documented that firm size dispersion, which generally coincides with productivity heterogeneity, is larger in rich countries, and has increased over time for U.S. firms.

3.3. Market Selection and Allocation Effects

We have just shown that competition in intra-national markets leads firms to be more heterogeneous in terms of productivity. This produces a reinforcing effect of market selection and makes allocation more efficient. To see this, (12) can be also written to:

$$Q_a = \frac{\sigma-1}{\sigma^{\frac{\sigma}{\sigma-1}}}\frac{L}{\tau_d}\left(\frac{L}{f_d}\right)^{\frac{1}{\sigma-1}}\varphi^* = \frac{\sigma-1}{\sigma^{\frac{\sigma}{\sigma-1}}}\frac{L}{\tau_d}\left(\frac{L}{f_d}\right)^{\frac{1}{\sigma-1}}\left[\frac{(Z-1)f_d}{\delta f_e}\right]^{\frac{1}{\theta}} \tag{12.1}$$

Q_a is the optimal reduced-form solution, the country's output or welfare. At the transition stages, $\dot{Q}_a = F_1(\dot{\varphi}^*) = F_2(\dot{\theta})$, and (13) is just the steady-state value of Q_a shaped by the steady-state value of θ.

The cutoff productivity (φ^*) is the key determinant in Q_a, which measures the selection effect, or the extent to which the most productive and large-sized firms are favored and the less productive are eliminated. φ^*, in turn, is determined by

θ (directly or via Z). In other words, just as international trade has a generally recognized selection effect, so does intra-national trade. Without productivity heterogeneity, selection effect is absent (equivalent to $\theta \to \infty$ so that the term $\left[\frac{(Z-1) f_d}{\delta f_e}\right]^{\frac{1}{\theta}} \to 1$). The higher the heterogeneity, the higher the cutoff productivity. The higher the cutoff productivity, the higher the outputs and profits of high productivity firms, so the total output of the country is higher.

The selection effect of intra-national trade gives rise to an intra- and inter-regional allocation function. Since the selection effect is a function of productivity heterogeneity, according to (9) and (11), a higher selection effect implied by a smaller θ and a larger value of Z reduces the mass of firms, and the aggregate price. The selection effect and the price effect produce a surplus of labor somewhere, and this labor will move to higher productivity locations. Therefore, higher heterogeneity and a larger selection effect lead to greater labor mobility. Due to the selection effect, the average firm size: Q/M increases and larger economies of scale are achieved. Welfare is then improved.

Calibrating the parameters to key U.S. aggregate and firm statistics, Melitz & Redding (2013) find a quantitatively significant gain in aggregate welfare (up to a few percentage points of GDP) in a heterogeneous firm model relative to a homogeneous firm model. One of the explanations provided in this section is that it is only in a heterogeneous firm model that intra-national trade is able to exert the selection effect, and generate welfare gains.

3.4. How Does Intra-National Trade Shape International Trade?

The evolution of productivity heterogeneity induced by intra-national trade also affects international trade. To see this, using (11.1), and $Q = L/P$, we obtain:

$$Q_w = (1 + A)^{\frac{1}{\theta}} Q_a \tag{14}$$

Q_a and Q_w are outputs, as well as the volumes of trade before and after opening. Output after opening is first determined by the output of intra-national trade before opening. Q_a, shaped by the level of intra-national productivity heterogeneity, determines the size of its foreign trade. One interpretation could be that the development of the domestic market stimulates demand in the world market. The multiplier, $(1 + A)^{\frac{1}{\theta}}$, reflecting the country's supply capacity for the world market, is also a function of productivity heterogeneity. Recall that $A = n\frac{f_x}{f_d}\left[\left(\frac{f_x}{f_d}\right)^{\frac{1}{\sigma-1}}\left(\frac{\tau_x}{\tau_d}\right)\right]^{-\theta}$, it is present in (7.1), (8.1), (9.1), (10.1), (11.1), and (12.1). It amplifies or reduces the post-opening values of average intra-national profit, average income, cutoff productivity level, mass of surviving firms, aggregate price, and welfare relative to those before opening.

When the economy is open to foreign trade, the parameter θ measuring productivity heterogeneity takes on a comparative meaning across countries, and can be interpreted as equivalent to international comparative advantage. This equivalence can be easily verified with the classical assumption of symmetry between countries. In a two-country, two-sector Ricardo model, assuming the existence of intra-country productivity heterogeneity, country A having higher productivity in sector 1 than in sector 2 and country B being the opposite, they have a comparative advantage in sectors 1 and 2, respectively. In the case of n countries and m sectors with $n = m$, assuming that 1) within each country, all sectors are ranked differently in terms of productivity, and 2) the ranking of sectors is different between each pair of countries, then, each country has the highest comparative advantage in only one sector. In this case, intra-national productivity heterogeneity coincides with international comparative advantage. If $n > m$, or $n < m$, a number of countries have the highest comparative advantage in the same sector, or each country has the highest comparative advantage in several sectors. Intra-national productivity heterogeneity is always correlated with international comparative advantage. In a more complicated world of asymmetric countries, intuitively, an indicator of the average productivity heterogeneity of these countries should be positively associated with the level of international comparative advantage.

In sum, through intra-national trade, market competition increases the heterogeneity of productivity, which is tantamount to strengthening the comparative advantage that drives the international division of labor, and thus international trade.

It can also be shown that in the Melitz model, the indicator reflecting intra-national productivity heterogeneity plays a similar role to the indicator reflecting inter-national comparative advantage in the Eaton & Korntum (2002) model. In the Melitz model, in the reduced-form equations for the price index (11) in the closed economy case and (11.1) in the open economy case, the indicator θ reflecting intra-national productivity heterogeneity has a positive impact on the price index, or, $\partial P / \partial \theta > 0$. In the Eaton & Korntum (2002) model, a similar objective function and supply behavior of firms is assumed, and the indicator θ this time reflects cross-national comparative advantage using a Fréchet productivity distribution. We also obtain $\partial P / \partial \theta > 0$.[2]

Having examined the relationship between output before and after opening, another angle is to inspect the relationship between domestic trade and exports after opening. By definition, $Q_w = Q_d + Q_x = \frac{M_w r_d(\tilde{\varphi}_d)}{P_w} + \frac{n P_x M_w r_x(\tilde{\varphi}_x)}{P_w}$. M_w and P_w are obtained from (9.1) and (11.1), respectively. Given $r_d(\tilde{\varphi}_d) = Z\sigma f_d$, $r_x(\tilde{\varphi}_x) = Z\sigma f_x$, and using the definitions of P_x in Section II, we obtain:

[2]The resulting reduced form of the price index is, $P = \left[\Gamma\left(\frac{\theta-(\sigma-1)}{\theta}\right)\right]^{1/(1-\sigma)} \left[\sum_{i=1}^{N} T_i (c_i d_{ni})^{-\theta}\right]^{-1/\theta}$, where Γ is the Gamma function and T_i reflects the absolute advantage of country i, both are specific to the Fréchet distribution, c_i is the production cost of country i, d_{ni} is the geographical barrier between country n and i. It is easy to show that, also, $\partial P / \partial \theta$.

$$\frac{Q_x}{Q_d} = A \tag{15}$$

The ratio of exports to domestic consumption is simply expressed as A. Using the definition of A, by differentiation, we have $\frac{\partial\left(\frac{Q_x}{Q_d}\right)}{\partial n} > 0$, $\frac{\partial\left(\frac{Q_x}{Q_d}\right)}{\partial\left(\frac{f_x}{f_d}\right)} < 0$, $\frac{\partial\left(\frac{Q_x}{Q_d}\right)}{\partial\left(\frac{\tau_x}{\tau_d}\right)} < 0$, $\frac{\partial\left(\frac{Q_x}{Q_d}\right)}{\partial \theta} < 0$. With them, we are able to identify the channels through which international trade is amplified.

International trade benefits first from market expansion as measured by n, the number of foreign countries. It is also affected by the relative fixed and variable costs of trade. The second and third inequalities mean that the relative trade costs between intra- and international trade negatively determine the relative sizes of the two markets. There is a clear substitution relationship between the two trade markets based on relative trade costs. Higher relative trade costs of international trade favor intra-national trade and vice versa. This topic has already been the subject of a number of empirical investigations in North American and European countries (McCallum, 1995; Wei, 1996; Anderson & Wincoop, 2003).

The last inequality implies that, to the extent that the market is open to international trade, the increase in productivity heterogeneity contributes to a greater extent to the increase in international trade than to intra-national trade. This is an extraordinary result. While the productivity heterogeneity generated by intra-national trade promotes both intra- and international trade, it makes the world economy more globalized.

3.5. Why Is Intra-National Trade Absent in the Leading Trade Models?

In a classical Ricardian model, consumer preference for variety and product substitution are not defined, but heterogeneity in productivity between countries is exogenously assumed. Without the possibility of product substitution, all varieties must be retained. Corresponding to (1), output reduces to $Q \equiv U = \int_0^{\Omega} q(\varphi)\mathrm{d}\varphi = L\tilde{\varphi}$, where $q(\varphi) = l\varphi$, $\tilde{\varphi} = \int_1^{\infty} \varphi g(\varphi)\mathrm{d}\varphi$, $g(\varphi) = \theta\varphi^{-(\theta+1)}$ is the common distribution following Pareto distribution, θ is the indicator of productivity heterogeneity assumed in the model, and the lowest productivity is normalized to 1. By integration, we obtain $Q = L\left(\frac{\theta}{\theta-1}\right)$. Intra-national trade is absent in the sense that it is reduced to an invariant outcome shaped only by some exogenously determined parameters. The model is unable to identify any selection effect. Without economies of scale as a driving force for firms, no dynamic increase in productivity heterogeneity occurs.

Another major trade model is the homogeneous firm model. Its cost function, $l = f + cq$ (with f as a fixed cost and c a constant reflecting uniform productivity), implies the existence of economies of scale at the firm level, but without the ability to exploit them given unallowed productivity heterogeneity. Considering it from the point of view of the heterogeneous film model, The main reduced-form solutions of a homogeneous firm model can be derived from the heterogeneous firm model as a special case of it with $\theta \to \infty$, or $Z \to 1$, so that from (9) and (10), $M_a \to \frac{L}{\sigma f_d}$ is a constant, and $\varphi_a^* = \left[\frac{f_d(Z-1)}{\delta f_e}\right]^{\frac{1}{\theta}} \to 1$, implying an absent selection effect. Corresponding to (7), (8), (11), and (12.1), $\pi \to 0$ for all firms, $r \to \sigma f_d$, $P_a \to \frac{\sigma^{\frac{\sigma}{\sigma-1}}}{\sigma-1}\tau_d\left(\frac{f_d}{L}\right)^{\frac{1}{\sigma-1}}$, and $Q_a \to \frac{\sigma-1}{\sigma^{\frac{\sigma}{\sigma-1}}}\frac{L}{\tau_d}\left(\frac{L}{f_d}\right)^{\frac{1}{\sigma-1}}$.[3] Therefore, in a homogeneous firm model, Q_a also reduces to several exogenous parameters. Thus, without heterogeneity in productivity and its evolution, intra-national trade per se has no influence on international trade.

Because of this difference in the role of intra-national-trade between the homogeneous and heterogeneous firm models, an important consequence is their difference in driving force of international trade. From $A = n\frac{f_x}{f_d}\left[\left(\frac{f_x}{f_d}\right)^{\frac{1}{\sigma-1}}\left(\frac{\tau_x}{\tau_d}\right)\right]^{-\theta}$, and ignoring the effects of relative trade costs by assuming $\frac{f_x}{f_d} = \frac{\tau_x}{\tau_d} = 1$, we get $A = n$. Then in a heterogeneous firm model, from (14), $Q_w = (n+1)^{\frac{1}{\theta}} Q_a$. This equation can be shown to be $Q_w = (n+1)^{\frac{1}{\sigma-1}} Q_a$ in a homogeneous firm model.[4] In other words, since $(n + 1)$ is the number of countries in the world, by comparing $(n+1)^{\frac{1}{\theta}}$ vs $(n+1)^{\frac{1}{\sigma-1}}$, we can conclude that while internationalization is amplified by increasing returns to variety in a homogenous firm model, it is driven by comparative advantage reflected by productivity heterogeneity in a heterogeneous firm model.

In summary, on the basis of a Melitz model, intra-national trade can be modeled due to the existence of both increasing returns to variety and productivity heterogeneity. The former is absent from a typical Ricardian model, and the latter from a homogeneous firm model.

[3]As can be easily shown, all these results with unadmitted heterogeneity derived from a heterogeneous firm model correspond perfectly to those derived from a benchmark homogeneous firm model (cf. (Fujita et al., 1999: chapter 4)).

[4]To see this, ignoring relative trade costs by setting them to be unity, the Equation (4.15) in Fujita et al. (1999: chapter 4) about international price index can be rewritten, in accordance with the terminology used in Melitz (2003), as $P_w = \{(1+n)M_a p^{1-\sigma}\}^{1/(1-\sigma)}$ and intra-national price index as $P_a = \{M_a p^{1-\sigma}\}^{1/(1-\sigma)}$. Using $Q = L/P$, we get the result.

4. Conclusion, Implication, and Limitation

A theory of trade based on intra-national trade has been constructed to show that intra-national trade pushes the most efficient firms to increase their productivities further, leading to increasing heterogeneity in productivity and comparative advantage over time; it exerts selection and allocation effects and improves national welfare, and enhances international trade. Thus, intra- and international trade are two intrinsic parts of an overall process.

This approach could yield rich empirical insights into the link between trade and development process. To explain the cause of international trade, existing trade theories focus on the driving forces external to each country: increasing returns to variety and economies of scale prevailing in the international market, as well as the comparative advantage existing between countries. In this approach, a country's intra-national trade shapes its international trade. For most underdeveloped countries, the lack of openness may stem not just only from their motivation and policies, but also from the lack of the development of intra-national trade market itself. Therefore, priorities should be placed on reducing the costs of intra-national trade associated with their transport and other infrastructure constraints, and on improving their industrial organization structures to expand the size of intra-national trade market. To illustrate more antagonistically, when facing with an African country with no comparative advantage in any industrial sector, existing trade theories tend to recommend that it abandon all industries and specialize in mining and agriculture, whereas our model, which views comparative advantage as a dynamic process, not a given, will recommend that it first develop its industrial system before embarking on international specialization, so that over time not only that country, but the world economy as well, will be better off.

This study focused on the "intra-national" cause of international trade, and left aside another no less important issue: the effect of international trade on intra-national trade. It has shown how comparative advantage evolves as a function of the operation of intra-national trade. Intuitively, the evolution of comparative advantage among countries is also subject to the development of international trade. This issue is not covered in this study, and requires other modeling structures.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Anderson, J. E., & van Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *The American Economic Review, 93,* 170-192. https://doi.org/10.1257/000282803321455214

Chaney, T. (2008). Distorted Gravity: The Intensive and Extensive Margins of International Trade. *The American Economic Review, 98,* 1707-1721. https://doi.org/10.1257/aer.98.4.1707

Combes, P.-P., Mayer, T., & Thisse, J.-F. (2008). *Economic Geography: The Integration of Regions and Nations*. Princeton University Press.

Dixit, A. K., & Sitlitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *The American Economic Review, 67,* 297-308.

Eaton, J., & Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica, 70,* 1741-1779. https://doi.org/10.1111/1468-0262.00352

Fujita, M., Krugman, P., &Venables, A. J. (1999). *The Spatial Economy: Cities, Regions, and International Trade.* The MIT Press. https://doi.org/10.7551/mitpress/6389.001.0001

He, Y. (2017). Intra-National Trade as Channels of Spillovers in Developing Countries. *Journal of Economic Integration, 32,* 358-399. https://doi.org/10.11130/jei.2017.32.2.358

Helpman, E., Melitz, M. J., & Yeaple, S. R., (2004). Export versus FDI with Heterogeneous Firms. *The American Economic Review, 94,* 300-316. https://doi.org/10.1257/000282804322970814

Hopenhayn, H. A. (1992). Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica, 60,* 1127-1150. https://doi.org/10.2307/2951541

Krugman, P. R. (1979). Increasing Returns, Monopolistic Competition, and International Trade. *Journal of International Economics, 9,* 469-479. https://doi.org/10.1016/0022-1996(79)90017-5

Krugman, P. R. (2015). Interregional and International Trade: Different Causes, Different Trends? In P. Nijkamp, A. Rose, & K. Kourtit (Eds.), *Regional Science Matters* (pp. 27-34). Springer. https://doi.org/10.1007/978-3-319-07305-7_3

McCallum, J. (1995). National Borders Matter: Canada-U.S. Regional Trade Patterns. *The American Economic Review, 85,* 615-623.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica, 71,* 1695-1725. https://doi.org/10.1111/1468-0262.00467

Melitz, M. J., & Redding, S. J. (2013). *Firm Heterogeneity and Aggregate Welfare.* CEP Discussion Paper No 1200. Centre for Economic Performance, London School of Economics and Political Science (LSE).

Melitz, M. J., & Redding, S. J. (2014). Heterogeneous Firms and Trade. In G. Gopinath, E. Helpman, & K. Rogoff (Eds.), *Handbook of International Trade* (Vol. 4, pp. 1-54). Elsevier. https://doi.org/10.1016/B978-0-444-54314-1.00001-X

Ohlin, B. (1933). *Interregional and International Trade.* Harvard University Press.

Okubo, T., & Tomiura, E. (2010). *Productivity Distribution, Firm Heterogeneity, and Agglomeration: Evidence from Firm-Level Data.* RIETI Discussion Paper Series. Research Institute for Economics and Business Administration ((RIETI).

Poncet, S. (2005). A Fragmented China: Measure and Determinants of Chinese Domestic Market Disintegration. *Review of International Economics, 13,* 409-430. https://doi.org/10.1111/j.1467-9396.2005.00514.x

Poschke, M. (2018). The Firm Size Distribution across Countries and Skill-Biased Change in Entrepreneurial Technology. *American Economic Journal: Macroeconomics, 10,* 1-41. https://doi.org/10.1257/mac.20140181

Sun, G.-Z. (2012). *The Division of Labor in Economics: A History.* Routledge.

Venables, A. J. (2005). Spatial Disparities in Developing Countries: Cities, Regions and International Trade. *Journal of Economic Geography, 5,* 3-21. https://doi.org/10.1093/jnlecg/lbh051

Wei, S.-J. (1996). *Intra-National versus International Trade: How Stubborn Are Nations in Global Integration?* NBER Working Papers No. 5531. National Bureau of Economic Research (NBER). https://doi.org/10.3386/w5531

Appendixes

Appendix 1: Derivation of Equation (6)

By definition, the average productivity is equal $\tilde{\varphi}\left(\varphi^{*}\right)=\left[\int_{\varphi^{*}}^{\infty}\varphi^{\sigma-1}\mu(\varphi)\mathrm{d}\varphi\right]^{\frac{1}{\sigma-1}}$, where the conditional distribution of $g(\varphi)$ over $\left(\varphi^{*},\infty\right)$ is $\mu(\varphi)=\frac{g(\varphi)}{1-G\left(\varphi^{*}\right)}$ if $\varphi\geq\varphi^{*}$; $\mu(\varphi)=0$ otherwise. $g(\varphi)$ is the common distribution and $G\left(\varphi^{*}\right)$ the continuous cumulative distribution on $\left(\varphi^{*},\infty\right)$. Using the definition of the Pareto distribution, $1-G\left(\varphi^{*}\right)=\varphi^{*^{-\theta}}$ and by integration, we obtain (6).

Appendix 2: Finding the FE Condition

Following a stochastic framework due to Hopenhayn (1992), before entering the market, firms must first make an initial investment, modeled as a fixed entry cost $f_e>0$ (measured in units of labor). Firms then draw their initial productivity parameter φ from a common distribution $g(\varphi)$ that has a continuous cumulative distribution $G(\varphi)$.

Define V_e as the net value of entry:

$$V_e=Pin\bar{V}-f_e=\frac{1-G\left(\varphi^{*}\right)}{\delta}\pi(\tilde{\varphi})-f_e \tag{A1}$$

where $Pin=1-G\left(\varphi^{*}\right)$ is the ex-ante probability of successful entry, δ is a constant probability in each period of a bad shock that would force the firm to exit. $\bar{V}=\frac{1}{\delta}\pi(\tilde{\varphi})$ is the average value of firms.

Free entry drives $V_e=0$, so that:

$$\pi(\tilde{\varphi})=\frac{\delta f_e}{1-G\left(\varphi^{*}\right)}=\frac{\delta f_e}{\varphi^{*^{-\theta}}} \tag{A2}$$

(A2) is the FE condition.

Appendix 3: Finding $p_a(\tilde{\varphi})$

By definition, $p_a(\tilde{\varphi})=\left[\int_{\varphi_a^{*}}^{\infty}p_a(\varphi)^{1-\sigma}\mu(\varphi)\mathrm{d}\varphi\right]^{\frac{1}{1-\sigma}}$, where $\mu(\varphi)$ is defined in appendix (1). Since aggregate prices before and after the opening are affected by their variable trade costs, we fix τ_d and τ_x. τ_d is the variable intra-national trade cost (following Melitz and Redding 2014) and τ_x the variable international trade cost. Trade costs per unit are modeled using the standard iceberg formulation, whereby τ_d (or τ_x) units (>1) of a good are shipped for one unit to arrive at its destination. Accordingly, the domestic and exporting pricing rules are modified as $p_a(\varphi)=\frac{\sigma}{\sigma-1}\frac{\tau_d}{\varphi}$ and $p_x(\varphi)=\frac{\sigma}{\sigma-1}\frac{\tau_x}{\varphi}$.

Using $p_a(\varphi)=\frac{\sigma}{\sigma-1}\frac{\tau_d}{\varphi}$ and by integration, $p_a(\tilde{\varphi})=\left(\frac{\sigma}{\sigma-1}\right)Z^{\frac{1}{1-\sigma}}\frac{\tau_d}{\varphi_a^*}$.

Appendix 4: Finding P_x

Using the definition of Pareto distribution, $P_x=\frac{1-G(\varphi_x^*)}{1-G(\varphi_d^*)}=\left(\frac{\varphi_x^*}{\varphi_d^*}\right)^{-\theta}$.

Using (3), $r_x(\varphi)=R\left[\frac{p_x(\varphi)}{P}\right]^{1-\sigma}$, and the domestic and exporting pricing rules defined in appendix (3), we get $\frac{\varphi_x^*}{\varphi_d^*}=\left(\frac{r_x(\varphi_x^*)}{r_d(\varphi_d^*)}\right)^{\frac{1}{\sigma-1}}\left(\frac{\tau_x}{\tau_d}\right)$. Using $\pi_x(\varphi_x^*)=\frac{r_x(\varphi_x^*)}{\sigma}-f_x=0$ and $\pi_d(\varphi_d^*)=\frac{r_d(\varphi_d^*)}{\sigma}-f_d=0$, we get $\frac{r_x(\varphi_x^*)}{r_d(\varphi_d^*)}=\frac{f_x}{f_d}$.

Thus, $P_x=\left(\frac{f_x}{f_d}\right)^{-\frac{\theta}{\sigma-1}}\left(\frac{\tau_x}{\tau_d}\right)^{-\theta}$.

Appendix 5: Finding $p(\tilde{\varphi}_t)$

By definition, $p(\tilde{\varphi}_t)=\frac{1}{\rho\tilde{\varphi}_t}$, and following Melitz (2003, p1710),

$$\tilde{\varphi}_t=\left\{\frac{1}{M_t}\left[M_w\left(\tau_d^{-1}\tilde{\varphi}_d\right)^{\sigma-1}+nM_x\left(\tau_x^{-1}\tilde{\varphi}_x\right)^{\sigma-1}\right]\right\}^{\frac{1}{\sigma-1}}.$$

From $\left(\frac{\tilde{\varphi}_d}{\varphi_d^*}\right)^{\sigma-1}=z$, $\tilde{\varphi}_d=Z^{\frac{1}{\sigma-1}}\varphi_d^*$. From $\left(\frac{\tilde{\varphi}_x}{\varphi_x^*}\right)^{\sigma-1}=z$, $\tilde{\varphi}_x=Z^{\frac{1}{\sigma-1}}\varphi_x^*$. From appendix (4), $\varphi_x^*=\left(\frac{f_x}{f_d}\right)^{\frac{1}{\sigma-1}}\left(\frac{\tau_x}{\tau_d}\right)\varphi_d^*$. Furthermore,

$M_x=P_xM_w=\left(\frac{f_x}{f_d}\right)^{-\frac{\theta}{\sigma-1}}\left(\frac{\tau_x}{\tau_d}\right)^{-\theta}M_w$. Putting them in Melitz's equation about $\tilde{\varphi}_t$,

and finally, using (9) and (9.1), $p(\tilde{\varphi}_t)=\frac{\sigma^{\frac{\sigma}{\sigma-1}}}{\sigma-1}M_t^{\frac{1}{\sigma-1}}\tau_d\left(\frac{f_d}{L}\right)^{\frac{1}{\sigma-1}}{\varphi_d^*}^{-1}$.

Appendix 6: Finding the Steady-State Value of Average Profit: $\bar{\pi}(\tilde{\varphi})$

With the average profit = $(Z-1)f_d$, and $Z=\theta/(\theta-(\sigma-1))$, the smaller $\theta-(\sigma-1)$ is, the larger Z will be. Through competition, intra-national market could make θ closer to $\sigma-1$. Intuitively, the Pareto distribution constraint: $\theta>(\sigma-1)$ can be interpreted as limiting productivity heterogeneity by the extent of scale economies implied by the elasticity of substitution. Mathematically, $\lim_{\theta\to(\sigma-1)}Z=+\infty$. In the real world, however, the increase in Z is limited because market concentration faces a number of constraints due to its monopolistic nature and other sociopolitical and cultural resistance. The underdevelopment of financial institutions may also be a constraint. Denoting γ to reflect the degree of these constraints, and assuming $\gamma \le \theta - (\sigma - 1)$, and using this relation in (7), $\frac{\sigma - 1}{\gamma} fd$ is the steady-state value of average profit.