Multi-Product Firms and Gravity

Ruo Shangguan*

Waseda University

January 2, 2021

Abstract

Using data on the universe of Chinese export transactions in 2006, we document that, conditional on the number of products each firm exports, lower-ranking products are more sensitive to distance. This fact is inconsistent with standard models but can be explained in a model with increasing marginal cost of production. This assumption implies that the market share of productive firms is lower, and more unproductive firms enter. Moderate variable trade costs consistent with the observed level of tariffs generate a much larger and non-constant trade elasticity, even when productivity is distributed according to an untruncated Pareto distribution. We find that if expanding the production scale by one log point increases marginal cost by 25%, then welfare gains from trade are 2%, compared to 12% under the constant marginal cost assumption.

Keywords: Gravity equation, multi-product firm, trade cost

JEL classification: D21, D22, F12, F13, F14, L1

*Institute for Empirical Research in Organizational Economics. Email: roy1312@gmail.com. I am grateful to my advisor Hidehiko Ichimura, and committee members Hideo Owan, Konstantin Kucheryavyy, Taiji Furusawa, Yasutora Watanabe for their guidance, encouragement, and support. I also thank Michal Fabinger, Pol Antràs, Davin Chor, and Kala Krishna for useful discussions.

1 Introduction

Multi-product firms dominate world trade[1]. Therefore, understanding how multi-product firms allocate their resources across products will provide a better understanding of international trade patterns. On the other hand, gravity equations have been the workhorse empirical specification in understanding trade patterns over the past decades (Head and Mayer, 2014). Given the importance of these two facts, many authors have proposed theoretical models that explain the salient features of multi-product firms in the data (e.g. Bernard et al. 2011; Mayer et al. 2014; Arkolakis et al. 2019), and under specific functional form assumptions, these models generate gravity equations that describe bilateral product-level trade flows.

Using Chinese Customs data, this paper documents that, conditional on the total number of exported products, the entry decisions of lower-ranking products are more sensitive to distance. This documented fact suggests that trade costs increase faster for lower-ranking products, therefore showing a particular form of interdependence across products within a firm. A small difference in product efficiency can be amplified via this mechanism and translates into a big difference in sales. Therefore, any attempt to infer trade costs from observed trade volume should not ignore this interdependence.

The phenomenon of changing distance elasticity holds true in both the sample of homogeneous goods and differentiated goods. Therefore, it cannot be simply explained by the fact that, 1) higher-ranking products tend to have higher quality and are therefore more suitable for long-distance trade (e.g. Hummels and Skiba 2004), or 2) top-ranking products are newly developed products, which have high variance in the production technology (e.g. Fieler 2011). Moreover, this phenomenon cannot be easily explained by the aforementioned models of multi-product firms. To see why, consider the gravity equations generated from those models. If the underlying primitive parameters (i.e., the demand elasticity and the shape parameter of the productivity distribution) are the same across products of different ranks, lowering the product rank is equivalent to shifting the location parameter of the productivity distribution downward. Thus, the distance-trade profile is predicted to have a parallel shift, in contrast to the changing slope found in the data. In order to explain the changing distance sensitivity in those frameworks, one could impose the assumption that the underlying elasticities of goods of different ranks differ systematically. Such an attempt would be complicated, however, by the fact that product rank is an endogenous choice by firms.

To give an explanation of the changing distance elasticity, instead of imposing ar-

[1] According to Bernard et al. 2007, multi-product firms accounted for 99.6% of U.S. export value in 2000.

bitrary restrictions on the underlying primitive parameters across products of different ranks, this paper studies a heterogeneous firm trade model *à la* Melitz (2003) that allows the marginal cost of production to increase with quantity. The increasing marginal cost of production provides a natural explanation for the changing distance elasticity phenomenon. Given that the firm is already exporting high-rank products, which increases the marginal cost of production compared to the case in which production is fully idle, introducing secondary products is costlier. Thus, the equal increase in the trade costs results in a higher hurdle for the entrance of secondary products. This mechanism also predicts that the distance elasticity of the intensive margin of the secondary products will not be more sensitive to distance. The prediction is supported by the data.

In most trade models, the marginal production cost is assumed to be constant. Thanks to this assumption, except for sharing a common productivity measure, the entry decisions into different product-destinations within each firm are separated from each other, which makes the analysis tractable. An immediate consequence of this assumption is that every firm, regardless of productivity, has the capacity to supply the whole world wherever there is demand. In reality, however, it can be costly to expand production. For example, in the short run, even profitable firms in good standing can have difficulty getting financed. And in the long-run, managerial attention has to be diluted when the scale of production grows. Allowing increasing marginal cost of production essentially imposes a limit on the firm's production capacity.

We explore the implications of increasing marginal cost in a symmetric two-country world, where we can derive analytical solutions. When it is costly to expand production, the advantage of productive firms is diminished: their market share is lower, and more unproductive firms enter. However, the effect of more unproductive firms on welfare is unclear: in a world where expansion is costly, the increase in the mass of firms dominates the increase in price. Therefore, the overall welfare could be the same as in a world where expansion is less costly. The gains from trade formula under the assumption of increasing marginal cost of production is explicitly derived, which extends the formula in Melitz and Redding (2015).

The gains from trade crucially depend on the elasticity of marginal cost. In the simulation of a two-country world, it is shown that a moderate deviation from the constant marginal cost case leads to a dramatic drop in gains from trade from 12% to 2%. It is then verified in a simulation of multiple countries and multiple products, that the model with increasing marginal production cost can reproduce the changing distance elasticity observed in the data.

In an influential paper, Eckel and Neary (2010) formalize the idea the each firm has a product of "core competence" that has the highest efficiency and secondary products with lower efficiency. Recently Arkolakis et al. (2019) document stylized facts of multi-product firms using Brazil data, and estimate the trade costs, in particular the market-access costs, faced by exporters at product level. Their model builds on the idea of Eckel and Neary (2010) and others, but leaves the within-firm entry decisions separable across products. The empirical fact documented in this paper complements those shown in Arkolakis et al. (2019), and the current paper also uses the idea of that efficiency differs across products within a firm, but the focus here is to study the implications when the operational decisions across products depend on each other.

In a closely related paper, Nocke and Yeaple (2014) study the implications of imposing capacity constraint on firms. In their model, firms explicitly allocate capital resources to symmetric products, which lowers the marginal cost of production. Therefore introduction of more products results in a higher marginal cost for each product. Their purpose is to analyze the relationship between firm scope and various firm performance measures. The assumption of increasing marginal production cost imposed in the current paper could be interpreted as an alternative way of modeling firm's capacity constraint. However, our focus is on the aggregate trade patterns related to product asymmetry, which is not allowed in their analysis.

In two important papers on the welfare gains from trade, Arkolakis et al. (2012, ACR) and Melitz and Redding (2015) discuss how to derive welfare changes given trade cost changes in a large class of trade models. In symmetric two-country case, our analysis of welfare implications generalizes their results. While their focus is on comparing a heterogeneous firms model with a homogeneous firms model, we are interested in the implications of relaxing the assumption of constant marginal costs.

This paper contributes to the literature studying how does trade volume respond to distance. A well known puzzle is that distance elasticity is often found to be increasing over time, despite the significant improvements in transportation and communication technologies in the past decades, as documented in Disdier and Head (2008); Berthelon and Freund (2008). Increasing marginal cost of production provides another source of friction, therefore could potentially explain why the distance elasticity is consistently high, as argued in Fabinger and Weyl (2018). This paper extends the idea of Fabinger and Weyl (2018) to explain behavior of multi-product firms, explores the heterogeneity of distance elasticity across products within firm, and provides support for such argument.

Finally, Bernard et al. (2010), Timoshenko (2015) and Sheveleva and Krishna (2018) study the dynamics of product level entry and exit. Although not incorporated explicitly in this paper, the interdependence across products has natural implications to the dynamic questions studied in those papers.

2 The Changing Distance Elasticity

In this section, we document that the entry decisions of lower ranking products are more distance-sensitive using the data on the universe of Chinese international export transactions in the year 2006. In order to compare our distance elasticities with the vast literature on the gravity equation, we report results using aggregated trade flows. And in order to reconcile our results with the vast literature on heterogeneous firms models, we also report results from firm-level regressions.

To estimate the regression on the aggregated trade flows data, total exports from province i (in China) to destination j (that is, a foreign country) are partitioned into sub-samples of firm-products defined by the product scope of a firm, $G_{ij}(\omega)$, and the product rank within a firm, $g_{ij}(\omega)$. Distance elasticity is estimated for each sub-sample from the regression

$$\ln X_{ijgG} = \beta_{0,gG} + \beta_{1,gG} \ln Distance_{ijgG} + \phi_{i,gG} + \phi_{j,gG} + u_{ijgG}, \tag{1}$$

where X_{ijgG} is total export value from province i to destination j for firms that export (in total) G products and for product ranking g within each firm. E.g., X_{ij12} is total export value of the top products for those firms that export 2 products from province i to destination j. Using the firm-level data, we can further decompose the total export value into extensive and intensive margins:

$$X_{ijgG} = N_{ijgG} \frac{X_{ijgG}}{N_{ijgG}}.$$

Table 1 (Table 2) shows estimates of $\beta_{1,gG}$ for extensive (intensive) margin. Due to space limitation, we report estimates for samples up to firms scope equal to 16. It is visually clear that the magnitude of the elasticities is increasing in product rank, conditional on firm scope for extensive margin, while the same is not true for the intensive margin. In order to test whether the change of distance elasticity is significant, we run the following OLS regression:

$$\beta^M_{1,gG} = \alpha_0 + \alpha_1 g + \alpha_2 G + \alpha_3 g \times G + u_{gG} \tag{2}$$

where α_1 captures the relation between product ranking and sensitivity to distance, α_2 captures relation between firm scope and sensitivity to distance, α_3 tests whether the slope changes. Table 3 verifies that the increasing distance sensitivity for lower-ranked products in the extensive margin. Interestingly, the effects of product scope, conditional on product rank, are significant and opposite for the intensive and extensive margins: the products of firms with higher scope are less sensitive to distance at the extensive margin, but more sensitive to distance at the intensive margin.

We also explore whether the phenomenon exists at the firm level by estimating the regression:

$$\ln X_{fgj} = \gamma_0 + \gamma_1 \ln Distance_{ij} + \gamma_2 g_{fj} + \gamma_3 \ln Distance_{ij} \times g_{fj} + \phi_f + \phi_j + u_{fgj}, \tag{3}$$

where X_{fgj} is export value to destination j of firm f's product with rank g, g_{fj} is the product's rank within firm f. Firm and destination fixed effects are controlled. As shown in Table 4, the interaction effect is negative and significant, consistent with the aggregate-level evidence.

The patterns documented in Table 3 are not easily explained by the standard models of multi-product firms (e.g., Arkolakis et al. (2019); Mayer et al. (2014); Bernard et al. (2011)). To see this, we can derive gravity equations, in particular for the bilateral extensive margin, in these models for each combination of product scope and rank. The trade elasticities in all these gravity equations will be the same, unless goods of different ranks differ systematically in the underlying parameters. But the question is then why products differ. One potential explanation is that higher ranking products tend to have higher quality, and as argued in the literature on the Alchian-Allen effect (e.g., Hummels and Skiba (2004)), the presence of additive trade costs makes it more attractive to export high quality goods to further destinations. Another potential explanation from the product cycle theory is that top ranking products are newly developed products, and therefore may have high variance in the production technology, which will be reflected in a lower distance elasticity in a Eaton and Kortum (2002) type model (e.g. Fieler (2011)). In order to assess both explanations, we repeat the above exercise for different groups of goods classified according to the Rauch classification. Because both arguments depend on the vertical differentiation of products, we would expect that homogeneous products do not exhibit the same pattern if these explanations were sufficient. However, as shown in Table 5 and Table 6, the estimate comparable to the α_1 in the previous regression is negative and significant for both samples of homogeneous and differentiated products. Interestingly, the intensive margin of lower ranking homogeneous products are significantly less sensitive to distance. In either way, the changing distance elasticity contradicts the assumption that trade elasticity is constant. The interaction effect of the firm-level regression is also negative, while the magnitude is much smaller. Overall, these evidence shows that it is unlikely that the systematic difference across products in terms of vertical differentiation explains the changing distance elasticity.

3 A Trade Model with Increasing Marginal Cost of Production

In this section we build a heterogeneous firms trade model with increasing marginal cost of production and show that this model can generate the changing distance elasticity consistent with the empirical evidence of the previous section. We also explore the theoretical implications of introducing increasing marginal cost of production. In order to make the intuition clearer, we start with a closed economy where each firm only produces one product. Then we explore the behavior of single-product firms in a symmetric two-country world. Finally, we construct a general model of multi-product firms in an asymmetric multi-country world.

3.1 Closed Economy and Single Product Firms

We first consider the closed economy case where each firm produces only one product. In this and next sub-sections, we normalize wage to 1. The setting of demand side is the same as the standard Melitz model. Consumers are homogeneous, each has a constant elasticity of substitution (CES) utility function

$$U = \left[\int_{\omega \in \Omega} q\left(\omega\right)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}},$$

where Ω is the set of varieties available, σ is the elasticity of substitution between goods. The corresponding price index is

$$P = \left[\int_{\omega \in \Omega} p\left(\omega\right)^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}}.$$

Given the demand function, each firm's revenue function can be written as

$$X_d\left(q\right) = \kappa \nu^{-1} q^{\nu}, \tag{4}$$

where $\nu = 1 - \frac{1}{\sigma}$, the demand shifter $\kappa = \nu P^{\nu} L^{1-\nu}$ captures the country size L and the price index P, q is the quantity produced by the firm. The total variable cost of producing q is

$$VC(q) = a\rho^{-1}q^{\rho}, \tag{5}$$

where a is the measure of firm's efficiency that is drawn from some distribution $Q(a)$, ρ is the parameter that controls how fast variable costs increase with total production. With $\rho = 1$, marginal cost of production will be constant and equal to a. In order to produce a positive quantity, the firm also needs to incur a fixed cost f_d. This paper focuses on the implications when $\rho > 1$, namely when the marginal cost of production is increasing in the quantity produced. In the short run, this structure serves as a convenient way of modeling capacity constraint or financial constraint that prevents firms from quick expansion. In the long run, this specification mainly captures the increasingly diluted attention of the top managers that accompanies the expansion of scales of production.

The profit function is therefore

$$\max_{q} \pi_d(q) = \kappa\nu^{-1}q^{\nu} - a\rho^{-1}q^{\rho} - f_d, \tag{6}$$

which implies the first-order condition

$$q_d(a) = \left(\frac{\kappa}{a}\right)^{\frac{1}{\rho-\nu}}. \tag{7}$$

Substituting this expression into the revenue function, we get

$$X_d(a) = \nu^{-1}\kappa^{\frac{\rho}{\rho-\nu}}a^{-\frac{\nu}{\rho-\nu}} \tag{8}$$

and

$$p_d(a) = \frac{X_d(a)}{q_d(a)} = \nu^{-1}\kappa^{\frac{\rho-1}{\rho-\nu}}a^{\frac{1-\nu}{\rho-\nu}}. \tag{9}$$

When $\rho > 1$, because the endogenous demand shifter κ also enters the price equation, the firm level price also depends on the firm entry, which affects average efficiency. It is easy to verify that the markup in this model is constant over the marginal variable costs:

$$\mu = \frac{p_d - VC'(q_d)}{p_d} = 1 - \nu.$$

The model can be solved by examining the zero-profit condition and the free entry condition. The zero-profit cutoff in this closed economy a^A is defined by

$$\pi_d\left(q_d\left(a^A\right)\right)=0, \tag{10}$$

from which we can derive

$$X_d\left(a^A\right)=\frac{f_d}{(1-\nu\rho^{-1})}. \tag{11}$$

The free entry condition is

$$Q\left(a^A\right)\bar{\pi}=f_e, \tag{12}$$

where f_e is the entry cost, and $\bar{\pi}$ is the expected profit conditional on that the draw of marginal cost is smaller or equal to a^A. Using Equation (11), we can derive

$$f_d\int_{a_{min}}^{a^A}\left(\left(\frac{a}{a^A}\right)^{-\frac{\nu}{\rho-\nu}}-1\right)dQ\left(a\right)=f_e, \tag{13}$$

which can be used to solve for a^A as a function of exogenous parameters of the model. It is convenient to express all other endogenous parameters of the model as the zero profit cutoff a^A. To make the expressions more compact, define the average productivity measure as

$$\tilde{a}^A\left(b,c\right)\equiv\left[\int_b^c\left(a\right)^{-\frac{\nu}{\rho-\nu}}\frac{dQ\left(a\right)}{Q\left(a^A\right)}\right]^{-\frac{\rho-\nu}{\nu}}. \tag{14}$$

One can show that the Equation (13) is equivalent to

$$\left(\frac{\tilde{a}^A\left(a_{min},a^A\right)}{a^A}\right)^{-\frac{\nu}{\rho-\nu}}=\frac{f_e}{f_dQ\left(a^A\right)}+1. \tag{15}$$

Using this relation, we can solve for the mass of firms in this economy as a function of a^A and other parameters:

$$N^A=\left(1-\nu\rho^{-1}\right)\frac{L}{F^A}, \tag{16}$$

where $F^A\equiv\frac{f_e}{Q\left(a^A\right)}+f_d$. Finally, the price index can be shown to be given by

$$P^A=\nu^{-\frac{1}{\rho}}L^{-\frac{1-\nu}{\nu}}\left[\frac{\left(1-\nu\rho^{-1}\right)}{F^A}\left[\tilde{a}^A\left(a_{min},a^A\right)\right]^{-\frac{\nu}{\rho-\nu}}\right]^{-\frac{\rho-\nu}{\nu\rho}}, \tag{17}$$

and therefore the demand shifter in the autarky is equal to

$$\kappa^A \equiv \nu^{\frac{\rho-\nu}{\rho}} \left[\frac{\left(1-\nu\rho^{-1}\right)}{F^A} \left[\tilde{a}^A \left(a_{min}, a^A \right) \right]^{-\frac{\nu}{\rho-\nu}} \right]^{-\frac{\rho-\nu}{\rho}}.$$

As an important benchmark, we provide further results under the assumption that the distribution of marginal costs follows an untruncated Pareto distribution:

Proposition 1. *If $\frac{1}{a}$ follows an untruncated Pareto distribution, namely*

$$Q(a) = \left(\frac{a}{a_{max}} \right)^{\theta}, \tag{18}$$

where a_{max} is the parameter that specifies the marginal cost of the least efficient firm, then

$$a^A = a_{max} \left(\frac{f_e}{f_d} \right)^{\frac{1}{\theta}} \left[\frac{\theta(\rho-\nu)-\nu}{\nu} \right]^{\frac{1}{\theta}},$$

$$N^A = \left(\frac{\rho-\nu}{\rho} \right) \left(-\frac{\nu}{\rho-\nu} + \theta \right) \frac{L}{\theta f_d}.$$

Moreover, $\frac{dN^A}{d\rho} > 0$, and $\frac{da^A}{d\rho} > 0$.

The proof of Proposition 1 is straightforward. At first glance one might expect that increasing ρ will reduce welfare because the production becomes more costly. But the expressions in Proposition 1 show that as ρ increases, there are two opposing forces that affect welfare: higher ρ reduces the advantage of more productive firms, and therefore induces more entry, which increases welfare through more variety. On the other hand, with higher ρ less productive firms get larger market share, and, therefore, the average prices increase, which reduces welfare.

3.2 Symmetric Two Countries and Single Product Firms

In this sub-section we consider a world with two symmetric countries. As we will show below, the entry decisions into the domestic and foreign markets depend on each other. To emphasize this, we write the entry decision explicitly as a length-2 vector

$$D = (1\{q_d > 0\}, 1\{q_x > 0\}),$$

where the first (second) dummy is 1 if the firm enters domestic (foreign) market. Under the symmetry assumption, the domestic market will always be more profitable than the foreign market because of lower trade costs, therefore the firm only needs to evaluate profit under $D = (1,0)$ and $D = (1,1)$. If both choices generate negative profit, the firm chooses not to produce. The assumptions on demand and production technology are the same as the autarky case, except that export is subject to an iceberg trade costs τ, meaning that τ units of goods are needed in order for one unit of good to arrive in the foreign market. The profit function is

$$\pi(\boldsymbol{q}, D) = \begin{cases} \kappa\nu^{-1}\left(q_d^\nu + q_x^\nu\right) - a\rho^{-1}q_f^\rho - (f_d + f_x) & D = (1,1) \\ \kappa\nu^{-1}q_d^\nu - a\rho^{-1}q_f^\rho - f_d & D = (1,0) \end{cases}$$

where $q_f(\boldsymbol{q}, D) = D' \begin{pmatrix} q_d \\ \tau q_x \end{pmatrix}$ is the total quantity sold. Using the first order condition of profit maximization, the revenue functions can be written as

$$X_d(a, D) = \begin{cases} \nu^{-1}a^{-\frac{\nu}{\rho-\nu}}(\kappa)^{\frac{\rho}{\rho-\nu}}\, t & D = (1,1) \\ \nu^{-1}a^{-\frac{\nu}{\rho-\nu}}(\kappa)^{\frac{\rho}{\rho-\nu}} & D = (1,0) \end{cases}$$

$$X_x(a, D) = \begin{cases} \nu^{-1}a^{-\frac{\nu}{\rho-\nu}}(\kappa)^{\frac{\rho}{\rho-\nu}}\tau^{-\frac{\nu}{1-\nu}}t & D = (1,1) \\ 0 & D = (1,0) \end{cases},$$

where the term $t = \left(1 + \tau^{-\frac{\nu}{1-\nu}}\right)^{-\frac{\nu(\rho-1)}{\rho-\nu}}$ shows the effect of exporting on the domestic sales. When $\rho > 1$, $t < 1$, therefore the presence of t introduces a negative externality of exporting on the domestic sales. This in turn implies a subtler entrance decisions. On one hand, the entrance into the foreign market brings new source of revenue, but it also raises the marginal cost of production for goods sold domestically. The profit function can be written as

$$\pi(a, D) = \begin{cases} \left[X_d(a, D) + X_x(a, D)\right]\left[1 - \nu\rho^{-1}\right] - (f_d + f_x) & D = (1,1) \\ X_d(a, D)\left[1 - \nu\rho^{-1}\right] - f_d & D = (1,0) \end{cases}$$

The profit is increasing in productivity $\frac{1}{a}$. The solution of the model is characterized by two zero profit cutoffs. The zero profit cutoff of domestic market a_d^T is defined by

$$\pi\left(a_d^T, (1,0)\right) = X_d\left(a_d^T, (1,0)\right)\left[1 - \nu\rho^{-1}\right] - f_d = 0,$$

while the zero profit cutoff for exporting a_x^T will be defined by the condition that entering both markets is equally profitable as only selling in the domestic market:

$$\pi\left(a_x^T,(1,1)\right)=\pi\left(a_x^T,(1,0)\right),$$

which is equivalent to the condition

$$\left[X_d\left(a_x^T,(1,1)\right)+X_x\left(a_x^T,(1,1)\right)\right]\left(1-\nu\rho^{-1}\right)-(f_d+f_x)$$

$$=X_d\left(a_x^T,(1,0)\right)\left(1-\nu\rho^{-1}\right)-f_d.$$

Note that when $\rho>1$, we cannot cancel the terms related to sales in the domestic market from the both sides because $X_d\left(a_x^T,(1,1)\right)\neq X_d\left(a_x^T,(1,0)\right)$. From the definition, it can be shown that the ratio of the two cutoffs satisfies

$$\left(\frac{a_x^T}{a_d^T}\right)=\left(\frac{f_d}{f_x}\right)^{\frac{\rho-\nu}{\nu}}\left[\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)t-1\right]^{\frac{\rho-\nu}{\nu}}. \tag{19}$$

The free entry condition is similarly defined as

$$Q\left(a_d^T\right)\bar{\pi}=f_e, \tag{20}$$

from which we can derive

$$f_d\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)^{\frac{(1-\nu)\rho}{\rho-\nu}}\left(\frac{\tilde{a}^T\left(a_{min},a_x^T\right)}{a_d^T}\right)^{-\frac{\nu}{\rho-\nu}}+f_d\left(\frac{\tilde{a}^T\left(a_x^T,a_d^T\right)}{a_d^T}\right)^{-\frac{\nu}{\rho-\nu}}$$

$$=f_x\frac{Q\left(a_x^T\right)}{Q\left(a_d^T\right)}+f_d+\frac{f_e}{Q\left(a_d^T\right)},$$

where

$$\tilde{a}^T(b,c)=\left[\int_b^c\left(a^{-\frac{\nu}{\rho-\nu}}\right)d\frac{Q(a)}{Q\left(a_d^T\right)}\right]^{-\frac{\rho-\nu}{\nu}}.$$

The Equations (19) and (20) determine the zero profit cutoffs as functions of exogenous parameters. Similar to the autarky case, we can express the mass of firms by the cutoffs and other parameters as

$$N^T=\left(1-\nu\rho^{-1}\right)\frac{L}{F^T}, \tag{21}$$

where $F^T \equiv f_x \frac{Q\left(a_x^T\right)}{Q\left(a_d^T\right)} + f_d + \frac{f_e}{Q\left(a_d^T\right)}$. Finally, the price index is given by

$$P^T = \nu^{-\frac{1}{\rho}} L^{-\frac{1-\nu}{\nu}} \left[\frac{\left(1 - \nu\rho^{-1}\right)}{F^T} \left(\tilde{a}_t^T\right)^{-\frac{\nu}{\rho-\nu}} \right]^{-\frac{\rho-\nu}{\nu\rho}} \tag{22}$$

where

$$\left(\tilde{a}_t^T\right)^{-\frac{\nu}{\rho-\nu}} = \left[\left(1 + \tau^{-\frac{\nu}{1-\nu}}\right)^{\frac{(1-\nu)\rho}{\rho-\nu}} \tilde{a}^T \left(a_{min}, a_x^T\right)^{-\frac{\nu}{\rho-\nu}} + \tilde{a}^T \left(a_x^T, a_d^T\right)^{-\frac{\nu}{\rho-\nu}} \right].$$

3.2.1 Gains From Trade Under Increasing Marginal Cost of Production

Next we derive changes in welfare due to changes in trade costs, under a general value of ρ. The expenditure share on domestic products is

$$\lambda = \frac{\int_{a_{min}}^{a_x^T} X_d\left(a, (1,1)\right) dQ\left(a\right) + \int_{a_x^T}^{a_d^T} X_d\left(a, (1,0)\right) dQ\left(a\right)}{\int_{a_{min}}^{a_x^T} X_d\left(a, (1,1)\right) dQ\left(a\right) + \int_{a_x^T}^{a_d^T} X_d\left(a, (1,0)\right) dQ\left(a\right) + \int_{a_{min}}^{a_x^T} X_x\left(a, (1,1)\right) dQ\left(a\right)},$$

which can be written as

$$\lambda = \frac{1}{1 + \tau^{-\frac{\nu}{1-\nu}} \Lambda}, \tag{23}$$

where

$$\Lambda \equiv \frac{t \int_{a_{min}}^{a_x^T} a^{-\frac{\nu}{\rho-\nu}} dQ\left(a\right)}{t \int_{a_{min}}^{a_x^T} a^{-\frac{\nu}{\rho-\nu}} dQ\left(a\right) + \int_{a_x^T}^{a_d^T} a^{-\frac{\nu}{\rho-\nu}} dQ\left(a\right)}$$

represents the relative sales of exporters in the domestic market. Denote

$$\delta\left(c\right) = t \int_{a_{min}}^{a_x^T} a^{-\frac{\nu}{\rho-\nu}} dQ\left(a\right) + \int_{a_x^T}^{c} a^{-\frac{\nu}{\rho-\nu}} dQ\left(a\right),$$

then $\Lambda \equiv \frac{\delta\left(a_x^T\right)}{\delta\left(a_d^T\right)}$, and $\delta\left(a_d^T\right)$ is proportional to the sales of domestic firms in the domestic market, $\delta\left(a_x^T\right)$ is proportional to the sales of foreign firms in the domestic market. We follow Melitz and Redding (2015) to classify the effects of changes in trade costs into the direct and the indirect effect. The direct effect refers to the impact of trade cost changes on a_x^T conditional on a_d^T, which can be seen from Equation (19). The indirect effect refers to the effect via changes of a_d^T. When $\rho > 1$, an immediate consequence of the presence of negative externality t is that the direct effect becomes more complicated, because domestic sales are affected by trade costs even conditional on a_d^T. This fact translates into a more complicated expression for partial trade elasticity, which is defined as

$$\epsilon \equiv -\frac{d\ln\left(\frac{1-\lambda}{\lambda}\right)}{d\ln\tau}\Bigg|_{a_d^T} = \frac{\nu}{1-\nu} - \frac{d\ln\Lambda}{d\ln a_x^T}\frac{d\ln a_x^T}{d\ln\tau}\Bigg|_{a_d^T}.$$

The definition is consistent with typical estimation strategy using a gravity equation, which holds the domestic zero profit cutoffs constant. When $\rho > 1$, $\frac{d\ln a_x^T}{d\ln\tau}\Big|_{a_d^T} = \rho\frac{t\tau^{-\frac{\nu}{1-\nu}}}{\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)t-1}$ instead of being 1 as in the case with $\rho = 1$. Moreover, we see that when $\rho > 1$, the derivative $\frac{\partial\ln\delta\left(a_d^T\right)}{\partial\ln a_x^T}\Big|_{a_d^T}$, which measures how the cumulative market share responses to a change in the zero profit cutoff, is no longer 0.

The welfare is measured using the real consumption per capita W^T, which is equal to the inverse of the price index, because the wage is normalized to 1. Using equations (21), (22), (23), one can show that

$$d\ln W^T = \frac{\rho-\nu}{\nu\rho}\left[d\ln N^e - d\ln\lambda + d\ln\delta\left(a_d^T\right)\right],$$

and using $d\ln W^T = -\frac{1}{\rho}d\ln\left(a_d^T\right)$, which can be derived from the definition of zero profit cutoff and Equation (19), we get

$$d\ln W^T = \frac{d\ln N^e - d\ln\lambda}{\epsilon - \rho\left[\frac{t\tau^{-\frac{\nu}{1-\nu}}}{\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)t-1}\frac{d\left(\ln\delta\left(a_x^T\right)-\ln\delta\left(a_d^T\right)\right)}{d\ln a_x^T}\Bigg|_{a_d^T} - \frac{d\ln\delta\left(a_d^T\right)}{d\ln a_d^T}\right]+C},$$

where $C = -\frac{\nu}{1-\nu} + \frac{\nu\rho}{\rho-\nu}$. When $\rho = 1$, the constant term C is zero, and $\frac{t\tau^{-\frac{\nu}{1-\nu}}}{\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)t-1}$ becomes 1, and the term proportional to total domestic sales $\delta\left(a_d^T\right)$ does not depend on a_x^T, therefore the formula reduces to the one derived in Melitz and Redding (2015). This extended formula shows how negative externality t affects welfare implication of trade cost changes. Firstly, it enters directly through the term $\frac{t\tau^{-\frac{\nu}{1-\nu}}}{\left[1+\tau^{-\frac{\nu}{1-\nu}}\right]t-1}$, secondly it enters through the derivatives of $\ln\delta\left(a_d^T\right)$. To see it more clearly, assume that the productivity follows the Pareto distribution, $Q(a) = \frac{a^{\theta} - a_{min}^{\theta}}{a_{max}^{\theta} - a_{min}^{\theta}}$, then one can show that

$$\frac{\partial \ln \delta\left(a_d^T\right)}{\partial \ln a_d^T} = \frac{\left(-\frac{\nu}{\rho-\nu} + \theta\right)\left[(t-1)\left(a_x^T\right)^{-\frac{\nu}{\rho-\nu}+\theta} + \left(a_d^T\right)^{-\frac{\nu}{\rho-\nu}+\theta}\right]}{\left[(t-1)\left(a_x^T\right)^{-\frac{\nu}{\rho-\nu}+\theta} - t\left(a_{min}\right)^{-\frac{\nu}{\rho-\nu}+\theta} + \left(a_d^T\right)^{-\frac{\nu}{\rho-\nu}+\theta}\right]}.$$

Therefore when $t < 1$, the cutoff of export a_x^T also affects the value of the derivative. Because of this feature, even under untruncated Pareto distribution (namely when $a_{min} = 0$) this derivative is not a constant. And, as a result, the partial trade elasticity is not a constant either.

3.3 Changing distance elasticity explained

In this sub-section, we use a simple extension of the symmetric two country model that allows each firm to produce up to two products to explain the changing distance elasticity documented in Section 2. The demand side is very similar to the previous cases except that an additional layer of summation is added into the consumer's utility function (more details can be found in Appendix A). On the supply side, the total variable cost of production is assumed to be

$$VC\left(a, q_f\right) = a\rho^{-1} q_f^{\rho},$$

where q_f is the effective aggregate quantity across all varieties within the firm given by

$$q_f = \sum_{g=1}^{G_d(a)} \left(\frac{q_{dg}(a, D)}{z(g)}\right) + \tau\left[\sum_{g=1}^{G_x(\omega)} \left(\frac{q_{xg}(a, D)}{z(g)}\right)\right], \tag{24}$$

where $g = 1, 2$ is the rank of products, $z(g)$ measures the efficiency of product g, $G_d(a)$ and $G_x(a)$ are the numbers of products selling in the domestic and the foreign markets respectively. This specification implicitly assumes that different products are perfectly substitutable in terms of occupying the firm's capacity, except for the efficiency difference embedded in $z(g)$. Admittedly, different product lines may be produced in different factories, and occupy different resources, therefore a more satisfying specification should take into account the distance between products in the competition for the firm-wise resources. But this simple setting delivers a clearer intuition of how the increasing marginal cost of production affects the product entry decisions. Under this setting, the firm-level profits are

$$\pi(a,D)=\frac{\left(\nu^{-1}-\rho^{-1}\right)}{a^{\frac{\nu}{\rho-\nu}}}\kappa^{\frac{\rho}{\rho-\nu}}\left\{\left[\sum_{g=1}^{G_d(\omega)}z^{\frac{\nu}{1-\nu}}(g)\right]+\tau^{\frac{\nu}{\nu-1}}\left[\sum_{g=1}^{G_x(\omega)}z^{\frac{\nu}{1-\nu}}(g)\right]\right\}^{\frac{(1-\nu)\rho}{\rho-\nu}}$$
$$-\left(\sum_{g=1}^{G_d(\omega)}f_{dg}+\sum_{g=1}^{G_x(\omega)}f_{xg}\right),$$

where the entry decision D is now a 2×2 matrix, with the first row denoting dummies representing entry decisions of two products in the domestic market, the second row denoting dummies representing entry decisions in the foreign market. Since destinations are symmetric, and the domestic market has lower trade costs, any product will be first introduced in the domestic market. Without loss of generality, we assume that product 1 is more efficient. In addition, for exposition simplicity, we assume that the difference in efficiency is big enough that a firm producing product 2 will always export the product 1 to the foreign market. In this case, the criterion of exporting the product 1 is

$$\pi\left(a,D=\begin{pmatrix}1&1\\0&0\end{pmatrix}\right)\geq\pi\left(a,D=\begin{pmatrix}1&0\\0&0\end{pmatrix}\right),$$

while the criterion for exporting the product 2 is

$$\pi\left(a,D=\begin{pmatrix}1&1\\1&1\end{pmatrix}\right)\geq\pi\left(a,D=\begin{pmatrix}1&1\\1&0\end{pmatrix}\right).$$

To see why the second product is more sensitive to trade costs, calculate the difference

$$\pi\left(a,D=\begin{pmatrix}1&1\\1&1\end{pmatrix}\right)-\pi\left(a,D=\begin{pmatrix}1&1\\1&0\end{pmatrix}\right)$$

$$=\frac{\left(\nu^{-1}-\rho^{-1}\right)}{a^{\frac{\nu}{\rho-\nu}}}\kappa^{\frac{\rho}{\rho-\nu}}\left\{\left\{\sum_{g=1}^{2}z^{\frac{\nu}{1-\nu}}(g)+\tau^{-\frac{\nu}{1-\nu}}\sum_{g=1}^{2}z^{\frac{\nu}{1-\nu}}(g)\right\}^{\frac{1-\nu}{1-\frac{\nu}{\rho}}}-\left\{\sum_{g=1}^{2}z^{\frac{\nu}{1-\nu}}(g)+\tau^{-\frac{\nu}{1-\nu}}\left[z^{\frac{\nu}{1-\nu}}(1)\right]\right\}^{\frac{1-\nu}{1-\frac{\nu}{\rho}}}\right\}-f_{x2}.$$

When $\rho>1$, we have that $\frac{1-\nu}{1-\frac{\nu}{\rho}}<1$, and therefore the function $h(x)=x^{\frac{1-\nu}{1-\frac{\nu}{\rho}}}$ is concave. As a result, net increase in profit will be lower when the sales of the existing products are higher. That means the second product faces higher friction upon introduction even the trade costs are the same. Intuitively, the higher friction comes from the negative externality imposed on the first product if the second product is introduced. In addition, as shown in Appendix A, given both products are exported, the optimal quantity sold to the foreign market is given by

$$q_{xg}\left(a\right)=a^{-\frac{1}{\rho-\nu}}\left(\frac{\kappa z\left(g\right)}{\tau}\right)^{\frac{1}{1-\nu}}\left\{\kappa^{\frac{1}{1-\nu}}\left[\sum_{g=1}^{2}\left(z\left(g\right)\right)^{\frac{\nu}{1-\nu}}\right]\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)\right\}^{-\frac{\rho-1}{\rho-\nu}},$$

so that trade costs affect the volume exported in a similar way for either product 1 or product 2. Through the mechanism discussed above, our model is able to explain the fact that extensive margin of trade of a lower-ranked product is more sensitive to distance, while the intensive margin is not.

4 Quantitative Evidence

4.1 Welfare Implications in a Symmetric Two-country World

In this sub-section, we show the welfare implication of increasing marginal cost of production in a symmetric two-country world. We assume that the productivity parameter $\frac{1}{a}$ follows an untruncated Pareto distribution, or $Q\left(a\right)=\frac{a^{\theta}}{a_{max}^{\theta}}$. In this case we can solve for the zero profit cutoffs in autarky as

$$a^{A}=a_{max}\left(\frac{f_e}{f_d}\right)^{\frac{1}{\theta}}\left(\frac{\theta\left(\rho-\nu\right)-\nu}{\nu}\right)^{\frac{1}{\theta}},$$

and the zero profit cutoff in trade as

$$a_d^{T}=a^{A}\left\{\left(\frac{f_d}{f_x}\right)^{\left(-1+\theta\frac{\rho-\nu}{\nu}\right)}\left[\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)t-1\right]^{\left(\theta\frac{\rho-\nu}{\nu}\right)}+1\right\}^{-\frac{1}{\theta}}.$$

Then the zero profit cutoff of foreign market in trade equilibrium can be derived using Equation (19). The welfare gains from going from autarky to a particular level of openness are

$$\frac{W^{T}}{W^{A}}=\left\{\left(\frac{f_d}{f_x}\right)^{\left(-1+\theta\frac{\rho-\nu}{\nu}\right)}\left[\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)t-1\right]^{\left(\theta\frac{\rho-\nu}{\nu}\right)}+1\right\}^{\frac{1}{\theta\rho}}. \tag{25}$$

The domestic trade share has the form

$$\lambda = \frac{t + \left(\frac{f_d}{f_x}\right)^{\left(1-\theta\frac{\rho-\nu}{\nu}\right)} \left[\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)t-1\right]^{\left(1-\theta\frac{\rho-\nu}{\nu}\right)} - 1}{\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)t + \left(\frac{f_d}{f_x}\right)^{\left(1-\theta\frac{\rho-\nu}{\nu}\right)} \left[\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)t-1\right]^{\left(1-\theta\frac{\rho-\nu}{\nu}\right)} - 1}.$$

Therefore when $\rho = 1$, the term inside the bracket of Equation (25) reduces to the domestic trade share $\lambda|_{\rho=1}$, and in this case the trade elasticity is a constant determined by the shape of the distribution θ, therefore the gains from trade are given by the ACR formula. However, when $\rho > 1$, the expression of λ is different from the term inside the bracket of Equation (25). Moreover, the partial trade elasticity is given by:

$$\epsilon = \frac{\nu}{1-\nu} + \frac{\rho\left(-\frac{\nu}{\rho-\nu}+\theta\right)t\tau^{-\frac{\nu}{1-\nu}}}{(t-1)\left(\left(\frac{f_d}{f_x}\right)\left[\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)t-1\right]\right)^{\left(\frac{\rho-\nu}{\nu}\theta\right)} + \left[\left(1+\tau^{-\frac{\nu}{1-\nu}}\right)t-1\right]}. \tag{26}$$

It is easy to verify that this expression reduces to θ when $\rho = 1$. When $\rho > 1$, it is a function of variable trade cost τ and the relative fixed costs $\frac{f_d}{f_x}$.

Table 7 shows the parameter values used in this simulation. We follow the trade literature to choose elasticity of substitution to be 5. In later exercise of exploring the effect of changing trade costs on welfare, we follow the calibration result in Fabinger and Weyl (2018) and fix $\rho = 1.25$, which implies that expanding scale by one log point would roughly increase marginal cost by 0.25 log pint, or roughly 25%. $\theta = 5$ is consistent with the estimates in the firm size distribution literature. The choices of fixed costs are made such that the variation of quantity of interest is driven by the variation in τ and ρ and their interaction effects.

Figure 1 compares the model solutions under different values of ρ. There are three patterns worth noting. Firstly, it shows that both a^A and a_d^T are increasing in ρ. Intuitively, higher speed at which marginal cost is increasing limits the advantage of high productivity firms, therefore less productive firms can also survive in the market. This is also reflected in the second panel, which shows that the mass of firm is increasing in ρ. The second pattern is that while $a_d^T < a^A$, which is due to increasing demand for labor when opening up to trade, the difference decreases as ρ gets larger. This is related to the third pattern in the first panel, that a_x^T is decreasing in ρ. Intuitively, as expanding production becomes more costly, exporting into a new destination becomes more demanding. This is also reflected in the forth panel, which shows that gains from trade is close to zero as ρ becomes high enough. Moreover, the magnitude of gains from trade is very sensitive to ρ: increase ρ from 1 to 1.2 makes gains from trade drop from 12% to a merely 2%. Finally, note that increasing ρ has a non-monotonic effect on welfare, as shown in the third panel. This shows two effects of entrance of less productive firms: less productive firms charge higher prices, which reduces welfare, but larger mass of firms increase welfare. As ρ increases, initially the price effects dominate, but then the mass effects become stronger.

Next we fix ρ, and explore the interaction between a positive ρ and variable trade costs. Under the untruncated Pareto distribution, the trade elasticity is a constant when $\rho = 1$, as shown in the blue circle dots plotted in the left panel of Figure 2. However, when ρ increases to 1.25, there are two interesting changes: firstly, the magnitude of trade elasticity increases dramatically. That means a small increase in τ leads to a much larger decrease in trade. Secondly, the trade elasticity is not a constant any more, instead it first decreases and then increases in variable trade costs. The right panel shows that conditional on the expenditure share on domestic goods, a much smaller value of τ is implied by the model when $\rho > 1$. It implies that trade model with increasing marginal cost of production has the potential to solve the puzzle that inferred trade costs is much higher than observed tariff and transportation costs.

4.2 Explaining the Changing Distance Elasticity

Next we use a numerical simulation to demonstrate that our model is able to deliver the changing distance elasticity observed in the data. We simulate a world with 3 countries and 3 products. Countries are symmetric, therefore $\kappa_j = \kappa$, and we use labor as the numeraire so that $w_i = 1$. Products are ordered by the efficiency $z(g)$, which is assumed to have the form $z(g) = g^{-\beta}$, where $\beta \geq 0$.The matrix of iceberg trade costs, whose elements are τ_{ij}, is equal to

$$\tau = \begin{pmatrix} 1 & 1.05 & 1.05^2 \\ 1.05 & 1 & 1.05 \\ 1.05^2 & 1.05 & 1 \end{pmatrix}.$$

All other parameter values used are listed in Table 8. $\beta = 0.5$ indicates that lower ranking products are less efficient, and the magnitude is consistent with the estimates in Arkolakis et al. (2019). As in the previous exercise, the productivity follows untruncated Pareto distribution. Fixed cost of export is a constant equal to 1. For the purpose of showing that our model is capable of generating the changing distance elasticity, we choose $\kappa = 4$ instead of solving it by general equilibrium conditions.

The simulation algorithm is as follows. First take $N = 10000$ draws are taken from

$Q(a)$ for each origin country. Given a, optimal entry choice $G_{ij}(a)$ is solved by numerically evaluating all possibilities. Note that under symmetry assumption, we can order the destinations and products by potential profits, which greatly reduces the number of necessary evaluations. Then other firm level variables are calculated given the optimal entry choices. Figure 3 shows how firm's responses change with a. The left figure plots the relation between the total number of product-destinations exported and a. As in standard models, more productive firms enter more markets. More interestingly, the right figure shows that, at the points where optimal number of entry decreases, there is a jump of export value for the top product-destination, which is due to the presence of negative externality when $\rho > 1$.

Figure 4 shows that our model can generate the changing distance elasticity. Analogous to the empirical exercise, we decompose the total sales of product 1 and 2, conditional on the firms that export 2 products, into extensive and intensive margins. The left plot shows that while exports of both products decrease due to increasing trade costs, the decrease of product 2 is faster, which implies a higher (in magnitude) distance elasticity. The rest two plots show that while the product 2 has a higher slope for the extensive margin (middle plot), it has a lower slope for the intensive margin (right plot). The variation of the extensive margin is bigger and dominates the variation of the intensive margin. The intuition is clear: when increasing production is costly, the scarce capacity introduces an additional source of friction, so the same amount of increase in trade costs will lead to a larger drop of less efficient products. These are consistent with the patterns that we document in Section 2.

5 Conclusion

In this paper, we document that, conditional on the number of products exported, the distance elasticity of trade increases for products with lower ranks. The pattern is not consistent with the standard multi-product firm models. Nor can it be explained by the arguments using systematic differences of vertical differentiation across products.

We show that a standard heterogeneous firm trade model that allows for increasing marginal cost of production can explain this pattern. The intuition is that the increasing marginal cost introduces additional friction for secondary products, even conditional on the same trade costs. Theoretical implications of increasing marginal costs of production are also explored. Under this assumption, the advantage of productive firms are lower, and more unproductive firms enter the market. We show how to incorporate increasing marginal cost of production into the analysis of welfare gains from trade. Even mild deviation from constant marginal cost of production reduces welfare gains from trade substantially.

When marginal cost of production is increasing, entry decisions into different product-destinations become correlated with each other. This is likely to have meaningful impact on the interpretation of gravity regressions, which take them as independent. Overall, our discussion points out that the magnitude of ρ is very important in evaluating the effect of trade policy.

References

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare, "New trade models, same old gains?," *American Economic Review*, 2012, *102* (1), 94–130.

— **, Sharat Ganapati, and Marc-Andreas Muendler**, "The Extensive Margin of Exporting Products: A Firm-Level Analysis," *SSRN Electronic Journal*, 2019, *1* (858).

Bernard, Andrew B, J Bradford Jensen, Stephen J Redding, and Peter K Schott, "Firms in International Trade," *Journal of Economic Perspectives*, jul 2007, *21* (3), 105–130.

— **, Stephen J Redding, and Peter K Schott**, "Multiple-Product Firms and Product Switching," *American Economic Review*, mar 2010, *100* (1), 70–97.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott, "Multiproduct firms and trade liberalization," *Quarterly Journal of Economics*, 2011, *126* (3), 1271–1318.

Berthelon, Matias and Caroline Freund, "On the conservation of distance in international trade," *Journal of International Economics*, jul 2008, *75* (2), 310–320.

Disdier, Anne-Célia and Keith Head, "The Puzzling Persistence of the Distance Effect on Bilateral Trade," *Review of Economics and Statistics*, 2008, *90* (1), 37–48.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, sep 2002, *70* (5), 1741–1779.

Eckel, Carsten and J Peter Neary, "Multi-Product Firms and Flexible Manufacturing in the Global Economy," *Source: The Review of Economic Studies*, 2010, *77* (1), 188–217.

Fabinger, Michal and E Glen Weyl, "Functional Forms for Tractable Economic Models and the Cost Structure of International Trade," nov 2018, pp. 1–99.

Fieler, Ana Cecília, "Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation," *Econometrica*, jul 2011, *79* (4), 1069–1101.

Head, Keith and Thierry Mayer, "Gravity Equations: Workhorse,Toolkit, and Cookbook," in "Handbook of International Economics," Vol. 4, Elsevier B.V., 2014, pp. 131–195.

Hummels, David and Alexandre Skiba, "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture," *Journal of Political Economy*, 2004, *112* (6), 1384–1402.

Mayer, Thierry, Marc J. Melitz, and Gianmarco I P Ottaviano, "Market size, competition, and the product mix of exporters," *American Economic Review*, 2014, *104* (2), 495–536.

Melitz, Marc J, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

Melitz, Marc J. and Stephen J. Redding, "New Trade Models, New Welfare Implications," *American Economic Review*, mar 2015, *105* (3), 1105–1146.

Nocke, Volker and Stephen Yeaple, "Globalization and multiproduct firms," *International Economic Review*, 2014, *55* (4), 993–1018.

Sheveleva, Lena and Kala Krishna, "Multiproduct Exporters: Learning and Knowing," 2018.

Timoshenko, Olga A., "Product switching in a model of learning," *Journal of International Economics*, 2015, *95* (2), 233–249.

	Extensive margin	Standard Error	Intensive margin	Standard Error
g	-0.016^{***}	(0.002)	0.0078	(0.007)
G	0.003^{***}	(0.000)	-0.004^{***}	(0.001)
$g \times G$	0.000	(0.000)	-0.000	(0.000)
No. Obs	1409		1409	
Adj. R^2	0.315		0.021	

Note: The table shows the results of regression equation (2). It shows the change of distance elasticity across different samples, defined by number of products exported and rank of product. Statistical significance that 1% level based on a two-tailed test is indicated by ***.

Table 3: Test whether distance elasticity changes

	$\ln X_{fgj}$	Standard Error
$\ln Distance_{ij}$	-0.038^{***}	(0.005)
g_{fj}	-0.004^{***}	(0.001)
$\ln Distance_{ij} \times g_{fj}$	-0.001^{***}	(0.000)
No. Obs	2848823	
Adj. R^2	0.038	

Note: The table shows the results of regression equation (3). It shows the change of distance elasticity at firm level. Standard errors are clustered at province level. Firm and destination fixed effects are controlled. Statistical significance at the 1% levels based on two-tailed tests is indicated by ***.

Table 4: Changing distance elasticity: firm level regression

no. prod.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
prod. rank																
1	-0.97	-0.79	-0.78	-0.88	-0.81	-0.81	-0.90	-0.66	-0.63	-0.67	-0.54	-0.49	-0.69	-0.61	-0.68	-0.65
2		-1.24	-1.02	-1.14	-1.00	-0.88	-0.98	-0.74	-0.73	-0.85	-0.74	-0.69	-0.75	-0.66	-0.68	-0.69
3			-1.25	-1.21	-1.05	-1.04	-1.05	-0.94	-0.87	-0.70	-0.76	-0.74	-0.90	-0.73	-0.81	-0.89
4				-1.38	-1.23	-1.17	-1.18	-1.12	-1.01	-0.87	-0.85	-0.88	-1.00	-0.90	-0.73	-0.82
5					-1.29	-1.23	-1.16	-1.04	-1.02	-0.88	-0.88	-0.85	-0.94	-1.04	-0.64	-0.91
6						-1.16	-1.15	-1.12	-1.14	-0.91	-1.00	-1.02	-1.01	-1.11	-0.69	-0.84
7							-1.27	-1.17	-1.08	-1.10	-0.91	-0.87	-1.20	-1.02	-0.93	-1.06
8								-1.13	-1.06	-1.00	-0.99	-0.92	-0.99	-1.26	-0.89	-1.05
9									-1.03	-1.06	-0.86	-1.12	-1.16	-0.93	-0.87	-0.95
10										-1.11	-1.15	-0.96	-1.24	-1.10	-1.13	-0.92
11											-1.07	-1.20	-0.97	-1.01	-0.77	-1.10
12												-1.11	-1.13	-1.05	-0.76	-1.07
13													-1.14	-1.30	-0.97	-1.21
14														-1.18	-1.15	-1.24
15															-1.00	-1.02
16																-0.91

Note: The table shows distance elasticity of extensive margin, the trade flow is aggregated conditional on number of products firm export and the rank of product. Origin and destination fixed effect are controlled.

Table 1: Estimates of β_1 for extensive margin

no. prod.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
prod. rank																
1	-0.51	-0.51	-0.29	-0.53	-0.61	-0.65	-0.91	-0.85	-0.41	-0.30	-0.08	-0.79	-0.97	-0.74	-0.66	-0.70
2		-0.24	-0.14	-0.41	-0.32	-0.26	-0.74	-0.84	-0.06	-0.61	-0.17	-0.66	-0.63	-1.05	-0.74	-0.67
3			-0.05	-0.37	-0.14	-0.56	-0.23	-0.73	-0.19	-0.37	-0.01	0.04	-0.15	-0.75	-0.44	-0.80
4				-0.12	-0.37	-0.27	-0.62	-0.57	-0.29	-0.26	-0.05	0.02	-0.57	-0.64	0.01	-0.83
5					-0.11	-0.06	0.17	-0.55	-0.19	-0.52	0.23	0.11	-0.56	-0.62	0.03	-0.12
6						-0.06	-0.68	-0.85	-0.30	-0.10	0.38	-0.01	-0.34	-0.79	0.41	-0.59
7							-0.37	-0.52	-0.49	0.08	0.09	0.15	-0.33	-0.29	0.16	-0.69
8								-0.53	-0.04	0.47	-0.27	0.50	-0.43	-0.87	-0.04	-0.27
9									-0.71	-0.78	0.33	0.31	-0.33	-1.13	-0.06	-0.60
10										-0.48	-0.37	-0.35	-0.78	-0.58	-0.15	-0.32
11											0.07	-0.10	-0.15	-1.03	0.44	-0.37
12												0.04	-0.22	-0.41	0.34	-0.68
13													-0.72	-0.48	0.03	-1.44
14														-0.75	0.34	-1.09
15															0.70	-0.35
16																-0.94

Note: The table shows distance elasticity of intensive margin, the trade flow is aggregated conditional on number of products firm export and the rank of product. Origin and destination fixed effect are controlled.

Table 2: Estimates of β_1 for intensive margin

	Extensive	Std. Error	Intensive	Std. Error	Extensive	Std. Error	Intensive	Std. Error
g	-0.026^{**}	(0.011)	0.172^{***}	(0.053)	-0.020^{***}	(0.002)	0.004	(0.008)
G	0.024^{***}	(0.002)	0.038^{***}	(0.012)	0.003^{***}	(0.000)	-0.004^{**}	(0.002)
$g \times G$	0.001	(0.001)	-0.012^{***}	(0.003)	-0.000	(0.000)	-0.000	(0.000)
No. Obs	218		218		1236		1236	
Adj. R^2	0.618		0.044		0.413		0.014	
Sample	Homogeneous		Homogeneous		Differentiated		Differentiated	

Note: The table shows the results of regression equation (2), separately for the samples of homogeneous goods and differentiated goods. Homogeneous products are Rauch categories 'goods traded on an organized exchange' or 'reference priced'. Statistical significance at the 5% and 1% levels based on two-tailed tests is indicated by **,***.

Table 5: Changing distance elasticity for differentiated/homogeneous products. Aggregate level.

	Estimate	Standard Error	Estimate	Standard Error
$\ln Distance_{ij}$	-0.032	(0.021)	-0.039^{***}	(0.005)
g_{fj}	-0.003^{***}	(0.001)	-0.004^{***}	(0.001)
$\ln Distance_{ij} \times g_{fj}$	-0.001^{***}	(0.000)	-0.001^{***}	(0.000)
No. Obs	295828		2410969	
Adj. R^2	0.027		0.039	
Sample	Homogeneous		Differentiated	

The table shows the results of regression equation (3) separately for the samples of homogeneous goods and differentiated goods. It shows the change of distance elasticity at firm level. Standard errors are clustered at province level. Firm and destination fixed effects are controlled. Homogeneous products are Rauch categories 'goods traded on an organized exchange' or 'reference priced'. Statistical significance at the 1% levels based on two-tailed tests is indicated by ***.

Table 6: Changing distance elasticity for differentiated/homogeneous products. Firm level.

$\frac{1}{1-\nu}$ (σ)	θ	ρ	τ	
5	5	1.25	1.05	
a_{min}	a_{max}	f_d	f_x	f_e
0	3	1	1	0.1

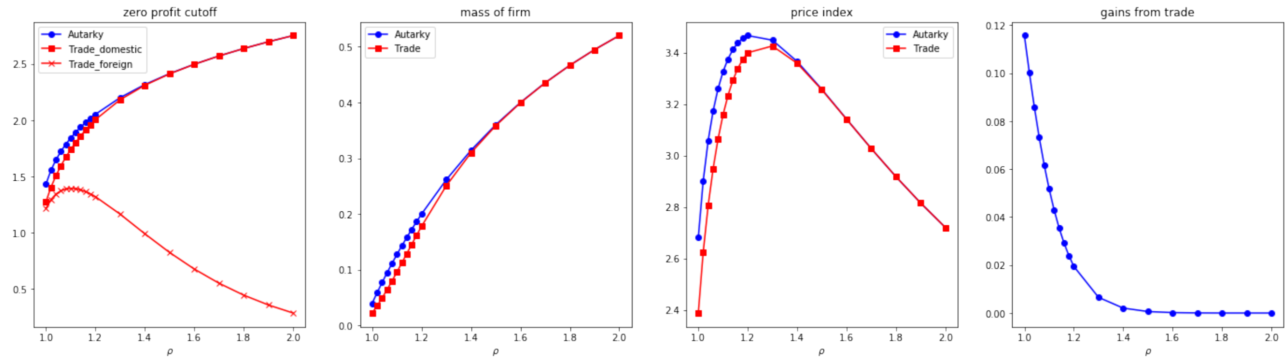
Note: The table shows the parameters values used in the quantitative exercise.

Table 7: Parameter values

$\frac{1}{1-\nu}$ (σ)	θ	ρ	β
5	5	1.25	0.5
a_{min}	a_{max}	f^x_{igj}	κ
0	3	1	4

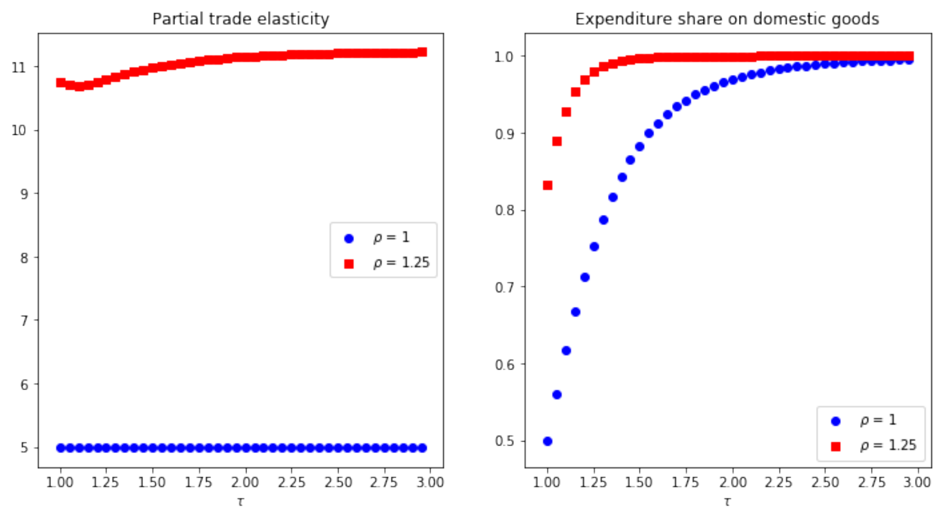
Note: The table shows the parameters values used in the quantitative exercise.

Table 8: Parameter values



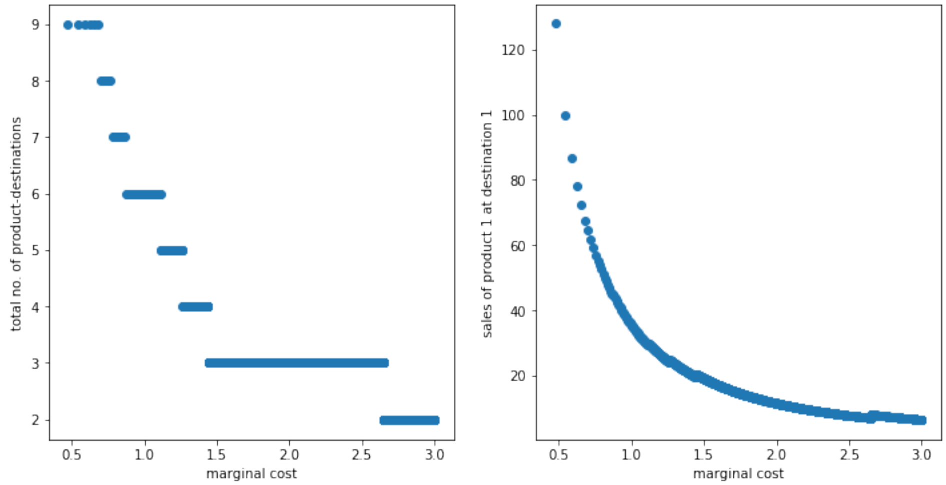
Note: The figures show the model solutions under different values of ρ, which is plotted as the horizontal axis. The definition of vertical axis is shown in the title of each plot.

Figure 1: Solutions under difference ρ, both autarky and trade



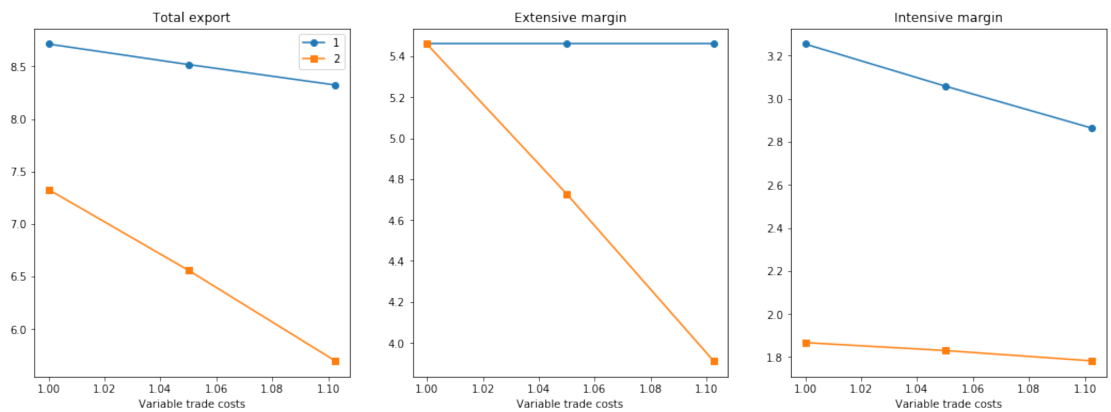
Note: The figures show the values of partial trade elasticity (left panel) and expenditure share on domestic goods (right panel) under different values of variable trade costs τ.

Figure 2: Trade elasticity and domestic trade share



Note: The figures show the optimal firm choices as a function of marginal cost. Left panel shows optimal number of total entries of product-destinations pairs, right panel shows the sales of product 1 at destination 1.

Figure 3: Correlation of exports



Note: The figure shows the cross sectional pattern of aggregated trade flows of product 1 and 2. The left panel shows how does total export value respond to destinations with different level of trade costs, the middle panel shows that the pattern of extensive margin, the right panel shows the pattern of intensive margin.

Figure 4: Changing distance elasticity

Appendix

A Multiple Asymmetric Country and Multi Product Firm

In this appendix, we generalize the environment to consider a world with J asymmetric countries and each firm is allowed to produce multiple products. Considering the implications of increasing marginal cost on multiple product firms is a natural extension because when $\rho > 1$, every additional variety would have similar negative externality on existing production, regardless of whether it is about a different destination or product. In this economy, each country is endowed with L_j $(j = 1, 2, ..., J)$ effective labor. A representative consumer in country j has the CES utility function of the form

$$U_j = \left[\sum_{i=1}^{J} \int_{\omega \in \Omega_{ij}} q_{ij}^{\nu} (\omega) \, d\omega \right]^{\frac{1}{\nu}},$$

where Ω_{ij} is the set of firms that export from country i to country j, ν is a constant that determines elasticity of substitution across varieties, and

$$q_{ij} (\omega) = \left[\sum_{g=1}^{G_{ij}(\omega)} q_{igj}^{\nu} (\omega) \right]^{\frac{1}{\nu}}$$

is the firm level aggregate across different products. $G_{ij} (\omega)$ is the number of products that firm ω from i exports to j. The demand is given by

$$q_{igj} (\omega) = \frac{p_{igj}^{\frac{1}{\nu-1}} (\omega)}{P_j^{\frac{\nu}{\nu-1}}} I_j, \tag{27}$$

where I_j is the total expenditure of country j, and price index is defined as

$$P_j = \left[\sum_{i=1}^{J} \int_{\omega \in \Omega_{ij}} P_{ij}^{\frac{\nu}{\nu-1}} (\omega) \, d\omega \right]^{\frac{\nu-1}{\nu}},$$

where the firm level price index is given by

$$P_{ij}(\omega) = \left[\sum_{g=1}^{G_{ij}(\omega)} p_{igj}^{\frac{\nu}{\nu-1}}(\omega) \right]^{\frac{\nu-1}{\nu}}.$$

From equation (27) we can drive the revenue function as

$$X_{igj}(\omega) = \kappa_j \nu^{-1} q_{igj}^{\nu}(\omega),$$

where $\kappa_j = \nu P_j^{\nu} I_j^{1-\nu}$.

Now we turn to profit maximization problem of firms from origin i. The total variable costs of production is assumed to be

$$VC\left(a, q_f\right) = a w_i \rho^{-1} q_f^{\rho},$$

where a (or $a(\omega)$) is a firm specific productivity measure, w_i is wage in country i, q_f (or $q_f(\omega)$) is the effective aggregate quantity produced by the firm, ρ is the elasticity that measures how fast variable cost increases with the effective aggregate quantity. In order to make sure that the cost function is convex, we assume $\rho \geq 1$. The effective aggregate quantity is the aggregate across all varieties within the firm by the function

$$q_f = \sum_j \tau_{ij} \left[\sum_{g=1}^{G_{ij}(\omega)} \left(\frac{q_{igj}(\omega)}{z(g)} \right) \right], \tag{28}$$

where $\tau_{ij} \geq 1$ represents the iceberg trade cost, $z(g)$ measures the efficiency of product g in terms of adding the effective aggregate quantity. Under this setting, the partial derivative of total variable with respect to quantity of each variety is

$$\begin{aligned} \frac{\partial VC\left(a, q_f\right)}{\partial q_{igj}(\omega)} &= a w_i q_f^{\rho-1} \frac{\partial q_f}{\partial q_{igj}(\omega)} \\ &= a w_i q_f^{\rho-1} \frac{\tau_{ij}}{z(g)}, \end{aligned}$$

Therefore the marginal variable cost of production of any product-destination depends on the total quantity q_f. Because of this feature, the optimal combination of product-destinations to export becomes an integrated decision: introducing any new product-destination raises marginal variable costs for all current product-destinations, and higher the current production, higher the marginal variable cost of introduction new product-destination.

The firm's profit function is

$$\pi = \sum_j \sum_{g=1}^{G_{ij}(\omega)} \left[X_{igj}(\omega) - f_{igj}^x \right] - VC\left(a, q_f\right)$$

$$= \sum_j \sum_{g=1}^{G_{ij}(\omega)} \left[\kappa_j \nu^{-1} q_{igj}^{\nu}(\omega) - f_{igj}^x \right] - a w_i \rho^{-1} q_f^{\rho}.$$

In order to solve for profit maximization problem, we first solve for optimal quantities $q_{igj}(\omega)$ taken $G_{ij}(\omega)$ as given, then find optimal $G_{ij}(\omega)$. The first order condition is

$$\frac{\partial \pi}{\partial q_{igj}(\omega)} = \kappa_j q_{fgj}^{\nu-1} - a w_i \tau_{ij} z^{-1}(g)\, q_f^{\rho-1}$$

therefore we get the relation

$$q_{igj}(a) = \left(\frac{a}{\kappa_j} w_i \tau_{ij} z^{-1}(g)\, q_f^{\rho-1} \right)^{\frac{1}{\nu-1}}. \tag{29}$$

Plug in this expression into Equation (28), we get

$$q_f(a) = (w_i a)^{-\frac{1}{\rho-\nu}} \left\{ \sum_j \left(\kappa_j\right)^{\frac{1}{1-\nu}} \left(\tau_{ij}\right)^{\frac{\nu}{\nu-1}} \left[\sum_{g=1}^{G_{ij}(\omega)} \left(z(g)\right)^{\frac{\nu}{1-\nu}} \right] \right\}^{\frac{1-\nu}{\rho-\nu}}, \tag{30}$$

and

$$q_{igj}(a) = (w_i a)^{-\frac{1}{\rho-\nu}} \left(\frac{\kappa_j z(g)}{\tau_{ij}} \right)^{\frac{1}{1-\nu}} \left\{ \sum_j \left(\kappa_j\right)^{\frac{1}{1-\nu}} \left(\tau_{ij}\right)^{\frac{\nu}{\nu-1}} \left[\sum_{g=1}^{G_{ij}(\omega)} \left(z(g)\right)^{\frac{\nu}{1-\nu}} \right] \right\}^{-\frac{\rho-1}{\rho-\nu_R}} \tag{31}$$

Then we can get the expression of firm level profit as

$$\pi = \left[\nu^{-1} - \rho^{-1} \right] (w_i a)^{-\frac{\nu}{\rho-\nu}} \left\{ \sum_j \kappa_j^{\frac{1}{1-\nu}} \tau_{ij}^{\frac{\nu}{\nu-1}} \left[\sum_{g=1}^{G_{ij}(\omega)} z^{\frac{\nu}{1-\nu}}(g) \right] \right\}^{\frac{(1-\nu)\rho}{\rho-\nu}} \tag{32}$$

$$-\sum_{j}\sum_{g=1}^{G_{ij}(\omega)} f_{igj}^{x}$$

Several points are worth noting. Firstly, as long as the elasticity of substitution is higher than 1, we have $0 < \nu \leq 1 \leq \rho$. When $\rho = 1$, the profit function becomes additively separable across product-destinations, as is commonly assumed. Secondly, the profit is increasing in firm's productivity $\frac{1}{a}$.Therefore more productive firms have the higher ability to cover fixed costs and enter more product-destinations. Thirdly, because $\frac{(1-\nu)\rho}{\rho-\nu} = 1 - \frac{\nu(\rho-1)}{\rho-\nu} \leq 1$, the profit function is submodular. In other words, the marginal return of introducing the same product-destination will be lower if the number of existing product-destinations are higher. Because of this property, choosing $G_{ij}(\omega)$ for each destination j is difficult.