A Tractable Model of Trade with Flexible Cost Structure *

Michal Fabinger
Tokyo Data Science

Konstantin Kucheryavyy†
U of Tokyo

Ruo Shangguan‡
Waseda University

December 18, 2020

Abstract

We introduce variable marginal costs of exporting in a heterogeneous firms trade model *à la* Melitz (2003) to match data on the frequency of shipment of exporters, which, potentially, is a source of additional information on the costs of trade. In our setup, the marginal costs of exporting depend on the quantity shipped in addition to the standard iceberg trade costs. This costs structure can be micro-found through a firm's inventory management problem. Under a Pareto distribution of firms' productivities, our model implies a tractable gravity equation as well as an expression for welfare gains from trade, for which the regular gravity equation and the Arkolakis *et al.* (2012, ACR) formula for gains from trade are special cases. Relative to the ACR formula, our expression for the gains from trade additionally depends on the elasticity of trade costs with respect to traded quantity and implies lower gains from trade. We apply a version of our model with a log-normal distribution of firms' productivities to data on transactions of Chinese exporters. Compared to a model with constant marginal costs of exporting, our model implies a lower distance elasticity of iceberg trade costs but a higher distance elasticity of fixed exporting costs.

*We are grateful to Hidehiko Ichimura, Taiji Furusawa, Peter Schott, Sam Kortum for insightful discussions.

†Graduate School of Public Policy, University of Tokyo. Email: kucher@pp.u-tokyo.ac.jp

‡Institute for Empirical Research in Organizational Economics, Waseda University. Email: roy1312@gmail.com

1 Introduction

The structure and magnitude of trade costs have important implications to various questions in the field of international trade. To name a few: how firms make entry decisions, how much they sell in each destination, and what price they charge for the products exported. Answers to these questions, in their turn, have important implications for the magnitudes of welfare gains from trade. In this paper, we study how a richer trade costs structure introduced into a heterogeneous firms trade model *à la* Melitz (2003) affects our understanding of these questions. Our main departure is the introduction of a new logistics costs component, which is a power function of the quantities traded. As a result, marginal trade costs depend on the quantity being traded.

The first contribution of this paper is to derive the formula for welfare changes under our richer trade cost structure. In a symmetric two-country world, we show that the welfare changes critically depends on the elasticity of cumulative market share of domestic firms with respect to the zero profit cutoff quantity. Importantly, this elasticity is in general different from the elasticity of relative expenditure share with respect to iceberg trade costs, or the so-called trade elasticity. Because in this model iceberg trade cost is no longer the only variable trade cost, it is natural to expect the elasticity of logistics costs to play a significant role. To see it more clearly, we show further results under the assumption that productivity parameters across firms follow a Pareto distribution. Three macro restrictions proposed in Arkolakis *et al.* (2012, ACR) are satisfied in this case, in particular we derive a gravity equation. Using the gravity structure, we show that the welfare changes predicted by this model depend on the expenditure on domestic goods and a combination of three elasticities, including the elasticity of relative import with respect to iceberg trade costs, the quantity elasticity of the logistics costs, and the demand elasticity. Our formula includes the ACR formula for gains from trade as a special case.

The logistics costs are micro-found through the inventory management problem. In this problem, each firm chooses an optimal frequency of shipment that balances the transaction costs and the inventory costs. If shipment frequency is high, the total transaction cost, which is the sum of transaction costs across shipments, will be high, while the inventory level can be kept at a low level and therefore storage cost will be low. This micro foundation of logistics costs yields three implications. Firstly, the inventory management problem gives intuitive interpretation on different values of the trade costs elasticity. When the trade costs elasticity is zero, the total logistics costs become constant, and the model reduces to the standard Melitz model. When trade costs elasticity is one half, it implies that the cost of an additional shipment is constant. When the trade costs elasticity is one, the cost of shipping an additional unit of a good is constant, therefore total logistics costs is proportional to total traded quantity. Secondly, the inventory management problem shows that conditional on the inventory costs and quantity traded, the shipment frequency is informative on the underlying magnitude of logistics costs. Thirdly, the inventory management problem also shows that trade cost elasticity can be estimated via a linear regression of log shipment frequency on log trade volume. If the shipment frequency is higher conditional on the same export volume, the model interprets that logistics costs increase at a slower speed.

Using the data on the universe of international trade transactions of Chinese exporters, we find the trade cost elasticity to be around 0.6 to 0.8 across products, which show a clear rejection of zero trade cost elasticity assumed in the standard model. Using Rauch classification to define product categories, we find that homogeneous products tend to have a lower trade costs elasticity, indicating a lower marginal cost of transportation. With the estimated trade costs elasticity, we compare the prediction of gains from trade, which is change in real income associated with moving to autarky, between the ACR formula and our formula. The data on domestic expenditure share is obtained from WIOD. We set the elasticity of substitution equal to five across models, in consensus with the literature. On average, we observe a 11.5% lower welfare gains from trade.

In order to further evaluate the impact of a different trade cost structure, we calibrate the trade costs parameters from oberved trade flows. We show that under the assumption that productivity follows Pareto distribution trade costs parameters can be identified through trade flows including bilateral shipment frequency. This result extends that in the Eaton *et al.* (2011). However, we also show that the analytical benchmark provided under Pareto provides a strange interpretation of the data: the logistics cost parameters decrease with distance. In order to avoid this problem, we calibrate the trade costs parameters with the assumption that productivity parameters across firms follow a log-normal distribution, using a method inspired by Fernandes *et al.* (2019) and Head *et al.* (2014). More specifically, we match the firm-level distributions of export sales and shipment frequency across origin-destination pairs. The information about the firm-level shipment frequency helps us to identify the additional parameter that governs the magnitude of logistics costs. The introduction of logistics costs has significant implications on the magnitude of inferred trade costs from trade flows. The estimation results show that the distance elasticity of iceberg trade costs declines, while the distance elasticity of fixed export costs increases.

The model provides a straightforward way to infer the logistics costs incurred by each firm. Under the assumption of 30% ad valerom inventory costs following Alessandria *et al.* (2010), we estimate the weighted average logistics costs is 6.2% of export value, out of which 3.7% is attributed to coordination costs. Under a milder assumption that inventory cost is product-destination specific, we use panel data to show that the trade cost parameters estimated using past data successfully explains the shipment frequency in the future in a way that is consistent with the model.

We provide additional evidence that suggests shipment frequency is informative about trade costs. First, we find that intermediary firms, conditional on narrowly defined product category and destination, have lower frequency of shipments. This is consistent with the idea emphasized in previous studies that intermediate firms pay higher coordination costs per transaction (e.g., Bai *et al.* (2017)). Second, we estimate a gravity style regression model and show that, within each firm and narrowly defined products, frequency of shipments tends to be higher to destinations having a Chinese ethnic group and/or sharing a common border, and frequency of shipments tends to be lower to destinations having higher business startup costs.

The model studied in this paper is a simplified version of the model studied in Fabinger and Weyl (2018). In contrast to Fabinger and Weyl (2018), we assume that the marginal cost of production is constant, which allows us to obtain explicit model solutions under a very general environment. On the empirical side, this paper also provides a richer micro-evidence regarding the frequency of shipment.

Many authors have explored the implications of a trade costs structure beyond the simple iceberg and fixed cost combination. For example, Alessandria *et al.* (2010) use cost per shipment to explain lumpiness in international trade transactions. Hummels and Skiba (2004) use costs per unit to explain why high quality goods tend to be shipped further away. Kropf and Sauré (2014) estimates the magnitude of fixed cost per shipment using an extension of the standard Melitz model. The effort of analyzing the implications of a richer cost structure, however, is limited by the difficulty that models very quickly become intractable once additional costs are added. On one hand, our model's flexible trade cost structure is capable of explaining the patterns emphasized in that literature. On the other, we keep the tractability under the commonly used assumption of Pareto distributed productivity.

Our paper contributes to the studies of trade costs per shipment by providing a tractable benchmark with variable marginal cost of trade. The ability of deriving a gravity equation allows a more transparent analysis of how trade costs structure influences aggregate trade flows. Our result on welfare implications extends the formula proposed in Arkolakis *et al.* (2012), who show that trade elasticity is the only elasticity needed to predict welfare changes in a large class of models. Different from papers assuming exponential decay of inventory (e.g., Blum *et al.* 2019; Kropf and Sauré 2014), we can derive a reduced-form formula of the optimal shipment frequency for each firm, which facilitates the empirical analysis.

The current paper is also related to the literature on inventory management (e.g. Alessandria *et al.* (2010)). A key difference is that we model the inventory problem on the exporter side instead of importers, with the emphasis of the productivity differences of producers. Additionally, in order to characterize the general equilibrium explicitly, we abstract from considerations of uncertainty of demand and delay in transportation.

2 Model setup

Our basic setup is similar to the standard Melitz (2003) model with the only modification that the marginal costs of serving markets depend on the amounts shipped (as opposed to being given by iceberg trade costs in the standard Melitz model).

The economy consists of J countries, indexed by i and j, each endowed with effective labor L_i. Labor is immobile across countries, but perfectly mobile between different uses within a country. Each country i can potentially produce a infinite set of varieties Ω_i indexed by ω. Only an endogenously determined subset Ω_{ij} of Ω_i is available in any destination country j. Utility of the representative consumer in country j is given by

$$U_j = \left(\sum_i \int_{\Omega_{ij}} q_{ij} \left(\omega\right)^{\nu_R} d\omega \right)^{\frac{1}{\nu_R}}, \tag{1}$$

where $q_{ij}\left(\omega\right)$ is the quantity of variety $\omega \in \Omega_{ij}$ from origin i consumed in destination j, and $0 < \nu_R < 1$ is related to the elasticity of substitution between varieties σ through the relation $\nu_R = \frac{\sigma-1}{\sigma}$. Utility maximization gives the inverse demand curve

$$p_{ij}\left(\omega\right) = A_j q_{ij}\left(\omega\right)^{\nu_R - 1}, \tag{2}$$

where the country-specific parameter $A_j = P_j^{\nu_R} I_j^{1-\nu_R}$ summarizes the market condition with I_j and P_j representing the total expenditure and price index in destination j. This gives the relationship between the value and quantity of each firm's export,

$$X_{ij}\left(\omega\right) = A_j q_{ij}\left(\omega\right)^{\nu_R}. \tag{3}$$

In what follows, in order to simplify notation, we also sometimes use the notation q_{fj} instead of $q_{ij}\left(\omega\right)$, with the understanding that firm f from origin i produces variety ω.

Technology of production of varieties features constant return to scale. Each firm in country i has marginal cost of production $a > 0$ that is drawn from a known distribution $G_i\left(a\right)$. The cost of serving any destination j has three components. The first component is the standard fixed costs: any firm f from origin i needs pay a cost of f_{ij}^x in terms of country

i's labor in order to enter market j. The second component is also standard. It is the usual iceberg trade costs: delivering one unit of any variety from origin i to destination j requires shipping $\tau_{ij} \geq 1$ units of this variety. Finally, the third component constitutes our innovation relative to the standard Melitz model and captures the idea of the "logistic costs": in order to deliver q_{fj} units of its variety, firm f from origin i needs to pay the cost $B_{ij}q_{fj}^{\nu_{LT}}$ measured in units of country i's labor, where $\nu_{LT} > 0$ is the trade costs elasticity with respect to the quantity shipped. In Section 3.3 we provide a micro-foundation of the logistic costs component via an inventory management problem. We make the following assumption about ν_{LT}:

Assumption 1. $0 < \nu_{LT} < \nu_R$.

Given the above specifications of technology of production and costs of serving markets, the profit country i's firm f from serving destination j is

$$\pi_{fj} = A_j q_{fj}^{\nu_R} - w_i B_{ij} q_{fj}^{\nu_{LT}} - a w_i \tau_{ij} q_{fj} - w_i f_{ij}^{x}. \tag{4}$$

Observe that, when $\nu_{LT} = 0$, logistics costs reduce to a constant and merge with the fixed costs. Hence, the standard Melitz model is a particular case of our setup for $\nu_{LT} = 0$. The first-order condition for the firm's profit maximization problem is

$$\frac{\partial \pi_{fj}}{\partial q_{fj}} = A_j \nu_R q_{fj}^{\nu_R - 1} - w_i B_{ij} \nu_{LT} q_{fj}^{\nu_{LT} - 1} - a w_i \tau_{ij} = 0, \tag{5}$$

and the second-order condition is

$$\frac{\partial^2 \pi_{fj}}{\partial q_{fj}^2} = A_j \nu_R \left(\nu_R - 1\right) q_{fj}^{\nu_R - 2} - w_i B_{ij} \nu_{LT} \left(\nu_{LT} - 1\right) q_{fj}^{\nu_{LT} - 2} < 0. \tag{6}$$

We can use the first-order condition (5) to express[1]

$$a = \frac{A_j \nu_R q_{fj}^{\nu_R - 1} - w_i B_{ij} \nu_{LT} q_{fj}^{\nu_{LT} - 1}}{w_i \tau_{ij}}. \tag{7}$$

[1]Observe that if $A_j \nu_R q_{fj}^{\nu_R - 1} - w_i B_{ij} \nu_{LT} q_{fj}^{\nu_{LT} - 1} < 0$, then, given that $a \geq 0$, we would have that $\partial \pi_{fj} / \partial q_{fj} < 0$, and so q_{fj} cannot be a profit-maximizing quantity.

This expression implies that a can be thought of as a function of q_{fj}, i.e. $a\left(q_{fj}\right)$. This is one of the key insights that significantly simplifies the theoretical analysis that follows.

Combining the expression (7) for a together with the condition that firm f should earn non-negative profits, gives

$$A_j\left(1-\nu_R\right)q_{fj}^{\nu_R}-w_iB_{ij}\left(1-\nu_{LT}\right)q_{fj}^{\nu_{LT}}\geq w_if_{ij}^x.$$

In order to make our theoretical analysis tractable, we impose a further restriction that $f_{ij}^x=0$.[2] Under this restriction, we can find that firm's profits are nonnegative if and only if $q_{fj}\geq q_{ij}^*$, where

$$q_{ij}^*\equiv\left[\frac{w_iB_{ij}\left(1-\nu_{LT}\right)}{A_j\left(1-\nu_R\right)}\right]^{\frac{1}{\nu_R-\nu_{LT}}}, \tag{8}$$

is the zero-profit quantity, and where we used our assumption that $\nu_{LT}<\nu_R$.

Next, the second-order condition (6) holds if and only if

$$q_{fj}^{\nu_R-\nu_{LT}}>\left(q_{ij}^*\right)^{\nu_R-\nu_{LT}}\frac{\nu_{LT}}{\nu_R},$$

which holds for any $q_{fj}\geq q_{ij}^*$ under our assumption that $\nu_{LT}<\nu_R$.

Finally, it is straightforward to check that $a'\left(q_{fj}\right)<0$ if and only if the second-order condition (6) holds. Hence, $a\left(q_{fj}\right)$ is a monotone function defined for all $q_{fj}\geq q_{ij}^*$ that provides a one-to-one mapping between a and q_{fj}. This, in turn, implies that there is a unique profit-maximizing quantity q_{fj} for each cost $a\in\left(0,a_{ij}^*\right)$, where $a_{ij}^*\equiv a\left(q_{fj}^*\right)$.

The introduction of positive ν_{LT} also impacts the relative prices between firms having different productivity. Consider two firms with $a_1<a_2$, namely firm 1 has higher productivity. By (7), at some market j,

$$\frac{a_1}{a_2}>\left(\frac{q_{f_1j}}{q_{f_2j}}\right)^{\nu_R-1},$$

[2]When we perform numerical analysis in Section 5, we remove the restriction $f_{ij}^x=0$ to match the extensive margin of trade flows observed in the data. Also, note that, even under the restriction $f_{ij}^x=0$, the Melitz model is still a particular case of our setup for $\nu_{LT}=0$.

where the term of the right hand side is equal to the relative prices at destination j, namely $\frac{p_{f_1 j}}{p_{f_2 j}} = \left(\frac{q_{f_1 j}}{q_{f_2 j}}\right)^{\nu_R - 1}$. It is well known that relative price in the standard Melitz model depends only on the relative productivity. When $\nu_{LT} > 0$, however, the same proportion of advantage in marginal cost translates into a bigger difference in prices, due to the ability of taking advantage of the scale of economy in the transportation technology by productive firms[3].

3 Gains from trade with more flexible trade costs

The goal of this section is to derive the implications of introducing a more flexible cost structure to welfare gains from trade liberalization in a standard heterogeneous trade model.

3.1 Two symmetric countries

In order to get a clearer intuition, we start from a world with only two countries. We use d to denote quantity related to the domestic market, and x to denote quantity related to the foreign market. To solve the model, we first use (8) for the domestic and the foreign market to get

$$\left(\frac{q_d^*}{q_x^*}\right)^{\nu_R - \nu_{LT}} = \frac{B_d}{B_x}, \tag{9}$$

then together with the free entry condition

$$\int_0^{a_d^*} \pi_d dG\left(a\right) + \int_0^{a_x^*} \pi_x dG\left(a\right) = f^e, \tag{10}$$

[3] On the other hand, observe that the markup is equal to

$$\frac{p_{fj}}{VC'\left(q_{fj}\right)} = \frac{1}{\nu_R},$$

where $VC\left(q_{fj}\right)$ is the total variable cost that includes both logistics and production costs. This shows that the price has a constant markup over the marginal variable cost. The important difference from the standard model is that the variable cost now depends on the quantity produced.

we can solve for the zero profit cutoffs q_d^* and q_x^*. Note that the cutoffs of marginal costs a_d^* and a_x^* can be derived from cutoffs of quantities via the first order condition (5). Other aggregated variables of interest can be expressed by the cutoffs. The expenditure share of domestic goods is

$$\lambda = \frac{\int_0^{a_d^*} X_d(a)\, dG(a)}{\int_0^{a_d^*} X_d(a)\, dG(a) + \int_0^{a_x^*} X_x(a)\, dG(a)}$$

$$= \frac{1}{1 + \frac{\int_0^{a_x^*} q_x^{\nu_R} dG(a)}{\int_0^{a_d^*} q_d^{\nu_R} dG(a)}}. \tag{11}$$

The mass of firm $N = N^e G\left(a_d^*\right)$ is

$$N = \frac{(1-\nu_R)\, L \left(q_d^*\right)^{\nu_R - \nu_{LT}}}{(1-\nu_{LT})\, B_d \left[\int_0^{a_d^*} q_d^{\nu_R} d\frac{G(a)}{G\left(a_d^*\right)} + \int_0^{a_x^*} q_x^{\nu_R} d\frac{G(a)}{G\left(a_d^*\right)}\right]}. \tag{12}$$

Using (8) and (12), the real consumption can be derived as

$$W = \frac{1}{P} = \left(\frac{B_d\,(1-\nu_{LT})}{\left(q_d^*\right)^{\nu_R-\nu_{LT}}(1-\nu_R)}\right)^{-\frac{1}{\nu_R}} L^{\frac{1-\nu_R}{\nu_R}}, \tag{13}$$

since B_d and L do not change (we consider foreign shocks), the previous equation reduces to

$$d\ln W = \left(\frac{\nu_R - \nu_{LT}}{\nu_R}\right) d\ln q_d^*. \tag{14}$$

Note that different from the case $\nu_{LT} = 0$, the elasticity of real consumption with respect to the zero profit cutoff now depends on demand and trade cost elasticity. In order to compare with the existing results in the literature, we keep the expression of firm mass, and connect to the domestic expenditure share (11) and express the change in real consumption as

$$d\ln W = \frac{1}{\nu_R\left(1 - \frac{\gamma\left(q_d^*\right)}{\nu_R - \nu_{LT}}\right)}\left(d\ln N^e - d\ln\lambda\right),$$

where $\gamma\left(q_d^*\right) \equiv \frac{d \ln \int_0^{a_d^*} q_d^{\nu_R} dG(a)}{d \ln q_d^*}$. Because $\int_0^{a_d^*} q_d^{\nu_R} dG\left(a\right)$ is proportional to the market share of the domestic firms, $\gamma\left(q_d^*\right)$ measures the sensitivity of market share with respect to the domestic zero profit cutoff. Under further restrictions proposed in ACR, $d \ln N^e = 0$ and therefore the domestic expenditure share λ and the elasticity $\nu_R\left(1 - \frac{\gamma\left(q_d^*\right)}{\nu_R - \nu_{LT}}\right)$ form the two sufficient statistics for calculating welfare changes. The analysis highlights the potential for a different elasticity when trade costs structure is more flexible. In the next sections, we solve for the formula under more functional form assumption and then quantify the difference in the implied welfare gains from trade.

3.2 Gains from trade: Pareto distribution

3.2.1 Aggregation formula

In order to derive explicit solutions, we impose the assumption that the inverse of marginal cost follows Pareto distribution given by its cumulative density function

$$G_i\left(a\right) = \theta^{-1}\left(\kappa_{G,i} a\right)^{\theta}, \tag{15}$$

where $\theta > 0$ and $a \in \left[0, \kappa_{G,i}^{-1} \theta^{\frac{1}{\theta}}\right]$. Under this assumption, together with $f_{ij}^x = 0$, and that the export decision is monotonic, it can be shown that the aggregated optimal quantity of any power ν, subject to the condition $\frac{\nu + \nu_R \theta - \theta}{\nu_{LT} - \nu_R} > 0$, is given by

$$\int_0^{a_{ij}^*} q^{\nu} dG_i\left(a\right) = \kappa_{G,i}^{\theta} A_j^{-\frac{\nu + \nu_{LT}\theta - \theta}{\nu_R - \nu_{LT}}} \left(w_i \tau_{ij}\right)^{-\theta} \left(w_i B_{ij}\right)^{\frac{\nu + \nu_R \theta - \theta}{\nu_R - \nu_{LT}}} H\left(\nu, \nu_R, \nu_{LT}, \theta\right) \tag{16}$$

where a_{ij}^* is the zero profit cutoff, and $H\left(\nu, \nu_R, \nu_{LT}, \theta\right)$ is a (positive-valued) function that depends only on parameters of the model. In order to simplify exposition, we drop the later three arguments in the $H\left(\cdot\right)$ function and write $H\left(\nu\right)$ to represent $H\left(\nu, \nu_R, \nu_{LT}, \theta\right)$.

3.2.2 Gravity equation

Using the formula (16), we can calculate the aggregate trade flow as

$$
\begin{aligned}
X_{ij} &= N_i^e A_j \int_0^{a_{ij}^*} q_{fj}^{\nu_R} dG(a) \\
&= N_i^e \kappa_{G,i}^{\theta} \left(A_j\right)^{\frac{\nu_{LT}+\nu_{LT}\theta-\theta}{\nu_{LT}-\nu_R}} \left(w_i \tau_{ij}\right)^{-\theta} \left(w_i B_{ij}\right)^{-\frac{\nu_R+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}} H\left(\nu_R\right).
\end{aligned} \tag{17}
$$

Trade share is then

$$
\lambda_{ij} = \frac{X_{ij}}{\sum_l X_{lj}} = \frac{N_i^e \kappa_{G,i}^{\theta} \left(w_i \tau_{ij}\right)^{-\theta} \left(w_i B_{ij}\right)^{-\frac{\nu_R+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}}}{\sum_l N_l^e \kappa_{G,l}^{\theta} \left(w_l \tau_{lj}\right)^{-\theta} \left(w_l B_{lj}\right)^{-\frac{\nu_R+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}}}. \tag{18}
$$

Despite the flexible trade cost structure, our model generates a familiar gravity equation. The trade elasticity, defined as $\frac{\partial \ln\left(X_{ij}/X_{jj}\right)}{\partial \ln \tau_{ij}}$, is given by the shape parameter of the firm's productivity distribution θ. However, because B_{ij} works similarly to variable trade costs, it is natural to expect that the trade elasticity along is not sufficient for summarizing welfare changes given trade liberalization.

3.2.3 Mass of firms

Next we show how to solve for mass of firms N_i^e in each country. Note that N_i^e represents the total mass of potential entrants, which is related to the mass of firms producing positive quantity in each country N_{ii} by the relation $N_{ii} = N_i^e G\left(a_{ii}^*\right)$[4]. We assume the trade is balanced: for any country i,

$$
\sum_j X_{ij} = \sum_j X_{ji},
$$

where the left hand side is also equal to the total labor revenue $w_i L_i$, the right hand side is also equal to the total expenditure I_i. Labor is used for the following purposes: entry of firms, logistic costs, and production. The amount of labor used for entry is $L_i^e \equiv N_i^e f_i^e$,where f_i^e is the entry cost. The labor market clearing condition is

$$
L_i = N_i^e f_i^e + \sum_j L_{ij}^{lm},
$$

[4]Following the literature, we adopt a static setting and let probability of death equal to 1.

where L_{ij}^{lm} represents the amount of labor used in logistics and production for the exports to country j in origin i. Using first order condition (5) and aggregation formula (16), we can show that L_{ij}^{lm} is equal to

$$L_{ij}^{lm} = N_i^e \int_0^{a_{ij}^*} \left[\left(1 - \nu_{LT}\right) B_{ij} q_{fj}^{\nu_{LT}} + \frac{\nu_R A_j q_{fj}^{\nu_R}}{w_i} \right] dG_i\left(a\right)$$

$$= \nu_R \left[\left(\frac{1 - \nu_{LT}}{\nu_R} \right) \frac{H\left(\nu_{LT}\right)}{H\left(\nu_R\right)} + 1 \right] \frac{X_{ij}}{w_i}.$$

Then we can solve that

$$N_i^e = \left(1 - \nu_R - \left(1 - \nu_{LT}\right) \frac{H\left(\nu_{LT}\right)}{H\left(\nu_R\right)}\right) \frac{L_i}{f_i^e}.$$

From the above calculations, we can also see that the aggregate profit is a constant share of the total revenue:

$$\Pi_i = \left(1 - \nu_R - \left(1 - \nu_{LT}\right) \frac{H\left(\nu_{LT}\right)}{H\left(\nu_R\right)}\right) \sum_j X_{ij}. \tag{19}$$

3.2.4 Wage

Finally, using the goods market clearing condition, the total output in country i is equal to the absorption of country i's production around the world:

$$Y_i = \sum_j \lambda_{ij} I_j. \tag{20}$$

Since $Y_i = I_i = w_i L_i$ for any country i, and λ_{ij} given in (18) is a function of wages and exogenous parameters, (20) gives a system of equation that can be used to solve wages in each country.

3.2.5 Price index and Gains from Trade

Given the gravity equation (18) and the fact that aggregate profit is a constant share of aggregate revenue, as shown in (19), it is clear that the model satisfies all the macro restrictions proposed in ACR. However, in this section, we show that the the welfare gains from trade can not be inferred from domestic trade share and trade elasticity only. Instead, the magnitudes of ν_{LT} and ν_R also play an important role. In order to better understand the price index in a destination j, it is useful to first investigate all the prices imported from origin i, which has the form

$$P_{ij}^{-\frac{\nu_R}{1-\nu_R}} = N_i^e A_j^{-\frac{\nu_R}{1-\nu_R}} \int_0^{a_{ij}^*} q_{fj}^{\nu_R} dG\left(a\right)$$

$$= N_i^e \kappa_{G,i}^{\theta} \left(A_j\right)^{\theta - \frac{\nu_R+\nu_R\theta-\theta}{\nu_{LT}-\nu_R} - \frac{\nu_R}{1-\nu_R}} \left(w_i \tau_{ij}\right)^{-\theta + \frac{\nu_R+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}} \left[a_{ij}^*\right]^{\frac{\nu_R+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}} \left[\frac{\left(\nu_R^{-1}-1\right)}{2\left(\nu_{LT}^{-1}-1\right)}\right]^{-\frac{\nu_R+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}} H\left(\nu_R\right).$$

Except for the constant terms, the bilateral price index can be decomposed into three parts. The first terms $N_i^e \kappa_{G,i}^{\theta}$ represent the effect of entry in the origin market, as the mass of potential is an endogenous quantity, it will be generally different when we change ν_{LT}. The intensive margin term $\left(A_j\right)^{\theta - \frac{\nu_R+\nu_R\theta-\theta}{\nu_{LT}-\nu_R} - \frac{\nu_R}{1-\nu_R}} \left(w_i \tau_{ij}\right)^{-\theta + \frac{\nu_R+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}}$ affects individual prices that export from i to j. (see the formula for standard model in page 208 of Costinot and Rodríguez-Clare (2015)). The the zero profit marginal cost cutoff term $\left[a_{ij}^*\right]^{\frac{\nu_R+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}}$ represents the effect from changes from extensive margin. Note that not only the power depends on the value of ν_{LT}, the cutoffs a_{ij}^* are in general different given different ν_{LT}. Using the aggregation formula (16) and the definition $A_j = P_j^{\nu_R} I_j^{1-\nu_R}$, we can get

$$\left(P_j\right)^{\frac{-\nu_R\left(\nu_{LT}+\theta\nu_{LT}-\theta\right)}{\nu_{LT}-\nu_R}} = C_P \sum_i N_i^e \kappa_{G,i}^{\theta} \left(w_i \tau_{ij}\right)^{-\theta} \left(w_i B_{ij}\right)^{-\frac{\nu_R+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}}, \tag{21}$$

where

$$C_P = \left(I_j\right)^{-\nu_R + \frac{\left(1-\nu_R\right)\left(\theta\nu_{LT}+\nu_R-\theta\right)}{\nu_{LT}-\nu_R}} H\left(\nu_R\right),$$

and N_i^e has been solved above. Next we derive the expression of real wage as an expression of the observed moments in the data. Using (18) and (21), we can write domestic trade share as

$$\lambda_{ii} = C_P N_i^e \kappa_{G,i}^{\theta} (w_i)^{-\theta} (w_i B_{ii})^{-\frac{\nu_R + \nu_R\theta - \theta}{\nu_{LT} - \nu_R}} (P_i)^{\frac{\nu_R(\nu_{LT} + \theta\nu_{LT} - \theta)}{\nu_{LT} - \nu_R}},$$

from which we can derive the expression of the real consumption as

$$W_i \equiv \frac{w_i}{P_i} = \left[\frac{\lambda_{ii}}{C_P N_i^e \kappa_{G,i}^{\theta} (B_{ii})^{-\frac{\nu_R + \nu_R\theta - \theta}{\nu_{LT} - \nu_R}}} \right]^{-\frac{\nu_{LT} - \nu_R}{\nu_R(\nu_{LT} + \theta\nu_{LT} - \theta)}}.$$

Because the mass of firms N_i^e is a fixed portion of the labor endowment, and under autarky, $\lambda_{ii}^A = 1$, the welfare gains from opening to trade in autarky is given by

$$\frac{W_i}{W_i^A} = \lambda_{ii}^{-\frac{\nu_{LT} - \nu_R}{\nu_R(\nu_{LT} + \theta\nu_{LT} - \theta)}}. \tag{22}$$

As a widely known formula shown in Arkolakis *et al.* (2012), the expenditure share on the domestic goods and the trade elasticity are the only two sufficient statistics to calculate welfare gains from opening to trade. In contrast, our model gives a formula that combines trade elasticity, trade cost elasticity and demand elasticity. It is easy to verify that $-\frac{\nu_R(\nu_{LT} + \theta\nu_{LT} - \theta)}{\nu_{LT} - \nu_R}$ is equal to $-\theta$ when $\nu_{LT} = 0$, therefore this formula generalizes the results in ACR. As discussed in the previous section, the trade is balanced, the aggregate profit is a constant share of the total revenue, and the gravity equation is derived. Therefore our only departure from the class of models studied in ACR comes from a different micro structure, in particular the trade costs structure that we are considering.

3.3 Micro foundation of logistics costs

In this sub-section, we propose a simple logistics costs minimization problem to micro-found the logistics costs introduced into the profit function of firms. The demand of a variety during a given period of time is predicted without uncertainty. The problem is then how to deliver the required quantity q. The trade-off is to balance inventory cost and cost per shipment: since production is not instantaneous, the producer has to store the products and incur inventory costs.[5] Since the amount of inventory is proportional to the quantity per shipment, in order to save inventory cost, firms tend to ship frequently. But since there is a cost related to each shipment (e.g. paper work and other coordination with trade partner), too frequent shipments will be very costly. Therefore the firm optimally chooses a shipment frequency that balances these two costs.

Formally, we assume that firms choose a constant quantity per each shipment, which is denoted by q_s. The associated inventory cost is assumed to have the form $C_I(q_s) = \kappa_I q_s$, since the amount of goods that need to be stored is proportional to q_s. And for each shipment, firms need to incur coordination cost $C_T(q_s) = \kappa_T q_s^{\alpha}$, where α measures how fast the coordination cost changes with quantity. When $\alpha < 1$, there is a return to scale in coordination: marginal cost of coordination is decreasing in quantity per shipment. And if $\alpha < 0$, the coordination cost $C_T(q_s)$ decreases with quantity per shipment. Finally we add a fixed cost component f^x that does not depend on frequency of shipment. The logistics problem is therefore

$$\min_{q_s} C_I(q_s) + \frac{q}{q_s} C_T(q_s) + f^x.$$

Three special cases are worth mentioning. When $\alpha = 0$, the coordination cost is constant per shipment, the problem becomes the same as the economic order quantity model. When $\alpha = 1$, total coordination cost $\frac{q}{q_s} C_T(q_s)$ will be constant, it is therefore optimal to set q_s as small as possible, and the variable part of logistics cost will be $\kappa_T q$. When $\alpha = -\infty$, coordination cost is zero if q_s is slightly above 1. By choosing q_s near 1, the total logistics cost can be made arbitrarily close to $\kappa_I + f^x$, which is fixed regardless of trade quantity. The first-order condition of the above problem gives

$$q_s^{2-\alpha} = \frac{(1-\alpha)\kappa_T}{\kappa_I} q$$

With $\alpha < 1$, then second-order condition[6] will be positive. Plug in the above solution into

[5] As an alternative way to justify inventory cost, if we assume demand is uniformly distributed across time, the goods shipped but not consumed immediately must be stored, and the producer can be assumed to share this inventory costs.

[6] Second-order condition is $q\kappa_T(\alpha-1)(\alpha-2) q_s^{\alpha-3}$

the cost function, the optimized logistics cost has the form

$$C_{LT}\left(q\right)=\left(1-\alpha\right)^{-\frac{1-\alpha}{2-\alpha}}\left(2-\alpha\right)\kappa_{I}^{\frac{\alpha-1}{\alpha-2}}\kappa_{T}^{\frac{1}{2-\alpha}}q^{\frac{1}{2-\alpha}}+f^{x}=Bq^{\nu_{LT}}+f^{x}, \tag{23}$$

where for the ease of exposition, we let

$$\nu_{LT}=\frac{1}{2-\alpha},$$

$$B=\frac{1}{\nu_{LT}^{\nu_{LT}}\left(1-\nu_{LT}\right)^{1-\nu_{LT}}}\kappa_{I}^{1-\nu_{LT}}\kappa_{T}^{\nu_{LT}}.$$

Since $\lim_{\nu_{LT}\to 0}Bq^{\nu_{LT}}=\kappa_{I}$, algebraically the variable part of logistic costs reduces to a constant. On the other hand, $\lim_{\nu_{LT}\to 1}Bq^{\nu_{LT}}=\kappa_{T}q$, in which case the marginal logistics cost is equal to κ_{T}. The model also gives an explicit expression for the optimal frequency of shipment:

$$F=\left(\frac{1}{\nu_{LT}}-1\right)^{-\nu_{LT}}\kappa_{I}^{\nu_{LT}}\kappa_{T}^{-\nu_{LT}}q^{(1-\nu_{LT})}=\frac{1}{1-\nu_{LT}}\kappa_{I}B^{-1}q^{(1-\nu_{LT})}. \tag{24}$$

It is also worth emphasizing that the logistics problem is separated from other demand or supply side assumptions. The relationship shown in (24) also provides a straightforward way to estimate ν_{LT} using linear regression. The value of ν_{LT} is an easy test to differentiate different models. When $\nu_{LT}=0$, as is often assumed in the standard heterogeneous firm model, the frequency of shipment is equal to export quantity, namely $F=q$. When the coordination cost per shipment is a constant, $\nu_{LT}=\frac{1}{2}$. Finally, when $\nu_{LT}=1$, coordination cost is proportional to the export quantity, the current formula does not apply, but the model predicts an infinitely frequent shipment to reduce inventory cost as low as possible.

3.4 Firm's optimization

The equation (5) cannot be solved analytically for arbitrary values of ν_{R} and ν_{LT}. In order to discuss more properties of the model explicitly, we assume that $\nu_{LT}=2\nu_{R}-1$. Although this assumption restricts the generality of our analysis, it facilitates the later analysis. Moreover, the restriction is not stronger than in the standard model, which implicitly imposes the assumption $\nu_{LT} = 0$. With this assumption, the optimal quantity maximizing firm's profit is

$$q_{fj}(a) = \left[\frac{\left(\nu_R A_j + \sqrt{(\nu_R A_j)^2 - 4 (a w_i \tau_{ij}) (w_i \nu_{LT} B_{ij})} \right)}{2 a w_i \tau_{ij}} \right]^{\frac{1}{1-\nu_R}}. \tag{25}$$

Each firm charges a price given by

$$p_{fj}(a) = \frac{2 a w_i \tau_{ij} A_j}{\left(\nu_R A_j + \sqrt{(\nu_R A_j)^2 - 4 (a w_i \tau_{ij}) (w_i \nu_{LT} B_{ij})} \right)}. \tag{26}$$

From here we see that price depends not only on the marginal cost of production, $a w_i \tau_{ij}$, but also on the market conditions A_j and logistics costs parameter B_{ij}.

It is also straightforward to see that the model features Alchian-Allen effect (see, for example, Hummels and Skiba (2004)). Fix marginal cost parameter a, add the standard quality measure into the utility function, we can calculate the relative demand for goods as

$$\frac{q_{fj}^H}{q_{fj}^L} = \left[\frac{\left(Q^H \nu_R A_j + \sqrt{(Q^H \nu_R A_j)^2 - 4 (a w_i \tau_{ij}) (w_i \nu_{LT} B_{ij})} \right)}{\left(Q^L \nu_R A_j + \sqrt{(Q^L \nu_R A_j)^2 - 4 (a w_i \tau_{ij}) (w_i \nu_{LT} B_{ij})} \right)} \right]^{\frac{1}{1-\nu_R}},$$

where $Q^H > Q^L$ represent high and low level of quality. The Alchian-Allen effect means that relative demand for high quality good is increasing in trade costs, which is true here because $\frac{\partial q_{fj}^H / q_{fj}^L}{\partial B_{ij}} > 0$.

Another interesting feature of this model is that a variety will not be exported to every destination even with zero fixed export cost. Indeed, we can see from equation (25) that first order condition can not be satisfied when the marginal cost is too high that $(\nu_R A_j)^2 - 4 (a w_i \tau_{ij}) (w_i \nu_{LT} B_{ij}) < 0$. Define a_{ij}^* as the zero profit cutoff that satisfies $\pi_{fj}\left(a_{ij}^*\right) = 0$. Export generates positive profit if and only if the marginal cost is lower than the zero profit cutoff, namely $a \leq a_{ij}^*$, where, using $\nu_{LT} = 2\nu_R - 1$,

$$a_{ij}^* = \frac{\left(A_j\right)^2}{4\left(w_i \tau_{ij}\right)\left(w_i B_{ij}\right)}. \tag{27}$$

It is easy to verify that $a \leq a_{ij}^*$ implies that first order condition is satisfied. Therefore there is a simple entry rule for this model: for each destination, if $a \leq a_{ij}^*$, firm will export; otherwise it is better not to trade. When $\nu_{LT} = 0$, the zero profit cutoff is equal to

$$a_{ij}^*|_{\nu_{LT}=0} = \left(\frac{1-\nu_R}{\nu_R}\right)^{\frac{1-\nu_R}{\nu_R}} \frac{\left(\nu_R A_j\right)^{\frac{1}{\nu_R}}}{\left(w_i \kappa_{I,ij}\right)^{\frac{1-\nu_R}{\nu_R}}\left(w_i \tau_{ij}\right)}. \tag{28}$$

On the path $\nu_{LT} = 2\nu_R - 1$, when $\nu_{LT} = 0$, $\nu_R = \frac{1}{2}$. In this case, (27) and (28) are the same. But when $\nu_{LT} > 0$, conditional on the same value of ν_R, we see that two expressions differ. The presence of cost per shipment implies that inventory costs alone do not determine the trade costs. Instead, the combination of coordination cost, inventory cost, and iceberg trade cost decides the threshold of export.

4 Data

We use Chinese custom data to document facts that support our model. The data contains the universe of international trade transactions of Chinese firms from 2000 to 2006. Table 1 shows number of exporters, products measured by HS8 codes, number of destinations, and total number of transactions in each year.

year	No. firm in trade data	No. products	No. countries	No. shipments
2000	62750	6734	222	5193140
2001	68459	6722	226	5900901
2002	78530	6889	228	7352879
2003	95610	7009	229	9215403
2004	120515	7014	229	11228228
2005	143895	7125	234	13674674
2006	208312	7416	236	25623829

Note: Summary statistics for each year. Products are defined by HS8 codes.

Table 1: Summary statistics

5 Empirical analysis

5.1 Estimation of ν_{LT}

The inventory management problem provides a straightforward way to estimate ν_{LT}. After taking log, the equation (24) becomes:

$$\ln F = \ln \left(\frac{1}{\nu_{LT}} - 1 \right)^{-\nu_{LT}} \kappa_I^{\nu_{LT}} \kappa_T^{-\nu_{LT}} + (1 - \nu_{LT}) \ln q, \tag{29}$$

therefore we can estimate ν_{LT} by simply regressing log shipment frequency on log quantity. We first estimate ν_{LT} using the whole sample. Since the data contains a 7-year time series, we include quadratic function of the number of years of positive trade for each firm-product-destination to control for experience effect. More specifically, we run the regression

$$\ln F_{fojt} = \beta_0 + (1 - \nu_{LT}) \ln q_{fojt} + \beta_2 Exp_{fojt} + \beta_3 \left(Exp_{fojt} \right)^2 + \phi_{foj} + u_{fojt}, \tag{30}$$

where F_{fojt} is the number of shipment frequency firm f, product o, export to destination j in year t, q_{fojt} is trade volume, Exp_{fojt} is a measure of past trade experience, and ϕ_{foj} is the firm-product-destination fixed effects. We add the experience measure and fixed effects in order to control for the potential heterogeneity in $\kappa_I^{\nu_{LT}} \kappa_T^{-\nu_{LT}}$. As shown in Table 2, the estimate of ν_{LT} is close to 0.6. Importantly, the estimated value is significantly difference from either 0 or 1, showing that commonly assumed trade cost structure that imposes $\nu_{LT} = 0$ is not consistent with the data.

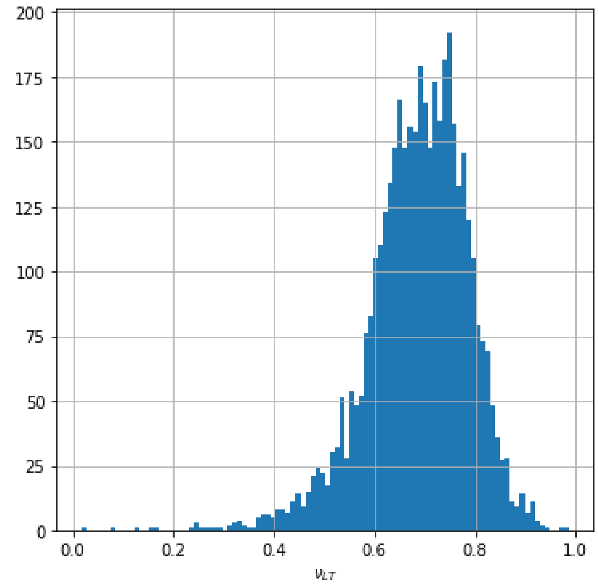
Variable	Estimate
ν_{LT}	0.6165^{***}
	(0.001)
β_2	0.0318^{***}
	(0.001)
β_3	-0.0163^{***}
	(0.000)
No. Obs.	11623454
R^2	0.495

Note: Estimate of ν_{LT} using the whole sample. Firm-product-destination fixed effects are controlled. Standard errors are reported in the brackets.

Table 2: Estimation of ν_{LT}

To have a better idea of potential heterogeneity of coordination costs across different sub-samples, using the year 2006 data, we run the frequency regression (29) for each product defined by the HS8 codes, controlling for the destination fixed effects. Figure 1 shows the distribution of estimated ν_{LT} across products, where we exclude products that have less than 30 observations. Note that consistent with the estimate using the whole sample, we clearly rejects the hypothesis that $\nu_{LT} = 0$. While $\nu_{LT} = 0.5$ or $\nu_{LT} = 1$ is closer to the data, the mean of this distribution is 0.66, with a small standard deviation equal to 0.09 [7].

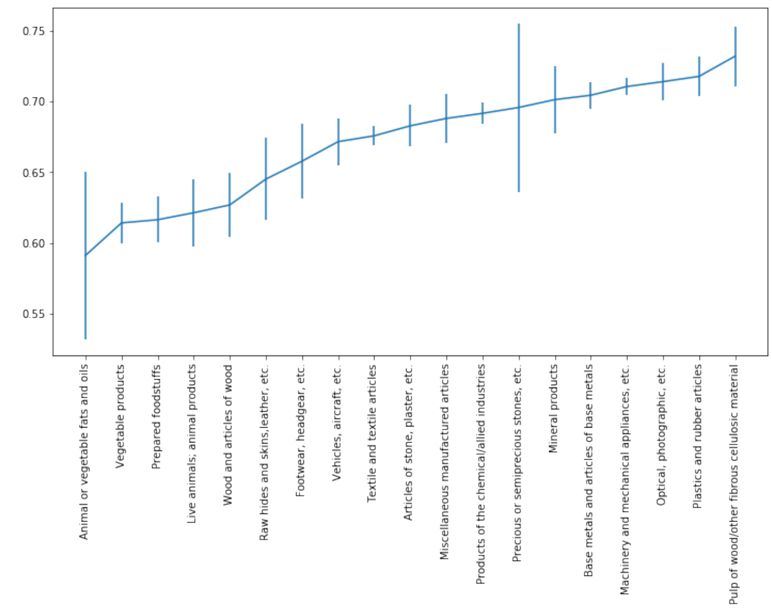
[7]Note that in contrast with theory, the frequency of shipment is measured as number of transactions, and therefore is a discrete variable in the data. To test whether this issue significantly affects our result, we run a Poisson regression instead of ordinary linear regression, and find that the results only change slightly.



Note: Distribution of estimated ν_{LT} across products. Products that have less than 30 observations are excluded. Mean of the estimates is 0.66, standard deviation is 0.09.

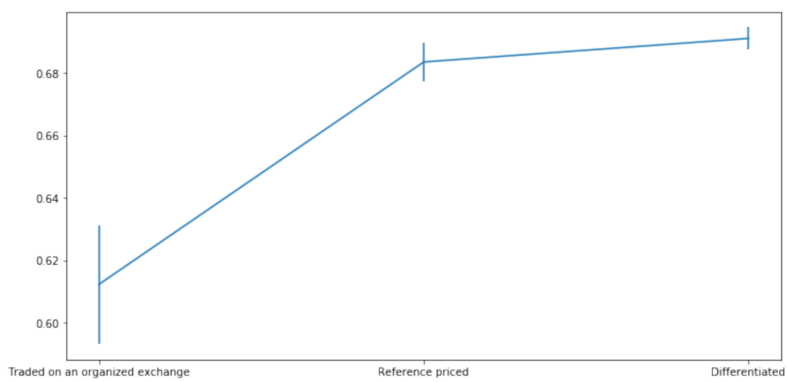
Figure 1: Distribution of estimated ν_{LT} across products

In order to understand what drives the heterogeneity of ν_{LT} across products, we regress the estimated product level ν_{LT} on the sector dummy. Figure 2 shows that the average value of ν_{LT} ranges from slightly above 0.55 to about 0.7. The vertical lines shows the 95 confidence interval of ν_{LT} in each sector. Sectors are ordered by the estimated value of ν_{LT}. Products that seem to be more homogeneous, like animal fat and prepared foodstuffs, tend to have lower ν_{LT} and therefore a lower logistics costs conditional on the quantity shipped. On the other hand, the more complex products like machinery tend to have a higher value of ν_{LT}. Consistent with this intuition, Figure 3 shows that when we regress estimated ν_{LT} across categories defined by the Rauch classification, a clear sorting pattern emerges: the differentiated products have a significantly higher value of ν_{LT} than the products having a reference price, which has ν_{LT} higher than the products traded in the organized exchange.



Note: Estimated ν_{LT} across sectors. The vertical lines shows the 95 confidence intervals of estimated ν_{LT} in each sector.

Figure 2: Estimated ν_{LT} across sectors of products



Note: Estimated ν_{LT} across Rauch categories. The vertical lines shows the 95 confidence intervals of estimated ν_{LT} in each category.

Figure 3: Estimated ν_{LT} across Rauch classifications

5.2 Welfare gains from trade

With estimated ν_{LT}, we can compare gains from trade across models. In Table 3, we show the gains from trade that move from the autarky to the observed level of expenditure share on domestic goods for a number of countries. We compare the gains from trade under the assumption that $\nu_{LT} = 0$ and $\nu_{LT} = 0.6$. ν_R is held to be 0.8, which is equivalent to elasticity of substitution equal to 5, consistent with the large literature estimating this parameter. The data of expenditure share comes from the WIOD in 2008 (see Timmer *et al.* (2015)). The model under $\nu_{LT} = 0.6$ gives a lower welfare gains from trade. The difference is higher for countries like Hungary where the share of expenditure on domestic goods is lower. Overall, we observe a 11.5% lower welfare gains from trade on average.

	Gains from Trade, %			Gains from Trade, %	
	$\nu_{LT} = 0$	$\nu_{LT} = 0.6$		$\nu_{LT} = 0$	$\nu_{LT} = 0.6$
Country	(1)	(2)	Country	(1)	(2)
AUS	2.32	2.08	IRL	8.04	7.21
AUT	5.65	5.06	ITA	2.89	2.58
BEL	7.49	6.72	JPN	1.69	1.51
BRA	1.50	1.34	KOR	4.30	3.85
CAN	3.77	3.37	MEX	3.30	2.95
CHN	2.65	2.37	NLD	6.16	5.52
CZE	6.00	5.37	POL	4.36	3.90
DEU	4.47	4.00	PRT	4.40	3.94
DNK	5.75	5.15	ROM	4.46	3.99
ESP	3.10	2.77	RUS	2.41	2.15
FIN	4.40	3.94	SVK	7.63	6.84
FRA	2.99	2.67	SVN	6.83	6.13
GBR	3.23	2.89	SWE	5.06	4.53
GRC	4.20	3.76	TUR	2.87	2.57
HUN	8.08	7.25	TWN	6.11	5.47
IDN	2.90	2.60	USA	1.77	1.58
IND	2.37	2.12	ROW	5.23	4.68
			Average	4.36	3.91

Notes: Average values in the last row are calculated based on the full set of countries.

Table 3: Gains from Trade

5.3 Model calibration

In this section, we show how to calibrate other model parameters, in particular other parameters related to trade costs, using trade data. The goal of the calibration exercise is to compare the implications in terms of the inferred trade costs necessary to rationalize observed trade flows. Since we can only identify the product between wage and other parameter, throughout this section, we normalize wage to one. Since we don't have data on inventory costs, we follow the literature on inventory management (e.g. Alessandria *et al.* (2010)) and set κ_I equal to 0.15, which is equivalent to a annual inventory costs of 30% of the stored quantities. ν_{LT} is set to be 0.6, consistent with the estimate using the whole sample shown in Table 2. The demand side parameter ν_R is set to be 0.8, implying the elasticity of substitution is equal to 5. This value is consistent with the literature that estimates this parameter using different data and approaches (see e.g. Anderson and van Wincoop (2004)). Note that under this assumption $\nu_{LT} = 2\nu_R - 1$ is satisfied. Next we first show the intuition of our identification strategy under the assumption that productivity follows Pareto distribution. However, the combination of Melitz model and Pareto distributed productivity yields predictions contradicting with the data regarding the intensive margin, as pointed out by Fernandes *et al.* (2019). This point is crucial for our analysis, as we use variations in the intensive margin to identify country pairwise trade costs. We therefore perform the calibration under the assumption that productivity follows log-normal distribution.

5.3.1 Calibration under Pareto

We first illustrate the idea of identifying trade costs parameters using flows under the assumption of $G(a)$ being Pareto distribution and $f_{ij}^x = 0$. Under these assumptions, we can apply the aggregation formula (16) to get the closed form solution for the bilateral trade flows. The number of firms that export from i to j is

$$N_{ij} = N_i^e \int_0^{a_{ij}^*} dG(a) = N_i^e G\left(a_{ij}^*\right)$$

$$= H(0) \left(A_j\right)^{\frac{\nu_{LT}\theta-\theta}{\nu_{LT}-\nu_R}} Z_{ij} \left(w_i B_{ij}\right)^{-\frac{\theta(1-\nu_R)}{\nu_R-\nu_{LT}}}, \tag{31}$$

where $Z_{ij} = N_i^e \kappa_{G,i}^{\theta} \left(w_i \tau_{ij}\right)^{-\theta}$. Therefore together with the expression on bilateral trade flow (17), the intensive margin is

$$\frac{X_{ij}}{N_{ij}} = \frac{H(\nu_R)}{H(0)} \left(A_j\right)^{\frac{\nu_{LT}}{\nu_{LT}-\nu_R}} \left(w_i B_{ij}\right)^{\frac{\nu_R}{\nu_R-\nu_{LT}}}. \tag{32}$$

The total frequency of shipments is

$$F_{ij} = N_i^e \int_0^{a_{ij}^*} F_{fj} dG(a) = N_i^e \int_0^{a_{ij}^*} \frac{1}{1-\nu_{LT}} \kappa_I B_{ij}^{-1} q_{fj}^{(1-\nu_{LT})} dG(a)$$

$$= \frac{H(1-\nu_{LT})}{1-\nu_{LT}} \kappa_I \left(A_j\right)^{\frac{(1-\nu_{LT})+\nu_{LT}\theta-\theta}{\nu_{LT}-\nu_R}} Z_{ij} \left(w_i B_{ij}\right)^{-\frac{1-\nu_R+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}}.$$

Average shipment frequency is

$$\frac{F_{ij}}{N_{ij}} = \frac{H(1-\nu_{LT})}{H(0)(1-\nu_{LT})} \kappa_I \left(A_j\right)^{\frac{(1-\nu_{LT})}{\nu_{LT}-\nu_R}} \left(w_i B_{ij}\right)^{\frac{1-\nu_R}{\nu_R-\nu_{LT}}} \tag{33}$$

Given that we observe X_{ij}, N_{ij} and F_{ij} in the data, and ν_{LT} can be estimated separately, we can solve for A_j, B_{ij}, Z_{ij} up to a level of wage and the shape parameter of marginal cost distribution θ. Since we observe firm level sales, we can identify θ from the shape of the sales distribution.

However, despite the transparency in identification, the assumption of Pareto generates peculiar predictions regarding the intensive margins of trade. Using Chinese Custom data in 2006, we estimate the gravity equation using exports from each province to each country, restrict to textile products to control for variation in inventory costs, which we hold to be constant. As shown in Table 4, we see that both average frequency of shipment and value per shipment decreases significantly with distance. Distance elasticity changes substantially from -1.90 to -1.17 when we include shipment frequency into the regression. However, because $\frac{1-\nu_R}{\nu_R-\nu_{LT}} > 0$, the model interprets the negative distance elasticity of $\frac{F_{ij}}{N_{ij}}$ as B_{ij} decreases with distance, which means that we need an even larger magnitude of τ_{ij} to rationalize the observed trade flows. We could not think of a plausible theoretical reason for this to hold true. Fortunately, as suggested in Fernandes *et al.* (2019), when the marginal costs follow log-normal distribution, the Melitz model is able to match the patterns of intensive margin. Therefore in the next sub-section, we calibrate the model using the same logic but with the assumption that the marginal costs follow log-normal distribution.

	$\ln\frac{X_{ij}}{F_{ij}}$	t-value	$\ln\frac{F_{ij}}{N_{ij}}$	t-value	$\ln X_{ij}$	t-value	$\ln X_{ij}$	t-value
$\ln Dist_{ij}$	-0.3541	-2.748	-0.4332	-5.4264	-1.9022	-9.2988	-1.1663	-7.5610
$\ln F_{ij}$							1.6989	40.6196
No. Obs	2235		2235		2235		2235	
Adj. R^2	0.260		0.364		0.698		0.831	

Note: estimation results from the gravity equation of shipment frequency. Origin and destination fixed effects are controlled.

Table 4: Gravity estimates

5.3.2 Calibration under log-normal

In this subsection, we calibrate trade costs parameters using the observed trade flows for models with $\nu_{LT} = 0$ and with $\nu_{LT} = 0.6$. For computational reason, we use firm level trade data from top 10 Chinese provinces[8] to top 10 destinations[9]. In order to avoid the problem of decreasing trade costs with distance, we assume that marginal cost follows the log-normal distribution,

$$G_i(a) = \Phi\left(\frac{\ln a - \mu_{a,i}}{\sigma_a}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable. The distribution assumption is also motivated partly by the fact that log normal assumption explains the whole distribution of export sales very well, as shown in Head *et al.* (2014).

In order to match the main empirical patterns, we extend the model such that each firm takes a draw of fixed export cost in each destination independently. Namely the profit function now becomes

$$\pi_{fj} = A_j q_{fj}^{\nu_R} - w_i B_{ij} q_{fj}^{\nu_{LT}} - a w_i \tau_{ij} q_{fj} - w_i \xi_{fj} f_{ij}^{x},$$

where $\ln \xi_{fj}$ follows a normal distribution with zero mean and variance σ_ξ^2. Since either

[8] Provinces include Guangdong, Jiangsu, Shanghai, Zhejiang, Shandong, Fujian, Tianjin, Beijing, Liaoning, Hebei

[9] Destinations include the US, Hong Kong, Japan, Korea, Germany, Netherlands, Singapore, UK, Taiwan and Canada

the first order condition nor the logistics costs minimization changes, the firm's choice of quantity q_{fj} or shipment frequency F_{fj} do not change. However, with heterogeneous fixed costs, the zero profit quantity $q^*_{ij}(\xi)$ now depends on the value of fixed cost shock. This feature allows the model to explain the presence of small firms in each destination, because even unproductive firms, as long as $w_i \xi_{fj} f^x_{ij}$ is low enough, may earn positive profits and enter.

Our first set of moments target percentiles of firm level sales and shipment frequency distributions for every province-destination pair. Compared with the literature estimating trade costs, we utilize the information from firm level shipment frequency. More specifically, we target $10th$, $15th$, ..., $95th$ percentiles of sales distribution, and $20th$, $35th$, ..., $95th$, percentiles of shipment frequency distribution. The targets of shipment frequency start from $20th$ percentile because in the data, even tiny export volume must have at least one shipment, and the firm level shipment frequency at low percentiles hit this lower bound and are therefore less informative about trade costs.

The shape of the distributions are informative about the variance of marginal costs σ^2_a, because within the province-destination, the marginal cost a is the only source of firm heterogeneity in sales or shipment frequency. The observed firm level sales and shipment frequency distribution represent mix of entrants across different levels of marginal cost, thus the shapes of the distributions are also informative about the variance σ^2_ξ. To see how we identify other parameters, the mean of the sales distribution

$$\frac{X_{ij}}{N_{ij}} = A_j \int \int_0^{a^*_{ij}} q^{\nu_R}_{fj} d \frac{G(a)}{G\left(a^*_{ij}\right)} dG_S(\xi),$$

for example, is helpful to identify A_j, τ_{ij} and B_{ij}. Similarly, the mean of shipment frequency distribution

$$\frac{F_{ij}}{N_{ij}} = \frac{1}{1-\nu_{LT}} \kappa_I B^{-1}_{ij} \int \int_0^{a^*_{ij}} q^{(1-\nu_{LT})}_{fj} d \frac{G(a)}{G\left(a^*_{ij}\right)} dG_S(\xi)$$

is helpful to identify B_{ij} and τ_{ij}. Moments other than the mean will also contribute to the identification of the parameters.

The extensive margin in this extended model can be expressed as:

$$N_{ij} = N_i^e \int \int_0^{a_{ij}^*(\xi)} dG\left(a\right) dG_S\left(\xi\right) = N_i^e \int G\left(a_{ij}^*\left(\xi\right)\right) dG_S\left(\xi\right),$$

where $G\left(a_{ij}^*\left(\xi\right)\right)$ is the share of firms find it profitable to enter given ξ, $\int G\left(a_{ij}^*\left(\xi\right)\right) dG_S\left(\xi\right)$ is the share of firms find it profitable to enter. We use the the relative extensive margin between different destinations within a province as the second set of target moments, which is

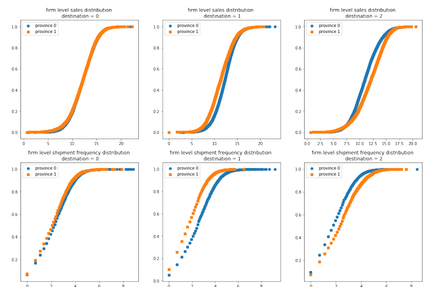
$$\frac{N_{ij}}{N_{i1}} = \frac{\int G\left(a_{ij}^*\left(\xi\right)\right) dG_S\left(\xi\right)}{\int G\left(a_{i1}^*\left(\xi\right)\right) dG_S\left(\xi\right)}.$$

By changing the f_{ij}^x , we can adjust N_{ij} to fit the relative extensive margin. Of course, the rise in B_{ij} or τ_{ij} also decrease N_{ij}, but for such a change, we should observe also a corresponding change in firm level sales or shipment frequency. Therefore the variation of relative extensive margin given sales and shipment frequency distributions identifies f_{ij}^x.

We make several identification assumptions in the above calibration exercise. Since $\mu_{a,i}$ is not going to be identified separately from τ_{ij}, we set it to zero. Similarly, wage is normalized to be one across provinces. Importantly, the relative extensive margin only identifies relative levels of f_{ij}^x within a province i. This can be seen from the fact that if f_{ij}^x are increased for all destinations such that N_{ij} decrease the same portion for all destinations, then the relative extensive margins do not change. These assumptions will not drive our main empirical findings, because we are interested in comparing distance elasticity controlling for province and destination effects across model. The province effects should absorb the effects of wage of province level productivity differences, and taking difference across models makes these assumptions even less likely to affect our results.

Table 4 show the empirical distributions of firm sales and shipment frequency across two provinces and three destinations. The first row shows the six sales distributions. The upper left figure shows that Province 0 (Guandong) and Province 1 (Jiangsu) have a similar sales to Destination 0 (US). The upper middle and upper right figures show that median firm from Province 0 has a higher sales to destination 1, but lower sales to

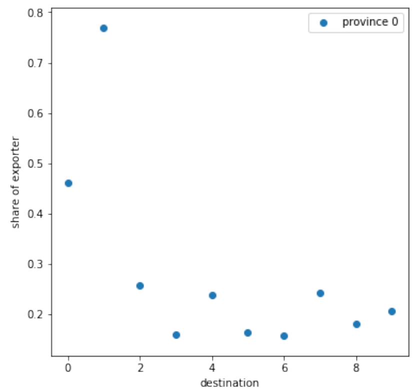
Destination 2 (Japan). This reflects the fact that Province 0 is closer to Destination 1 (Hong Kong) but further away from Destination 2. The distributions of shipment frequency at the second row exhibit similar patterns. These cross-province variations in the firm distributions conditional on the destination play a significant role in our identification strategy.



Note: The figures plot the empirical distributions of sales (the first row) and shipment frequency (the second row) from two provinces to three destinations. Province 0 is plotted in blue, and Province 1 is plotted in orange. Each column represents a different destination.

Figure 4: Empirical distributions

Figure 5 shows the share of exporters that export to each destination (X-axis shows the label of each destination) for Province 0. Note there are substantial variation across destinations. For example, over 70% of exporters export to Destination 1, but less than 20% exporters export to Destinations 3. We target share of exporters because we do not have a good estimate for the potential mass of entrants and therefore do not identify the level of fixed costs. To see this, if we observe a low number of exporters, it could be either due to a small mass of entrants, or a high level of fixed costs.



Note: The figure plots the share of firms from Province 0 that export to each of the ten destinations.

Figure 5: Extensive margins

Compared with the model in Eaton *et al.* 2011, this model abstracts from idiosyncratic sales across destinations within a firm. It does not allow for a big firm in one destination to sell little in another destination. This model also does not allow for idiosyncratic shipment frequency across firms conditional on sales. That is, it predicts that a perfect correlation between sales and shipment frequency. All these problems can be addressed by introducing more firm-destination specific shocks on demand and logistics costs. Since introducing such features do not contribute to our main goal of inferring province-destination level trade costs, we choose not to further complicate the model.

We calibrate the model by the method of simulated moments. The simulation algorithm is summarized as follows:

- For each province, we simulate 10,000 firms, indexed by s. Each simulated firm takes a productivity draw $a_i(s)$, and for each destination a fixed export cost draw $\xi_{ij}(s)$
- Given other parameters, we solve for f_{ij}^x by matching the relative extensive margin
- In the outer loop, search over σ_ξ, σ_a, A_j, B_{ij} , τ_{ij} to match firm level sales and shipment frequency distribution in each destination

With the restriction $\nu_{LT} = 0$, we repeat the procedure to calibrate trade costs except for B_{ij} without using the information from shipment frequency. The calibration result of σ_ξ

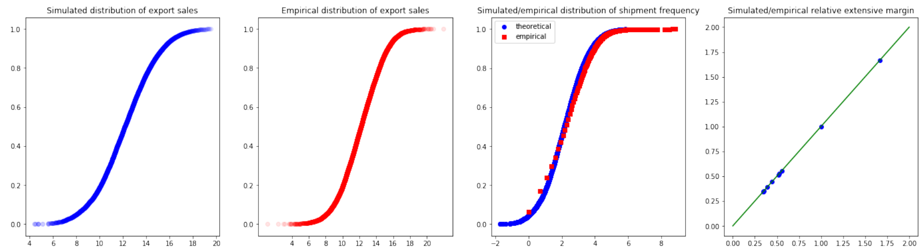
and σ_a is shown in Table 5. We run the algorithm from different random initial points, and report the result having the lowest error. The standard deviation of calibrated σ_ξ is 0.8, the standard deviation of calibrated σ_a is 0.03. The inferred σ_ξ is higher under model with $\nu_{LT} = 0.6$. It comes from the fact that productive firms have higher advantage when $\nu_{LT} > 0$, therefore in order to rationalize the presence of small firms we need more extreme shocks on fixed costs.

	σ_ξ	σ_a
$\nu_{LT} = 0.6$	3.92	0.71
	(0.80)	(0.03)
$\nu_{LT} = 0$	3.11	0.72
	(0.27)	(0.04)

Note: Estimates reported are the ones with lowest error from different random initial points. The standard deviations across calibrated parameters are reported in the bracket under the parameter value.

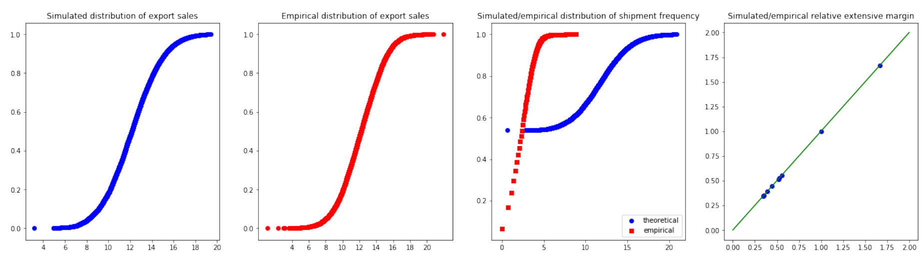
Table 5: Estimates of σ_ξ, σ_a

Figure 6 shows the model fit of model with $\nu_{LT} = 0.6$ for trade margins from Guangdong Province to the US. The first panel shows the simulated distribution of log export sales, and the second panel shows the corresponding empirical distribution. The fit is good for the whole distribution. Similarly, the third panels shows that the model with $\nu_{LT} = 0.6$ fits well the shipment frequency distribution. In the fourth panel, the simulated ratio of $\frac{N_{ij}}{N_{iUS}}$ is plotted against its empirical counterpart. All points lie well on the 45 degree line. Figure 7 shows the model fit of model with $\nu_{LT} = 0$. The standard model explains the firm level export sales equally well, but has little power in terms of explaining the distribution of shipment frequency.



Note: The first panel shows the simulated distribution of log export sales from Guangdong Province to the US. The second panel shows the corresponding empirical distribution. The third panel shows the simulated and empirical distribution of shipment frequency. The fourth panel plots the simulated $\frac{N_{ij}}{N_{iUS}}$ against the corresponding empirical relative extensive margin for Guangdong Province.

Figure 6: Model fit, $\nu_{LT} = 0.6$



Note: The first panel shows the simulated distribution of export log sales from Guangdong Province to US. The second panel shows the corresponding empirical distribution. The third panel shows the simulated and empirical distribution of shipment frequency. The fourth panel plots the simulated $\frac{N_{ij}}{N_{iUS}}$ against the corresponding empirical relative extensive margin for Guangdong Province.

Figure 7: Model fit, $\nu_{LT} = 0$

Our main interest is to compare the distance elasticity of trade costs across models. To do this, we run the regression

$$\ln y_{ij} = \beta_0 + \beta_1 \ln Dist_{ij} + \phi_i + \psi_j + u_{ij},$$

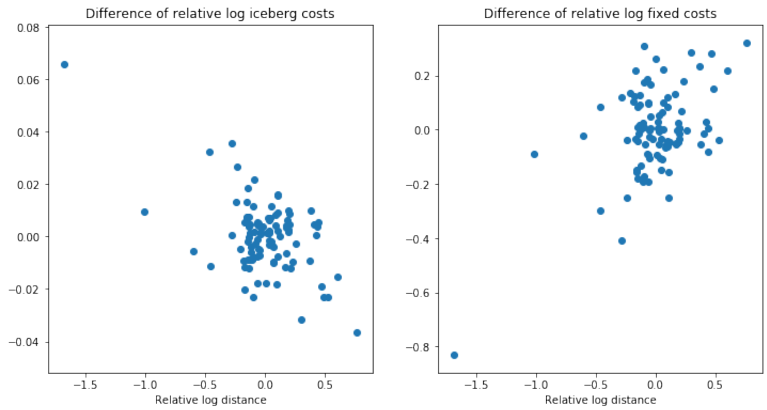
where $y_{ij} = B_{ij}^{LT}$, τ_{ij}^{LT}, $f_{ij}^{x,LT}$ are calibrated parameters under $\nu_{LT} = 0.6$ and $y_{ij} = \tau_{ij}^{MEL}$, $f_{ij}^{x,MEL}$ are parameters calibrated under $\nu_{LT} = 0$. The estimation results of both models are shown in Table 6. The first four rows show the correlation between distance and estimated trade cost parameters in the extended model, after controlling for province and destination fixed effects. And the lower four rows show the same estimates for the restricted model. We see that when $\nu_{LT} = 0.6$, B_{ij}, τ_{ij} and f_{ij}^x all increase with distance. B_{ij} has the lowest distance elasticity. Comparing with the standard model, the distance elasticity of τ_{ij} decreases from 0.26 to about 0.24, but the distance elasticity of f_{ij}^x increases from 0.97 to 1.22.

To see the difference more clearly, we first regress both log distance and log trade costs parameters ($\ln \tau_{ij}^{LT}$, $\ln f_{ij}^{x,LT}$ and $\ln \tau_{ij}^{MEL}$, $\ln f_{ij}^{x,MEL}$) to province and destination fixed effects, then calculate the difference of residuals of trade costs parameters. Figure 8 plots the residual of distance and the difference of residuals of trade costs parameters. The left figure shows that at closer destinations, the model with $\nu_{LT} = 0.6$ requires a higher normalized τ_{ij} to rationalize the observed trade flows. This is because the model with $\nu_{LT} = 0.6$ infers from the shipment frequency data that B_{ij} is increasing in distance, therefore at the closer destinations, B_{ij} is relatively lower compared to the reference destination, which implies relatively a higher τ_{ij} is needed to rationalize the same trade flow. While in the model with $\nu_{LT} = 0$, B_{ij} is held equal to the constant κ_I, therefore there is no such a force that pushes τ_{ij} to be higher. The similar logic applies to the further destination. On the other hand, the model with $\nu_{LT} = 0.6$ indicates a steeper curve of f_{ij}^x with distance. This is due to the scale of economy in the transportation technology, which, for instance, leads to a higher fixed export costs in the further destination in order to match the patterns in the extensive margin. Overall, this result shows that correctly incorporating the information in the shipment frequency systematically changes our interpretation of the observed trade patterns.

$\nu_{LT} = 0.6$	$\ln B_{ij}$	t-value	$\ln \tau_{ij}$	t-value	$\ln f_{ij}^x$	t-value
$\ln Dist_{ij}$	0.1364	4.3392	0.2374	6.9384	1.2210	7.8046
No. Obs	100		100		100	
adj. R^2	0.70		0.76		0.69	
$\nu_{LT} = 0$			$\ln \tau_{ij}$	t-value	$\ln f_{ij}^x$	t-value
$\ln Dist_{ij}$			0.2610	6.8978	0.9695	8.0976
No. Obs			100		100	
adj. R^2			0.63		0.73	

Note: Correlation between estimated trade costs and distance, controlling for province and destination fixed effects.

Table 6: Correlation between estimated trade costs and distance



Note: The left panel plots the log difference of normalized iceberg trade costs between the two models and the normalized log distance. The right panel plots the log difference of normalized fixed export costs between the two models and the normalized log distance.

Figure 8: Comparison between estimated trade costs and distance

5.4 How big is the trade costs per shipment?

In this sub-section, we show how to get a sense of magnitude of trade costs per shipment in our model. To ease the exposition, we specify the trade costs in ad valerom terms. The inventory cost and coordination cost are now specified as

$$C_I\left(X_s\right) = \kappa_I X_s,$$

$$C_T\left(X_s\right) = \kappa_T X_s^{\alpha},$$

where X_s is the value, instead of quantity, per shipment. Specifying the model in this way does not change the basic idea, but allows for an easy way to quantify the magnitude of trade costs. Following the same procedure, we can obtain the logistics costs as

$$C_{LT}\left(X\right) = B_{ij} X^{\nu_{LT}} = \frac{1}{\left(1-\nu_{LT}\right)} \kappa_I F^{-1} X, \tag{34}$$

where F is the shipment frequency, and X is the total export value. Fix the product as "Men's and boys'shirts of cotton, without specially made collar (HS code 62052000)", and set $\kappa_I = 0.175$, $\nu_{LT} = 0.6$, we calculate how big are the trade costs. We fix the product so that the assumption of constant inventory cost is of less concern. The chosen product is a popular export product with annual export value at 1.18 billion US dollar. Because the average inventory level across the year is $\frac{X_2}{2}$, $\kappa_I = 0.175$ implies that annual cost of inventory is 35% of the average inventory level. This is comparable to the 35% depreciation rate assumed in Alessandria *et al.* (2010). In these case we get a export value weighted average logistics costs equal to 5.37% of the total export value, out of which 2.15% is attributed to inventory costs, and 3.22% is attributed to coordination costs. The magnitude of coordination costs is close to the 3.6% reported in Alessandria *et al.* (2010). The results also show a substantial heterogeneity of trade costs across firms. For about 41% of the observations, the trade is done in one shipment, implying the share of logistics costs out of export value is 43.75%, while the observation at the 25% quantile has only to pay 8.75% of export value. The fact that overall average is even lower indicates that most of the trade is done by big firms that ship very frequently.

5.5 Dynamics of shipment frequency

The assumption on the inventory costs allow us to get interesting results in the previous subsection, but admittedly it is a strong assumption. In this sub-section, we exploit the dynamic aspects of our data to test the model under a milder assumption on the inventory cost. Note equation (24) shows that the proper ratio of shipment frequency and export quantity reflects the trade off between inventory cost and coordination cost:

$$\ln \frac{\kappa_T}{\kappa_I} = \ln \left(\frac{\nu_{LT}}{1-\nu_{LT}} \right) F^{-\frac{1}{\nu_{LT}}} q^{\frac{1-\nu_{LT}}{\nu_{LT}}}.$$

We are going to assume that coordination cost denoted by $\kappa_{T,foj}$ is firm-product-destination specific, while inventory cost parameter $\kappa_{I,oj}$ is product-destination specific. Under this assumption, conditional on the product-destination, we can identify the relative coordination cost across firms as

$$\ln \kappa_{T,foj} - \frac{1}{N_{oj}} \sum_f \ln \kappa_{T,foj} = \ln F_{foj}^{-\frac{1}{\nu_{LT}}} q_{foj}^{\frac{1-\nu_{LT}}{\nu_{LT}}} - \frac{1}{N_{oj}} \sum_f \ln F_{foj}^{-\frac{1}{\nu_{LT}}} q_{foj}^{\frac{1-\nu_{LT}}{\nu_{LT}}},$$

where N_{oj} is the number of firms that export a positive quantity of product o to destination j. We utilize the dynamic aspect of Chinese custom data to construct the relative coordination cost estimates using 2000-2005 data, then use them to explain the shipment frequency in year 2006. More concretely, we run the OLS regressions (30) using 2000-2005 data for each product-destination, selecting product-destination with more than 100 firms, and firm-product-destination with at least 3 years' observations, using number of past years with positive trade as a measure of experience. The experience measure is included to handle the possibility that coordination costs vary across time given a firm-product. For example, trade partners may get more familiar with each other as the relationship gets longer, or the inspection cost may decrease as the buyer collects more information about the producer. Comparing with (24), we see that the fixed effects capture the coordination costs we would like to estimate:

$$\phi_{foj} = \nu_{LT} \ln \kappa_{I,oj} - \nu_{LT} \ln \kappa_{T,foj}.$$

We can obtain an estimate of the coordination costs relative to the mean as

$$-\frac{1}{\nu_{LT}} \left(\hat{\phi}_{foj} - \frac{1}{N_{oj}} \sum_{f} \hat{\phi}_{foj} \right) = \ln \kappa_{T,foj} - \frac{1}{N_{oj}} \sum_{f} \ln \kappa_{T,foj}. \tag{35}$$

After obtaining this estimate, we divide the firm-product-destinations in year 2006 into 4 equally sized groups according to total export value, then within each group, we divide observations to 10 equally sized bins according to estimated coordination costs using 2000-2005 data. For each bin, we then calculate the average frequency of shipment and value per shipment, also demeaned at the product-destination level. In Figure 9, the title of each figure shows the range of log total export value for each group, the X-axis represents mean of relative coordination cost in each bin, the Y-axis represents (log) relative frequency of shipment or value per shipment, with the number of the bin with smallest coordination cost is normalized to 0. We see clearly that as coordination costs increase, the value per shipment increases, while frequency of shipment decreases, as predicted by the model.

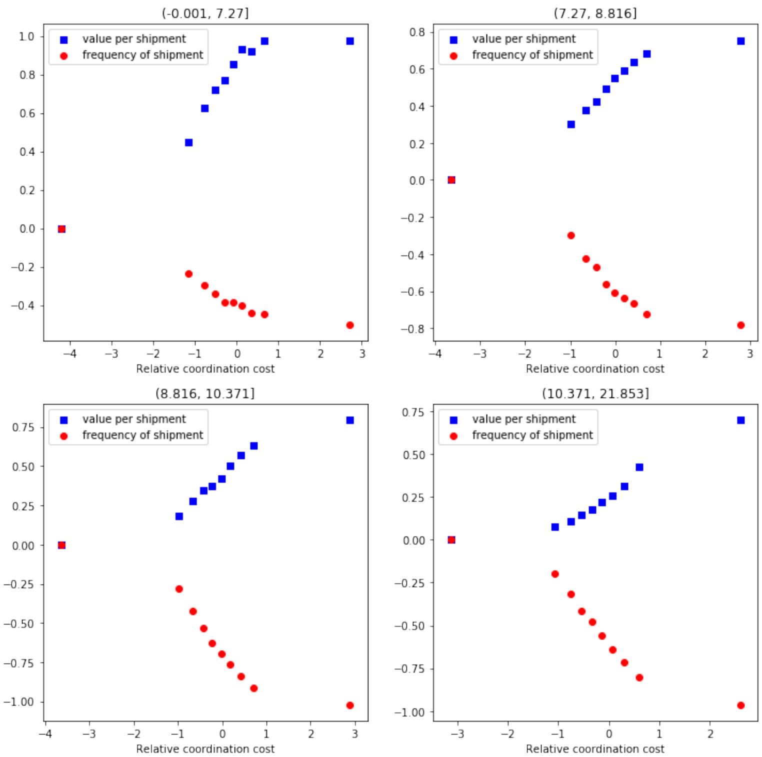


Figure 9: Relative coordination cost and shipment arrangement

5.6 Other evidence

In this subsection we use the sample of Guangdong province to show evidence that supports the idea how firms arrange shipment may reflect their underlying trade costs. Using data from a single province controls for the potential origin effects. Guangdong Province is the largest province in terms of trade value, accounting for about 30% of total trade value in China. We first examine the difference in shipment frequency between intermediaries and other firms. The literature studies intermediaries (e.g. Bai *et al.* (2017)) point out that firms export via intermediaries pay less entry cost, but incur higher variable cost, because producers have to pay a margin to intermediaries, and they also lose the chance of directly interact with buyers , thus are less likely to build long-term trust with buyers. This argument shows that trade via intermediaries is likely to have higher coordination cost and thus lower frequency of shipment. The regression we run to test this is

$$\log F_{foj} = \gamma_0 + (1 - \nu_{LT}) \log q_{foj} + \phi_{oj} + u_{foj},$$

where ϕ_{oj} is product-country fixed effect. The regressions are run for both the sub-sample of intermediary firms and non-intermediary firms. table 7 shows that the sub-sample of intermediaries not only has a lower value of intercept, but also has a higher estimated ν_{LT}. If the inventory costs are product specific and therefore on average the same across intermediaries and non-intermediaries, the lower intercept implies a higher coordination cost parameter κ_T for the intermediaries. Moreover, the higher estimated ν_{LT} shows that the logistics costs increase at a faster speed for the intermediaries.

	Intermediary	Non-intermediary
Intercept	-0.0657^{***}	0.0794^{***}
t-value	-153.959	133.821
ν_{LT}	0.838^{***}	0.787^{***}
t-value	704.982	752.8732
No. Obs	1321772	1022501
Adj. R^2	0.27	0.36

Note: Standard deviation is reported in the bracket, product-country fixed effect is controlled.

Table 7: Shipment frequency and intermediary

As another evidence of the associate between shipment frequency and trade costs, we run a gravity style regression and test whether the usual proxies of trade barriers are good predictors of shipment frequency. In the regression, we put annual number of shipment for each firm-product-destination on the left hand side, controls for value of shipment and gravity variables on the right hand side. For the gravity variables, besides GDP and GDP per ca pita, we also add proxy for trade costs including contiguous or not, sharing common language or not, and the cost of business startup in the destination country. The gravity variables are obtained from CEPII. To use the variation across destinations, firm-product fixed effect is controlled. Inference is clustered at firm level. Table 8 shows that contiguity and common language indeed associate higher frequency of shipment, and a higher business startup costs is associated with lower frequency of shipment.

	Estimation result	t-value
$\log X_{foj}$	0.305	214.057
$\log GDP_j$	0.055	78.143
$\log GDPpca_j$	0.016	14.858
$Contig_{ij}$	0.014	17.181
$Enthno_{ij}$	0.042	27.028
$BusiCost_{ij}$	-0.002	-1.956
No. Obs	2491115	
Adj. R^2	0.56	

Note: Firm-product fixed effect is controlled, standard deviation clustered at firm level.

Table 8: Shipment frequency and gravity variables

6 Conclusion

In this paper we show that tractable solutions can be obtained for heterogeneous firm trade model with flexible structure of trade costs, under a simple restriction on the parameter space. Our results show that a more realistic trade cost structure can have substantial impact on evaluation of trade policy. When trade costs depend on the quantity being traded, the trade elasticity is no longer a sufficient statistics for calculating welfare changes given changes in trade costs. An immediate question is that how about the models on the path other than $\nu_{LT} = 2\nu_R - 1$ or $\nu_{LT} = 0$, can we also get solutions by interpolating the solutions on between the lines? Answering the question can significantly expand the class of models that can be used to study international trade.

References

ALESSANDRIA, G., KABOSKI, J. P. and MIDRIGAN, V. (2010). Inventories, lumpy trade, and large devaluations. *American Economic Review*, **100** (5), 2304–2339.

ANDERSON, J. E. and VAN WINCOOP, E. (2004). Trade Costs. *Journal of Economic Literature*, **42** (3), 691–751.

ARKOLAKIS, C., COSTINOT, A. and RODRÍGUEZ-CLARE, A. (2012). New trade models, same old gains? *American Economic Review*, **102** (1), 94–130.

BAI, X., KRISHNA, K. and MA, H. (2017). How you export matters: Export mode, learning and productivity in China. *Journal of International Economics*, **104**, 122–137.

BLUM, B. S., CLARO, S., DASGUPTA, K. and HORSTMANN, I. J. (2019). Inventory management, product quality, and cross country income differences. *American Economic Journal: Macroeconomics*, **11** (1), 338–388.

COSTINOT, A. and RODRÍGUEZ-CLARE, A. (2015). *Trade Theory with Numbers: Quantifying the Consequences of Globalization*, vol. 4. Elsevier B.V.

EATON, J., KORTUM, S. S. and KRAMARZ, F. (2011). An Anatomy of International Trade: Evidence From French Firms. *Econometrica*, **79** (5), 1453–1498.

FABINGER, M. and WEYL, E. G. (2018). Functional Forms for Tractable Economic Models and the Cost Structure of International Trade. pp. 1–99.

FERNANDES, A., KLENOW, P., MELESHCHUK, S., DENISSE PIEROLA, M. and RODRIGUEZ-CLARE, A. (2019). The Intensive Margin in Trade. *IMF Working Papers*, **18** (259), 1.

HEAD, K., MAYER, T. and THOENIG, M. (2014). Welfare and trade without pareto. *American Economic Review*, **104** (5), 310–316.

HUMMELS, D. and SKIBA, A. (2004). Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture. *Journal of Political Economy*, **112** (6), 1384–1402.

KROPF, A. and SAURÉ, P. (2014). Fixed costs per shipment. *Journal of International Economics*, **92**, 166–184.

MELITZ, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, **71** (6), 1695–1725.

TIMMER, M. P., DIETZENBACHER, E., LOS, B., STEHRER, R. and DE VRIES, G. J. (2015). An Illustrated User Guide to the World Input-Output Database: The Case of Global Automotive Production. *Review of International Economics*, **23** (3), 575–605.

A Derivation of aggregation formula

Let us start with calculating the integral $\int_0^{a_{ij}^*} q^\nu dG_i(a)$ for some $\nu > 0$. Using expression (7) for a, and that $G_i(a) = \theta^{-1}(\kappa_{G,i}a)^\theta$, we can derive

$$\int_0^{a_{ij}^*} q^\nu dG_i(a) = -\int_{q_{ij}^*}^\infty q^\nu dG\left(\frac{\nu_R A_j q^{\nu_R-1} - \nu_{LT} w_i B_{ij} q^{\nu_{LT}-1}}{w_i \tau_{ij}}\right)$$

$$= \int_{q_{ij}^*}^\infty \left\{ q^\nu \frac{(1-\nu_R)\nu_R A_j q^{\nu_R-2} - (1-\nu_{LT})\nu_{LT} w_i B_{ij} q^{\nu_{LT}-2}}{w_i \tau_{ij}} \right.$$
$$\left. \times\, \kappa_{G,i}^\theta \left(\frac{\nu_R A_j q^{\nu_R-1} - \nu_{LT} w_i B_{ij} q^{\nu_{LT}-1}}{w_i \tau_{ij}}\right)^{\theta-1} \right\} dq$$

$$= \kappa_{G,i}^\theta (w_i\tau_{ij})^{-\theta}(1-\nu_R)(\nu_R A_j)^\theta \int_{q_{ij}^*}^\infty q^{\nu-1+\nu_R\theta-\theta}\left(1-\frac{\nu_{LT} w_i B_{ij}}{\nu_R A_j} q^{\nu_{LT}-\nu_R}\right)^{\theta-1} dq$$
$$- \left\{ \kappa_{G,i}^\theta (w_i\tau_{ij})^{-\theta}(1-\nu_{LT})\nu_{LT}(\nu_R A_j)^{\theta-1} w_i B_{ij} \right.$$
$$\left. \times \int_{q_{ij}^*}^\infty q^{\nu+\nu_{LT}-2+(\nu_R-1)(\theta-1)}\left(1-\frac{\nu_{LT} w_i B_{ij}}{\nu_R A_j} q^{\nu_{LT}-\nu_R}\right)^{\theta-1} dq \right\}.$$

Introduce the change of variables $x = \left(q/q_{ij}^*\right)^{\nu_{LT}-\nu_R}$. Then

$$\int_0^{a_{ij}^*} q^\nu dG_i(a) = \left\{ \frac{1-\nu_R}{\nu_R-\nu_{LT}} \kappa_{G,i}^\theta (w_i\tau_{ij})^{-\theta}(\nu_R A_j)^\theta \left(q_{ij}^*\right)^{\nu+\nu_R\theta-\theta} \right.$$
$$\left. \times \int_0^1 x^{\frac{\nu+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}-1}\left(1-\frac{\nu_{LT} w_i B_{ij}}{\nu_R A_j}\left(q_{ij}^*\right)^{\nu_{LT}-\nu_R} x\right)^{\theta-1} dx \right\}$$
$$- \left\{ \frac{(1-\nu_{LT})\nu_{LT}}{\nu_R-\nu_{LT}} \kappa_{G,i}^\theta (w_i\tau_{ij})^{-\theta}(\nu_R A_j)^{\theta-1} w_i B_{ij} \left(q_{ij}^*\right)^{\nu+\nu_{LT}-1+(\nu_R-1)(\theta-1)} \right.$$
$$\left. \times \int_0^1 x^{\frac{\nu+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}}\left(1-\frac{\nu_{LT} w_i B_{ij}}{\nu_R A_j}\left(q_{ij}^*\right)^{\nu_{LT}-\nu_R} x\right)^{\theta-1} dx \right\}.$$

Using the fact that

$$q_{ij}^{*}=\left[\frac{w_{i}B_{ij}\left(1-\nu_{LT}\right)}{A_{j}\left(1-\nu_{R}\right)}\right]^{\frac{1}{\nu_{R}-\nu_{LT}}},$$

we get

$$\begin{aligned}\int_{0}^{a_{ij}^{*}}q^{\nu}dG_{i}\left(a\right)=&\left\{\frac{1-\nu_{R}}{\nu_{R}-\nu_{LT}}\kappa_{G,i}^{\theta}\left(w_{i}\tau_{ij}\right)^{-\theta}\left(\nu_{R}A_{j}\right)^{\theta}\left[\frac{w_{i}B_{ij}\left(1-\nu_{LT}\right)}{A_{j}\left(1-\nu_{R}\right)}\right]^{\frac{\nu+\nu_{R}\theta-\theta}{\nu_{R}-\nu_{LT}}}\right.\\&\left.\times\int_{0}^{1}x^{\frac{\nu+\nu_{R}\theta-\theta}{\nu_{LT}-\nu_{R}}-1}\left(1-\frac{\left(1-\nu_{R}\right)\nu_{LT}}{\left(1-\nu_{LT}\right)\nu_{R}}x\right)^{\theta-1}dx\right\}\\&-\left\{\frac{\left(1-\nu_{LT}\right)\nu_{LT}}{\nu_{R}-\nu_{LT}}\kappa_{G,i}^{\theta}\left(w_{i}\tau_{ij}\right)^{-\theta}\left(\nu_{R}A_{j}\right)^{\theta-1}w_{i}B_{ij}\left[\frac{w_{i}B_{ij}\left(1-\nu_{LT}\right)}{A_{j}\left(1-\nu_{R}\right)}\right]^{\frac{\nu+\nu_{LT}-1+\left(\nu_{R}-1\right)\left(\theta-1\right)}{\nu_{R}-\nu_{LT}}}\right.\\&\left.\times\int_{0}^{1}x^{\frac{\nu+\nu_{R}\theta-\theta}{\nu_{LT}-\nu_{R}}}\left(1-\frac{\left(1-\nu_{R}\right)\nu_{LT}}{\left(1-\nu_{LT}\right)\nu_{R}}x\right)^{\theta-1}dx\right\}.\end{aligned}$$

At this point we are going to use the hyper-geometric function ${}_{2}F_{1}\left(a,b;c;z\right)$ defined by

$$B\left(b,c-b\right){}_{2}F_{1}\left(a,b;c;z\right)=\int_{0}^{1}x^{b-1}\left(1-x\right)^{c-b-1}\left(1-zx\right)^{-a}dx,$$

where $B\left(b,c-b\right)$ is the beta function. The integral in the expression for $B\left(b,c-b\right){}_{2}F_{1}\left(a,b;c;z\right)$ is defined only if $|z|<1$ and $c>b>0$. We have

$$\int_{0}^{1}x^{\frac{\nu+\nu_{R}\theta-\theta}{\nu_{LT}-\nu_{R}}-1}\left(1-\frac{\left(1-\nu_{R}\right)\nu_{LT}}{\left(1-\nu_{LT}\right)\nu_{R}}x\right)^{\theta-1}dx=B\left(\gamma_{1}\left(\nu\right),1\right){}_{2}F_{1}\left(1-\theta,\gamma_{1}\left(\nu\right);\gamma_{1}\left(\nu\right)+1;\gamma_{2}\right),$$

where

$$\gamma_{1}\left(\nu\right)\equiv\frac{\nu+\nu_{R}\theta-\theta}{\nu_{LT}-\nu_{R}}\quad\text{and}\quad\gamma_{2}\equiv\frac{\left(1-\nu_{R}\right)\nu_{LT}}{\left(1-\nu_{LT}\right)\nu_{R}},$$

and where we need to have $\gamma_{1}\left(\nu\right)>0$ and $\gamma_{2}<1$. The last inequality holds under our assumptions that $0<\nu_{LT}<\nu_{R}<1$.

We have

$$B\left(\gamma_{1}\left(\nu\right),1\right)=\int_{0}^{1}t^{\gamma_{1}\left(\nu\right)-1}dt=\frac{1}{\gamma_{1}\left(\nu\right)}t^{\gamma_{1}\left(\nu\right)}\Big|_{0}^{1}=\frac{1}{\gamma_{1}\left(\nu\right)}.$$

Then

$$\int_0^1 x^{\frac{\nu+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}-1}\left(1-\frac{(1-\nu_R)\,\nu_{LT}}{(1-\nu_{LT})\,\nu_R}x\right)^{\theta-1}dx=\frac{1}{\gamma_1(\nu)}\times {}_2F_1\left(1-\theta,\gamma_1(\nu);\gamma_1(\nu)+1;\gamma_2\right).$$

Next,

$$\begin{aligned}\int_0^1 x^{\frac{\nu+\nu_R\theta-\theta}{\nu_{LT}-\nu_R}}\left(1-\frac{(1-\nu_R)\,\nu_{LT}}{(1-\nu_{LT})\,\nu_R}x\right)^{\theta-1}dx&=B\left(\gamma_1(\nu)+1,1\right){}_2F_1\left(1-\theta,\gamma_1(\nu)+1;\gamma_1(\nu)+2;\gamma_2\right)\\&=\frac{1}{\gamma_1(\nu)+1}\times {}_2F_1\left(1-\theta,\gamma_1(\nu)+1;\gamma_1(\nu)+2;\gamma_2\right).\end{aligned}$$

Thus, we get

$$\int_0^{a_{ij}^*} q^\nu dG_i(a)=\kappa_{G,i}^{\theta}\left(w_i\tau_{ij}\right)^{-\theta}A_j^{-\frac{\nu+\nu_{LT}\theta-\theta}{\nu_R-\nu_{LT}}}\left(w_iB_{ij}\right)^{\frac{\nu+\nu_R\theta-\theta}{\nu_R-\nu_{LT}}}H\left(\nu,\nu_R,\nu_{LT},\theta\right),$$

where

$$\begin{aligned}H\left(\nu,\nu_R,\nu_{LT},\theta\right)\equiv{}&\frac{1-\nu_R}{\nu_R-\nu_{LT}}\left[\nu_R\right]^{\theta}\left[\frac{1-\nu_{LT}}{1-\nu_R}\right]^{\frac{\nu+\nu_R\theta-\theta}{\nu_R-\nu_{LT}}}\frac{1}{\gamma_1(\nu)}\times {}_2F_1\left(1-\theta,\gamma_1(\nu);\gamma_1(\nu)+1;\gamma_2\right)\\&-\left\{\frac{(1-\nu_{LT})\,\nu_{LT}}{\nu_R-\nu_{LT}}\left[\nu_R\right]^{\theta-1}\left[\frac{1-\nu_{LT}}{1-\nu_R}\right]^{\frac{\nu+\nu_{LT}-1+(\nu_R-1)(\theta-1)}{\nu_R-\nu_{LT}}}\right.\\&\left.\qquad\times\frac{1}{\gamma_1(\nu)+1}\times {}_2F_1\left(1-\theta,\gamma_1(\nu)+1;\gamma_1(\nu)+2;\gamma_2\right)\right\}.\end{aligned}$$