

Nonparametric Identification and Estimation of Productivity Distributions and Trade Costs*

AYŞE ÖZGÜR PEHLİVAN

Department of Economics, Bilkent University

aysepehlivan@bilkent.edu.tr

QUANG VUONG

Department of Economics, New York University

qv1@nyu.edu

April 2015

**Incomplete.* We would like to thank David G. Abler, Daron Acemoglu, Ufuk Akcigit, Costas Arkolakis, Gaurab Aryal, Ivan Canay, Thomas Chaney, Kerem Coşar, Jonathan Eaton, Kevin Hasker, Hugo Hopenhayn, Peter Neary, Whitney Newey, Isabelle Perrigne, Robert Porter, Andrés Rodríguez-Clare, Elie Tamer, Alexander Torgovitsky, Neil Wallace, Taner Yiğit and seminar participants at Northwestern University and Bilkent University for their helpful comments.

Abstract

This paper studies the nonparametric identification and estimation of productivity distributions and trade costs in an Eaton and Kortum (2002) type Ricardian trade model. Our identification and estimation strategy gains insights from the empirical auction literature, however, our methodology is novel since we face additional problems resulting from the nature of the trade data. Our methodology does not require data on prices which are usually quite hard to obtain and manages to identify the underlying structure by using disaggregated simple bilateral trade data consisting only of trade values (expenditures) and traded quantities. We recover destination-source-sector specific productivity distributions and trade costs nonparametrically. The fact that these productivity distributions and trade costs are both country and sector specific provides important insights about not only cross country differences in productivity distributions and trade costs but also differences across sectors. Moreover, in Eaton and Kortum (2002) and variants, the productivity distributions of countries are assumed to come from a certain parametric family, Fréchet, and it has now become a common tradition in models of international trade to use either Fréchet or Pareto distributions to represent the distribution of productivities. These parametrizations provide great analytical convenience; however, recent studies show that gains from trade estimates are very sensitive to these parametrizations. In order to quantify the welfare gains from trade and answer related policy questions, checking the validity of these parametrizations and analyzing how productivity distributions behave is very important. In view of this, using a flexible structure for productivity distributions and trade costs we can perform similar counterfactuals as in Eaton and Kortum (2002) and variants, and compare the welfare results to those of the current literature.

1 Introduction

This paper studies the nonparametric identification and estimation of the productivity distributions and trade costs in an Eaton and Kortum (2002) type Ricardian trade model. Following the seminal papers of Eaton and Kortum (2002) and Melitz (2003), international trade literature has been dominated by Ricardian and heterogeneous firm models using probabilistic representation of technologies. In this setup, countries or firms draw their productivities from some productivity distribution. In Eaton and Kortum (2002) the productivity distributions of countries are assumed to come from a certain parametric family, Fréchet, and it has now become a common tradition in models of international trade to use either Fréchet or Pareto distributions to represent the distribution of productivities. These parametrizations provide great analytical convenience, however, recent studies show that gains from trade estimates are very sensitive to these parametrizations. Arkolakis, Costinot and Rodriguez-Clare (2012) argue that one parameter of the productivity distribution is actually one of the only two statistics governing gains from trade in most of those models. Simonovska and Waugh (2013) show that incorrect estimates of that parameter causes the gains from trade to be estimated half of what it should be. Therefore, if we would like to quantify the welfare gains from trade and answer related policy questions, checking the validity of these parametrizations is very important. Once the productivity distributions are recovered and their nonparametric estimates are obtained; we can not only check the validity of the parametrizations but also perform similar counterfactuals as in Eaton and Kortum (2002) and variants, which enables us to compare the estimated welfare results to those of the current literature.

In view of all these, in an Eaton and Kortum (2002) type Ricardian model of trade, we keep a flexible structure for productivity distributions of countries and recover them nonparametrically. We are also able to recover the destination-source-good specific trade costs nonparametrically, which is an important contribution to the literature as these costs can never be fully observed in the data but are very crucial in welfare analysis.

Our identification and estimation strategy gains insights from the empirical auction literature since the setup is very similar. However, our identification and estimation methodology will be novel in the sense that we face additional problems resulting from the nature of the trade data. For a survey of empirical models of auctions see Athey and Haile (2006) and for a survey of nonparametric methods used in auctions see Athey and Haile (2007). In Eaton and Kortum (2002) there is perfect competition among the countries and whichever country provides the good at the lowest price will win the competition and be the supplier of a given good. An immediate analogy to an asymmetric first price auction can be made in the sense that the bidder who bids the highest will be the winner of the auction. Just as bidders draw their valuations from their individual specific distribution of valuations, in our case countries draw their productivities from their country specific productivity distributions. These productivities are independent across countries similar to a first price auction with independent private valuations (IPV). In a first price auction when all the bids are observed in the data, Guerre, Perrigne and Vuong (2000) show the nonparametric identification of underlying distribution of valuations for symmetric IPV case. The extension for the asymmetric case is straightforward, see Campo, Perrigne and Vuong (2003). From the observables (bids) they first recover what they call the pseudo valuations and then the underlying distribution(s) of valuations which is (are) unobservable. In other words, they can recover the unobserved distribution(s) of valuations from the observable bid data. In a similar fashion, if we have data on prices offered by each country for any particular good, from these observables (prices) we can recover the productivity distribution(s) which is (are) unobservable. Actually, in that case our job would have been much easier than the case of a first price auction. This would be due to the fact that since there is perfect competition, there will be marginal cost pricing and since there is no private information, the equilibrium strategy is much simpler (which is basically price equals to marginal cost), hence the problem we face compared to Guerre, Perrigne and Vuong (2000). However, due to the nature of the disaggregated bilateral trade data we face two additional problems:

Our first problem is the fact that we do not observe all the prices that are offered by the countries but only the winning price. In an asymmetric first price auction with IPV, as long as the winning bid and the identity of the winner are known, nonparametric identification of the underlying distributions of valuations follows from the identification of competing risk models¹. In our case, the identity of the winning country is not observed either, yet we show that the underlying distributions can still be identified from the observables. Our second problem is that data on actual prices usually do not exist and the disaggregated simple bilateral trade data available usually consist only of trade values (expenditures) and traded quantities. Our identification and estimation methodology is novel in the sense that it does not require data on actual prices and manages to identify the underlying structure by using only the disaggregated simple bilateral trade data.

The perfect competition homogenous good setup in Eaton and Kortum (2002) implies that for a given product category the lowest cost supplier should supply the whole market; however, this is not what we see in bilateral trade data. Even in the least aggregated level, we see multiple sellers. In Eaton and Kortum (2002) they use aggregate trade flows in their analysis, so they do not need to address this issue. Our efforts in this paper can also be seen as a complementary to their seminal work in the following sense: We take advantage of the disaggregated information available in bilateral trade data and also try to reconcile their perfect competition homogenous good setup with the existence of multiple sellers.

As we mentioned above, it has now been very popular in models of international trade to use either Fréchet or Pareto distribution to represent the distribution of productivities. One might wonder whether any justification for such distributional assumptions other than analytical convenience has been provided in the literature. In Eaton and Kortum (2002)

¹A typical example of a competing risks model is the following: Suppose there are two causes of death. Let T_1 be the lifetime of the individual exposed to cause 1 alone and T_2 be the lifetime of the individual exposed to cause 2 alone. It is usually the case that we cannot observe T_1 and T_2 but we can observe when the individual dies, i.e., $T = \min\{T_1, T_2\}$. In addition, it is usually the case that the reason of death whether cause 1 or cause 2 is known. Call $\delta = \arg\min\{T_1, T_2\}$. It is known from the competing risks literature that the distributions of T_1 and T_2 can be identified if the joint distribution of (T, δ) is known when the distributions of T_1 and T_2 are independent. See Kalbfleisch and Prentice (2002), Rao (1992), Heckman and Honoré (1989).

and Bernard, Eaton, Jensen and Kortum (2003), the productivity distributions are assumed to be Fréchet. In others such as Helpman, Melitz and Yeaple (2004), Melitz and Ottaviano (2008), Helpman, Melitz and Rubinstein (2008), Eaton, Kortum and Kramarz (2010) and many others the productivity distribution is assumed to be Pareto. Fréchet distribution is also called the Type II extreme value distribution since it is related to the asymptotic distribution of the largest value. Eaton and Kortum (2002) refers to Kortum (1997) and Eaton and Kortum (1999) where they show how certain processes of innovation and diffusion give rise to this type of distribution. They argue that while producing any good, the actual technique that would ever be used in a country represents the best discovered one to date so it is reasonable to represent technology with an extreme value distribution. The models in Eaton and Kortum (2002) and Bernard, Eaton, Jensen and Kortum (2003), are mainly concerned with the best producers of a country for each good, which is considered as the explanation of the choice of an extreme value type for a country's productivity distribution. On the other hand, in Helpman, Melitz and Yeaple (2004), Melitz and Ottaviano (2008), Helpman, Melitz and Rubinstein (2008), Eaton, Kortum and Kramarz (2010) the main concern is not necessarily the best producers, which justifies the choice of an "non-extreme" value type of distribution such as Pareto. Helpman, Melitz and Yeaple (2004) refer to Axtell (2001) for justification, where he shows that US firm size distribution closely follows a Pareto distribution. This, however, does not say much about the underlying productivity distribution.² Both distributional assumptions provide great analytical convenience in these models.

The rest of this paper is organized as follows: Section 2 introduces the model. Section 3 establishes the nonparametric identification of the productivity distributions and trade costs. Section 4 introduces nonparametric estimation methodology and estimates. Section 5 concludes.

²It is worth mentioning the link between Pareto and Fréchet: The limiting distribution of the maximum of independent random variables having Pareto distribution is Frechet which suggests that when all firms draw from Pareto the distribution of the best can be represented as Fréchet.

2 Model

We follow the multi-country Ricardian model of trade introduced by Eaton and Kortum (2002). We consider a world with N countries indexed by $n, i = 1, 2, \dots, N$ and a finite number, J , of goods indexed by $j = 1, 2, \dots, J$.³ Throughout all the paper, index i will refer to the source country whereas index n will refer to the destination country.

Following the probabilistic representation of technologies in Eaton and Kortum (2002), country i 's productivity in producing good j for destination/market n , z_{ni}^j , is the realization of a random variable Z_{ni}^j which is drawn from distribution, $F_{Z_{ni}^j}(\cdot)$. This specification of productivity distributions, which is destination-source-good specific, is a more general specification than the one in Eaton and Kortum (2002). In their case, it is only source country specific meaning that for any destination n and for any good j countries draw their productivities from these source country specific distributions, which is obviously a special case of our specification. Moreover, in Eaton and Kortum (2002) and variants, these productivity distributions are assumed to come from a certain parametric family, Fréchet. Here, we do not restrict the productivity distributions to come from a certain parametric family. In addition to the parametrization, in those models, there is also an additional restriction: The comparative advantage parameter θ , which governs the shape of the Fréchet distribution, is restricted to be the same across countries. Here, we relax that assumption as well.

Labor is the only factor of production.⁴ As in Eaton and Kortum (2002), there is constant returns to scale. Also, as in Eaton and Kortum (2002), there is perfect competition. The total demand for good j in country n is denoted by Q_n^j .⁵ Before describing equilibrium prices and trade flows we would like to make the following remark: Due to perfect competition assumption, there will be marginal cost pricing and the source with the lowest marginal

³In Eaton and Kortum (2002) there is a continuum of goods, however we assume a finite number of goods due to some technical issues.

⁴In Eaton and Kortum (2002) there are also intermediate goods which can be traded and used as inputs, however in our model we will only have tradable final goods.

⁵Note that in Eaton and Kortum (2002) there is the CES demand structure. Here, we do not model the demand explicitly to simplify our nonparametric identification but we will account for the dependency between the total demand and the productivity draws of all countries for any given market n and good j .

cost should supply the entire market for good j since marginal costs are constant. Suppose country i' is the one with the lowest marginal cost, then it implies $Q_{ni'}^j = Q_n^j$ and $Q_{ni}^j = 0$ for all $i \neq i'$ where Q_{ni}^j is the quantity of good j supplied by country i to country n . Similarly, in terms of trade flows we have $X_{ni'}^j = X_n^j$ and $X_{ni}^j = 0$ for all $i \neq i'$ where X_n^j is the total expenditure of country n on good j , of which X_{ni}^j is spent on good j bought from country i . When we look at the bilateral trade data on a disaggregated level, however, we do not see such pattern; instead we see positive amounts for more than one exporter. In Eaton and Kortum (2002) they do not have to deal with this issue since they perform their analysis using aggregated trade flows. In our case, however, we use the disaggregated data. In order to reconcile the perfect competition homogenous good setup with the fact that in the data we observe multiple sellers for a given product, we provide the following interpretation:

Suppose for every good j , these goods are sold in batches and for every good j there exists a $\lambda^j > 0$ such that the size of the batch is $1/\lambda^j$ and $\lambda^j Q_n^j$ denotes how many batches are sold. Suppose there is competition for supplying each batch of good j and hence, countries compete $\lambda^j Q_n^j$ times in order to supply $1/\lambda^j$ units of good j in country n . One may also think that there are $\lambda^j Q_n^j$ separate auctions and in each auction there is competition to supply each batch of good j . For each time or for each auction t , $t = 1, 2, \dots, \lambda^j Q_n^j$, suppose country i 's productivity draw, $Z_{n,i,t}^j$, comes from $F_{Z_{ni}^j|Q_n^j}(\cdot|\cdot)$. One may think of this case as each time or in each auction, a different firm from country i is entering the competition and hence it will be a different draw. These draws are independent across t . Thus, we make the following assumption:

Assumption A1 :

(i) For any given n and j , $\{Z_{n,i,t}^j; i = 1, \dots, N; t = 1, \dots, \lambda^j Q_n^j\}$ are mutually independent across i and t given Q_n^j .

(ii) For any given n, j and i , $\{Z_{n,i,t}^j; t = 1, \dots, \lambda^j Q_n^j\}$ are identically distributed from $F_{Z_{ni}^j|Q_n^j}(\cdot|\cdot)$ given Q_n^j .

Assumption A1(i) states that for a given destination n and a given good j , productivities are mutually independent across countries and across those auctions given the total demand Q_n^j . The last part, Assumption A1(ii), guarantees that for a given destination n , a given good j , and a given country i , for each t , $Z_{n,i,t}^j$ draws come from the same distribution given the total demand Q_n^j .

Constant returns to scale implies that in country i cost of producing good j for destination n in the t th auction is $w_i/z_{n,i,t}^j$ where w_i is the wage rate in the economy.⁶ Trade costs are assumed to be of iceberg form, i.e., delivering a batch of good j to destination n requires $d_{ni}^j \geq 1$ units of good j to be produced in i . Moreover, we allow for trade costs to depend on the total demand for that good in that destination and have $d_{ni}^j = d_{ni}^j(q)$ for a given $Q_n^j = q$.⁷ Therefore, in the t th auction for country i supplying a batch of good j to country n costs $w_i d_{ni}^j / z_{n,i,t}^j$.

Due to perfect competition, country i offers marginal cost pricing, hence the price country i offers in the t th auction is:

$$P_{n,i,t}^j = \frac{w_i d_{ni}^j}{Z_{n,i,t}^j} \quad (1)$$

Denote the distribution of $P_{n,i,t}^j$ given Q_n^j by $G_{ni}^j(\cdot|\cdot)$ which is the distribution of the price that is offered by country i to country n in the t th auction of good j . Note that $G_{ni}^j(p|q) \equiv$

⁶Note that cost of producing one batch of good j in country i for destination n is $w_i/z_{n,i,t}^j$ and for one unit of good j it is $w_i \lambda^j / z_{n,i,t}^j$.

⁷This specification of iceberg trade costs is also more general than the specification in Eaton and Kortum (2002) model. In Eaton and Kortum (2002), they have $d_{ni}^j(q) = d_{ni}$ for all j and for any given $Q_n^j = q$, i.e., iceberg trade costs only depend on the destination and the source. Given a particular destination and source it is the same for every good. In our case, it will not only depend on the destination and the source but also be good specific and depend on the level of total demand in that destination.

$\Pr [P_{n,i,t}^j \leq p | Q_n^j = q] = \Pr \left[\frac{w_i d_{ni}^j}{Z_{n,i,t}^j} \leq p | Q_n^j = q \right]$. Hence,

$$G_{ni}^j(p|q) \equiv 1 - F_{Z_{ni}^j|Q_n^j} \left(\frac{w_i d_{ni}^j}{p} | q \right) \quad (2)$$

Consumers in country n buy good j from the lowest-cost supplier in the t th auction, so the price they end up paying in the t th auction is:

$$P_{n,t}^j = \min_i \{ P_{n,i,t}^j \} \quad (3)$$

where $i = 1, 2, \dots, N$.

For notational convenience define $A_n^j = \lambda^j Q_n^j$. Let Q_{ni}^j denote the quantity of good j supplied by country i to country n . Similarly, define $A_{ni}^j = \lambda^j Q_{ni}^j$. Note that depending on its productivity draw in each auction, each country can now be the lowest cost supplier for certain auctions, which means A_{ni}^j , hence, Q_{ni}^j can now be positive for multiple sellers: A_{ni}^j can also be interpreted as how many auctions out of a total of A_n^j auctions country i actually wins, i.e., for how many auctions country i turns out to be the lowest cost supplier for good j . Hence, we can write the following expression for A_{ni}^j :

$$A_{ni}^j = \lambda^j Q_{ni}^j = \sum_{t=1}^{\lambda^j Q_n^j} 1 (P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i) \quad (4)$$

where $1 (P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i)$ is basically the indicator that country i is the winner in the t th auction of good j where $s = 1, \dots, N$.

The total expenditure X_n^j of country n on good j can be written as:

$$X_n^j = \sum_{t=1}^{\lambda^j Q_n^j} P_{n,t}^j \quad (5)$$

Basically, the total expenditure of country n on good j is the sum of the winning prices,

i.e., sum of prices that country n pays in each auction t for good j . The expenditure X_{ni}^j of country n spent on good j coming from country i can be expressed as:

$$X_{ni}^j = \sum_{t=1}^{\lambda^j Q_n^j} P_{n,i,t}^j 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i) \quad (6)$$

Notice that country n 's spending X_{ni}^j on good j coming from country i is equal to the sum of the prices offered by country i for the particular auctions of good j that country i is the winner. Given Q_n^j , denote the distribution of the random variable $P_{n,i,t}^j 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i)$ by

$$H_{ni}^j(p|q) \equiv \Pr [P_{n,i,t}^j 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i) \leq p | Q_n^j = q] \quad (7)$$

Note that $H_{ni}^j(p|q)$ does not depend on t , hence there is no t subscript. This follows from Assumption A1(ii). We will use this distribution in the identification section.

For identification purposes, which is the topic of the next section; we also make the following assumption, Assumption A2. We use Assumption A2 in the identification of $G_{ni}^j(\cdot|\cdot)$, the distribution of the offered price $P_{n,i,t}^j$ given Q_n^j , and in the identification of d_{ni}^j , iceberg trade costs.

Now, let $\underline{z}_{ni}^j(q)$ be the lower bound and $\bar{z}_{ni}^j(q)$ be the upper bound of the support of $F_{Z_{ni}^j|Q_n^j}(\cdot|q)$ for any given n, i, j , and for all q , i.e., $F_{Z_{ni}^j|Q_n^j}(\underline{z}_{ni}^j(q)|q) = 0$ and $F_{Z_{ni}^j|Q_n^j}(\bar{z}_{ni}^j(q)|q) = 1$.

Assumption A2 :

- (i) For any given n and j , we have $\underline{z}_{ni}^j(q) = 0$ for all q and for all i .
- (ii) For any given n and j , we have $\bar{z}_{ni}^j(q) = \bar{z}_{nn}^j(q) < \infty$ for all q and for all i .
- (iii) For any given j , we have $d_{nn}^j(q) = 1$ for all q and for all n .

Assumption A2(i) basically states that for any given destination n and any given good j , the lower bound of the productivity distribution is zero for all countries and for any level of total

demand q in that destination. Assumption A2(ii) says that for a given good j and for any destination n , the upper bound of the productivity distribution is the same irrespective of the source country and the level of total demand q in that destination. Assumption A2(iii) is related to the usual iceberg trade cost assumption, in which the iceberg trade costs are assumed to be 1 when the country buys from itself. This is true for any good j and for any level of total demand q in that destination. We use Assumption A2(i) to identify $G_{ni}^j(\cdot|\cdot)$, the distribution of the offered price $P_{n,i,t}^j$ given Q_n^j and use Assumption A2(ii)-A2(iii) to identify the iceberg trade costs, d_{ni}^j .

Lastly, before moving to the identification section, following Eaton and Kortum (2002) we introduce the labor market and since the empirical implementation of this model will be to the manufacturing goods, specify how manufactures fit into the entire economy. In our case since labor is the only factor of production, manufacturing labor income in country i is equal to the total manufacturing income. Total manufacturing income is:

$$w_i L_i = \sum_{n=1}^N \pi_{ni} X_n \quad (8)$$

where L_i is manufacturing workers, X_n is the total spending on manufactures by country n , and π_{ni} is the share of country n 's manufacturing spending on goods from country i , i.e., $\pi_{ni} = X_{ni}/X_n$ where X_{ni} is the manufacturing spending of country n on goods from country i . Denote total income in country n by Y_n and α is the share of total income spent on manufactures hence

$$X_n = \alpha Y_n \quad (9)$$

Y_n consists of income generated in manufacturing $Y_n^M = w_n L_n$ and income generated in nonmanufacturing Y_n^O . As in Eaton and Kortum (2002) we assume nonmanufacturing output can be traded costlessly and use it as numeraire.

We consider the mobile labor case in which labor is mobile between manufacturing and nonmanufacturing sectors. The wage w_n is pinned down by the productivity in the nonman-

ufacturing sector and total income Y_n is exogenous. Combining (8) and (9) gives

$$w_i L_i = \sum_{n=1}^N \pi_{ni} \alpha Y_n \quad (10)$$

and for all i (10) determines manufacturing employment L_i .

3 Nonparametric Identification

In this section we investigate whether we can recover uniquely the unobserved productivity distribution $F_{Z_{ni}^j}(\cdot)$ and iceberg trade costs d_{ni}^j from the observables for each n , i and j . The observables in our data are the bilateral trade quantities and expenditures between countries $\{Q_{ni}^j, X_{ni}^j\}_{j=1}^J$ for each n and i . We also observe w_i for each i . In this section, for any good j , we take the joint distribution $F_{Q_{ni}^j, (Q_{ns}^j)_{s \neq i}, X_{ni}^j, (X_{ns}^j)_{s \neq i}, Q_n^j}(\cdot, \dots, \cdot)$ of the observables as known. We will discuss how to estimate such a distribution from the observables in our data in the estimation section.

Identification is achieved in four steps: First, we identify the distribution $H_{ni}^j(\cdot|\cdot)$ of the random variable $P_{n,i,t}^j 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i)$ given Q_n^j . This random variable can be interpreted as the expenditure of country n on good j coming from country i in the t th auction, which will be zero when country i is not the winner and will be equal to price $P_{n,i,t}^j$ offered by country i when country i is the winner. We do not observe either $P_{n,i,t}^j 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i)$'s or $P_{n,i,t}^j$'s. We do, however, observe X_{ni}^j 's which are basically the sum of independent $P_{n,i,t}^j 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i)$'s over t . We will use this relation to identify $H_{ni}^j(\cdot|\cdot)$. Once we have the distribution $H_{ni}^j(\cdot|\cdot)$ of $P_{n,i,t}^j 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i)$ given Q_n^j , in the second step we identify the distribution $G_{ni}^j(\cdot|\cdot)$ of the random variable $P_{n,i,t}^j$ given Q_n^j . Identification of the distribution $G_{ni}^j(\cdot|\cdot)$ of $P_{n,i,t}^j$ given Q_n^j follows from the identification of competing risks models. In the third step, for any given $Q_n^j = q$, we identify d_{ni}^j as a function of q for each n , i and j . Finally, we establish the identification of the productivity distribution $F_{Z_{ni}^j|Q_n^j}(\cdot|\cdot)$ for each n , i , and j . After identifying the distribution $F_{Z_{ni}^j|Q_n^j}(\cdot|\cdot)$ for each n , i ,

and j , identification of $F_{Z_{ni}^j}(\cdot)$ is trivial since $F_{Q_n^j}(\cdot)$ is known from the observables.

3.1 Identification of $H_{ni}^j(\cdot|\cdot)$

For notational convenience let $R_{n,i,t}^j = P_{n,i,t}^j 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i)$. Hence, $X_{ni}^j = \sum_{t=1}^{\lambda^j Q_n^j} R_{n,i,t}^j$ and $R_{n,i,t}^j \sim H_{ni}^j(\cdot|\cdot)$ for any t given Q_n^j . Note that across t , $R_{n,i,t}^j$'s are independent given Q_n^j . This follows from Assumption A1(i), i.e., the independence of $Z_{n,i,t}^j$'s across t given Q_n^j .

Using definition of the characteristic functions, note that we can write

$$\varphi_{X_{ni}^j|Q_n^j}(u|q) = E[e^{iuX_{ni}^j}|Q_n^j = q] = E\left[e^{iu\sum_{t=1}^{\lambda^j q} R_{n,i,t}^j}|Q_n^j = q\right] \quad (11)$$

$$= E\left[\prod_{t=1}^{\lambda^j q} e^{iuR_{n,i,t}^j}|Q_n^j = q\right] = \left[\varphi_{R_{ni}^j|Q_n^j}(u|q)\right]^{\lambda^j q} \quad (12)$$

where $\varphi_{R_{ni}^j|Q_n^j}(\cdot|q)$ is the characteristic function of $R_{n,i,t}^j$ given $Q_n^j = q$. The fourth equality above follows from the fact that $R_{n,i,t}^j$'s are independent across t given Q_n^j . Using (12), for any n, i , and j we can write

$$\varphi_{R_{ni}^j|Q_n^j}(u|q) = \left[\varphi_{X_{ni}^j|Q_n^j}(u|q)\right]^{1/\lambda^j q} \quad (13)$$

for any q . We can write $\varphi_{X_{ni}^j|Q_n^j}(u|q)$ as $\varphi_{X_{ni}^j|Q_n^j}(u|q) = r_{X_{ni}^j|Q_n^j}(u|q)e^{i\theta_{X_{ni}^j|Q_n^j}(u|q)}$ where $r_{X_{ni}^j|Q_n^j}(u|q)$ is the modulus and $\theta_{X_{ni}^j|Q_n^j}(u|q)$ is the argument of $\varphi_{X_{ni}^j|Q_n^j}(u|q)$. Since $\varphi_{X_{ni}^j|Q_n^j}(0|q) = 1$ for any q , we have $r_{X_{ni}^j|Q_n^j}(0|q) = 1$ and $\theta_{X_{ni}^j|Q_n^j}(0|q) = 0$ for any q . There are $\lambda^j q$ solutions for $\left[\varphi_{X_{ni}^j|Q_n^j}(u|q)\right]^{1/\lambda^j q}$ given by

$$\left[\varphi_{X_{ni}^j|Q_n^j}(u|q)\right]^{1/\lambda^j q} = \left[r_{X_{ni}^j|Q_n^j}(u|q)\right]^{1/\lambda^j q} e^{i\left(\frac{\theta_{X_{ni}^j|Q_n^j}(u|q) + 2k\pi}{\lambda^j q}\right)} \quad (14)$$

where for any q , k is any integer such that $0 \leq k \leq \lambda^j q - 1$. Note that for any q , only the one with $k = 0$, i.e., $\left[r_{X_{ni}^j|Q_n^j}(u|q)\right]^{1/\lambda^j q} e^{i\left(\frac{\theta_{X_{ni}^j|Q_n^j}(u|q)}{\lambda^j q}\right)}$ satisfies the condition $\varphi_{R_{ni}^j|Q_n^j}(0|q) = 1$ for

any q since $\theta_{X_{ni}^j|Q_n^j}(0|q) = 0$ for any q . Therefore, $\varphi_{R_{ni}^j|Q_n^j}(u|q) = \left[r_{X_{ni}^j|Q_n^j}(u|q) \right]^{1/\lambda^j q} e^{i \left(\frac{\theta_{X_{ni}^j|Q_n^j}(u|q)}{\lambda^j q} \right)}$. Hence, for any given good j , $\varphi_{R_{ni}^j|Q_n^j}(\cdot|q)$ is identified for any n and i and for any q . Characteristic function uniquely determines the distribution, thus we establish the following lemma.

Lemma 1: *For any j , $\varphi_{R_{ni}^j|Q_n^j}(\cdot|q)$ is identified for all n and i and for any q . Hence, for any j , $H_{ni}^j(\cdot|q)$ is identified for any n and i and for any q*

Note that identification of $H_{ni}^j(\cdot|Q_n^j)$ is achieved using only trade quantities and expenditures.

3.2 Identification of $G_{ni}^j(\cdot|\cdot)$

Using the definition of $H_{ni}^j(\cdot|\cdot)$ and further decomposing the probability on the RHS below, we have

$$\begin{aligned} H_{ni}^j(p|q) &\equiv \Pr [P_{n,i,t}^j 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i) \leq p | Q_n^j = q] \\ &= \Pr [P_{n,i,t}^j 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i) \leq p \text{ and } P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q] \\ &\quad + \Pr [P_{n,i,t}^j 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i) \leq p \text{ and } (P_{n,i,t}^j > P_{n,s,t}^j, \text{ for some } s \neq i) | Q_n^j = q] \\ &= \Pr [P_{n,i,t}^j \leq p \text{ and } P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q] \tag{15} \\ &\quad + \Pr [P_{n,i,t}^j > P_{n,s,t}^j, \text{ for some } s \neq i | Q_n^j = q] \tag{16} \end{aligned}$$

where the last equality follows from the fact that $1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i) = 0$ when $P_{n,i,t}^j > P_{n,s,t}^j$, for some $s \neq i$. In (15) note that the second term on the RHS is equal to $1 - \Pr [P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q]$. The probability $\Pr [P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q]$ that country i is the lowest cost supplier for any unit t of a given good j for a given $Q_n^j = q$ can be written as

$$\Pr [P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q] = \frac{E[A_{ni}^j | Q_n^j = q]}{\lambda^j q} \tag{17}$$

and notice that everything on the RHS of (17) can be observed from the data.⁸ Recall that by Lemma 1, for any j , $H_{ni}^j(\cdot|q)$ can be obtained from the observables for any n and i and for any q . Now that we can obtain $\Pr [P_{n,i,t}^j > P_{n,s,t}^j, \text{ for some } s \neq i | Q_n^j = q]$, the probability $\Pr [P_{n,i,t}^j \leq p \text{ and } P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q]$ in (15), which is the probability that the price $P_{n,i,t}^j$ country i offers to country n for t th unit of good j is the lowest and $P_{n,i,t}^j$ being less than some given p for a given Q_n^j , can also be obtained. Denote $\Pr [P_{n,i,t}^j \leq p \text{ and } P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q]$ by $\tilde{H}_{ni}^j(p|q)$.

From the literature on competing risks models, it is known that the class of functions $\tilde{H}_{ni}^j(\cdot|q)$, $i = 1, 2, \dots, N$ uniquely determine the distribution functions $G_{ni}^j(\cdot|q)$, $i = 1, 2, \dots, N$, for $P_{n,i,t}^j$ for any given $Q_n^j = q$. The next lemma establishes the identification of $G_{ni}^j(\cdot|q)$. In the Appendix we provide the proof which follows Rao (1992).

Lemma 2: *The distribution $G_{ni}^j(\cdot|q)$ is identified by*

$$G_{ni}^j(p|q) = 1 - \exp \left\{ - \int_0^p \left[1 - \sum_{s=1}^N \tilde{H}_{ns}^j(\nu|q) \right]^{-1} d\tilde{H}_{ni}^j(\nu|q) \right\} \quad (18)$$

for any given j and for all n and i and for all p and q where $s = 1, \dots, N$.

3.3 Identification of d_{ni}^j

Using (2) we can write

$$F_{Z_{ni}^j|Q_n^j}(z|q) = 1 - G_{ni}^j\left(\frac{w_i d_{ni}^j}{z} | q\right) \quad (19)$$

⁸This follows from the fact that $A_{ni}^j = \sum_{t=1}^{\lambda^j Q_n^j} 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i)$. We can write $E[A_{ni}^j | Q_n^j = q] = E\left[\sum_{t=1}^{\lambda^j q} 1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i) | Q_n^j = q\right] = \sum_{t=1}^{\lambda^j q} E\left[1(P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i) | Q_n^j = q\right] = \sum_{t=1}^{\lambda^j q} \Pr [P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q] = \lambda^j q \Pr [P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q]$. The last equality follows from the fact that for a given level of total demand Q_n^j , for any unit t of good j , the probability that country i is the winner is the same since $Z_{n,i,t}^j$'s are identically distributed across t given Q_n^j by Assumption A1(iii).

for all n and i and for any given j and for all z and q . For a given n , take $i = n$, then we have

$$F_{Z_{nn}|Q_n^j}(z|q) = 1 - G_{nn}^j\left(\frac{w_n}{z}|q\right) \quad (20)$$

which follows from the fact that $d_{nn}^j = 1$ for any j and for any level of q . Let $\underline{p}_{ni}^j(q)$ be the lower bound of the support of $P_{n,i,t}^j$ given Q_n^j for a given q level of Q_n^j . By Lemma 2, $G_{ni}^j(\cdot|q)$ is identified, hence the lower bound $\underline{p}_{ni}^j(q)$ is known for all n and i and for any given j and for all q . Therefore, $\underline{p}_{nn}^j(q)$ is known. Hence, $\bar{z}_{nn}^j(q)$ is identified for all n and q for a given j by the following equation:

$$\bar{z}_{nn}^j(q) = \frac{w_n}{\underline{p}_{nn}^j(q)} \quad (21)$$

Since by Assumption A2, $\bar{z}_{ni}^j(q) = \bar{z}_{nn}^j(q)$, for a given j and n , $\bar{z}_{ni}^j(q)$ is also identified for all i and q using (21). Note that since now both $\underline{p}_{ni}^j(q)$ and $\bar{z}_{ni}^j(q)$ are known d_{ni}^j as a function of q , $d_{ni}^j(q)$, for any given j , for all n and i and for all q is identified by the following equation:

$$\begin{aligned} d_{ni}^j &= \frac{\underline{p}_{ni}^j(q)\bar{z}_{ni}^j(q)}{w_i} \\ &= \frac{w_n \underline{p}_{ni}^j(q)}{w_i \underline{p}_{nn}^j(q)} \end{aligned}$$

3.4 Identification of $F_{Z_{ni}|Q_n^j}(\cdot|\cdot)$

Once we identify $d_{ni}^j(q)$, for any given j , for all n and i and for all q , identification of $F_{Z_{ni}|Q_n^j}(\cdot|\cdot)$ is trivial. Using (19) we have

$$F_{Z_{ni}|Q_n^j}(z|q) = 1 - G_{ni}^j\left(\frac{w_i d_{ni}^j(q)}{z}|q\right) \quad (22)$$

Since $G_{ni}^j(\cdot|q)$ and $d_{ni}^j(q)$ are both identified, now, $F_{Z_{ni}|Q_n^j}(\cdot|q)$ is identified for all n and i and for any given j and for all q . Hence, $F_{Z_{ni}}^j(\cdot)$ is identified for all n and i and for any given j since $F_{Q_n^j}(\cdot)$ is known from the observables. The next proposition states this result and concludes this section.

Proposition 1: *For all n and i and for any given j , $F_{Z_{ni}}^j(\cdot)$ is identified.*

4 Nonparametric Estimation

To be completed...

5 Conclusion

This paper establishes the nonparametric identification and estimation of productivity distributions and trade costs in an Eaton and Kortum (2002) type Ricardian trade model. This enables us to check the validity of the common distributional assumptions on productivity distributions in the literature and to assess how the results of the current literature will be affected if these distributional assumptions are relaxed. We can replicate the counterfactuals in the current literature such as what the welfare gains are from moving to a zero-gravity world with no trade costs or what the welfare gains are due to raising trade costs to their autarky levels and see how much of the results change. Moreover, the productivity distributions we recover are sector specific which provides important insights about sectoral productivities. Similarly, our methodology enables us to obtain the trade costs nonparametrically. Trade costs are never fully observed in the data yet they are very important for welfare analysis. In addition to being destination-source country specific, trade costs we recover are also sector specific which enables us to understand the heterogeneity in trade costs across sectors. *[To be completed]*

APPENDIX

Proof of Lemma 2

Recall that $\tilde{H}_{ni}^j(p|q) = \Pr [P_{n,i,t}^j \leq p \text{ and } P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q]$ where $s = 1, \dots, N$. Since $P_{n,s,t}^j$'s are independent across s by Assumption A1(ii), $\tilde{H}_{ni}^j(p|q)$ can be written as

$$\tilde{H}_{ni}^j(p|q) = \int_0^p \prod_{\substack{s \neq i \\ s=1, \dots, N}} [1 - G_{ns}^j(\nu|q)] dG_{ni}^j(\nu|q) \quad (23)$$

$$\begin{aligned} &= \int_0^p \left(\frac{\prod_{s=1}^N [1 - G_{ns}^j(\nu|q)]}{1 - G_{ni}^j(\nu|q)} \right) dG_{ni}^j(\nu|q) \\ &= - \int_0^p \left[\prod_{s=1}^N [1 - G_{ns}^j(\nu|q)] \right] d \ln [1 - G_{ni}^j(\nu|q)] \end{aligned} \quad (24)$$

where the last equality follows from the fact that $[1/(1-G_{ni}^j(\nu|q))]dG_{ni}^j(\nu|q) = -d \ln [1 - G_{ni}^j(\nu|q)]$.

Now, let $Y = \min_{s=1, \dots, N} \{P_{n,s,t}^j\}$. Note that we can write $\Pr[Y \leq \nu | Q_n^j = q] = \Pr[P_{n,1,t}^j \leq \nu \text{ and } P_{n,1,t}^j \leq P_{n,s,t}^j, s \neq 1 | Q_n^j = q] + \dots + \Pr[P_{n,i,t}^j \leq \nu \text{ and } P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q] + \dots + \Pr[P_{n,N,t}^j \leq \nu \text{ and } P_{n,N,t}^j \leq P_{n,s,t}^j, s \neq N | Q_n^j = q]$. This follows from the fact that the probability that the minimum price is less than or equal to some ν can be rewritten as the sum of the individual probabilities of each country being the winner and winning with a price less than or equal to ν . Recall that these probabilities are basically $\tilde{H}_{ns}^j(\nu|q)$'s where $s = 1, \dots, N$. Hence, we can write

$$\Pr[Y \leq \nu | Q_n^j = q] = \sum_{s=1}^N \tilde{H}_{ns}^j(\nu|q) \quad (25)$$

$\Pr[Y \leq \nu | Q_n^j = q]$ can also be written in terms of $G_{ns}^j(\nu|q)$'s such that $\Pr[Y \leq \nu | Q_n^j = q] = 1 - \Pr[Y > \nu | Q_n^j = q] = 1 - \prod_{s=1}^N \Pr[P_{n,s,t}^j > \nu | Q_n^j = q] = 1 - \left[\prod_{s=1}^N [1 - G_{ns}^j(\nu|q)] \right]$. Combining

this with (25), we obtain

$$\left[\prod_{s=1}^N [1 - G_{ns}^j(\nu|q)] \right] = 1 - \sum_{s=1}^N \tilde{H}_{ns}^j(\nu|q) \quad (26)$$

Putting (26) into (24) we get

$$\tilde{H}_{ni}^j(p|q) = - \int_{-\infty}^p \left[1 - \sum_{s=1}^N \tilde{H}_{ns}^j(\nu|q) \right] d \ln [1 - G_{ni}^j(\nu|q)]$$

Now, note that $d\tilde{H}_{ni}^j(p|q) = -[1 - \sum_{s=1}^N \tilde{H}_{ns}^j(p|q)] d \ln [1 - G_{ni}^j(p|q)]$. Thus, $d \ln [1 - G_{ni}^j(p|q)] = -[1 - \sum_{s=1}^N \tilde{H}_{ns}^j(p|q)]^{-1} d\tilde{H}_{ni}^j(p|q)$. Solving for $G_{ni}^j(p|q)$ we obtain (18), i.e.,

$$G_{ni}^j(p|q) = 1 - \exp \left\{ - \int_0^p \left[1 - \sum_{s=1}^N \tilde{H}_{ns}^j(\nu|q) \right]^{-1} d\tilde{H}_{ni}^j(\nu|q) \right\}$$

in Lemma 2. $\square$

REFERENCES

- Arkolakis, Costas, Arnaud Costinot and Andrés Rodríguez-Clare, 2012. "New Trade Models, Same Old Gains?," *American Economic Review*, 102(1), 94-130.
- Athey, Susan and Phil Haile, 2007. "Nonparametric Approaches to Auctions," *Handbook of Econometrics*, 6A, Elsevier, 3847-3965.
- Athey, Susan and Phil Haile, 2006. "Empirical Models of Auctions," *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress, Volume II*, Cambridge University Press, 1-45.
- Atkeson, Andrew and Ariel Burstein, 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 98(5), 1998-2031.
- Axtell, Robert L. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science*, 293, 1818–1820.
- Baldwin, Richard and James Harrigan, 2010. "Zeros, Quality and Space: Trade Theory and Trade Evidence," *American Economic Journal: Microeconomics*, forthcoming.
- Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen and Samuel Kortum, 2003. "Plants and Productivity in International Trade," *American Economic Review*, 93(4), 1268-1290.
- Campo, Sandra, Isabelle Perrigne and Quang Vuong, 2003. "Asymmetry in First-Price Auctions with Affiliated Private Values", *Journal of Applied Econometrics*, 18, 197-207.
- Chaney, Thomas 2008. "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98(4), 1707-1721.
- Das, Sanghamitra, Mark J. Roberts and James R. Tybout, 2007. "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica*, 75(3), 837-873.

Del Gatto, Massimo, G. Mion and GIP. Ottaviano, 2007. "Trade Integration, Firm Selection and the Costs of Non-Europe," Working Paper CRENoS 200703, Centre for North South Economic Research, University of Cagliari and Sassari, Sardinia.

Eaton, Jonathan and Samuel Kortum, 1999. "International Technology Diffusion: Theory and Measurement", *International Economic Review*, 40, 537-570.

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

Eaton, Jonathan, Samuel Kortum and Francis Kramarz, 2010. "An Anatomy of International Trade: Evidence from French Firms," NBER Working Paper 14610.

Feenstra, Robert C., Robert E. Lipsey and Harry P. Bowen, 1997. "World Trade Flows, 1970-1992, with Production and Tariff Data," NBER Working Paper 5910.

Guerre, Emmanuel, Isabelle Perrigne and Quang Vuong, 2000. "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68(3), 525-574.

Heckman, James J., and Bo. E. Honoré, 1989. "The Identifiability of the Competing Risks Model," *Biometrika*, 76, 325-330.

Helpman, Elhanan, Marc J. Melitz and Stephen R. Yeaple, 2004. "Export Versus FDI with Heterogeneous Firms," *American Economic Review*, 94(1), 300-316.

Helpman, Elhanan, Marc J. Melitz and Yona Rubinstein, 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, 123(2), 441-487.

Kalbfleisch, Jack D. and Ross L. Prentice, 2002. *The Statistical Analysis of Failure Time Data*, Wiley, New York.

Kortum, Samuel, 1997. "Research, Patenting and Technological Change," *Econometrica*, 65(6), 1389-1419.

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

Melitz, Marc J. and Giancarlo I. P. Ottaviano, 2008. "Market Size, Trade, and Productivity," *Review of Economic Studies*, 75(1), 295-316.

Pagan, Adrian and Aman Ullah, 1999. *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

Pehlivan, Ayse O. and Quang Vuong, 2011. "International Trade and Supply Function Competition" Working Paper, Pennsylvania State University.

Pehlivan, Ayse O. and Quang Vuong, 2010. "Estimation of the Market Clearing Price Under Discriminatory Pricing," Working Paper, Pennsylvania State University.

Rao, Prakasa B.L.S., 1992. *Identifiability in Stochastic Models: Characterization of Probability Distributions*, Academic Press, Cambridge

Simonovska, Ina, 2010. "Income Differences and Prices of Tradables," Working Papers 1015, University of California, Davis, Department of Economics.

Simonovska, Ina and Michael E. Waugh, 2013. "The Elasticity of Trade: Estimates and Evidence," *Journal of International Economics*, forthcoming.

Waugh, Michael E., 2009. "International Trade and Income Differences," Staff Report 435, Federal Reserve Bank of Minneapolis.