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

First Version: March 2012
This Version: June 2018

*We would like to thank David G. Abler, Daron Acemoglu, Ufuk Akcigit, Costas Arkolakis, Gaurab Aryal, Ivan Canay, Thomas Chaney, Jonathan Eaton, Hugo Hopenhayn, Peter Neary, Whitney Newey, Isabelle Perrigne, Rob Porter, Andrés Rodríguez-Clare, Elie Tamer, Alexander Torgovitsky, Neil Wallace, Taner Yiğit and seminar participants at Northwestern University, Bilkent University, Econometric Society North American Winter Meetings in Philadelphia, Society for Economic Dynamics Meetings at Toulouse School of Economics, IIOC in Philadelphia, Bilkent University Summer Workshop in Econometric Theory and Applications 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 provides a novel way to reconcile Eaton and Kortum (2002) type models with disaggregated product level 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-product 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.
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 (2014) 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.

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 liter-
ature 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
models\textsuperscript{1}. 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 mar-
ket; 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 ‡ ows
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 ad-
vantage 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. Pehlivan and Vuong (2017) studies a supply function competition model in which
exporters compete in supply functions and recovers destination-source speci…c productivity
distributions and marginal costs. Those productivity distributions and marginal costs,
however, are not product speci…c. Here, introducing a completely different methodology,
we recover destination-source-product speci…c productivity distributions and trade costs. In

\textsuperscript{1}A 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 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 \( (T, \delta) \) is known when the distributions of \( T_1 \) and \( T_2 \) are independent. See Kalbfleisch and Prentice (2002), Rao (1992), Heckman and Honoré (1989).
Pehlivan and Vuong (2017), they recover marginal costs but they cannot identify trade costs separately whereas in this paper we recover destination-source-product specific trade costs separately. Moreover, in Pehlivan and Vuong (2017), demand is assumed to be exogenous whereas in this paper our methodology allows for demand not being exogenous.

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) 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 (2011) 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 (2011) 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. 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.

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 $n, i = 1, 2, ..., N$ and a finite number, $J$, of goods indexed by $j = 1, 2, ..., J$.

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, $F_{Z_{ni}^j} (·)$.

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 $\theta$, which governs the shape of the Fréchet distribution, is

---

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.

3In Eaton and Kortum (2002) there is a continuum of goods, however we assume a finite number of goods due to some technical issues.
restricted to be the same across countries. Here, we relax that assumption as well.

Labor is the only factor of production. 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 \( Q_{jn}^j \). 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 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 \( Q_{nj'}^j = Q_n^j \) 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 \( X_{nj'}^j = X_n^j \) 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 for \( \lambda_j Q_n^j \) 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 \( 1/\lambda_j \) units of good \( j \). For each time or for each auction \( t, t = 1, 2, ..., \lambda_j Q_n^j \), suppose country \( i \)'s productivity draw, \( Z_{nj,i,t}^j \), comes from \( F_{Z_{nj,i}^j|Q_n^j} (\cdot | \cdot) \). One may think of this case as

---

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

(ii) For any given \( n, j \) and \( i \), \( \{Z_{n,i,t}^j; t = 1, ..., \lambda_i^j Q_n^j\} \) are identically distributed from \( F_{Z_{n,t}^j | Q_n^j} (\cdot | \cdot) \) 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 or good \( j \) to destination \( n \) requires \( d_{ni}^j \) dollars worth 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:

\[
P_{n,i,t}^j = \frac{w_i d_{ni}^j}{Z_{n,i,t}^j} \tag{1}
\]

Denote the distribution of \( P_{n,i,t}^j \) given \( Q_n^j \) by \( G_{ni}^j (\cdot | Q_n^j) \) 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

\(^6\)Note 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 \( w_i \lambda_i^j / z_{n,i,t}^j \).

\(^7\)This 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.
$$G_{n}^{j}(p|q) \equiv \Pr \left[ P_{n,i,t}^{j} \leq p|Q_{n}^{j} = q \right] = \Pr \left[ \frac{w_{i}q_{n}^{j}}{z_{n,i,t}^{j}} \leq p|Q_{n}^{j} = q \right].$$ Hence,

$$G_{n}^{j}(p|q) \equiv 1 - F_{Z_{n}^{j}|Q_{n}^{j}}\left(\frac{w_{i}q_{n}^{j}}{p}\right)$$ (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:

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

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

For notational convenience define $A_{n}^{j} = \lambda^{j}Q_{n}^{j}$. Let $Q_{n}^{j}$ denote the quantity of good $j$ supplied by country $i$ to country $n$. Similarly, define $A_{n}^{j} = \lambda^{j}Q_{n}^{j}$. Note that depending on its productivity draw in each auction, each country can now be the lowest cost supplier for certain auctions, which means $A_{n}^{j}$, hence, $Q_{n}^{j}$, can now be positive for multiple sellers $A_{n}^{j}$ 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 $A_{n}^{j}$:

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

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

The total expenditure $X_{n}^{j}$ of country $n$ on good $j$ can be written as:

$$X_{n}^{j} = \sum_{t=1}^{\lambda^{j}Q_{n}^{j}} P_{n,t}^{j}$$ (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:

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

(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 $P_{n,i,t}^j 1 \left( P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i \right)$ by

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

(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) and (1). 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.

Assumption A2:

(i) For any given $n$ and $j$, we have $z_{ni}^j(q) = 0$ for all $q$ and for all $i$.

(ii) For any given $n$ and $j$, we have $z_{ni}^j(q) = \bar{z}^j_n(q) < \infty$ 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
destination $n$, the upper bound of the productivity distribution is the same irrespective of
the source country $i$ 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 $G_{ni}^j(\cdot)$, 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:

$$w_i L_i = \sum_{n=1}^{N} \pi_{ni} X_n$$

(8)

where $L_i$ is manufacturing workers, $X_n$ is the total spending on manufactures by country $n$,
and $\pi_{ni}$ 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 $\alpha$ is the share of total income spent on
manufactures hence

$$X_n = \alpha Y_n$$

(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

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

(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 $F_{Z_n}^j(\cdot)$ 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 $\{Q_n^j, X_n^j\}_{j=1}^J$ 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 $F_{Q_n^j, (Q_n^s)_{s \neq i}, (X_n^j)_{s \neq i}, Q_n^j, (\cdots)}$ 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 $H_{ni}^j(\cdot | Q_n^j)$ 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 $H_{ni}^j(\cdot | Q_n^j)$. Once we have the distribution $H_{ni}^j(\cdot | Q_n^j)$ 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 $G_{ni}^j(\cdot | Q_n^j)$ of the random variable $P_{n,i,t}^j$ given $Q_n^j$. Identification of the distribution $G_{ni}^j(\cdot | Q_n^j)$ 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 $F_{Z_n}^j(\cdot | Q_n^j)$ for each $n$, $i$, and $j$. After identifying the
distribution $F_{Z_{ni}^j | Q_n^j} (\cdot | Q_n^j)$ 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 $H_{ni}^j (\cdot | Q_n^j)$

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

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

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

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

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 $\left[ \varphi_{X_{ni}^j | Q_n^j} (u|q) \right]^{1/\lambda^j}$ given by

$$\left[ \varphi_{R_{ni}^j | Q_n^j} (u|q) \right]^{1/\lambda^j} = \left[ r_{X_{ni}^j | Q_n^j} (u|q) \right]^{1/\lambda^j} e^{i \frac{\theta_{X_{ni}^j | Q_n^j} (u|q) + 2k\pi}{\lambda^j}} \tag{14}$$
where for any \( q, k \) is any integer such that \( 0 \leq k \leq \lambda^n q - 1 \). Note that for any \( q \), only the one with \( k = 0 \), i.e., \( r_{X_{ni}^j|Q_n^j}(u|q) \) 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, \( \varphi_{R_{ni}^j|Q_n^j}(u|q) = \left[ r_{X_{ni}^j|Q_n^j}(u|q) \right]^{1/\lambda^n q} e \left( \frac{\theta_{X_{ni}^j|Q_n^j}(u|q)}{\lambda^n q} \right) \) for any \( q \) since \( \theta_{X_{ni}^j|Q_n^j}(0|q) = 0 \) for any \( q \). 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 \( G_{ni}^j(\cdot|Q_n^j) \)

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

\[
H_{ni}^j(p|q) = \Pr \left[ P_{n,i,t}^j \left( P_{n,i,t}^j \leq p_{n,s,t}^j, s \neq i \right) \leq p | Q_n^j = q \right] \\
= \Pr \left[ P_{n,i,t}^j \leq p_{n,s,t}^j, s \neq i \right] p \left( P_{n,i,t}^j \leq p_{n,s,t}^j, s \neq i | Q_n^j = q \right) \\
+ \Pr \left[ P_{n,i,t}^j \leq p_{n,s,t}^j, s \neq i \right] p \left( P_{n,i,t}^j > p_{n,s,t}^j, s \neq i | Q_n^j = q \right)
\]

\[
= \Pr \left[ P_{n,i,t}^j \leq p \right] p \left( P_{n,i,t}^j \leq p_{n,s,t}^j, s \neq i | Q_n^j = q \right) \\
+ \Pr \left[ P_{n,i,t}^j > p \right] p \left( P_{n,i,t}^j \leq p_{n,s,t}^j, s \neq i | Q_n^j = q \right)
\]

\[
\text{(15)}
\]

where the last equality follows from the fact that \( 1 \left( P_{n,i,t}^j \leq p_{n,s,t}^j, s \neq i \right) = 0 \) 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 \( 1 - \Pr \left[ P_{n,i,t}^j \leq p_{n,s,t}^j, s \neq i | Q_n^j = q \right] \). The probability \( \Pr \left[ P_{n,i,t}^j \leq p_{n,s,t}^j, s \neq i | Q_n^j = q \right] \) 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

\[ \Pr \left[ P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q \right] = \frac{E \left[ A_{ni}^j Q_n^j = q \right]}{\lambda^j q} \quad (17) \]

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 \( \Pr \left[ P_{n,i,t}^j > P_{n,s,t}^j, \text{ for some } s \neq i | Q_n^j = q \right] \), the probability \( \Pr \left[ P_{n,i,t}^j \leq p \text{ and } P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q \right] \) 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 \( \Pr \left[ P_{n,i,t}^j \leq p \text{ and } P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q \right] \) by \( \tilde{H}_{ni}^j (p|q) \).

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

In the Appendix we provide the proof which follows Rao (1992).

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

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

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

---

\(^8\)This follows from the fact that \( A_{ni}^j = \sum_{t=1}^{N \lambda q} (p_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i) \). We can write \( E [A_{ni}^j Q_n^j = q] = E \left[ \sum_{t=1}^{N \lambda q} 1(\text{for } P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i) | Q_n^j = q \right] = \sum_{t=1}^{N \lambda q} E \left[ 1(\text{for } P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i) | Q_n^j = q \right] = \sum_{t=1}^{N \lambda q} \Pr [P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q] = \lambda^j q \Pr [P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n^j = q] \). The last equality follows from the fact that for a given level of total demand \( Q_n^j \), for any auction or batch \( t \) 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(ii).
3.3 Identification of \( d_{ni}^j \)

Using (2) we can write

\[
F_{Z_{ni}^j|Q_n^j}(z|q) = 1 - G_{ni}^j\left(\frac{w_{d_{ni}^j}}{z}|q\right)
\]

(19)

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

\[
F_{Z_{nn}^j|Q_n^j}(z|q) = 1 - G_{nn}^j\left(\frac{w_{d_{nn}^j}}{z}|q\right)
\]

(20)

Let \( p_{ni}^j(q) \) be the lower bound of the support of \( P_{ni,i}^j \) 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 \( p_{ni}^j(q) \) is known for all \( n \) and \( i \) and for any given \( j \) and for all \( q \). Therefore, \( p_{ni}^j(q) \) is known. Noting that

\[
p_{ni}^j(q) = \frac{w_{d_{ni}^j}(q)}{\pi_{ni}^j(q)}
\]

(21)

and

\[
p_{nn}^j(q) = \frac{w_{d_{nn}^j}(q)}{\pi_{nn}^j(q)}
\]

(22)

and since \( d_{nn}^j(q) = 1 \) for any \( j \) and for any level of \( q \) by Assumption A2(iii) and \( \pi_{ni}^j(q) = \pi_{n}^j(q) \), for all \( i \), for a given \( n \) and \( j \) by Assumption A2(ii) we can identify \( d_{ni}^j(q) \) for any given \( j \), for all \( n \) and \( i \) and for all \( q \), using the following equation:

\[
d_{ni}^j(q) = \frac{w_{n} p_{ni}^j(q)}{w_{i} p_{nn}^j(q)}
\]

(23)

Moreover, \( \pi_{n}^j(q) \) is identified for all \( q \) for a given \( j \) from (22).
3.4 Identification of $F_{Z_{ni}j|Q_ni}^j(\cdot|Q_ni)$

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

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

(24)

Since $G_{ni}^j(\cdot|q)$ and $d_{ni}^j(q)$ are both identified, now, $F_{Z_{ni}j|Q_ni}^j(\cdot|q)$ is identified for all $n$ and $i$ and for any given $j$ and for all $q$ Hence, $F_{Z_{ni}j}^j(\cdot)$ is identified for all $n$ and $i$ and for any given $j$ since $F_{Q_ni}^j(\cdot)$ 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 $j$, $F_{Z_{ni}j}^j(\cdot)$ is identified.

4 Data

Our source of trade data is the United Nations Commodity Trade Statistics Database (COMTRADE) which is accessible online. As in Eaton and Kortum (2002) we use bilateral trade data for manufacturing imports in 1990. Bilateral trade data consists of traded quantity and trade values (in dollars) between countries and is available on disaggregated level according to different classifications. The disaggregation level is indicated by digits and higher digits correspond to more disaggregated product categories. Eaton and Kortum (2002) look at 19 OECD countries and use 4-digit SITC (Standard International Trade Classification) Revision2 data. To determine which product categories correspond to the Bureau of Economic Analysis (BEA) manufacturing industry codes they use Maskus (1991) concordance. They, then aggregate to obtain total manufacturing imports between countries.

---

9Those 19 countries are: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, United Kingdom, United States. Also, Eaton and Kortum (2002) uses United Nations-Statistics Canada data (World Trade Database) as explained by Feenstra, Lipsey and Bowen (1997) which is a slightly modified version of United Nations Commodity Trade Statistics Data.
<table>
<thead>
<tr>
<th>Country</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>0.0024</td>
</tr>
<tr>
<td>Austria</td>
<td>0.0668</td>
</tr>
<tr>
<td>Belgium-Luxembourg</td>
<td>0.1047</td>
</tr>
<tr>
<td>Canada</td>
<td>0.0099</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.0203</td>
</tr>
<tr>
<td>Finland</td>
<td>0.0165</td>
</tr>
<tr>
<td>France</td>
<td>0.1665</td>
</tr>
<tr>
<td>Greece</td>
<td>0.0078</td>
</tr>
<tr>
<td>Italy</td>
<td>0.1402</td>
</tr>
<tr>
<td>Japan</td>
<td>0.0985</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.0946</td>
</tr>
<tr>
<td>New Zealand</td>
<td>0.0003</td>
</tr>
<tr>
<td>Norway</td>
<td>0.0101</td>
</tr>
<tr>
<td>Portugal</td>
<td>0.0132</td>
</tr>
<tr>
<td>Spain</td>
<td>0.0316</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.0352</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.0875</td>
</tr>
<tr>
<td>USA</td>
<td>0.0940</td>
</tr>
</tbody>
</table>

In contrast to Eaton and Kortum (2002), we use the data at the disaggregated level, in particular, 5-digit SITC Revision 2, and the categories corresponding to BEA manufacturing industries according to the concordance provided by Feenstra, Lipsey and Bowen (1997).\textsuperscript{10-11}

For this particular application we look at the German market, i.e. $n = \text{Germany}$.\textsuperscript{12} We have $J = 1268$ products hence we have $\{Q_{ni}^j, X_{ni}^j\}_{j=1}^J$, where $n = \text{Germany}$, $i = 1, ..., 19$. Those 19 OECD countries and their respective percentages in the total value of manufacturing imports of Germany is given in Table 1.

\textsuperscript{10}Whenever 5 digit level is available we use data at this level. For some categories no further classification was available after 4 digits. In those cases we use the 4 digit level.

\textsuperscript{11}Eaton and Kortum (2002) mention that using the concordance by Feenstra, Lipsey and Bowen (1997) made no difference in their case.

\textsuperscript{12}This is actually the Federal Republic of Germany according to the United Nations Commodity Trade Statistics Database. Also, Belgium in their sample is actually Belgium and Luxembourg since they were still reporting together. We, however, follow the convention in Eaton and Kortum (2002) and call them Germany and Belgium, respectively.
5 Nonparametric Estimation

If we had many observations for a given good \( j \) then we could estimate the joint distribution \( F_{Q^{j}_{ni},(Q^{j}_{ns})_{x_{ni}^{j},x_{ns}^{j}}}(\cdot,\cdot,\cdot) \) from the observables immediately. This is not the case with our data however, since we have only a single observation for a given good \( j \). Thus, we make the following assumptions.

Assumption B1 :

(i) For any given \( n \), the arrays \( \{Z^{j}_{n,i,t}; i = 1, ..., N; t = 1, ..., \lambda^{j}Q^{j}_{n}\} \) are mutually independent across \( j \) given \( (Q^{1}_{n}, .., Q^{\bar{j}}_{n}, .., Q^{J}_{n}) \).

(ii) For any given \( n \) and for every \( j \), we have \( \{Z^{j}_{n,i,t}; i = 1, ..., N; t = 1, ..., \lambda^{j}Q^{j}_{n}\} \) independent of \( (Q^{1}_{n}, .., Q^{j-1}_{n}, Q^{j+1}_{n}, .., Q^{J}_{n}) \) given \( Q^{j}_{n} \).

(iii) For any given \( n \), \( (Q^{1}_{n}, .., Q^{\bar{j}}_{n}, .., Q^{J}_{n}) \) are mutually independent.

Assumption B1(i) basically states that for any given destination \( n \), all the productivities for good \( j \) are independent of all the productivities for good \( j' \) given the total demands for all the goods. Assumption B1(ii) says that the productivities for good \( j \) only depend on the total demand for good \( j \) but not on the total demands for the other goods. In Eaton and Kortum (2002) productivities are assumed to be mutually independent across goods as well. Assumption B1(iii) says for any given destination \( n \) total demands for all the goods are mutually independent.

Assumption B2 :

(i) For any given \( n \) and \( j \), we have \( F_{Q^{j}_{n}}(q) = F_{A_{n}}(\lambda^{j}q) \) for all \( q \).

(ii) For any given \( n \), \( j \) and \( i \), we have \( F_{Z^{j}_{ni}|Q^{j}_{n}}(\cdot|q) = F_{Z^{j}_{ni}|A_{n}}(\cdot|\lambda^{j}q) \) for all \( q \).

Assumption B2(i) implies that on average number of auctions/batches, \( A^{j}_{n} \), is the same for
all goods since $A^j_n = \lambda^j q \sim F_{A^j_n}(\cdot)$ for all $j$.

**Assumption B3:**

For any given $n$, $j$ and $i$, we have $d^j_{ni}(q) = d_{ni}(\lambda^j q)$ for all $q$.

### 5.1 Choosing $\lambda^j$

We mentioned before that 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 $\lambda^j Q^j_n$ denotes how many batches are sold. Suppose there is competition for supplying each batch of good $j$ and hence, countries compete for $\lambda^j Q^j_n$ times in order to supply $1/\lambda^j$ units of good $j$ in country $n$. An alternative interpretation would be that there are $\lambda^j Q^j_n$ separate auctions and in each auction there is competition to supply $1/\lambda^j$ units of good $j$. For our estimation, hereafter, we pick $\lambda^j = \frac{\sum\sum x^j_{ni}}{\sum\sum Q^j_h}$, where $\lambda^j$ can be considered as the average unit value of good $j$ across all import markets around the world. Thus, for any $j$, $\lambda^j$ is observed. As we defined previously $A^j_n = \lambda^j Q^j_n$ and $A^j_n$ can now also be interpreted as the value of total demand, $Q^j_n$, at the average world price. Also, we defined $A^j_{ni} = \lambda^j Q^j_{ni}$ and $A^j_{ni}$ can now be interpreted as the value of good $j$ supplied by country $i$ to country $n$ at the average world price. Table 2 provides summary statistics for $A^j_{ni}$ values for each exporter country across goods.

### 5.2 Estimation of $d^j_{ni}$

We will first estimate $d^j_{ni}(q) = d_{ni}(\lambda^j q)$. From (6) it can be seen that the lowest possible value $X^j_{ni}$ can get is zero meaning that country $i$ did not win any auction for good $j$ in destination $n$. In other words, the lower bound of the distribution of $X^j_{ni}$ given $Q^j_n$ is zero. Now consider the lower bound, $X^j_{ni}^*(q)$, of the distribution of $X^j_{ni}$ given $Q^j_n$ and when $X^j_{ni} > 0$. Basically, when we know that country $i$ won at least one auction for good $j$ such that $X^j_{ni} > 0$
<table>
<thead>
<tr>
<th>Country</th>
<th>Mean</th>
<th>St.Dev</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>437,944</td>
<td>5,136,399</td>
<td>109,418,056</td>
</tr>
<tr>
<td>Austria</td>
<td>10,396,672</td>
<td>33,053,050</td>
<td>618,476,908</td>
</tr>
<tr>
<td>Belgium-Luxembourg</td>
<td>17,417,264</td>
<td>121,943,091</td>
<td>4,054,870,696</td>
</tr>
<tr>
<td>Canada</td>
<td>1,346,670</td>
<td>17,192,883</td>
<td>583,119,315</td>
</tr>
<tr>
<td>Denmark</td>
<td>3,605,257</td>
<td>13,317,858</td>
<td>294,413,636</td>
</tr>
<tr>
<td>Finland</td>
<td>2,694,681</td>
<td>20,203,636</td>
<td>437,345,751</td>
</tr>
<tr>
<td>France</td>
<td>25,971,030</td>
<td>142,750,769</td>
<td>3,796,900,758</td>
</tr>
<tr>
<td>Greece</td>
<td>1,560,668</td>
<td>13,263,228</td>
<td>307,529,716</td>
</tr>
<tr>
<td>Italy</td>
<td>24,354,679</td>
<td>91,205,988</td>
<td>1,647,218,640</td>
</tr>
<tr>
<td>Japan</td>
<td>14,539,897</td>
<td>161,192,839</td>
<td>5,493,273,372</td>
</tr>
<tr>
<td>Netherlands</td>
<td>17,242,510</td>
<td>40,894,500</td>
<td>410,896,132</td>
</tr>
<tr>
<td>New Zealand</td>
<td>56,113</td>
<td>1,174,892</td>
<td>40,147,053</td>
</tr>
<tr>
<td>Norway</td>
<td>1,706,464</td>
<td>15,186,126</td>
<td>480,643,023</td>
</tr>
<tr>
<td>Portugal</td>
<td>2,436,756</td>
<td>16,670,855</td>
<td>394,348,129</td>
</tr>
<tr>
<td>Spain</td>
<td>5,814,978</td>
<td>48,843,963</td>
<td>1,435,100,606</td>
</tr>
<tr>
<td>Sweden</td>
<td>5,026,713</td>
<td>22,728,601</td>
<td>467,235,790</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>13,376,705</td>
<td>60,167,218</td>
<td>1,152,107,740</td>
</tr>
<tr>
<td>USA</td>
<td>11,262,286</td>
<td>68,225,713</td>
<td>1,535,926,065</td>
</tr>
</tbody>
</table>
what is the lowest possible value \( X_{ni}^j \) can get given \( Q_{nj}^i \)? The lowest possible value in this case will occur when country \( i \) wins only one auction and wins it with the lowest possible price it can offer which is \( p_{ni}^j(q) \). Therefore, \( X_{ni}^j = p_{ni}^j(q) \). Now, define \( \phi_{ni}^j(q) \) as

\[
\phi_{ni}^j(q) \equiv \frac{X_{ni}^j(Q_{nj}^i)}{w_i} = \frac{p_{ni}^j(q)}{w_i}
\]

for all \( n, i \) and \( j \). Using (25), we can rewrite (23) as

\[
d_{ni}^j(q) = \frac{\phi_{ni}^j(q)}{\phi_{nn}^j(q)} = \frac{\phi_{ni}(\lambda_j q)}{\phi_{nn}(\lambda_j q)} = d_{ni}(\lambda_j q)
\]

where the last two equalities follow from Assumption B1, B2 and B3. Since \( X_{ni}^j(Q_{nj}^i) \) is the lower bound we know that for all \( j \) such that \( X_{ni}^j > 0 \) given \( Q_{nj}^i \), we have

\[
\frac{X_{ni}^j}{w_i} \geq \frac{X_{ni}^j(Q_{nj}^i)}{w_i} = \phi_{ni}(Q_{nj}^i) = \phi_{ni}(\lambda_j Q_{nj}^i) = \phi_{ni}(A_{nj}^i)
\]

Hence, when we plot \( (X_{ni}^j/w_i, A_{nj}^i) \) observations for all \( j \) such that \( X_{ni}^j > 0 \) they should all lie above \( \phi_{ni}(A_{nj}^i) \) curve. We estimate this lower boundary curve, \( \phi_{ni}(\cdot) \), using uniform b-splines.

Define \( \tilde{\phi}_{ni}(\tilde{L}A_{nj}^i) = \phi_{ni}(A_{nj}^i) \) where \( \tilde{L}A_{nj}^i = \frac{\ln A_{nj}^i - \min(\ln A_{nj}^i)}{\max(\ln A_{nj}^i) - \min(\ln A_{nj}^i)} \). Note that \( \tilde{L}A_{nj}^i \in [0, 1] \) for all \( j \) hence we can use b-splines. Consider \( \tilde{\phi}_{ni}(\tilde{L}A_{nj}^i) = \sum_{k=0}^{K} \psi_k b_k(\tilde{L}A_{nj}^i) \) where \( b_k(\cdot) \) are uniform b-splines of degree \( p \) and we have \( K + 1 \) basis functions. Since \( \phi_{ni}(\cdot) \) is the lower boundary, the idea is that our estimator is the curve that maximizes the area under it while having all the \( (X_{ni}^j/w_i, A_{nj}^i) \) observations above, following Tsybakov (1993). Specifically, we solve the following problem:

\[
\begin{align*}
\max_{\{\psi_k\}_{k=0}^{K}} & \int_0^1 \tilde{\phi}_{ni}(\tilde{L}A_{nj}^i) d\tilde{L}A_{nj}^i \\
\text{s.t.} & \quad \frac{X_{ni}^j}{w_i} \geq \tilde{\phi}_{ni}(\tilde{L}A_{nj}^i) \text{ for all } j = 1, \ldots, J
\end{align*}
\]
Substituting the definition of $\tilde{\phi}_{ni}(\cdot)$ into the objective function and the constraint note that (27) becomes a linear programming problem since both the objective function and the constraint are linear in $\psi_k$'s.

5.3 Estimation of $\hat{H}^j_{ni}(\cdot|Q^j_n)$

First, we will estimate the characteristic function, $\varphi_{R^{j}_{ni}|Q^n}(\cdot|Q^j_n)$, of $R^{j}_{ni,i;t}$ given $Q^j_n$. Let $\varphi_{R^{j}_{ni}|A_n}(\cdot|A^j_n)$ be the characteristic function of $R^{j}_{ni,i;t}$ given $A^j_n$ and $\varphi_{X^{j}_{ni}|A_n}(\cdot|A^j_n)$ be the characteristic function of $X^{j}_{ni}$ given $A^j_n$. Note that from Assumption B1, B2, and B3, it can be easily shown that given $A^j_n$ both functions do not depend on $j$. Following similar arguments as in (14) it can be shown that:

$$[\varphi_{R^{j}_{ni}|A_n}(u|a)] = [r_{X^{j}_{ni}|A_n}(u|a)]^{1/2} e^{i\vartheta_{X^{j}_{ni}|A_n}(u|a)} 	ag{28}$$

where $r_{X^{j}_{ni}|A_n}(u|a)$ is the modulus and $\vartheta_{X^{j}_{ni}|A_n}(u|a)$ is the argument of $\varphi_{X^{j}_{ni}|A_n}(\cdot|a)$. Since we know

$$\varphi_{X^{j}_{ni}|A_n}(u|a) \equiv E[e^{i\vartheta_{X^{j}_{ni}|A_n}(u|a)}] = a 	ag{29}$$

consider the following nonparametric regression to get our estimator of $\varphi_{X^{j}_{ni}|A_n}(u|a)$:

$$\hat{\varphi}_{X^{j}_{ni}|A_n}(u|a) = \frac{1}{J} \sum_{j'=1}^{J} e^{iuX^{j'}_{ni}} K\left(\frac{\ln a - \ln A^j_n}{cs_{A_n} J^{1/5}}\right) 	ag{30}$$

where $s_{A_n}$ is the standard error of the random variable $\ln A_n$, $h_1$ is some bandwidth, and $K(\cdot)$ is some kernel with $K'(\cdot)$ its derivative. Using rule of thumb for $h_1$, we set $h_1 = cJ^{-1/5}$ where $c = 1.06$. We use a standard Gaussian kernel in our estimation.

Using $e^{iuX^{j'}_{ni}} = \cos(uX^{j'}_{ni}) + i \sin(uX^{j'}_{ni})$ and defining $a_{ni}(u, a)$ and $b_{ni}(u, a)$ as follows, (30)
From Assumption B1, B2, and B3, we know that can be written as

\[
\hat{\varphi}_{X_n|A_n}(u|a) = \frac{\sum_{j=1}^{J} \cos(uX_{ni}^j)K\left(\frac{ln a - \ln A_n^j}{c_s A_n^{j-1/5}}\right)}{\sum_{j=1}^{J} K\left(\frac{ln a - \ln A_n^j}{c_s A_n^{j-1/5}}\right) a_{ni}(u,a)} + i \frac{\sum_{j=1}^{J} \sin(uX_{ni}^j)K\left(\frac{ln a - \ln A_n^j}{c_s A_n^{j-1/5}}\right)}{\sum_{j=1}^{J} K\left(\frac{ln a - \ln A_n^j}{c_s A_n^{j-1/5}}\right) b_{ni}(u,a)}
\]

\[= a_{ni}(u,a) + i b_{ni}(u,a) \quad (31)\]

Using polar coordinate representation, (31) can be expressed as

\[\hat{\varphi}_{X_n|A_n}(u|a) = \hat{r}_{X_n|A_n}(u|a)e^{i\hat{\theta}_{X_n|A_n}(u|a)} \quad (32)\]

where \(\hat{r}_{X_n|A_n}(u|a) = \sqrt{a_{ni}(u,a)^2 + b_{ni}(u,a)^2}\) and

\[
\hat{\theta}_{X_n|A_n}(u|a) = \begin{cases} 
\arctan\left(\frac{b_{ni}(u,a)}{a_{ni}(u,a)}\right) & \text{if } a_{ni}(u,a) > 0 \\
\arctan\left(\frac{b_{ni}(u,a)}{a_{ni}(u,a)}\right) + \pi & \text{if } a_{ni}(u,a) < 0 \text{ and } b_{ni}(u,a) \geq 0 \\
\arctan\left(\frac{b_{ni}(u,a)}{a_{ni}(u,a)}\right) - \pi & \text{if } a_{ni}(u,a) < 0 \text{ and } b_{ni}(u,a) < 0 \\
\frac{\pi}{2} & \text{if } a_{ni}(u,a) = 0 \text{ and } b_{ni}(u,a) > 0 \\
-\frac{\pi}{2} & \text{if } a_{ni}(u,a) = 0 \text{ and } b_{ni}(u,a) < 0 \\
\text{indeterminate} & \text{if } a_{ni}(u,a) = 0 \text{ and } b_{ni}(u,a) = 0
\end{cases}
\]

Hence, we define our estimator of \(\hat{\varphi}_{R_{ni}|Q^i_n}(u|a)\) as

\[\hat{\varphi}_{R_{ni}|A_n}(u|a) = \left[\hat{r}_{X_n|A_n}(u|a)\right]^{1/a} e^{i\left(\frac{\hat{\theta}_{X_n|A_n}(u|a)}{a}\right)} \quad (33)\]

Now, define \(H_{ni}^A(m|A_n^i) \equiv \Pr\left[R_{n,i,t}^j \leq m|A_n^i = a\right] = \Pr\left[P_{n,i,t}^j \leq P_{n,s,t}^j \leq s \neq i \leq m|A_n^i = a\right].\)

From Assumption B1, B2, and B3, we know that \(H_{ni}^j(\cdot|Q_n^i) = H_{ni}^A(\cdot|A_n^i) = H_{ni}^A(\cdot|A_n^j),\) i.e., \(R_{n,i,t}^j \mid A_n^i \sim H_{ni}^A(\cdot|A_n^j).\) Following Levy’s Inversion Formula, we define our estimator of
\[ H_{n_i}^A(m|A_n^i = a) \]

\[
\hat{H}_{n_i}^A(m|A_n^i = a) = \frac{1}{2\pi} \int_{-T}^{T} \frac{1 - e^{-itm}}{it} d_T(t) \varphi_{R_{n_i}|A_n}^A(t|a) dt \tag{34}
\]

where

\[
d_T(t) = \begin{cases} 
1 - \frac{|t|}{T} & \text{if } |t| \leq T \\
0 & \text{otherwise}
\end{cases}
\tag{35}
\]

is the damping factor and \( t \in [-T, T] \). Denote the derivative of \( H_{n_i}^A(m|A_n^i) \) with respect to \( m \) as \( h_{n_i}^A(m|A_n^i) \). We use the following estimator for \( h_{n_i}^A(m|A_n^i) \)

\[
\hat{h}_{n_i}^A(m|A_n^i = a) = \frac{1}{2\pi} \int_{-T}^{T} e^{-itm} d_T(t) \varphi_{R_{n_i}|A_n}^A(t|a) dt
\tag{36}
\]

where \( d_T(t) \) is the damping factor defined in (35).

Lastly, we define our estimator of \( H_{n_i}^j(m|Q_n^j = q) \) as

\[
\hat{H}_{n_i}^j(m|Q_n^j = q) = \hat{H}_{n_i}^A(m|\lambda^j Q_n^j = \lambda^j q)
\tag{37}
\]

### 5.4 Estimation of \( f_{Z_{n_i}|Q_{n_i}} \)

Let us define \( G_{n_i}^A(p|A_n^j) \equiv \Pr [P_{n,i,t}^j \leq p|A_n^j = a] \). From Assumption B1, B2, and B3, we know that \( G_{n_i}^j(\cdot|Q_n^j) = G_{n_i}^A(\cdot|\lambda^j Q_n^j) = G_{n_i}^A(\cdot|A_n^j) \), i.e., \( P_{n,i,t}^j | A_n^j \sim G_{n_i}^A(\cdot|A_n^j) \). Note that using Assumption B1, B2, and B3, we can write

\[
G_{n_i}^A(p|a) \equiv \Pr [P_{n,i,t}^j \leq p|A_n^j = a] = \Pr \left[ \frac{w_id_{ni}^j}{z_{n,i,t}^j} \leq p|A_n^j = a \right] = 1 - \Pr \left[ z_{n,i,t}^j \leq \frac{w_id_{ni}^j}{p}|A_n^j = a \right] \tag{38}
\]

\[
= 1 - F_{Z_{n_i}|A_n} \left( \frac{w_id_{ni}^j}{p}|a \right)
\]
From (38), we have $F_{z_{ni}|A_n}(z|a) = 1 - G_{ni}^A\left(\frac{w_{d_{ni}}(a)}{z}|a\right)$, taking derivative of both sides with respect to $z$; we have

$$f_{Z_{ni}|A_n}(z|a) = g_{ni}^A \left(\frac{w_{d_{ni}}(a)}{z}|a\right) \frac{w_{d_{ni}}(a)}{z^2}$$  \hspace{1cm} (39)$$

where $f_{Z_{ni}|A_n}(\cdot|a)$ and $g_{ni}^A(\cdot|a)$ are the derivatives of $F_{Z_{ni}|A_n}(\cdot|a)$ and $G_{ni}^A(\cdot|a)$, respectively with respect to the first arguments. Using (38) and (39) we can write

$$f_{Z_{ni}|A_n}(z|a) = \frac{w_{d_{ni}}(a)}{z^2} \frac{g_{ni}^A \left(\frac{w_{d_{ni}}(a)}{z}|a\right)}{1 - G_{ni}^A\left(\frac{w_{d_{ni}}(a)}{z}|a\right)}$$  \hspace{1cm} (40)$$

Following Lemma 2 and Assumption B1, B2, and B3, we have

$$G_{ni}^A(p|a) = 1 - \exp \left\{ - \int_0^p \left[ 1 - \sum_{n=1}^N \tilde{H}_{ns}^A(v|a) \right]^{-1} d\tilde{H}_{ni}^A(v|a) \right\}$$  \hspace{1cm} (41)$$

where $\tilde{H}_{ni}^A(p|q) = \Pr \left[ P_{n,i,t}^j \leq p \text{ and } P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i|A_n = a \right]$ similar to $\tilde{H}_{ni}^j(p|q)$ and $s = 1, ..., N$. Taking derivative of both sides in (41) with respect to $p$ we have

$$g_{ni}^A(p|a) = \exp \left\{ - \int_0^p \left[ 1 - \sum_{n=1}^N \tilde{H}_{ns}^A(v|a) \right]^{-1} d\tilde{H}_{ni}^A(v|a) \right\} \left[ 1 - \sum_{n=1}^N \tilde{H}_{ns}^A(p|a) \right]^{-1} \tilde{h}_{ni}^A(p|a)$$  \hspace{1cm} (42)$$

where $\tilde{h}_{ni}^A(\cdot|a)$ is the derivative of $\tilde{H}_{ni}^A(\cdot|q)$ with respect to the first argument. From (41) and (42) we can write

$$\frac{g_{ni}^A(p|a)}{1 - G_{ni}^A(p|a)} = \frac{\tilde{h}_{ni}^A(p|a)}{1 - \sum_{n=1}^N \tilde{H}_{ns}^A(p|a)}$$  \hspace{1cm} (43)$$

where $s = 1, ..., N$. Similar to (15) and (17) we can write

$$\tilde{H}_{ns}^A(p|a) = H_{ns}^A(p|a) - \left[ 1 - \frac{E \left[ A_{ni}^j | A_n = a \right]}{a} \right]$$  \hspace{1cm} (44)$$
for any $s$ where $s = 1, ..., N$. Summing over all $s$ in (44), we get

$$\sum_{s = 1, n \neq s}^{N} \tilde{H}_{ns}^{A}(p|a) = \sum_{s = 1, n \neq s}^{N} H_{ns}^{A}(p|a) - N + 2$$ (45)

where $s = 1, ..., N$. Taking derivative of both sides with respect to $p$ in (44) we have

$$\tilde{h}_{ns}^{A}(p|a) = h_{ns}^{A}(p|a)$$ (46)

since the second expression on the RHS in (44) does not depend on $p$. Now, using (45) and (46) we can rewrite (43) as

$$\frac{g_{ni}^{A}(p|a)}{1 - C_{ni}^{A}(p|a)} = \frac{h_{ni}^{A}(p|a)}{(N - 1) - \sum_{s = 1}^{N} H_{ns}^{A}(p|a)}$$ (47)

Finally, using (47) we can rewrite (40) as

$$\frac{f_{Z_{ni}|A_{n}(z|a)}}{F_{Z_{ni}|A_{n}(z|a)}} = \frac{w_{i}d_{ni}(a)}{z^{2}} \cdot \frac{h_{ni}^{A}\left(\frac{w_{i}d_{ni}(a)}{z}\right)}{(N - 1) - \sum_{s = 1}^{N} H_{ns}^{A}(\frac{w_{i}d_{ni}(a)}{z}|a)}$$ (48)

Note that

$$\ln F_{Z_{ni}|A_{n}(\varpi_{ni}(a)|a)} - \ln F_{Z_{ni}|A_{n}(z|a)} = \int_{z} \frac{f_{Z_{ni}|A_{n}(u|a)}}{F_{Z_{ni}|A_{n}(u|a)}} du$$

The first expression on the LHS above is zero so we can write

$$F_{Z_{ni}|A_{n}(z|a)} = \exp \left( - \int_{z} \frac{f_{Z_{ni}|A_{n}(u|a)}}{F_{Z_{ni}|A_{n}(u|a)}} du \right)$$

---

13This follows from the fact that $\sum_{s = 1}^{N} \tilde{H}_{ns}^{A}(p|a) = \sum_{s = 1}^{N} H_{ns}^{A}(p|a) - \sum_{s = 1}^{N} \left[ 1 - \frac{E[A_{ns}^{i}|A_{n}^{i} = a]}{a} \right]$. Note that

$$\sum_{s = 1}^{N} \left[ 1 - \frac{E[A_{ns}^{i}|A_{n}^{i} = a]}{a} \right] = N - \frac{E[\sum_{s = 1}^{N} A_{ns}^{i}|A_{n}^{i} = a]}{a} = N - \frac{E[A_{n}^{i}|A_{n}^{i} = a]}{a}.$$ Last equality follows from $\sum_{s = 1}^{N} A_{ns}^{i} = A_{n}^{i}$. Since $E[A_{n}^{i}|A_{n}^{i} = a] = a$ we have the second expression on RHS equal to 1.
Hence, we have
\[
f_{Z_n|A_n}(z|a) = \frac{f_{Z_n|A_n}(z|a)}{F_{Z_n|A_n}(z|a)} \exp \left( - \int z f_{Z_n|A_n}(u|a) \frac{du}{F_{Z_n|A_n}(u|a)} \right)
\]

Finally, we define the following estimator for \( f_{Z_n|A_n}(z|a) \) such that
\[
\hat{f}_{Z_n|A_n}(z|a) = \frac{w_id_{ni}(a)}{z^2} \sum_{s=1}^{N} \frac{\hat{h}_A^{\frac{w_id_{ni}(a)}{z}|a}}{(N-1) - \sum_{s=1}^{N} \hat{H}_A^{\frac{w_id_{ni}(a)}{z}|a}} \exp \left( - \int z f_{Z_n|A_n}(u|a) \frac{du}{F_{Z_n|A_n}(u|a)} \right)
\]
where \( \hat{H}_A^{\cdot|a} \), \( \hat{h}_A^{\cdot|a} \), and \( \hat{z}_{ni}(a) \) are our estimators of \( H_A^{\cdot|a} \), \( h_A^{\cdot|a} \), and \( z_{ni}(a) \), respectively from the previous sections.

Lastly, we define our estimator of \( f_{Z_n|Q_n^j}(z|q) \) as
\[
\hat{f}_{Z_n|Q_n^j}(z|Q_n^j = q) = \frac{\hat{f}_{Z_n|A_n}(z|\lambda^j Q_n^j = \lambda^j q)}{\hat{f}_{Z_n|A_n}(z|\lambda^j Q_n^j = \lambda^j q)}
\]

6 Empirical Results

As we mentioned before for this particular application, we look at the German market of manufacturing imports for 1990 for 19 OECD countries as in Eaton and Kortum (2002). Table 3 shows the estimated probabilities for country \( i \) to be the winner for a given value of total demand \( A_n^j \) in Germany. These are the nonparametric regression estimates of
\[
Pr \left[ P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | A_n^j = a \right] \text{ using }
\]
\[
Pr \left[ P_{n,i,t}^j \leq P_{n,s,t}^j, s \neq i | A_n^j = q \right] = \frac{E \left[ A_{ni}^j | A_n^j = a \right]}{a}
\]

These probabilities are highest for Belgium-Lux, France, Italy, UK, and Netherlands and lowest for Australia, New Zealand, and Greece but they also show different patterns across different \( A_n^j \) values as well. For instance Japan has high probability estimates for higher \( A_n^j \)
Table 3: Estimated Probabilities of Each Country Being Winner for $A_j^n$ Values (in dollars)

<table>
<thead>
<tr>
<th>$A_j^n$ Values</th>
<th>1 mil</th>
<th>10 mil</th>
<th>50 mil</th>
<th>100 mil</th>
<th>500 mil</th>
<th>1 bil</th>
<th>10 bil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>0.00940</td>
<td>0.00404</td>
<td>0.00421</td>
<td>0.00270</td>
<td>0.00362</td>
<td>0.00195</td>
<td>0.00011</td>
</tr>
<tr>
<td>Austria</td>
<td>0.05802</td>
<td>0.08162</td>
<td>0.06763</td>
<td>0.07648</td>
<td>0.07506</td>
<td>0.06997</td>
<td>0.02787</td>
</tr>
<tr>
<td>Belgium-Lux</td>
<td>0.11419</td>
<td>0.10803</td>
<td>0.11356</td>
<td>0.11020</td>
<td>0.11062</td>
<td>0.10473</td>
<td>0.16128</td>
</tr>
<tr>
<td>Canada</td>
<td>0.00282</td>
<td>0.00752</td>
<td>0.00814</td>
<td>0.00832</td>
<td>0.00643</td>
<td>0.00956</td>
<td>0.00192</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.02031</td>
<td>0.03235</td>
<td>0.03162</td>
<td>0.03210</td>
<td>0.02702</td>
<td>0.02378</td>
<td>0.00596</td>
</tr>
<tr>
<td>Finland</td>
<td>0.00820</td>
<td>0.00988</td>
<td>0.00990</td>
<td>0.01205</td>
<td>0.01837</td>
<td>0.02738</td>
<td>0.00124</td>
</tr>
<tr>
<td>France</td>
<td>0.14825</td>
<td>0.13826</td>
<td>0.16789</td>
<td>0.16437</td>
<td>0.13982</td>
<td>0.14268</td>
<td>0.24746</td>
</tr>
<tr>
<td>Greece</td>
<td>0.00955</td>
<td>0.00796</td>
<td>0.00761</td>
<td>0.01004</td>
<td>0.01426</td>
<td>0.01412</td>
<td>0.00011</td>
</tr>
<tr>
<td>Italy</td>
<td>0.15270</td>
<td>0.14328</td>
<td>0.14588</td>
<td>0.15703</td>
<td>0.16519</td>
<td>0.15843</td>
<td>0.10846</td>
</tr>
<tr>
<td>Japan</td>
<td>0.04516</td>
<td>0.05131</td>
<td>0.04858</td>
<td>0.05110</td>
<td>0.09102</td>
<td>0.08469</td>
<td>0.19090</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.18646</td>
<td>0.16170</td>
<td>0.15201</td>
<td>0.15044</td>
<td>0.12617</td>
<td>0.11199</td>
<td>0.02806</td>
</tr>
<tr>
<td>New Zealand</td>
<td>0.00058</td>
<td>0.00075</td>
<td>0.00210</td>
<td>0.00156</td>
<td>0.00004</td>
<td>0.00003</td>
<td>0.00001</td>
</tr>
<tr>
<td>Norway</td>
<td>0.01300</td>
<td>0.00984</td>
<td>0.01392</td>
<td>0.01332</td>
<td>0.01174</td>
<td>0.01633</td>
<td>0.00145</td>
</tr>
<tr>
<td>Portugal</td>
<td>0.01708</td>
<td>0.00871</td>
<td>0.01799</td>
<td>0.01587</td>
<td>0.01426</td>
<td>0.01746</td>
<td>0.00218</td>
</tr>
<tr>
<td>Spain</td>
<td>0.02800</td>
<td>0.03223</td>
<td>0.02562</td>
<td>0.02156</td>
<td>0.03206</td>
<td>0.02956</td>
<td>0.08006</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.01463</td>
<td>0.03079</td>
<td>0.03487</td>
<td>0.03272</td>
<td>0.03614</td>
<td>0.04304</td>
<td>0.00855</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.10282</td>
<td>0.09815</td>
<td>0.08415</td>
<td>0.07461</td>
<td>0.07418</td>
<td>0.09034</td>
<td>0.06446</td>
</tr>
<tr>
<td>USA</td>
<td>0.06882</td>
<td>0.07359</td>
<td>0.06431</td>
<td>0.06554</td>
<td>0.05400</td>
<td>0.05395</td>
<td>0.06991</td>
</tr>
</tbody>
</table>
levels, i.e., for high value of total demand.

Our results for trade costs by solving our linear programming problem in (27) using basis splines are presented in Table 4. Since we do not have home country, here Germany, data we assume Denmark is the base country, i.e., $d_{nDen}(q) = d_{nDen}(\lambda^j q) = d_{nDen}(a) = 1$ for all $j$ and $q$. Hence, the following results will be relative to Denmark’s iceberg trade cost. As we mentioned before we recover iceberg trade costs for every good in our sample. We represent the results for 3 different product categories out of 1268 having low, medium, high $A^j_n$ values. In Eaton and Kortum (2002) and variants trade cost estimates are not product specific, however according to our results they vary quite a bit. So one trade cost estimate for one country might not be very informative when you would like apply industry specific policies. For instance for Belgium we observe that for goods that have a low value of total demand (low $A^j_n$) Belgium has the lowest trade cost whereas for high values she loses this advantage. Moreover, when interpreting $d_{nDen}^j$, iceberg trade cost, estimates one should always be careful as they might also be capturing some features of market structure.

<table>
<thead>
<tr>
<th>Product</th>
<th>Wood charcoal, agglomerated or not</th>
<th>Inner tubes of tires</th>
<th>Printing paper and writing paper, in rolls or sheets; uncoated</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A^j_n$ Value</td>
<td>9,186,338</td>
<td>29,807,245</td>
<td>216,135,946.03</td>
</tr>
<tr>
<td>Australia</td>
<td>1.2458</td>
<td>1.3164</td>
<td>1.3684</td>
</tr>
<tr>
<td>Austria</td>
<td>1.293</td>
<td>1.1454</td>
<td>1.195</td>
</tr>
<tr>
<td>Belgium-Lux</td>
<td>0.0092</td>
<td>0.4199</td>
<td>1.7596</td>
</tr>
<tr>
<td>Canada</td>
<td>1.0285</td>
<td>0.9111</td>
<td>0.9506</td>
</tr>
<tr>
<td>Denmark</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Finland</td>
<td>0.8873</td>
<td>0.7861</td>
<td>0.8201</td>
</tr>
<tr>
<td>France</td>
<td>0.1675</td>
<td>1.4649</td>
<td>4.5499</td>
</tr>
<tr>
<td>Greece</td>
<td>2.2627</td>
<td>2.0045</td>
<td>2.0913</td>
</tr>
<tr>
<td>Italy</td>
<td>0.9457</td>
<td>0.107</td>
<td>4.946</td>
</tr>
<tr>
<td>Japan</td>
<td>1.1604</td>
<td>1.0277</td>
<td>1.0724</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.0521</td>
<td>2.6809</td>
<td>11.3556</td>
</tr>
<tr>
<td>New Zealand</td>
<td>1.0432</td>
<td>1.6748</td>
<td>1.7646</td>
</tr>
<tr>
<td>Norway</td>
<td>0.9142</td>
<td>0.8099</td>
<td>0.845</td>
</tr>
<tr>
<td>Portugal</td>
<td>3.9351</td>
<td>3.4846</td>
<td>3.637</td>
</tr>
<tr>
<td>Spain</td>
<td>1.6162</td>
<td>1.4318</td>
<td>1.4938</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.9417</td>
<td>0.8349</td>
<td>0.8295</td>
</tr>
<tr>
<td>UK</td>
<td>0.0908</td>
<td>0.5693</td>
<td>3.2624</td>
</tr>
<tr>
<td>US</td>
<td>0.9051</td>
<td>0.8018</td>
<td>0.8365</td>
</tr>
</tbody>
</table>
7 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 and trade costs we recover are sector specific which provides important insights about sectoral productivities and trade costs. Trade costs are never fully observed in the data and our methodology enables us to obtain the trade costs nonparametrically. In addition to being destination-source country specific, trade costs we recover are sector specific which enables us to understand the heterogeneity in trade costs across sectors. [To be completed]
APPENDIX

Proof of Lemma 2

Recall that \( H_{ni}(p|q) = \Pr \left[ P_{n,i,t}^j \leq p \text{ and } P_{n,s,t}^j \leq P_{n,s,t}^j, s \neq i | Q_n = q \right] \) where \( s = 1, \ldots, N \).

Since \( P_{n,s,t}^j \)'s are independent across \( s \) by Assumption A1(i), \( H_{ni}(p|q) \) can be written as

\[
\begin{align*}
\tilde{H}_{ni}^j(p|q) &= \int_0^p \prod_{s \neq i} \left[ 1 - G_{ns}^j(\nu|q) \right] dG_{ni}^j(\nu|q) \\
&= \int_0^p \left( \prod_{s=1}^N \left[ 1 - G_{ns}^j(\nu|q) \right] \right) \frac{dG_{ni}^j(\nu|q)}{1 - G_{ni}^j(\nu|q)} \\
&= - \int_0^p \left[ \prod_{s=1}^N \left[ 1 - G_{ns}^j(\nu|q) \right] \right] d\ln \left[ 1 - G_{ni}^j(\nu|q) \right]
\end{align*}
\]

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

Now, let \( Y = \min_{s=1,\ldots,N} \{ P_{n,s,t}^j \} \). Note that we can write \( \Pr[Y \leq \nu|Q_n^j = q] = \Pr[P_{n,1,t}^j \leq \nu \text{ and } P_{n,s,t}^j \leq P_{n,s,t}^j, s \neq 1 | Q_n = q] + \cdots + \Pr[P_{n,1,t}^j \leq \nu \text{ and } P_{n,s,t}^j \leq P_{n,s,t}^j, s \neq N | Q_n = q] \). This follows from the fact that the probability that the minimum price is less than or equal to some \( \nu \) 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 \( \nu \). Recall that these probabilities are basically \( \tilde{H}_{ns}^j(\nu|q) \)'s where \( s = 1, \ldots, N \). Hence, we can write

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

\( \Pr[Y \leq \nu|Q_n^j = q] \) 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 (53), we obtain

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

Putting (54) into (52) we get

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

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, $d \ln[1 - G_{ni}^j (p|q)] = -[1 - \sum_{s=1}^{N} \tilde{H}_{ns} (p|q)]^{-1} d\tilde{H}_{ni}^j (p|q)$. 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. □
REFERENCES


