Aggregate Productivity Shocks, Zeros and Gravity

Rachidi Kotchoni* Bruno Larue†

Abstract

We modify the Helpman-Melitz-Rubinstein (HMR) model to account for firm-level productivity shocks with aggregate effects. Firms' productivities are drawn from the same distribution ex-ante, but are then subject to independent firm-specific shocks when the number of firms is assumed finite or country-specific shocks when the number of firms is assumed infinite. We propose three gravity model specifications that account for aggregate productivity shocks. Each model entails different corrections at the estimation stage. The most intuitive econometric results are achieved by the specification based on a finite number of firms subjected to both random productivity shocks and random trading costs. We also found that over time, the selection effect is increasingly being driven by aggregate productivity shocks. This is not surprising given that Galaix (2011) recently found that firm-level shocks for the 100 largest US firms account for one-third of variations in aggregate output.

Keywords: Aggregate Productivity Shocks – Gravity Model – Heterogeneity – Random Trade Costs – Sample Selection – Trade Flows.

*Visiting scholar, Université de Montréal. Email for contact: rachidi.kotchoni@gmail.com

†Corresponding Author: CREATE /Pavillon Paul-Comtois /2425, rue de l'Agriculture, Local#4417 /Université Laval /Québec (QC) G1V 0A6 /Phone: 418 656 2131 poste 5098 /Fax: 418 656 7821 /E-mail: bruno.larue@eac.ulaval.ca

1 Introduction

Helpman, Melitz and Rubinstein (2008) (henceforth HMR) developed an insightful theoretical model capable of explaining three important stylized facts about international trade. First, as in Melitz's (2003) classic model, firms are heterogenous when it comes to their productivity and capacity to export. This allows for the possibility of no trade occurring between a given country i and another country j and it permits to identify the determinants of the selection of countries into trading relationships. Second, the model allows for asymmetric trade flows between country pairs. In particular, the model has the potential to explain why country i exports to, but does not import from country j. Third, the model generates a gravity equation in which distance and GNPs along with other variables condition positive trade flows. The resulting empirical gravity equation features a sample selection correction (involving the inverse Mills ratio) which is found to be less important than a correction for unobserved firm heterogeneity (HMR, p.471).

In the HMR model, each country j in the world is endowed with a continuum of firms of measure N_j. Firms are heterogenous in productivity, and their distribution across different productivity levels is described by a truncated Pareto measure. At equilibrium, the model implies that a fraction π_{ij} of the firms of country j export to country i while allowing for the possibility that $\pi_{ij} = 0$. This is an important feature of trade data that previous models with fully symmetric firms could not explain. Still, it also means that the number of exporting firms is infinite when $\pi_{ij} > 0$, in contrast with the fact that the total number of firms is not very large for the majority of industries within a given country. Several industries are characterized by natural entry barriers such as large sunk investments that limit the number of players. For instance, a large proportion of international trade in cars is dominated by a few companies and the same can be said about trade in aircraft, processed agricultural products and electronics. Two firms, Samsung and Hyundai, jointly account for 35% of South Korea's exports (Gabaix 2011, p.784). As pointed out by Eaton, Kortum and Sotelo (2012) "*it is hard to reconcile the small (sometimes zero) number of firms engaged in selling from one country to another with a continuum.*"

The HMR model posits a spatial distribution for firms productivity while being mute regarding their temporal behaviour. At first glance, this spatial distribution is consistent with two radically different temporal behaviour. In the first case, each firm is endowed with a constant productivity level so that when it exports, it does so consistently over time (everything else equal). In the second case, firms switch across different productivity levels randomly over time in a manner that keeps their spatial distribution unaltered. However, the latter interpretation may not allow for the occurrence of zero aggregate trade flows when the number of candidates exporting firms is infinite. We are thus left with the first interpretation, which implies a chain of competitive advantage such that if aggregate exports from country j to country i are consistently positive, it must be the case that the most productive firm in country j is exporting consistently. Failure to export on the part of this firm entails that no other firm from country j can export to country i. However, there is empirical evidence that firms' productivity vary over time and that there is much " ins and outs" at the firm level as well. Using establishments and enterprise data for Canada, Sabuhoro, Larue and Gervais (2006) found that for one-third of all establishments the length of an export episode does not exceed one month and that firms learn from past failures as the number of past exits increases export survival. They also found that the hazard of exiting foreign markets varies negatively and significantly with the relative size of the exporting establishment and the number of exported products" . The last point is also documented by Bernard, Redding and Schott (2011). Still, even very large exporting firms with a history of high productivity are not immune to negative productivity shocks and may be forced to exit. The US import refusal on September 4 of 2012 of a shipment of beef contaminated with E. coli from Canada's XL Foods led to the largest meat recall in Canada's history, the temporary closure of the plant and a highly publicized break in its exports.

If the number of candidate exporting firms is infinite, introducing some randomness in firms productivity is irrelevant for determining aggregate trade flows as long as firms behave independently. This is what Eaton, Kortum and Sotelo (2012) meant by saying that " *shocks to individual firms can never have an aggregate effect.*" For the shocks on individual firms' productivity to have observable implications on aggregate trade flows, one must either allow for the number of firms to be finite while subject to random productivity shocks or allow for a common shock to affect the productivity of all firms.[1] In either case, the result is an aggregate productivity shock affecting aggregate trade flows. This paper examines the theoretical and empirical implications of the presence of such aggregate productivity shocks in the HMR model.

We assume that each country in the world is endowed with N_j firms that are homogenous ex-ante with respect to the distribution of productivity, but heterogenous ex-post regarding realized productivity. Productivity shocks at the firm level may be justified by unforeseen equipment plants failures and human errors. We assume that the ex-ante distribution of the productivity index is given by the truncated Pareto measure as in the original HMR model for the spatial distribution of firms. Consequently, realized trade flows ($\widetilde{M}_{ij}$) between country pairs are random but expected trade flows ($E(\widetilde{M}_{ij}) = M_{ij}$) are equal to the expressions that one obtains when the number of firms per country is infinite (i.e., the one given in HMR, 2008). The realized export from country j to country i is given by $\widetilde{M}_{ij} = M_{ij}\widetilde{R}_{ij}$, where M_{ij} is the expected value of $\widetilde{M}_{ij}$ and $\widetilde{R}_{ij}$ is a unit mean aggregate shock implied by firm-level productivity shocks. The size of M_{ij} depends on the respective GNPs of i and j, the distance between i and j, the fixed and variable costs for exporting from j to i as well as some other country-specific fixed effects. HMR specified the export costs as random conditional on country-pair specific regressors, which further implies that $M_{ij} = E(M_{ij})\, R_{ij}$, where R_{ij} is another error term with unit mean. Hence, realized trade flows are of the form $\widetilde{M}_{ij} = E\left(\widetilde{M}_{ij}\right)\varepsilon_{ij}$ with $\varepsilon_{ij} = R_{ij}\widetilde{R}_{ij}$. An alternative approach to introduce an aggregate shock of type $\widetilde{R}_{ij}$ in the HMR model exposes the productivity of each firm to an aggregate shock common to all firms located in a given country (as in Eaton and Kortum, 2002) and an idiosyncratic shock specific to the firm. In both case, we attempt to assess how important the aggregate productivity shock $\widetilde{R}_{ij}$ is in determining the country pairs (i, j) that effectively trade.

Our theoretical framework supports three different model specifications. The first model assumes a finite number of firms subject to aggregate productivity shocks and random trading costs. This yields $\widetilde{M}_{ij} = E\left(\widetilde{M}_{ij}\right) R_{ij}\widetilde{R}_{ij}$, where the distributions of R_{ij} and $\widetilde{R}_{ij}$ have a point mass at zero. The second model assumes an infinite number of firms subject to aggregate productivity shocks and random trading costs. In this case, the distribution of $\widetilde{R}_{ij}$ lies on a strictly positive support so that the selection of countries into trading relationships is only driven by the shocks on trading costs (R_{ij}), as in HMR. Intuitively, if the most productive firm (among an infinite number of firms that behave independently) cannot export to a given destination, it must be that trading costs are prohibitive and the aggregate trade flow is zero. The third model assumes that the number of firms is finite and trading costs are non random so that $\widetilde{M}_{ij} = E\left(\widetilde{M}_{ij}\right)\widetilde{R}_{ij}$. An interesting feature of this case is that the sample selection and heterogeneity effect cancel each other out from the estimating equation. As a result, this model may be estimated without having to run a first-step Probit for the extensive margin.

Our empirical application relies on the HMR world trade data (available on Elhanan Helpman's website). We estimate the models separately for the years 1980 and 1989 in order to gauge the dynamics of elasticities. We find supportive evidence for the model that assumes a finite number of

[1] Gabaix (2011) contends that firm-specific shocks do not cancel each other and that volatility affecting the largest 100 US firms account for 33% of variations in US output. This suggests that firm-specific shocks can have an impact on aggregate trade flows.

firms, aggregate productivity shocks and random trading costs. The results suggest that on average, aggregate productivity shocks contributed up to 51% of the sample selection effect in 1980 and up to 57% in 1989. The model with finite number of firms, aggregate productivity shocks and random trading costs also delivers more intuitive results compared to competing models. According to this model, the negative effect of the log-distance and the disadvantage of island countries to trade have eroded while the advantage granted by a common land border and the disadvantage of landlocked countries to trade have remained stable during the 80's. Also, common currency union zones and free trade agreements have had bigger positive impacts on trade in 1989 compared to the beginning of the decade.

The remainder of the paper is organized as follows. Section 2 revisits the original Helpman-Melitz-Rubinstein model. A version of the HMR model with a finite number of firms is presented in Section 3. In Section 4, we discuss alternative approaches to introduce aggregate productivity shocks in trade flows without having to assume that the number of firms is finite. In Section 5, we develop three special cases from our theoretical framework that account for aggregate productivity shocks, HMR's heterogeneity and sample selection issues. In Section 6, we derive feasible estimating equations. Section 7 presents our empirical application and Section 8 concludes.

2 Revisiting the Helpman-Melitz-Rubinstein Model

HMR assumes that there are J countries in the world, and each country produces and consumes a continuum of products. The demand of country j for product l is given by:

$$x_j\left(l\right) = \frac{p_j\left(l\right)^{-\varepsilon} Y_j}{P_j^{1-\varepsilon}}, \tag{1}$$

where $p_j\left(l\right)$ is the price at which product l sells in country j, P_j and Y_j are respectively the ideal price index and the income of country j, and $\varepsilon > 1$ is the elasticity of substitution across products. The price index P_j is defined as:

$$P_j = \left[\int_{l\in B_j} p_j\left(l\right)^{1-\varepsilon} dl\right]^{1/(1-\varepsilon)}, \tag{2}$$

where B_j is the set of goods available for consumption in country j.

Each firm of the world economy produces a distinct product, and each country j has a continuum of firms of measure N_j. A country-j firm produces one unit of output with a combination of inputs whose value is $c_j a$, where a is the number of inputs used and c_j is the price of that input. The productivity, $\frac{1}{a}$, is firm-specific while the factor price, c_j, is country-specific. According to these assumptions, a country-j firm with productivity $\frac{1}{a}$ maximizes its profits by setting a domestic price equal to $q_{jj}\left(a\right) = \frac{c_j a}{\alpha}$, where $\alpha \equiv \frac{\varepsilon - 1}{\varepsilon}$. When the same firm exports from country j to country i, its price is $q_{ij}\left(a\right) = \tau_{ij}\frac{c_j a}{\alpha}$, where $\tau_{ij} > 1$ reflects all variable costs necessary to deliver the firm's product to the importing country. This variable cost may include transport cost as well as other costs induced by trade resistance factors. At the price q_{ij}, the demand of the remote market for a product from country j is given by:

$$x_{ij}\left(a\right) = \left(\tau_{ij}\frac{c_j a}{\alpha}\right)^{-\varepsilon} \frac{Y_i}{P_i^{1-\varepsilon}}. \tag{3}$$

There are also fixed costs $c_j f_{ij}$ incurred by an exporting firm for serving country i. Thus the profit from exporting a product from j to i is:

$$\pi_{ij}\left(a\right) = \left(1-\alpha\right)\left(\frac{\tau_{ij} c_j a}{\alpha P_i}\right)^{1-\varepsilon} Y_i - c_j f_{ij} ,$$

where the fixed export costs are defined by $c_j f_{ij}$. The minimum productivity required for a country-j firm to be able to export to country i is implicitly defined by $\pi_{ij}(a_{ij}) = 0$. Rearranging, we obtain:

$$a_{ij} = \frac{\alpha P_i}{c_j \tau_{ij}} \left(\frac{c_j f_{ij}}{(1-\alpha) Y_i} \right)^{1/(1-\varepsilon)}. \tag{4}$$

In order to determine the volume of importation of country i from country j, one needs to know the relative proportion of each type of firms within the exporting country. HMR assume that these proportions are the same for all countries and are described by a truncated Pareto distribution with support $[a_L, a_H]$. The cumulative distribution function of the heterogeneity index a is thus given by:

$$\Pr(a \le x) \equiv G(x) = \frac{x^k - a_L^k}{a_H^k - a_L^k}, a_L \le x \le a_H, \tag{5}$$

where $k > \varepsilon - 1$. However, these proportions must be multiplied by N_j in order to reflect the relative size of the different economies. Accordingly, the value of the imports of country i from country j is:

$$\begin{aligned} M_{ij} &\equiv N_j \int_{a_L}^{a_{ij}} x_{ij}(a)\, q_{ij}(a)\, dG(a) \\ &= N_j \left(\frac{\tau_{ij} c_j}{\alpha P_i} \right)^{1-\varepsilon} Y_i V_{ij}, \end{aligned} \tag{6}$$

where

$$V_{ij} = \begin{cases} \int_{a_L}^{a_{ij}} a^{1-\varepsilon} dG(a) & \text{for } a_{ij} > a_L \\ 0 & \text{otherwise.} \end{cases}$$

An explicit calculation of V_{ij} yields $V_{ij} = \frac{k a_L^{k-\varepsilon+1}}{(k-\varepsilon+1)\left(a_H^k - a_L^k\right)} W_{ij}$, where

$$W_{ij} = \max\left\{ \left(\frac{a_{ij}}{a_L} \right)^{k-\varepsilon+1} - 1, 0 \right\}. \tag{7}$$

Overall, the HMR model generalizes Anderson and van Wincoop's (2003) in two respects. First, it highlights the extensive margin of trade flows and its relationship to firm heterogeneity and fixed trade costs. Second, it accounts for the fact that the volume of exports from j to i is potentially different from the volume of exports from i to j. However, a maintained assumption in the HMR model is that each country j in the world has a continuum of firms of measure N_j. A direct consequence of this assumption is that the number of country-j firms exporting to country i is either zero (when no firm is qualified to export to i) or infinite (when a proportion $G(a) > 0$ of firms are qualified to export to i). But in the real world, there are at least two factors that cause most exporting industries not to involve a large number of players. First, large sunk investment playing as entry barriers often limits the number of candidates exporting firms. Second, a firm that is qualified to export to a remote market still faces a high probability of exit within a relatively short length of time (Sabuhoro, Larue and Gervais, 2006). In the next section, we accommodate these empirical facts by re-formulating the HMR model for a finite number of firms facing random productivity shocks.

3 The HMR Model with a Finite Number of Firms

The measure $G(a)$ that gives the spatial distribution of the productivity index in (5) puts nonzero weights on firms that do no export at all. An alternative approach to compute trade flows might account for the a priori knowledge that only firms for which $a \le a_{ij}$ are able to export. In this case, one would compute V_{ij} using the distribution of the index a within the subset of exporting firms, which is given by:

$$G_{ij}(x) \equiv \Pr\left(a \leq x | a_L \leq a \leq a_{ij}\right) = \frac{x^k - a_L^k}{a_{ij}^k - a_L^k}.$$

While $G(a)$ is the unconditional distribution of a within each country (common to all countries), $G_{ij}(x)$ is the distribution of a in country j conditional on the firm actually exporting to country i. Hence although countries are homogenous ex ante regarding the distribution $G(a)$, they are heterogenous ex post with respect to the conditional distribution $G_{ij}(x)$. Using $G_{ij}(x)$ to compute trade flows yields:

$$M_{ij} = N_{ij} \frac{1}{G\left(a_{ij}\right)} \left(\frac{\tau_{ij} c_j}{\alpha P_i}\right)^{1-\varepsilon} Y_i V_{ij}, \tag{8}$$

where N_{ij} is the measure of firms that export from j to i.

Equation (8) states that the aggregate trade flow M_{ij} is equal to the measure of the firms that export, N_{ij}, times the average value of exports of a representative exporting firm $\frac{1}{G(a_{ij})}\left(\frac{\tau_{ij} c_j}{\alpha P_i}\right)^{1-\varepsilon} Y_i V_{ij}$. In turn, this average value has two components. The first component, the ratio $\frac{1}{G(a_{ij})}$, is interpreted as the extensive margin component because it depends solely on the proportion of exporting firms $G\left(a_{ij}\right)$. The second component, $\left(\frac{\tau_{ij} c_j}{\alpha P_i}\right)^{1-\varepsilon} Y_i V_{ij}$, may be attributed to the intensive margin as it relates to the volume of trade of firms already involved in trade. The latter component is increasing in the proportion of firms that export because:

$$V_{ij} = \frac{k a_L^{k-\varepsilon+1}}{(k-\varepsilon+1)\left(a_H^k - a_L^k\right)} \max\left\{\left(\frac{a_H^k - a_L^k}{a_L^k} G\left(a_{ij}\right) + 1\right)^{\frac{k-\varepsilon+1}{k}} - 1, 0\right\}. \tag{9}$$

The expressions of M_{ij} in (6) and in (8) are reconciled if and only if $N_{ij} = N_j G\left(a_{ij}\right)$. The latter equality necessarily holds in the HMR model because firms are in infinite number while each firm is deterministically identified by its productivity level. As an implication from Equation (6), keeping $\left(\frac{\tau_{ij} c_j}{\alpha P_i}\right)^{1-\varepsilon}$ and Y_i constant, the only way for country j to increase its export to country i is to increase V_{ij} through the extensive margin $G\left(a_{ij}\right)$. This would be the case for example if the productivity of country-j firms had improved exogenously. In this case, an increase in trade volume arises only from the emergence of new trading relations at the firm level. By contrast, if $\left(\frac{\tau_{ij} c_j}{\alpha P_i}\right)^{1-\varepsilon}$ and $G\left(a_{ij}\right)$ are kept constant, an increase in trade volume may arise following an increase of the GNP Y_i of the importing country. In this case, a fixed number of country j exporting firms must produce more intensively in order to respond to a higher level of demand in country i.

The equation $N_{ij} = N_j G\left(a_{ij}\right)$ does not hold exactly in a context where the number of firms is finite and a firm's productivity is random. To see this, let each country j have a finite number N_j of firms. At the beginning of each period t, a firm perceives its own productivity as a random variable $1/a$, where a follows the truncated Pareto distribution (5), as in HMR. In mature industries, the technology is known to all and is readily available. However, even a mastered technology may give rise to random productivity because of uncertainties about the marginal product of certain inputs (absenteeism, equipment failures). Hence, we shall assume that each firm knows whether it is able to export or not only after observing the realization of its productivity a and the threshold a_{ij} at the end of the period. Under these assumptions, the observed aggregate trade flow from country j to country i is given by:

$$\widetilde{M}_{ij} = \sum_{k=1}^{N_j} x_{ij}\left(a_{(kj)}\right) q_{ij}\left(a_{(kj)}\right) 1\left(a_{(kj)} \in \left[a_L, a_{ij}\right]\right), \tag{10}$$

where quantities and prices $x_{ij}(a_k)$ and $q_{ij}(a_k)$ are defined as in Section 2, $1\left(a_{(kj)} \in [a_L, a_{ij}]\right) = 1$ if $a_{(kj)} \in [a_L, a_{ij}]$ and $1\left(a_{(kj)} \in [a_L, a_{ij}]\right) = 0$ otherwise.

The current modified HMR model has three main implications. First, the realized trade flow (10) is random and it satisfies

$$\begin{aligned} \widetilde{M}_{ij} &= Y_i \left(\frac{\tau_{ij} c_j}{\alpha P_i^{1-\varepsilon}}\right)^{1-\varepsilon} \sum_{k=1}^{N_j} a_{(kj)}^{1-\varepsilon} 1\left(a_{(kj)} \in [a_L, a_{ij}]\right) \text{ and} \\ E\left(\widetilde{M}_{ij}\right) &= N_j Y_i \left(\frac{\tau_{ij} c_j}{\alpha P_i}\right)^{1-\varepsilon} V_{ij} \equiv M_{ij}, \end{aligned}$$

where M_{ij} is given by Equation (6). Second, the realized number of exporting firms, $N_{ij} = \sum_{k=1}^{N_j} 1\left(a_{(kj)} \in [a_L, a_{ij}]\right)$, is also random and it satisfies $E(N_{ij}) = N_j G(a_{ij})$. Third, $G(a_{ij})$ is no longer the proportion of firms that trade. It now denotes the probability of trade at the firm level. This means that the ex-post number of firms involved in trade (i.e., the realization of N_{ij}) can be zero even though $G(a_{ij})$ is strictly positive. Firms that are expected to export may not do so because of adverse productivity shocks. Indeed, N_{ij} follows the Binomial distribution $B(N_j, G(a_{ij}))$ and the probability that country i imports from j is given by:

$$\Pr(N_{ij} > 0) = 1 - (1 - G(a_{ij}))^{N_j}. \tag{11}$$

The latter implication relates our framework to the balls-and-bins model of Armenter and Koren (2012). These authors propose a disaggregated trade model where shipments are treated as balls that are randomly assigned to bins. The balls are identified by the HS classification in the origin country while the bins are labelled by the classification in the destination country. Each ball has probability s_{ij} of originating from category i and landing into the category j. An interesting feature of Armenter and Koren's model is that a bin may end up being empty even though its ex-ante probability of receiving balls is strictly positive.

Eaton, Kortum and Sotelo (2012) considered a model where the number of country j firms achieving at least a given productivity level a is generated by a Poisson distribution with intensity $\mu_j(a)$, where $\mu_j(.)$ is a continuous measure on the support of a. Here, we follow a different route that is rather closer in spirit to the original HMR (2008) model.

Keeping in mind that $\widetilde{M}_{ij}$ is observed while its expectation M_{ij} is not, we define the multiplicative error $\widetilde{R}_{ij}$ as:

$$\widetilde{R}_{ij} \equiv \begin{cases} \frac{\widetilde{M}_{ij}}{M_{ij}} \text{ if } M_{ij} > 0, \text{ and} \\ 0, \text{ otherwise.} \end{cases}, \tag{12}$$

so that by construction, $\widetilde{M}_{ij} = M_{ij}\widetilde{R}_{ij}$. Replacing the expressions of $\widetilde{M}_{ij}$ and M_{ij} into (12) yields:

$$\widetilde{R}_{ij} = \begin{cases} \frac{1}{N_j V_{ij}} \sum_{k=1}^{N_j} a_{(kj)}^{1-\varepsilon} 1\left(a_{(kj)} \in [a_L, a_{ij}]\right) \text{ if } V_{ij} > 0, \text{ and} \\ 0, \text{ otherwise.} \end{cases}. \tag{13}$$

Hence, realized trade flows are random even when trading costs are constant over time. An important property of $\widetilde{R}_{ij}$ following from (11) is that

$$\lim_{N_j \to \infty} \Pr\left(\widetilde{R}_{ij} > 0\right) = \lim_{N_j \to \infty} \Pr(N_{ij} > 0) = 1,$$

which means that trade flows from j to i eventually become strictly positive as the economy of country j grows.

4 Other Sources of Randomness in Aggregate Trade Flows

Reformulating the HMR model with a finite number of firms that are subject to random productivity shocks allows us to introduce in an intuitive manner a multiplicative aggregate shock $\widetilde{R}_{ij}$ with unit mean in the gravity equation. However, if the productivity shocks are independent across firms, then the probability of $\widetilde{R}_{ij} = 0$ shrinks to zero as the number of candidate exporting firms increases to infinity. Furthermore, a law of large number applies such that:

$$\widetilde{R}_{ij} = \frac{1}{N_j V_{ij}} \sum_{k=1}^{N_j} a_{(kj)}^{1-\varepsilon} 1 \left(a_{(kj)} \in [a_L, a_{ij}] \right) \to 1 \text{ as } N_j \to \infty. \tag{14}$$

Hence, as N_j increases to infinity, $\widetilde{R}_{ij}$ vanishes as a multiplicative shock, as argued by Eaton, Kortum and Sotelo (2012). Based on this observation, one might argue that the aggregate productivity shock is irrelevant for large economies. However, other plausible factors may cause realized trade flows to be random even when firms are in infinite number.

First, note that the threshold productivity index a_{ij} depends on the price index and income of country i as well as on the export costs incurred by country j firms (see Equation 4). Hence, a_{ij} is directly affected by the macroeconomic fluctuations of the origin and destination countries. For instance, HMR derived their estimating equation by positing that trading costs are random. According to this assumption, a_{ij} and M_{ij} are also random and thus:

$$\widetilde{M}_{ij} = E\left(M_{ij}\right) R_{ij} \widetilde{R}_{ij}, \tag{15}$$

where $R_{ij} \equiv \frac{M_{ij}}{E(M_{ij})}$ and $\widetilde{R}_{ij} \equiv \frac{\widetilde{M}_{ij}}{M_{ij}}$. As N_j increases to infinity, $\widetilde{R}_{ij}$ collapses to one so that R_{ij} becomes the limiting error term. An equation like (15) can also be derived if we assume that the destination country i is subject to aggregate macroeconomic shocks so that its income Y_i is random.

Second, the productivity of the candidates exporting firms may also depend on domestic macroeconomic shocks. For example, Eaton and Kortum (2002) posited that country j's efficiency in producing good k is the realization of a random variable Z_j, where the distribution of Z_j is country-specific, independent and identically distributed across goods. This assumption could be formalized by specifying the productivity index of a country j firm as:

$$b_{(kj)} = a_{(j)}^{\rho} a_{(kj)}^{1-\rho}, \tag{16}$$

where $a_{(j)}$ and $a_{(kj)}$ are independent Pareto random variables with support $[a_L, a_H]$, $a_{(j)}$ is common to all firms of country j and $a_{(kj)}$ is specific to the firm k of country j. For trade to occur, $b_{(kj)}$ must lie between the bounds a_L and a_{ij}. The corresponding bounds for the firm-specific shock $a_{(kj)}$ can be derived for a given common productivity shock $a_{(j)}$. Thus, trade flows are given by:

$$\widetilde{M}_{ij} = a_{(j)}^{\rho(1-\varepsilon)} \sum_{k=1}^{N_j} x_{ij}\left(a_{(kj)}\right) q_{ij}\left(a_{(kj)}\right) 1\left(a_{(kj)} \in [lb, ub]\right), \tag{17}$$

with $lb = \left(a_L / a_{(j)}^{\rho}\right)^{1/(1-\rho)}$ and $ub = \left(a_{ij} / a_{(j)}^{\rho}\right)^{1/(1-\rho)}$. As depicted by the bounds of $a_{(kj)}$, a strong common shock is compensated by a lower probability of export conditional on the realization of the common shock.

Let us examine the implications of (16) when trading costs are non-random. Then, straightforward calculations show that $E\left(\widetilde{M}_{ij}|a_{(j)}\right) = M_{ij}a_{(j)}^{-\rho k/(1-\rho)}$, where:

$$
\begin{aligned}
M_{ij} &= N_j\left(\frac{\tau_{ij}c_j}{\alpha P_i}\right)^{1-\varepsilon} Y_i V_{ij}, && (18)\\
V_{ij} &= \frac{\frac{k}{1-\rho}\left(a_L\right)^{\frac{k}{1-\rho}+1-\varepsilon}}{\left(a_H^k - a_L^k\right)\left[\frac{k}{1-\rho}+1-\varepsilon\right]} W_{ij} \text{ and}\\
W_{ij} &= \max\left\{\left(\frac{a_{ij}}{a_L}\right)^{\frac{k}{1-\rho}+1-\varepsilon} - 1, 0\right\}. && (19)
\end{aligned}
$$

Hence, observed trade flows can be written as $\widetilde{M}_{ij} = E\left(\widetilde{M}_{ij}\right)\widetilde{R}_{ij}$, where:

$$
E\left(\widetilde{M}_{ij}\right) = \frac{a_H^{k-\rho k/(1-\rho)} - a_L^{k-\rho k/(1-\rho)}}{\left(a_H^k - a_L^k\right)\left(1-\rho/(1-\rho)\right)} M_{ij},
$$

and $\widetilde{R}_{ij} = \widetilde{R}_{1,ij}\widetilde{R}_{2,ij}$ with:

$$
\begin{aligned}
\widetilde{R}_{1,ij} &= \frac{\widetilde{M}_{ij}}{E\left(\widetilde{M}_{ij}|a_{(j)}\right)} = \begin{cases} \frac{a_{(j)}^{\rho(k/(1-\rho)+1-\varepsilon)}}{N_j V_{ij}} \sum_{k=1}^{N_j} a_{(kj)}^{(1-\varepsilon)(1-\rho)} 1\left(a_{(kj)} \in [lb, ub]\right) & \text{if } V_{ij} > 0, \\ 0, \text{ otherwise.} \end{cases} \text{ and} && (20)\\
\widetilde{R}_{2,ij} &\equiv \frac{E\left(\widetilde{M}_{ij}|a_{(j)}\right)}{E\left(\widetilde{M}_{ij}\right)} = \frac{\left(a_H^k - a_L^k\right)\left(1-\rho/(1-\rho)\right)}{a_H^{k-\rho k/(1-\rho)} - a_L^{k-\rho k/(1-\rho)}} a_{(j)}^{-\rho k/(1-\rho)}. && (21)
\end{aligned}
$$

This decomposition is of $\widetilde{R}_{ij}$ is rationalized on the ground that as the number of firms increases to infinity, $\widetilde{R}_{1,ij}$ converges to one in probability and $\widetilde{R}_{ij}$ reduces to $\widetilde{R}_{2,ij}$.

In summary, trade flows are likely to be described by Equation (15), where R_{ij} captures the joint effect of the randomness of trading costs and the background risk of the importing country while $\widetilde{R}_{ij}$ captures the joint effect of the finiteness of the number of firms and the background risk of the exporting country.

5 Heterogeneity and Sample Selection

From our previous discussion, the HMR model can be modified in three different directions, each leading to a particular econometric model.

5.1 Finite number of firms, non-random trading cost and aggregate productivity shocks

When trading costs are non-random, realized trade flows can be written as $\widetilde{M}_{ij} = M_{ij}\widetilde{R}_{ij}$,where $M_{ij} = E\left(\widetilde{M}_{ij}\right)$ and $\widetilde{R}_{ij} = \widetilde{R}_{1,ij}\widetilde{R}_{2,ij}$, with $\widetilde{R}_{1,ij}$ and $\widetilde{R}_{2,ij}$ given by (20) and (21). By construction, $E\left(\widetilde{R}_{ij}\right) = 1$.

If we had to estimate this model in multiplicative form by using only observations with strictly positive trade flows, then we would have to account for the fact that the average value of $\widetilde{R}_{ij}$ is inflated on this subsample. Indeed:

$$
E\left(\widetilde{M}_{ij}|\widetilde{R}_{ij} > 0\right) = M_{ij}E\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right).
$$

Hence, we have:

$$\widetilde{M}_{ij} = M_{ij} E\left(\widetilde{R}_{ij} | \widetilde{R}_{ij} > 0\right) \left(\frac{\widetilde{R}_{ij}}{E\left(\widetilde{R}_{ij} | \widetilde{R}_{ij} > 0\right)} \right), \tag{22}$$

where $\widetilde{R}_{ij}/E\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right)$ is an error term with unit mean on the subsample where the realizations of $\widetilde{R}_{ij}$ are strictly positive. From Equation (22), we see that $E\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right)$ is a multiplicative sample selection bias correction term. Note that:

$$E\left(\widetilde{R}_{ij}\right) = E\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right) \Pr\left(\widetilde{R}_{ij} > 0\right) = 1,$$

and hence, $E\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right) = \Pr\left(\widetilde{R}_{ij} > 0\right)^{-1}$.

Next, we consider applying the logarithm to both sides of Equation (22). We have:

$$\log \widetilde{M}_{ij} = \log\left(M_{ij} E\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right)\right) + \log\left(\widetilde{R}_{ij}/E\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right)\right).$$

The log-error term $\log\left(\widetilde{R}_{ij}/E\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right)\right)$ is not centered on the domain $\left\{\widetilde{R}_{ij} > 0\right\}$. A centered error term is obtained by rewriting:

$$\begin{aligned} \log \widetilde{M}_{ij} &= \log\left(M_{ij} E\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right)\right) \\ &\quad + E\left(\log \widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right) - \log E\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right) + \widetilde{r}_{ij}, \end{aligned} \tag{23}$$

where $\widetilde{r}_{ij} = \log \widetilde{R}_{ij} - E\left(\log \widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right)$ has zero mean and $E\left(\log \widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right) - \log E\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right)$ captures the residual heterogeneity. If $\widetilde{R}_{ij}$ were IID, then the latter term would be constant across (i, j) and its omission from the estimating equation would only affect the estimate of the intercept. But realistically, we cannot expect trade data to be homoscedastic for reasons that are discussed at length in Santos Silva and Tenreyro (2006). Hence, this residual heterogeneity term is likely to be country pair-specific and its omission from the estimating equation would be harmful for slopes coefficient estimates as well. In an hypothetical world where each country exports to all other countries, the sample selection effect vanishes and the residual heterogeneity reduces to $E\left(\log \widetilde{R}_{ij}\right)$.

From Equation (18), we have:

$$\log M_{ij} = \beta_0 + \chi_i + \widetilde{\chi}_j - (\varepsilon - 1) \log \tau_{ij} + \log W_{ij}, \tag{24}$$

where β_0 gathers all constant terms and:

$$\begin{aligned} \chi_i &= \log Y_i - (1 - \varepsilon) \log P_i, \\ \widetilde{\chi}_j &= \log N_j + (1 - \varepsilon) \log c_j \text{ and} \end{aligned}$$

The term $\log W_{ij}$ captures the heterogeneity in log-expected trade flows. In practice, the fixed effects $\left(\chi_i, \chi_j, \log \tau_{ij}\right)$ are proxied by a set of regressors x_{ij}. Replacing into the expression of log-realized trade flows yields a HMR-like trade flow equation even though trading cost are not random:

$$\log \widetilde{M}_{ij} = x_{ij}\beta + \log W_{ij} + E\left(\log \widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right) + \widetilde{r}_{ij}, \tag{25}$$

where x_{ij} includes importer and exporter dummy variables.

5.2 Finite number of firms, random trading cost and aggregate productivity shocks

When trading costs are random and aggregate productivity shocks are present, realized trade flows must be written as in (15), that is, $\widetilde{M}_{ij} = E(M_{ij}) R_{ij} \widetilde{R}_{ij}$ where the definitions of M_{ij} and $\widetilde{R}_{ij}$ are the same in the previous subsection and $R_{ij} = \frac{M_{ij}}{E(M_{ij})}$. In fact, M_{ij} inherits its randomness from trading costs.

HMR (p.453) assumed that $\tau_{ij}^{\varepsilon-1} = D_{ij}^{\gamma} \exp(-u_{ij})$, where D_{ij} measures the distance between countries i and j (including other trade resistance factors) and $u_{i,j} \sim N(0, \sigma_u^2)$. Substituting into Equation (18) yields:

$$M_{ij} = \frac{\widetilde{k} a_L^{\widetilde{k}+1-\varepsilon}}{\left(a_H^k - a_L^k\right)\left(\widetilde{k}+1-\varepsilon\right)} N_j Y_i D_{ij}^{\gamma} \left(\frac{c_j}{\alpha P_i}\right)^{1-\varepsilon} \exp(-u_{ij}) W_{ij},$$

where $\widetilde{k} = \frac{k}{1-\rho}$. HMR (p.455) further assumed that the fixed input quantity required to export is specified as $f_{ij} = \exp\left(\phi_{Ex,j} + \phi_{IM,i} + \kappa\phi_{ij} - v_{ij}\right)$, where $\phi_{Ex,j}$ and $\phi_{IM,i}$ are respectively related to export costs and trade barriers imposed by the importing country, ϕ_{ij} includes any additional fixed trade costs tied to the country pair i and j and $v_{i,j} \sim N(0, \sigma_v^2)$. This assumption implies that W_{ij} is also random through the zero profit condition to export:

$$\left(\frac{a_{ij}}{a_L}\right)^{\varepsilon-1} = \frac{(1-\alpha)}{c_j}\left(\frac{\alpha P_i}{a_L c_j}\right)^{\varepsilon-1} Y_i D_{ij}^{-\gamma} \exp\left(-\phi_{Ex,j} - \phi_{IM,i} - \kappa\phi_{ij} + u_{ij} + v_{ij}\right). \tag{26}$$

Assuming that u_{ij} and v_{ij} are independent of $\widetilde{R}_{ij}$, we have:

$$\widetilde{M}_{ij} = \exp(x_{ij}\beta) E\left[\exp(-u_{ij}) W_{ij}\right] \epsilon_{ij}, \tag{27}$$

where x_{ij} is a set of regressors that include a constant,

$$\begin{aligned}
\exp(x_{ij}\beta) &\equiv \frac{k a_L^{\widetilde{k}-\varepsilon+1}}{\left(a_H^k - a_L^k\right)\left(\widetilde{k}-\varepsilon+1\right)} N_j Y_i D_{ij}^{\gamma} \left(\frac{c_j}{\alpha P_i}\right)^{1-\varepsilon}, \\
R_{ij} &\equiv \frac{M_{ij}}{E(M_{ij})} = \frac{\exp(-u_{ij}) W_{ij}}{E\left[\exp(-u_{ij}) W_{ij}\right]} \text{ and } \epsilon_{ij} \equiv R_{ij}\widetilde{R}_{ij}.
\end{aligned}$$

If we had to estimate the model (27) in multiplicative form by relying only on observations with positive trade flows, then we would have to re-normalize the error term as follows:

$$\widetilde{M}_{ij} = \exp(x_{ij}\beta) E\left(\exp(-u_{ij}) W_{ij}\right) E\left(\epsilon_{ij} | \epsilon_{ij} > 0\right) \left(\frac{\epsilon_{ij}}{E\left(\epsilon_{ij} | \epsilon_{ij} > 0\right)}\right), \tag{28}$$

where $\frac{\epsilon_{ij}}{E(\epsilon_{ij}|\epsilon_{ij}>0)}$ is an error term with unit mean on the domain $\{\epsilon_{ij} > 0\}$ and $E(\epsilon_{ij}|\epsilon_{ij} > 0)$ is the multiplicative sample selection bias correction term. We have

$$E\left(\epsilon_{ij} | \epsilon_{ij} > 0\right) = \Pr\left(R_{ij} > 0\right)^{-1} \Pr\left(\widetilde{R}_{ij} > 0\right)^{-1}.$$

Note that $\Pr(R_{ij} > 0)$ depends on the realizations of trading costs and it gives the probability of export possibilities arising for country j firms to destination i. Likewise, $\Pr\left(\widetilde{R}_{ij} > 0\right)$ depends on the realizations of firms productivity and more precisely on the probability that the most productive firm in country j is able to export to destination i. Hence, a firm might not export either because trading costs are prohibitive or because the realization of its productivity is too low.

Let us consider log-linearizing (28). We have:

$$\log \widetilde{M}_{ij} = x_{ij}\beta + \log E\left(\exp\left(-u_{ij}\right)W_{ij}\right) + \log E\left(\epsilon_{ij}|\epsilon_{ij} > 0\right) + \log\frac{\epsilon_{ij}}{E\left(\epsilon_{ij}|\epsilon_{ij} > 0\right)}.$$

In turn, we write the last term as:

$$\log\frac{\epsilon_{ij}}{E\left(\epsilon_{ij}|\epsilon_{ij} > 0\right)} = E\left(\log\epsilon_{ij}|\epsilon_{ij} > 0\right) - \log E\left(\epsilon_{ij}|\epsilon_{ij} > 0\right) + \widetilde{\epsilon}_{ij},$$

where $\log E\left(\epsilon_{ij}|\epsilon_{ij} > 0\right)$ is the sample selection bias correction term, $E\left(\log\epsilon_{ij}|\epsilon_{ij} > 0\right) - \log E\left(\epsilon_{ij}|\epsilon_{ij} > 0\right)$ is the residual heterogeneity and $\widetilde{\epsilon}_{ij}$ is a zero mean error.

Equation 6 of Santos Silva and Tenreyro (2009) implies that $E\left[\exp\left(u_{ij}\right)W_{ij}|x_{ij}\right] = \overline{W}\left(z^*_{ij}, \delta, r\right)$, where:

$$\overline{W}\left(z^*_{ij}, \delta, r\right) = \exp\left(\sigma^2_u/2\right)\left[\exp\left(\delta^2/2 + \delta r + \delta z^*_{ij}\right)\Phi\left(z^*_{ij} + \delta + r\right) - \Phi\left(z^*_{ij} + r\right)\right], \tag{29}$$

Φ is the CDF of the standard normal distribution, z^*_{ij} is a latent variable such that $\Phi\left(z^*_{ij}\right) = \Pr\left(R_{ij} > 0\right)$ and (δ, r) are ancillary parameters. By substituting this into the expression of $\log\widetilde{M}_{ij}$, we obtain:

$$\log \widetilde{M}_{ij} = x_{ij}\beta + \log\overline{W}\left(z^*_{ij}, \delta, r\right) + E\left(\log\epsilon_{ij}|\epsilon_{ij} > 0\right) + \widetilde{\epsilon}_{ij}, \tag{30}$$

where $\log\overline{W}\left(z^*_{ij}, \delta, r\right)$ captures the heterogeneity in log-expected trade flows.

5.3 Infinite number of firms and random trading cost

Recall that the error term stemming from aggregate productivity shocks is $\widetilde{R}_{ij} = \widetilde{R}_{1,ij}\widetilde{R}_{2,ij}$, where $\widetilde{R}_{1,ij}$ and $\widetilde{R}_{2,ij}$ given by (20) and (21). As the number of firms N_j increases to infinity, a law of large number applies to (20) so that $\widetilde{R}_{1,ij}$ converges in probability to unity. Hence, the limiting error term $\widetilde{R}_{2,ij}$ depicted by (21) is strictly positive. This is consistent with the fact that the probability of at least one firm being productive enough to export to any destination converges to one as the number of firms increases without bound:

$$\lim_{N_j\to\infty}\Pr\left(\widetilde{R}_{ij} > 0\right) = 1.$$

Put differently, the sample selection phenomenon is not affected by the productivity shocks at the limit, that is:

$$\lim_{N_j\to\infty} E\left(\epsilon_{ij}|\epsilon_{ij} > 0\right) = \Pr\left(R_{ij} > 0\right)^{-1}.$$

From here, two different approaches may be considered. In the first one, aggregate productivity shock exists, but do not contribute to the sample selection effect. This leads to consider the same model as (30):

$$\log \widetilde{M}_{ij} = x_{ij}\beta + \log\overline{W}\left(z^*_{ij}, \delta, r\right) + E\left(\log\epsilon_{ij}|\epsilon_{ij} > 0\right) + r_{ij}, \tag{31}$$

with the notable difference that here $\Phi\left(z^*_{ij}\right) = \Pr\left(R_{ij} > 0\right) = \Pr\left(\epsilon_{ij} > 0\right)$ while ϵ_{ij} is a mean preserving spread of R_{ij}. In the second approach, one simply assumes that there is no aggregate productivity shock at all so that $\epsilon_{ij} = R_{ij}$. In our framework, Model (31) is the closest to the spirit of the original HMR model. As in HMR, it assumes an infinite number of firms, random trading costs and no aggregate productivity shock. Unlike in HMR, it assumed that firms are homogenous ex-ante but heterogenous ex-post. However, the latter assumption has no observable implication because the number of firms is assumed infinite.

6 Empirical Specifications and Estimation

The log-linearized models (25), (30) and (31) are restricted to the subsample of observations with strictly positive trade flows, which raises a sample selection issue. Furthermore, the random term $\log \epsilon_{ij}$ is heterogenous and not centered around zero. If ϵ_{ij} is Gaussian, then the expectation of $\log \epsilon_{ij}$ depends on the mean and variance of ϵ_{ij}. In general, $E(\log \epsilon_{ij})$ depends on all the moments of ϵ_{ij} and hence, it is heterogenous whenever ϵ_{ij} is not identically distributed across country pairs. The biases arising from sample selection and heterogeneity in ϵ_{ij} are jointly corrected by including the term $E(\log \epsilon_{ij}|\epsilon_{ij} > 0)$ in the estimating equations. The sample selection contributes up to $\log E(\epsilon_{ij}|\epsilon_{ij} > 0)$ and the remainder, $E(\log \epsilon_{ij}|\epsilon_{ij} > 0) - \log E(\epsilon_{ij}|\epsilon_{ij} > 0)$, is the residual heterogeneity.

If not dealt with, this residual heterogeneity may constitute an important source of bias when estimating parameters from the log-linearized gravity (Santos Silva and Tenreyro, 2006). In light of this, we derive estimation equations for each of the models (25), (30) and (31) below.

6.1 Finite number of firms, non-random trade cost and aggregate productivity shocks

The suitable estimating equation for this case must be deduced from Equation (25):

$$\log \widetilde{M}_{ij} = x_{ij}\beta + \log W_{ij} + E\left(\log \widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right) + \widetilde{r}_{ij}.$$

This equation can be brought to data only after suitable proxies are found for W_{ij} and $E\left(\log \widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right)$.

For this purpose, we shall assume that the distribution of $\widetilde{R}_{ij}$ conditional on $\widetilde{R}_{ij} > 0$ is lognormal:

$$\log \widetilde{R}_{ij}|\widetilde{R}_{ij} > 0 \sim N\left(\mu_{ij}, \sigma_{ij}^2\right),$$

Under this assumption, $E(R_{ij}) = \Phi\left(z_{ij}^*\right)\exp\left(\mu_{ij} + \sigma_{ij}^2/2\right) = 1$, where z_{ij}^* is a latent variable that satisfies $\Phi\left(z_{ij}^*\right) = \Pr\left(\widetilde{R}_{ij} > 0\right)$. Hence:

$$\mu_{ij} \equiv E(\log R_{ij}|R_{ij} > 0) = -\log \Phi\left(z_{ij}^*\right) - \sigma_{ij}^2/2, \tag{32}$$

We further specify σ_{ij}^2 as:

$$\sigma_{ij}^2 = \log\left(1 + \exp\left(\theta_0 + \theta_1\left(x_{ij}\beta\right)\right)\right), \tag{33}$$

where θ_0 and θ_1 are scalar ancillary parameters. The lognormality assumption implies that $\widetilde{r}_{ij} \sim N\left(0, \sigma_{ij}^2\right)$ on the domain $\widetilde{R}_{ij} > 0$ and:

$$\begin{aligned} E\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right) &= \Phi\left(z_{ij}^*\right)^{-1} \text{ and} \\ Var\left(\widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right) &= \Phi\left(z_{ij}^*\right)^{-2}\exp\left(\theta_0 + \theta_1\left(x_{ij}\beta\right)\right). \end{aligned}$$

Replacing $\mu_{ij} \equiv E\left(\log \widetilde{R}_{ij}|\widetilde{R}_{ij} > 0\right)$ into (25) yields:

$$\log \widetilde{M}_{ij} = x_{ij}\beta + \log W_{ij} - \log \Phi\left(\widehat{z}_{ij}^*\right) - \sigma_{ij}^2/2 + \widetilde{r}_{ij}, \tag{34}$$

where $W_{ij} = \left(\frac{a_{ij}}{a_L}\right)^{\frac{k}{1-\rho}+1-\varepsilon} - 1$ is strictly positive whenever $\widetilde{R}_{ij} > 0$ and:

$$\Phi\left(z_{ij}^*\right) = \frac{k a_L^{2k}\left(\left(\frac{a_H}{a_L}\right)^{k-\rho k/(1-\rho)} - 1\right)}{\left(a_H^k - a_L^k\right)^2\left(k - \rho k/(1-\rho)\right)}\left(\left(\frac{a_{ij}}{a_L}\right)^{k/(1-\rho)} - 1\right).$$

The term W_{ij} can be expressed as a function of $\Phi\left(z_{ij}^*\right)$ as:

$$W_{ij} = \left(1 + \frac{\left(a_H^k - a_L^k\right)^2 (k - \rho k/(1-\rho))}{k a_L^{2k}\left(\left(\frac{a_H}{a_L}\right)^{k-\rho k/(1-\rho)} - 1\right)} \Phi\left(z_{ij}^*\right)\right)^{\frac{k/(1-\rho)+1-\varepsilon}{k/(1-\rho)}} - 1.$$

Hence, W_{ij} is an infinite order polynomial of $\Phi\left(z_{ij}^*\right)$. The dominant term of this expansion is given by:

$$W_{ij} \simeq \frac{k/(1-\rho)+1-\varepsilon}{k/(1-\rho)} \frac{\left(a_H^k - a_L^k\right)^2 (k - \rho k/(1-\rho))}{k a_L^{2k}\left(\left(\frac{a_H}{a_L}\right)^{k-\rho k/(1-\rho)} - 1\right)} \Phi\left(z_{ij}^*\right).$$

As an implication,

$$\log W_{ij} - \log \Phi\left(\widehat{z}_{ij}^*\right) \simeq \log\left(\frac{k/(1-\rho)+1-\varepsilon}{k/(1-\rho)} \frac{\left(a_H^k - a_L^k\right)^2 (k - \rho k/(1-\rho))}{k a_L^{2k}\left(\left(\frac{a_H}{a_L}\right)^{k-\rho k/(1-\rho)} - 1\right)}\right), \tag{35}$$

where we note that the RHS is a constant.

Hence, a first order approximation of Model (25) is given by:

$$\log \widetilde{M}_{ij} = x_{ij}\beta - \sigma_{ij}^2/2 + \widetilde{r}_{ij}, \tag{36}$$

which is a log-linear model that ignores the sample selection and the heterogeneity in expected trade flows.

6.2 Finite number of firms, random trade cost and aggregate productivity shocks

We now consider the case where productivity and trading costs are both random. For this case, Equation (30) is used as starting point:

$$\log \widetilde{M}_{ij} = x_{ij}\beta + \log \overline{W}\left(z_{ij}^*, \delta, r\right) + E\left(\log \epsilon_{ij} | \epsilon_{ij} > 0\right) + \widetilde{\epsilon}_{ij},$$

where $\overline{W}\left(z_{ij}^*, \delta, r\right)$ is given by (29). For Equation (30) to be operational, we need to find proxies for $\overline{W}\left(z_{ij}^*, \delta, r\right)$ and $E\left(\log \epsilon_{ij} | \epsilon_{ij} > 0\right)$. As previously, we assume that ϵ_{ij} follows a log-normal distribution with a point-mass at zero:

$$\log \epsilon_{ij} | \epsilon_{ij} > 0 \sim N\left(\mu_{ij}, \sigma_{ij}^2\right),$$

where

$$\begin{aligned} \mu_{ij} &= -\log \Phi\left(\widetilde{z}_{ij}^*\right) - \sigma_{ij}^2/2, \\ \sigma_{ij}^2 &= \log\left(1 + \exp\left(\theta_0 + \theta_1\left(x_{ij}\beta\right)\right)\right), \end{aligned}$$

and $\widetilde{z}_{ij}^*$ is a latent variable that satisfies $\Phi\left(\widetilde{z}_{ij}^*\right) = \Pr\left(\epsilon_{ij} > 0\right)$.

A proxy $\widehat{\widetilde{z}}_{ij}^{*}$ can be designed for $\widetilde{z}_{ij}^{*}$ by fitting a Probit regression for the indicator of trade. Indeed, note that there is a positive trade flow from j to i if and only if $\frac{a_{ij}}{a_L} > 1$. Thus, we may define z_{ij} by:

$$\widetilde{z}_{ij}^{*} = \frac{1}{\sqrt{\sigma_u^2 + \sigma_v^2}} \log \left(\frac{a_{ij}}{a_L} \right)^{\varepsilon - 1},$$

so that $\widetilde{M}_{ij} > 0$ if and only if $\widetilde{z}_{ij}^{*} > 0$. Equation (26) allows us to write:

$$\widetilde{z}_{ij}^{*} = x_{ij}\gamma + \eta_{ij},$$

where $x_{ij}\gamma$ collects all fixed effects and $\eta_{ij} = \frac{u_{ij} + v_{ij}}{\sqrt{\sigma_u^2 + \sigma_v^2}}$ is standard normal. Hence:

$$\Pr\left(\widetilde{M}_{ij} > 0 | x_{ij}\right) \equiv \Phi\left(\widetilde{z}_{ij}^{*}\right) = \Phi\left(x_{ij}\gamma\right). \tag{37}$$

A proxy is obtained for $\widetilde{z}_{ij}^{*}$ as $\widehat{\widetilde{z}}_{ij}^{*} = x_{ij}\widehat{\gamma}$, where $\widehat{\gamma}$ is a Probit estimator of γ.

But we still need to find a proxy for the latent variable z_{ij}^{*} which enters in the expression of $\overline{W}\left(z_{ij}^{*}, \delta, r\right)$. For that purpose, note that:

$$\begin{aligned} \Phi\left(\widetilde{z}_{ij}^{*}\right) &= \Pr\left(\epsilon_{ij} > 0\right) = \Pr\left(R_{ij} > 0\right) \Pr\left(\widetilde{R}_{ij} > 0\right), \\ &\leq \Pr\left(R_{ij} > 0\right) = \Phi\left(z_{ij}^{*}\right). \end{aligned}$$

A simple approach to impose the restriction $\Phi\left(z_{ij}^{*}\right) \geq \Phi\left(\widetilde{z}_{ij}^{*}\right)$ is to assume that:

$$\Phi\left(z_{ij}^{*}\right) \equiv \frac{1}{2}\Phi\left(\widetilde{z}_{ij}^{*}\right) + \frac{1}{2}\Phi\left(\widetilde{z}_{ij}^{*} + \exp\left(\rho\right)\right),$$

where ρ is unrestricted. This leads us to proxy z_{ij}^{*} by:

$$\widehat{z}_{ij}^{*} = \Phi^{-1}\left(\frac{1}{2}\Phi\left(\widehat{\widetilde{z}}_{ij}^{*}\right) + \frac{1}{2}\Phi\left(\widehat{\widetilde{z}}_{ij}^{*} + \exp\left(\rho\right)\right)\right). \tag{38}$$

where $\widehat{\widetilde{z}}_{ij}^{*}$ is a proxy of $\widetilde{z}_{ij}^{*}$.

Finally, the estimating equation is:

$$\log \widetilde{M}_{ij} = x_{ij}\beta + \log \overline{W}\left(\widehat{z}_{ij}^{*}, \delta, r\right) - \log \Phi\left(\widetilde{z}_{ij}^{*}\right) - \sigma_{ij}^2/2 + \widetilde{\epsilon}_{ij}, \tag{39}$$

where $\widetilde{\epsilon}_{ij} \sim N\left(0, \sigma_{ij}^2\right)$ on the domain $\epsilon_{ij} > 0$. Note that $\widehat{z}_{ij}^{*}$ coincides with $\widehat{\widetilde{z}}_{ij}^{*}$ when $\rho = -\infty$ so that Equation (39) nests the one presented in the next subsection.

6.3 Infinite Number of Firms and Random Trade Costs

The starting point for deriving an estimating equation for this case is Equation (31), which is given by:

$$\log \widetilde{M}_{ij} = x_{ij}\beta + \log \overline{W}\left(z_{ij}^{*}, \delta, r\right) + E\left(\log \epsilon_{ij} | \epsilon_{ij} > 0\right) + r_{ij},$$

where $\epsilon_{ij} = R_{ij}\widetilde{R}_{ij}$ with $\widetilde{R}_{ij} > 0$. If we let z_{ij}^{*} and $\widetilde{z}_{ij}^{*}$ denote latent variables such that $\Pr\left(R_{ij} > 0\right) = \Phi\left(z_{ij}^{*}\right)$ and $\Pr\left(\epsilon_{ij} > 0\right) = \Phi\left(\widetilde{z}_{ij}^{*}\right)$, then $\epsilon_{ij} > 0$ if and only if $R_{ij} > 0$. It thus follows that the estimating equation for this case is obtained by specializing Equation (39) to $z_{ij}^{*} = \widetilde{z}_{ij}^{*}$, which yields:

$$\log \widetilde{M}_{ij} = x_{ij}\beta + \log \overline{W}\left(\widehat{z}_{ij}^{*}, \delta, r\right) - \log \Phi\left(\widehat{z}_{ij}^{*}\right) - \sigma_{ij}^2/2 + r_{ij}, \tag{40}$$

where $r_{ij} \sim N\left(0, \sigma_{ij}^2\right)$ and $\widehat{z}_{ij}^*$ is obtained as the proxy of the normalized latent variable of the Probit regression presented around Equation (37).

It might be useful to also consider the original estimating equation advocated by HMR to see how it performs relative to our Equation (40). HMR's estimating equation is given by (See HMR, 2008, Equation 14, pp. 456):

$$\log \widetilde{M}_{ij} = x_{ij}\beta + \widehat{w}_{ij} + \eta_2 \widehat{IMR}_{ij} + e_{ij}, \tag{41}$$

where $\widehat{w}_{ij} = \log\left\{\exp\left(\eta_1\left(\widehat{IMR}_{ij} + \widehat{z}_{ij}^*\right)\right) - 1\right\}$ is a proxy for $E\left(\log W_{ij} | \widetilde{R}_{ij} > 0\right)$ and $\widehat{IMR}_{ij} = \phi\left(\widehat{z}_{ij}^*\right) / \Phi\left(\widehat{z}_{ij}^*\right)$ is the inverse Mills ratio which proxies $E\left(\log \epsilon_{ij} | \epsilon_{ij} > 0\right)$ and e_{ij} is treated as an homoscedastic error. In (41), $\widehat{w}_{ij}$ captures the heterogeneity in trade flows while $\eta_2 \widehat{IMR}_{ij}$ corrects for sample selection bias à la Heckman (1979). In fact, HMR's empirical approach consists of first log-linearizing the expression of $\widetilde{M}_{ij}$ before correcting for the sample selection and heterogeneity biases. In particular, the proxy used for $E\left(\log W_{ij} | \widetilde{R}_{ij} > 0\right)$ amounts to approximating $E \log f(u)$ by $\log f\left(\widehat{E}u\right)$ for some random variable u. Santos Silva and Tenreyro (2009) suggest an alternative approach, which consists of first obtaining the exact expression of $E\left(\widetilde{M}_{ij} | \epsilon_{ij} > 0\right)$ as a function of $E\left(W_{ij} | \epsilon_{ij} > 0\right)$ and $E\left(\epsilon_{ij} | \epsilon_{ij} > 0\right)$ before applying the logarithm. The latter approach amounts to approximating $\log E f(u)$ by $\log \widehat{E} f(u)$ and it leads to our Equation (40).

Santos Silva and Tenreyro (2009) implemented their method empirically and their results suggest that both approaches yield qualitatively similar results. Nonetheless, the approach of Santos Silva and Tenreyro is advocated in this paper as it is theoretically more accurate. Note that beside the difference highlighted above, Equation (40) further extends the basic HMR model to account for the error heteroscedasticity.

7 Empirical Application

The data used for this application has been previously used by HMR (2008) and it describes trade flows between 158 countries during the 80's. We focus only on the years 1980 and 1989 and restrict to countries that import from at least one origin or that export to at least one destination. In total, this amounts to 24649 bilateral trade flows for each of these years, of which 10975 and 11203 are strictly positive respectively in 1980 and 1989.

For each of the two years, we estimate separate probit models that explain the determinants of the selection of countries into trading relationship. The following explanatory variables are used:

(i) Log-distance between the country pair;

(ii) Land border indicator: equals to 1 if the partners pair has a common land border;

(iii) Island indicator: equals 1 if at least one partners in the pair is an island;

(iv) Landlock indicator: equals 1 if at least one partners in the pair is landlocked;

(v) Legal system indicator: equals 1 if the two partners have same legal system;

(vi) Language indicator: equals 1 if the two partners have at least one common language;

(vii) Colonial ties indicator: equals 1 if the two partners have historical colonial ties;

(viii) Currency union indicator: equals 1 if the two partners are in a common currency union zone;

(ix) FTA indicator: equals 1 if the two partners have a free trade agreement;

(x) Religion indicator: an index that is increasing in the percentage of population sharing a common religion.

In addition to the regressors above, importer and exporter specific fixed effects are also included in the model. Table 1 shows the estimation results. Factors that negatively impact the probability of trade are the log-distance, a common land border and whether one of the partners is an island or landlocked. The negative impact of the existence of a common land border on the probability of trade is quite puzzling. HMR suggested that this may be reflecting the existence of border conflicts. Interestingly, the negative coefficients associated with the previous factors have decreased over the decade. In particular, the disadvantage of landlocked countries to trade creation is not significant for 1989.

All other explanatory variables positively impact trade creation, although the advantages to trade creation associated with a common legal system and the existence of colonial ties are not significant for 1980. Except for the religion coefficient, all positive elasticities have grown in magnitude between 1980 and 1989. Notably, FTAs contributed significantly more to trade creation in 1989 than they did in 1980.

Table 1: Probit Estimation Results (extensive margin)

	1980	1989
variables / parameters	Estimate	Estimate
Log distance	-0,740***	-0,548***
Land border	-0,408***	-0,227**
Island	-0,377***	-0,325***
Landlock	-0,269***	-0,139
Legal system	0,031	0,085***
Common language	0,253***	0,268***
Colonial ties	0,430	0,766**
Currency union	0,903***	0,913***
FTA	1,290***	1,837***
Religion	0,398***	0,320***

We now turn to the estimation of the intensive margin using only strictly positive trade flows. Subsequently, the label "Model 1" denotes the estimating equation for the model with finite number of firms, random trade costs and aggregate productivity shocks. "Model 2" is used for the case with infinite number of firms and random trade costs. "Model 3" pertains to the case with finite number of firms, non-random trading costs and aggregate productivity shocks. Finally, "Model 0" is HMR's estimating equation, which can be cast into the same theoretical framework as Model 2. Model 2 can be obtained from Model 1 by letting the ancillary parameter ρ diverge to $-\infty$. Except Model 0, all three other estimating equations assume heteroscedastic errors. Thus, models 1, 2 and 3 are estimated by maximum likelihood while Model 0 is estimated by nonlinear least squares.

Table 2 shows the estimation results for Models 0 and 2, which are our benchmark models. The negative effects of the log-distance on trade flows predicted by both estimating equations are of similar magnitudes and increasing over time, consistently with the well-known distance puzzle[2]. However, Models 0 and 2 disagree regarding the magnitude of the land border effect and its behavior over time. According to both models, a common land border increased trade by a factor of approximately $e^{0.8} = 2.225$ in 1980, all else equal. In 1989, Model 0 predicts that this effect has remained stable since the beginning of the decade while Model 2 predicts a large decrease to $e^{0.378} = 1.459$.

[2] According to Disdier and Head (2008), the estimated negative impact of distance on trade rose around the middle of the century and has remained persistently high since then. The fact that there has not been a decline in recent times is puzzling considering all of the advances in transport logistics.

Models 0 and 2 also disagree regarding the predictions of the disadvantage of island and landlocked countries to trade. According to Model 0, the disadvantage of being an island country has remained quite stable over time (trade is reduced by a factor of $e^{-0.447} = 0.639$ in 1980 versus $e^{-0.410} = 0.663$ in 1989) while that of being a landlocked country has slightly worsen over time ($e^{-0.546} = 0.579$ in 1980 versus $e^{-0.673} = 0.510$ in 1989). According to Model 2, the disadvantage of island countries to trade is less severe than predicted by Model 0 and it has eroded over time ($e^{-0.303} = 0.738$ in 1980 versus $e^{-0.199} = 0.819$ in 1989) while the disadvantage of being a landlocked country has remained stable over the same period. The predictions of Model 2 are somewhat consistent with an increased globalization of the world economy.

Models 0 and 2 are best understood by comparing equations (40) and (41). Both account for selection and heterogeneity. Because model 2, like model 0, assumes that the number of firms is infinite, random productivity shocks do not affect aggregate trade flows. Therefore, the difference between the two models mainly comes from whether heterosdasticity is accounted for or not.

Table 2 : Gravity estimation results: the benchmark models.

Model 0: HMR (2008)'s estimating equation with homoscedastic errors.

Model 2: Same model but with Santos Silva and Tenreyro (2006)'s correction and heteroscedastic errors.

Variables	1980		1989	
	Model 0	Model 2	Model 0	Model 2
Log distance	-0.766***	-0,757***	-0.956***	-0,981***
Land border	0.800***	0,833***	**0.790*** **	**0,378*** **
Island	**-0.447*** **	**-0,303** **	**-0.410*** **	**-0,199** **
Landlock	**-0.546*** **	**-0,504** **	**-0.673*** **	**-0,525*** **
Legal system	0.339***	0,376***	0.273***	0,319***
Common language	0.099*	0,027	0.238***	0,354***
Colonial ties	1.059***	1,094***	0.706***	0,769***
Currency union	1.013***	0,936***	1.696***	1,820***
FTA	0.982***	1,083***	0.461***	0,377***
Ancillary param.				
η_1 (heterogeneity)	0.637***	-	0.510***	-
η_2 (Selection)	0.472***	-	0.610***	-
θ_0 (heterosced.)	-	2,421***	-	4,673***
θ_1 (heterosced.)	-	0,122***	-	-0,192***
δ (heterogeneity)	-	0,806***	-	0,064
r (heterogeneity)	-	0,698***	-	0,920***

Subsequently, we use Model 2 as benchmark and compare is to Models 1 and 3. Table 2 shows the estimation results. We see that the parameter ρ of Model 1 is estimated to be positive ($\widehat{\rho} = 0.366$ for 1980 and $\widehat{\rho} = 0.608$ for 1989) while a large and negative estimate would have implies that Models 1 and 2 are roughly identical. This result suggests that aggregate productivity shocks determine the selection of countries into trading relationship in synergy with the randomness in trading costs. Model 1 also delivers elasticity estimates that are more consistent with intuition than Model 2. For example, Model 1 predicts that the negative effect of the log-distance has decrease between 1980 and 1989 while Model 2 predicts an increase.

We shall also favor Model 1 over Model 3 for a two reasons. First, Model 3 consistently predicts that the noise variance is a decreasing function of expected trade flows ($\widehat{\theta}_1 = -0.223$ in 1980 and $\widehat{\theta}_1 = -0,320$ in 1989). This is counter-intuitive because measurement errors are likely to increase with the magnitude of aggregate trade flows. Second, Model 3 predicts a much lower impact of FTA on trade volumes compared to Model 1, another counter-intuitive result which contrasts with the fact that more and more countries chose to be bound by trade agreements.

In order to assess how important is the aggregate productivity shocks, we consider decomposing the total sample selection effect into:

$$\log \Pr\left(\epsilon_{ij} > 0\right)^{-1} = \log \Pr\left(R_{ij} > 0\right)^{-1} + \log \Pr\left(\widetilde{R}_{ij} > 0\right)^{-1},$$

where $\log \Pr\left(R_{ij} > 0\right)^{-1}$ is attributable to trading costs and $\log \Pr\left(\widetilde{R}_{ij} > 0\right)^{-1}$ is attributable to productivity shocks. We measure the importance of aggregate productivity shocks by the following ratio:

$$\lambda = \frac{\sum_{ij} \log \Pr\left(\widetilde{R}_{ij} > 0\right)^{-1}}{\sum_{ij} \log \Pr\left(\epsilon_{ij} > 0\right)^{-1}} = 1 - \frac{\sum_{ij} \log \Pr\left(R_{ij} > 0\right)}{\sum_{ij} \log \Pr\left(\epsilon_{ij} > 0\right)}.$$

We find $\lambda = 51\%$ for 1980 and $\lambda = 57\%$ for 1989. This suggest that aggregate productivity shocks are more important than trade costs randomness in determining the selection of countries into trading relationships. However, the randomness in trading costs are still very important (49% and 43%) to justifie Model 1 compared to Model 3.

Table 3: Gravity estimation results.

Model 1: Finite number of firms, random trading costs and aggregate productivity shocks.

Model 2: Infinite number of firms and random trading costs.

Model 3: Finite number of firms, non-random trading costs and aggregate productivity shocks.

	1980			1989		
variables	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Log distance	**-1,191*****	**-0,757*****	-1,023***	**-1,001*****	**-0,981*****	-0,979***
Land border	0,586***	0,833***	0,281***	0,555***	0,378***	0,259***
Island	-0,594***	-0,303**	-0,421***	-0,205*	-0,199**	-0,490***
Landlock	-0,307*	-0,504**	-0,547***	-0,308*	-0,525***	-0,256*
Legal system	0,357***	0,376***	0,439***	0,386***	0,319***	0,344***
Common language	0,291***	0,027	0,160***	0,360***	0,354***	0,370***
Colonial ties	1,375***	1,094***	1,250***	1,241***	0,769***	0,929***
Currency union	1,258***	0,936***	1,369***	1,782***	1,820***	2,071***
FTA	**1,195*****	1,083***	**0,385*****	**1,353*****	0,377***	**0,218****
Ancillary param.						
θ_0 (heterosced.)	2,169***	2,421***	5,421***	1,286***	4,673***	5,573***
θ_1 (heterosced.)	**0,072*****	0,122***	**-0,223*****	**0,083*****	-0,192***	**-0,320*****
δ (heterogen.)	5×10^{-4}***	0,806***	-	1×10^{-4}***	0,064	-
r (heterogen.)	0,361	0,698***	-	-0,369*	0,920***	-
ρ	**0,366***	-	-	**0,608*****	-	-

8 Conclusion

We modify the Helpman-Melitz-Rubinstein (HMR) model to account for firm-level productivity shocks that translate into aggregate effects on trade flows, in the spirit of Gabaix (2011). First, we allow the number of candidate exporting firms to be finite and the productivity of each individual firm to be random. Hence, firms are homogenous ex-ante regarding the distribution of productivity, but heterogenous ex-post regarding realized productivity. Second, we allow candidates exporting firms of the same country to share a common productivity shock. Third, as in HMR (2008), trade costs are allowed to be random. The first assumption alone implies that firm-level shocks translate into aggregate productivity shocks unless the number of candidates exporting firms is infinite. The second assumption alone implies that trade flows are subject to aggregate productivity shocks even if the number of candidate exporting firms is infinite, but aggregate trade flows are strictly positive at the limit. Under the first and second assumptions, the realizations of aggregate productivity shocks determine whether a country is able to export to a given destination or not. Under all three assumptions, the selection of countries into trading relationships is determined by the aggregate productivity shocks in synergy with shocks with trade costs.

We consider three model specifications based on the number of firms and the randomness / non-randomness of trade costs that emerge from our theoretical framework. All three models are estimated by maximum likelihood and correct for hetoroskedastic errors as Poisson estimators advocated by Santos Silva and Tenreyro (2006) within the intuitively-pleasing 2-step structure of HMR. One model assumes a finite number of firms, aggregate productivity shocks and random trading costs. Another model assumes an finite number of firms, aggregate productivity shocks and random trading costs. Finally, the third alternative model assumes a finite number of firms, aggregate productivity shocks and non random trading costs. We estimate all three models and HMR's two-step estimator using bilateral world trade data for the year 1980 and 1989 and find supportive evidence for the presence of aggregate productivity shocks. On average,aggregate productivity shocks accounted for 51% of the sample selection effect in 1980 and about 57% in 1989.

Our model based on a finite number of firms, random trade costs and aggregate productivity shocks delivered elasticity estimates that are more consistent with intuition than the three other alternatives. It predicts that *(i)* the negative effect of the distance on trade flows and the disadvantage of island countries to trade has eroded over time, *(ii)* the positive effect of a common land border and the disadvantage of landlocked countries to trade have remained stable during the 1980's, and *(iii)* a common currency union zone and a free trade agreement have had larger positive impacts on the intensity of trade in 1989 compared to 1980.

References

[1] Anderson, J. E. and E. van Wincoop (2003) "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review* 93, 170–192.

[2] Armenter R. and M. Koren (2012) "A Balls-and-Bins Model of Trade," *Unpublished Manuscript*, available at *http://miklos.koren.hu/view/list/*

[3] Bernard, A. B., Redding, S. J. and Peter K. Schott (2011) "Multiproduct Firms and Trade Liberalization" *The Quarterly Journal of Economics* 126, 1271–1318.

[4] Besedes T. and T. J. Prusa (2006a) "Ins, Outs, and the Duration of Trade," *The Canadian Journal of Economics* 39:1, pp. 266-295.

[5] Besedes T. and T. J. Prusa (2006b) "Product Differentiation and Duration of US Import Trade," *Journal of International Economics* 70, 2, pp 339–358.

[6] Disdier, Head, K. 2008. The Puzzling Persistence of the Distance Effect on Bilateral Trade. Review of Economics and Statistics, 90,1:37-48.

[7] Gabaix, X. (2011) The Granular Origins of Aggregate Fluctuations. Econometrica, 79,733-772,

[8] Heckman, J. J. (1979) "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

[9] Helpman, E., Melitz, M. and Y. Rubinstein (2008), "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics* 123, 441-487.

[10] Eaton, J. and S. Kortum (2002) "Technology, Geography and Trade," *Econometrica*, 70:5, 1741–1779.

[11] Eaton, J., Kortum, S. and S. Sotelo (2012) "International Trade: Linking Micro and Macro," *NBER Working Paper Series*, Working Paper 17864. Available at http://www.nber.org/papers/w17864.

[12] Melitz, M. (2003) "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity," *Econometrica* 71, 1695-1725.

[13] Sabuhoro, J.B., Larue, B. and Y. Gervais (2006) "Factors Determining the Success or Failure of Canadian Establishments on Foreign Markets: A Survival Analysis Approach," *The International Trade Journal* 20:1, 33-73

[14] Santos Silva, J. M. C. and S. Tenreyro (2006) "The Log of Gravity," *Review of Economics and Statistics* 88, 641–658.

[15] Santos Silva, J. M. C. and S. Tenreyro (2009) "Trading Partners and Trading Volumes: Implementing the Helpman-Melitz-Rubinstein Model Empirically," *CEP Discussion Paper No 935*, Centre for Economic Performance, London School of Economics and Political Science.

[16] Tinbergen, J. (1962) "*Shaping the World Economy,*" (New York: The Twentieth Century Fund).