



ISSN 2279-6894  
**IMT LUCCA EIC WORKING**  
**PAPER SERIES 01**  
**January 2013**

RA **Economics and institutional change**

# Stochastic Trade Networks

Massimo Riccaboni  
Stefano Schiavo

ISSN 2279-6894  
IMT LUCCA EIC WORKING PAPER SERIES #01/2013  
© IMT Institute for Advanced Studies Lucca  
Piazza San Ponziano 6, 55100 Lucca

Research Area  
**Economics and institutional change**

# Stochastic Trade Networks

**Massimo Riccaboni**

IMT Institute for Advanced Studies Lucca

**Stefano Schiavo**

School of International Studies and Department of Economics and Management, University of Trento

# Stochastic Trade Networks

Massimo Riccaboni\*      Stefano Schiavo†

This version — January 2013

## Abstract

This paper develops a simple network model to describe the dynamic of the intensive and extensive margin of international trade flows. The result is achieved by means of the combination of two mechanisms of proportional growth: the first (discrete) determines the formation of trade links, the second (continuous) governs trade intensity. We show that our setup is able to simultaneously match a large number of empirical regularities, such as the fraction of zero trade flows across pairs of countries or the high concentration of trade with respect to both products and destinations. Our findings suggest that stylized facts are strongly interconnected across different levels of aggregation of trade data, so that a unifying explanation is called for. By incorporating stochastic elements into standard trade models we can improve their ability to explain relevant facts about world trade.

**JEL Codes:** F14, F43, O25

**Keywords:** international trade, networks, preferential attachment, urn models, proportionate growth

---

Financial support received through the research project ‘The international trade network: empirical analyses and theoretical models’ ([www.tradenetworks.it](http://www.tradenetworks.it)) funded by the Italian Ministry of Education, University and Research (Scientific Research Programs of National Relevance 2009) is gratefully acknowledged. The authors thank Marco Bee, Andrea Fracasso, Giorgio Fagiolo and the participants to the Trade NetWorkshops 2.0 (Lucca) and 3.0 (Trento), the 16th International Conference of the Society for Computational Economics (London), DEGIT XV (Frankfurt) and XVII (Milan) conferences, and the 1st Annual CIRANO Workshop on Networks in Trade and Finance (Montreal) for insightful comments on earlier drafts of the paper.

\*LIME, IMT – Institute for Advanced Studies, Lucca (Italy) and Department of Managerial Economics, Strategy and Innovation, Katholieke Universiteit Leuven, Leuven (Belgium). E-mail: [m.riccaboni@imtlucca.it](mailto:m.riccaboni@imtlucca.it)

†Corresponding author. School of International Studies and Department of Economics and Management, University of Trento (Italy) and OFCE-DRIC (France). E-mail: [stefano.schiavo@unitn.it](mailto:stefano.schiavo@unitn.it)

# 1 Introduction

The literature on international trade has advanced in the past decade through a fruitful interplay between new observations on exporting firms and new theories that have introduced firm heterogeneity into trade models. New stylized facts have been uncovered and new trade models have been developed to explain these facts. However, despite these results, the core aggregate relationships among different stylized facts have remained largely untouched (Eaton *et al.*, 2012; Arkolakis *et al.*, 2013).

This paper contributes to filling this gap by introducing a simple stochastic framework to describe bilateral trade flows among countries, and tests it using both simulations and real data. The result is a parsimonious representation of economic interaction in a multi-country world akin to the gravity model, defined at the level of exporting firms. In particular, we develop a network model (similar in spirit to Chaney, 2011) which combines two mechanisms of proportionate trade as the main determinants of the number of export links (extensive margin) and their value (intensive margin).

We assume that both (i) the formation of new (bilateral) trade links, and (ii) the value of trade flows evolve according to simple proportional rules. These two hypotheses yield a parsimonious setup capable of matching a large number of the regularities that characterize international trade data. Furthermore, whereas the model is defined at the firm level, we derive predictions for different levels of aggregation with no need to make any *ad hoc* changes to the basic structure. This aggregation property is one of the main strengths of our approach.

Using both simulations and real data we show that our simple framework is capable to match many different empirical regularities identified in the literature. The conclusion we draw from this exercise is that incorporating stochastic elements into standard economic models is a fruitful avenue of research to improve the ability of trade theory to explain empirical facts emerging from the data, similarly to what has been successfully done in other fields by Gabaix (1999), Klette and Kortum (2004), Sutton (1998) and Luttmer (2007) among others.

There is a strong parallel between our approach and the well-known gravity model of international trade. First, both are essentially empirical models, are parsimonious, and rely on simple proportionality assumptions to describe international trade flows. Second, they are modular, i.e. they apply to different scales of aggregation: the same setup can be used to explain both aggregate trade and its constituent parts in terms of goods, regions, or firms. Third, they both represent an empirical benchmark against which data can be confronted: significant departures from the benchmark are of primary interest in that they signal the presence of frictions (as

for the gravity model) or, more generally, the need for a theoretical explanation of the economic forces at play.

Given our aim to properly account for several stylized facts by means of a single model, we refer to the large body of literature that has documented the empirical regularities characterizing international trade, and how these stylized facts, especially the recent ones emerging from firm-level data, challenge existing theories. In particular, we consider the share of zeros among bilateral trade data, the high concentration of trade flows, and the role of the intensive and extensive margins in determining the different export performance of countries.

As new and richer datasets become available, they show that existing theoretical models are often at odds with empirical regularities: the baseline gravity specification, for instance, is unable to account for the large fraction of zeros in the data, and although extensions have been recently proposed to accommodate this problem (Helpman *et al.*, 2008; Baldwin and Harrigan, 2011), they focus on a single specific issue, but have little to say about other relevant stylized facts. Trying to overcome these limitations, Armenter and Koren (2010) have recently proposed a stochastic model that describes US exports as a series of balls randomly falling into bins of different size, each representing a product-destination pair. In spite of its simplicity, this setup matches the large number of zeros in product-destination and firm-level trade data. We take a similar route, but instead of focusing on a single country we take a global approach and propose a model that describes the structure and evolution of world trade flows. Similarly to Eaton *et al.* (2012), we do not assume a continuum of firms but rather aggregate over a finite number of goods (each produced by a distinct firm). With respect to Armenter and Koren (2010) our setup matches more facts (for instance describe both the extensive and the intensive margin), with higher precision and, what is more, we need a smaller set of assumptions to start with.

Although we share the view that competing economic theories should not be judged on the basis of their ability to match stylized facts that are well explained by simple stochastic interactions, we do not push the argument to consider a stochastic model only useful when missing an empirical fact (thus signaling the working of specific economic forces). Rather, we also aim to show that stochastic elements can increase the explanatory power of standard models of trade. This is not far from the use of the gravity model in international trade: the gravity equation is, at least in its origin, nothing more than a simple empirical model representing a reference point (Anderson, 2011). Departures from such a benchmark signal the presence of relevant frictions (or economic forces) that are brought into the picture by means of

additional variables to be included in the analysis.<sup>1</sup>

In a recent contribution, Chaney (2011) discusses how existing trade models based on heterogeneous firms (e.g. Melitz, 2003; Bernard *et al.*, 2003; Chaney, 2008) are unable to make any prediction on the cross-sectional distribution of the number of destinations served by each exporter. He then proposes a model based on social network theory that accurately matches the empirical features of the extensive margin of trade, although it is silent about the intensive margin.<sup>2</sup> Besides closely matching the cross-sectional distribution of the extensive margin of trade, our setup has the additional advantage that it also yields clear-cut predictions the behavior of the intensive margin. It accounts for zeros in the data very well while maintaining the good fit of the standard gravity equation among country pairs.

The paper is organized as follows: Section 2 takes a closer look at a number of stylized facts about international trade flows that are relevant to our work and will be addressed by the model. The model itself is presented in Section 3 alongside with its most important predictions; these are compared with the data and tested by means of simulations whose results are discussed in Section 4. Lastly, we lay down some conclusions and outline possible patterns for future research.

## 2 Empirical regularities

This section spells out the main empirical regularities about bilateral trade flows that are relevant to our work and that we set out to explain. They concern both the extensive and the intensive margin of trade, as well as the relationship between them. In the literature they are often identified by means of three main issues, namely (i) the large fraction of zeros among bilateral flows; (ii) the high concentration of trade values, and (iii) the contribution of intensive and extensive margins to trade growth.

The increasing availability of firm-level data has highlighted how these stylized facts tend to be invariant to the level of observation, i.e. they show up in the data irrespective of whether one looks at aggregate trade, product-specific figures, or firm-level data.

---

<sup>1</sup>Indeed, in the early days of the gravity model, Savage and Deutsch (1960) proposed a “null model” based on the gravity specification to predict bilateral import/export flows between country pairs, and this so-called probabilistic approach (as defined by Sen and Smith, 1995) represents one of the two fundamental paths along which the empirical literature on the gravity model has developed (Bergstrand and Egger, 2011).

<sup>2</sup>The reference to this network view of world trade is particularly relevant here because the model we present below has a straightforward network representation (Riccaboni and Schiavo, 2010). Therefore it represents a possible extension of the work by Chaney (2011), as it is complementary to models of international trade with heterogeneous firms such as Melitz (2003) or Bernard *et al.* (2003).

Many of the regularities have long been recognized and several attempts to account for them exist in the literature, although this is often done in a piecemeal way, tackling one or a small subset of the stylized facts at a time (Bernard *et al.*, 2007; Armenter and Koren, 2010; Baldwin and Harrigan, 2011).<sup>3</sup> On the contrary, the simple model outlined in Section 3 is capable of tackling them simultaneously.

Let us start from the empirical regularities concerning the extensive margin of trade, i.e. the number of trade relationships held by each agent, being it a country (when looking at aggregate trade), or a firm.

**Stylized fact 1.** About half of the country pairs do not trade with one another.

The fraction of zeros grows larger when we look at disaggregated (product- or firm-specific) data.

**Stylized fact 2.** The cross-sectional distribution of the number of trade relationships held by each country (or each firm) is highly skewed. This holds irrespective of whether trade relationships are defined as the number of markets served, the number of products shipped, or the number of product-destination pairs.

Concentration does not solely characterize the extensive margin, but is even more apparent when we look at the values of trade flows, i.e. at the intensive margin of trade.

**Stylized fact 3.** The most popular export products and/or destinations account for a disproportionate share of a country's (or a firm's) total export.

**Stylized fact 4.** The distribution of bilateral trade values is also highly skewed, with few destination markets (or commodities exported, or product-destination pairs, or exporters) making up a disproportionate share of total trade. Again, this holds at different levels of aggregation.

When looking at the relative importance of the intensive and the extensive margin in explaining the dynamics of trade flows, previous works have unveiled some further interesting patterns.

**Stylized fact 5.** The extensive margin is the main driver of cross-sectional differences in trade participation.

**Stylized fact 6.** The intensive margin, on the other hand, explains most of the dynamics of world trade over time.

---

<sup>3</sup>Appendix A provides the reader with a detailed discussion of the stylized facts, together with the list of relevant references.

These regularities are robust across data source, period, unit of analysis, and definition of trade relationship. In particular, strikingly similar results are obtained using aggregate data, the UN-NBER dataset containing SITC 4-digit commodity flows (documented in Feenstra *et al.*, 2005), US export shipments (used by Armenter and Koren, 2010), or information about French firm export behavior (described by Eaton *et al.*, 2004).

Furthermore, although the literature has generally tackled these stylized facts in isolation, one can argue they are related with each other (similar to what is done by Klette and Kortum, 2004, in their analysis of innovation and growth). Indeed, we show below that our stochastic framework implies some interactions among them, which strengthen the case for tackling them simultaneously.

### 3 A stochastic model of trade networks

This section briefly describes our stochastic setup. The purpose is to explain the extensive as well as the intensive margin of international trade based on simple mechanisms of proportional growth. As in Chaney (2011), we assume that individual firms enter foreign markets if and only if they have a contact in that market. Once a contact is established, the initial value of trade can be defined as in standard trade models à la Melitz (2003) and evolves according to a stochastic growth process (i.e. it is subject to random demand shocks). As for the extensive margin, we start from a simple preferential attachment rule of network growth (Ijiri and Simon, 1977; Barabási and Albert, 1999).

The formal set-up is as follows. Firms are distributed over a finite set of locations (countries). Once born, a firm never changes location. Time is discrete and at time  $t = 0$  there are  $n_i$  firms at location  $i$ . Firms form direct links with one another. New links are created and removed with constant Poisson probabilities: birth rate  $\lambda$  and death rate  $\gamma$  (with  $\lambda > \gamma$ ). Each firm has both an in-degree distribution (the set of inward or import contacts) and an out-degree distribution (the set of outward or export contacts). Differently from Chaney (2011), who takes the firm out-degree as fixed, we assume that both import and export contacts change according to a simple proportional growth rule. This is crucial in our setup to explain the distribution of zeros in international trade.

The total population of firms, as well as the number of contacts, grows at a constant rate. In the following we describe the evolution of firm out-degree distribution. We assume that the very same probabilistic process is at work for the acquisition of in-degree connections. Thus, the network growth process is symmetric and results



hold for the in-degree connectivity as well.<sup>4</sup>

At each time step, new trade opportunities arise in proportion to the current size of the economic network  $\lambda K(t)$  where  $\lambda$  is the growth rate and  $K(t)$  is the total number of contacts at time  $t$ . At the same time, contacts are removed with probability  $\gamma$ . Opportunities (in the form of trade contacts) are assigned to (or removed from) firm  $j$  with a probability  $p_j$  proportional to the number of contacts it already has:  $p_j = k_j(t)/K(t)$ .<sup>5</sup> Since there are  $N(t)_i$  firms at location  $i$ , the same proportional rule applies at the country level: the probability that a country captures a new export contact is proportional to the number of export contacts it has already in place. Moreover, there is a constant probability  $b$  that a new contact is assigned to a new firm.<sup>6</sup> Therefore there are  $bK(t)$  newly born firms (new goods traded). In each location, the entry rate of new firms is proportional to the total number of contacts. As a result, the number of contacts obeys the following differential equation

$$dK/dt = (\lambda - \gamma + b)K(t) \quad (1)$$

and the number of firms is given by

$$N(t) = (K(t) - K(0))b/(b + \lambda - \gamma) + N_0 \quad (2)$$

Thus, the system experiences an exponential growth if  $\lambda - \gamma + b > 0$ .

At time  $t$ , the value of export to a foreign contact is  $\nu(t)$ . The amount of goods shipped to each destination is independent of the number of contacts and depends on firm productivity and other variables—such as distance, trade barriers or market size—just as in a standard model.<sup>7</sup> At each time step, all trade relations experience a random (demand) shock  $\epsilon(t)$ , so that the value of trade is decreased or increased

---

<sup>4</sup>By differentiating the parameters of the model it is possible to generalize results for the asymmetric case.

<sup>5</sup>This assignment rule takes into consideration the relative attractiveness of each alternative origin/destination, and is therefore consistent with the need to account for “multilateral resistance”, as postulated by Anderson and van Wincoop (2003).

<sup>6</sup>The entry process described here is similar to the process of innovation modeled by Klette and Kortum (2004): indeed, new trading opportunities can be thought of as innovations. When applied at the country level, entry should not be taken literally as the creation of a new entity (clearly possible but quite rare), but rather as the entry of a country in a specific sub-market where it was not present before. The network would then be a multigraph formed by several layers, each one representing a different good, so that entry represents the commencement of operations in a new sector. This is in line with Melitz-type models where firms produce differentiated varieties of the same products, and the whole economy is made up by a collection of separate sectors.

<sup>7</sup>Indeed, one could easily assume that  $\nu(t)$  has the usual form taken by export revenues with CES preferences under monopolistic competition, i.e.  $\nu(t) = \tau^{1-\sigma} (\rho\varphi)^{\sigma-1} RP^{\sigma-1}$ , where  $\tau$  represents variable trade costs,  $\varphi$  is firm productivity,  $\sigma$  is the elasticity of substitution,  $\rho = (\sigma - 1)/\sigma$ ,  $R$  is income in the foreign country, and  $P$  summarizes the price of competing varieties.

by a random factor:  $\nu(t+1) = \nu(t)\epsilon(t)$ , where  $\epsilon(t)$  is a random variable with finite mean and variance.<sup>8</sup> This geometric Brownian motion (or Gibrat process) of growth for the value of trade yields a lognormal distribution of the intensive margin of trade at the level of single shipments. Upon aggregation, since the sum of lognormals is no more a lognormal, there can be some departures in the upper tail of the intensive margin for firms and countries.<sup>9</sup> However, for all practical means, the lognormal distribution remains a valid benchmark for the intensive margin of trade.

As for the extensive margin, Chaney (2011) has already demonstrated that the proportional growth process of contacts leads to a scale-free Pareto connectivity distribution.<sup>10</sup> In the continuum limit, the proposed growth mechanism gives rise to a master equation of  $p(k, t_j, t)$  which is the probability, for firm  $j$  born at time  $t_j$ , to have  $k$  contacts at step  $t$  (Yamasaki *et al.*, 2006):

$$\begin{aligned} \frac{\partial p(k, t_j, t)}{\partial t} = & \lambda \frac{(k-1)}{K(t)} p(k-1, t_j, t) + \\ & + \gamma \frac{(k+1)}{K(t)} p(k+1, t_j, t) - (\lambda + \gamma) \frac{k}{K(t)} p(k, t_j, t) \end{aligned} \quad (3)$$

where  $K(t) \equiv n + (\lambda - \gamma + b)t$  is the total number of contacts at time  $t$  and  $p(1, t_j, t_j) = 1$ .<sup>11</sup>

As discussed in Yamasaki *et al.* (2006), the analytical solution of Eq. (3) is given by

$$p(k, t) = \frac{n}{n + bt} p(k, 0, t) + \frac{b}{n + bt} \int_0^t dt_j p(k, t_j, t) \quad (4)$$

where the functional form of  $p(k, t_j, t)$  is given by

$$p(k, t_j, t) = (1 - \eta_{t_j, t}) \left(1 - \frac{\gamma}{\lambda} \eta_{t_j, t}\right) \eta_{t_j, t}^{k-1} \quad (5)$$

and

$$\eta_{t_j, t} = \frac{1 - \left(\frac{t_j + n(\lambda - \gamma + b)^{-1}}{t + n(\lambda - \gamma + b)^{-1}}\right)^{\frac{(\lambda - \gamma)}{(\lambda - \gamma + b)}}}{1 - \frac{\gamma}{\lambda} \left(\frac{t_j + n(\lambda - \gamma + b)^{-1}}{t + n(\lambda - \gamma + b)^{-1}}\right)^{\frac{(\lambda - \gamma)}{(\lambda - \gamma + b)}}} \quad (6)$$

---

<sup>8</sup>Hence, for what concerns the intensive margin we stick to the traditional way to account for trade intensity.

<sup>9</sup>For a discussion see Growiec *et al.* (2008)

<sup>10</sup>For the sake of simplicity, we do not consider random contacts in this version of the model. See Riccaboni and Schiavo (2010) for a generalized version of the model which includes random connections for both imports and exports. As in Chaney (2011), random contacts are responsible for an exponential lower tail of the connectivity distribution.

<sup>11</sup>Equation (3) is a generalization of the master equation of birth and death processes

We consider two limiting solutions of Eq. (3).

- *Case i* : No entry of new firms ( $b = 0$ ). The growth of the system is solely due to the preferential attachment of new contacts to the pre-existing firms. In this case

$$p(k) \sim e^{-\frac{kn_i}{t(\lambda-\gamma)}}. \quad (7)$$

This limiting case implies that the distribution of the number of contacts by firm is a pure exponential distribution.

- *Case ii* : At  $t = 0$  there is no firm  $n_i = 0$ , and new firms enter with probability  $b \neq 0$ . In this case, for large  $t$ ,  $p(k)$  is a Pareto distribution

$$p(k) \sim k^{-2}. \quad (8)$$

At  $t = 0$  the number of contacts for already-existing firms is  $g(t_0)$ . The growth rate of a new firm created at time  $t_0$  is proportional to  $1/g(t_0)$ . If we sort firms according to their connectivity, the rank  $R(k)$  of a firm  $j$  is proportional to the time of its creation ( $t_j$ ), i.e.  $R(k) \propto t_j$ . Thus  $k \sim 1/g(t_j) \sim 1/t_j \sim 1/R(t_j)$  and we arrive to the standard formulation of the Pareto law according to which the size of a firm  $k$  is inversely proportional to its rank. If we take into account the decrease of the growth rate with the influx of new firms, one can show after some tedious algebra that  $k \sim R^{-(\lambda-\gamma)/(\lambda-\gamma+b)}$ , which includes  $k \sim R^{-1}$  as a limiting case for  $b \rightarrow 0$ . Since  $R(k)$  is the number of firms whose connectivity is larger than  $k$ , we can write in the continuum limit  $R(k) \sim \int_k^\infty p(k)dk$  and hence  $p(k) \sim k^{-2-b/(\lambda-\gamma)}$  which is well approximated by a Pareto distribution when  $b$  is small. This limiting case can be considered as a different initial condition of the model where birth or death of firms are allowed starting from  $N = 0$ . This initial condition results in a pure Pareto distribution of the number of contacts for each firm. This case is identical to the Simon model (Ijiri and Simon, 1977) and leads to the same connectivity distribution as in Chaney (2011).

The full solution of Eq. (3), which is given by Eqs. (4)–(6), represents a power law with an exponential cutoff (see Yamasaki *et al.*, 2006), and can be interpreted as follows. We start with  $N_0$  firms and let newly born firms enter ( $b > 0$ ). Due to the preferential attachment rule, old firms have on average more links than new ones. Thus for large  $k$ ,  $p(k)$  is governed by the exponential distribution of the old firms (*Case i*) while for small  $k$ ,  $p(k)$  is governed by the Pareto distribution of the new ones (*Case ii*). In the limit of large  $t$ , the exponential cut-off vanishes away.

However, as in the case of Chaney (2011), the resulting distribution of the number of contacts is a mixture of two polar cases. Since each location has a different number of firms  $n_i$  to begin with, the connectivity distribution of firms at different locations will differ. Locations with a larger pool of local (old) firms exhibit a more pronounced exponential cut-off of the Pareto connectivity distribution. This feature of our model is crucial to correctly predict the actual distribution of international connection and to replicate the pattern of zeros observed in international trade data.

There are several alternative interpretations for the proposed model. The most literal one is that individual firms meet other firms, some of them are located in foreign markets as described in the model. In our model countries produce and consume a variety of products. Country  $i$  has a measure  $N_i$  of firms, each producing a distinct product, which is also different from the products exported by other countries. As in Melitz (2003), firm-level heterogeneity can be introduced with a stochastic technology generating process. However, as noticed by Chaney (2011), the Eaton *et al.* (2011) setup implies that the fraction of firms that export to exactly  $M$  markets inherits the Pareto distribution of productivities. Thus the distribution of market relationships comes from the Pareto distribution of productivity shocks. In our probabilistic setup a Pareto distribution of contacts is generated by means of a preferential attachment mechanism. Chaney (2011) provides some network-based argument in favor of a proportionate growth process. However, at least in principle, this process could be related to some idiosyncratic property such as a proportionate growth of firm level productivity.

## 4 Results

The model yields a number of analytic results that are consistent with the data. First of all, for instance, in the previous section we have shown how the model predicts the connectivity distribution (i.e. the extensive margin) follows a power-law with an exponential cut-off: Figure 1 confirms this prediction. The main plot displays the probability distribution in log-log scale, whereby the power-law is the straight line body, and the exponential cutoff is represented by the right tail. The inset presents the same distribution in semi-log scale: this time it is the exponential part of the distribution that becomes a straight line, so that with this trick we can magnify what happens to the probability distribution as  $K$  grows large.

In addition, we run a series of simulations of the model in order to collect additional evidence on the ability of the model to correctly predict empirical patterns

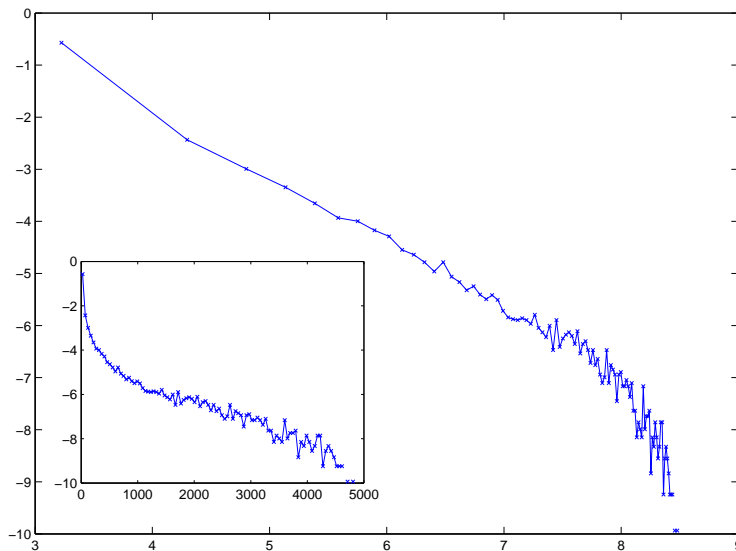


Figure 1: Complementary cumulative distribution of the number of trade links by country: main plot in double log scale, inset in semi-log scale. NBER-UN data for 1997.

emerging in the data.<sup>12 13</sup>

## 4.1 Extensive margin

The extensive margin of trade refers to the distribution of a set of  $n$  products into a discrete set of  $m$  economic entities (countries). In probabilistic theory the models of allocation of  $n$  objects into  $m$  sets are called urn models. For ease of exposition, we speak of  $n$  “balls” for products and  $m$  “bins” for countries (as in Armenter and Koren, 2010).

We are interested in the distribution of products by country obtained by counting the number of configurations of  $n$  randomly distributed balls in  $m$  bins. First, the number of different patterns crucially depends on whether balls and bins are distinguishable. In this context, the identity of balls does not matter. Thus, we

---

<sup>12</sup>Here we follow almost literally the idea that under monopolistic competition every firm produces and exports a single differentiate variety. As a result we can establish a proportionality between the number of products traded (at different levels of aggregation) and the number of contacts established. This allows us to circumvent the fact that detailed firm-level information on the number of trading partners are not available to us.

<sup>13</sup>Let us note in passing that the model captures other empirical regularities —such as the distribution of the growth rates of trade values, and the relationship between total trade and the variance of its growth rates (Riccaboni and Schiavo, 2010)— that, although not investigated in the economic literature, characterize trade data.

consider the case of indistinguishable units of trade.<sup>14</sup> This is equivalent to the partition of  $n$  objects into  $m$  sets. With indistinguishable balls two distributions are different only if the corresponding  $m$ -tuples  $(n)$  are not identical. Feller (1950, 38) shows that the number of distinguishable allocations is given by

$$A_{n,m} = \binom{n+m-1}{m-1}$$

while the number of distinguishable allocations in which no bin is empty is  $B_{n,m} = \binom{n-1}{m-1}$ . Thus,  $B_{n,m}/A_{n,m}$  tends to 1 by increasing  $n$  with constant  $m$ .

Armenter and Koren (2010) estimate from the data the distribution of the number of shipments and then generate predictions for the occurrence of zeros in the trade data. By contrast, instead of assuming this distribution to match the data, our model offers a theory to generate such a distribution. In this respect, our approach is similar in spirit to Chaney (2011). Here we compare the prediction of our model with the results presented in Armenter and Koren (2010) and Shi (2010) who use shipment data for US and Chinese trade, respectively.

We generate the number of zeros at different levels of aggregation from 10 to 2 digits of the Standard International Trade Classification. We treat any product-destination pair at the most disaggregated level of the classification as a specific connection in our network model. In simulating our model for the two cases we use only one free parameter. For the US and Chinese case, we must set the number of product classes in which firms are already active at the beginning of the simulation ( $t = 0$ ). This is equivalent to define the number of firms at a given location in our model ( $n_i$ ). Despite the lack of data does not allow us to test the model at the firm level, a nice property of our setup is that it can be rescaled to any level of aggregation of the system. By knowing the number of lower  $l$ -digit level sectors in which firms are present ( $n_i^l$ ) we can easily predict the number of sectors at the upper  $u$ -digit level and vice-versa by applying a the following probabilistic relationship:

$$n_i^u = 1 - (1 - n_i^l)^{p_l/p_u}. \quad (9)$$

Table 1 compares the real distribution of zeros with predictions of our model (RS) and the ones by Armenter and Koren (AK). Clearly our model provides a better prediction of the percentage of zeros than the AK model at all levels of aggregation and for both countries. Moreover, as shown in Table 2 our model reasonably approximate even the distribution of non-zero entries whereas the AK

---

<sup>14</sup>We may still suppose the shipments to be labeled, but focus our attention on events that are independent of the labeling.

Table 1: Comparing different stochastic models

country	level	products	destinations	zeros (%)			$N_0$ (%)
				data	AK <sup>†</sup>	RS <sup>‡</sup>	
USA	section	21	229	16	10	16	71.42*
	2-digit	97	229	36	23	36	15.46*
	4-digit	1244	229	66	52	66	1.21*
	6-digit	5182	229	79	68	79	0.29*
	10-digit	8867	229	82	72	82	0.18
China	2-digit	97	233	39	24	39	9.28*
	4-digit	1226	233	67	53	65	0.72*
	6-digit	4995	233	78	67	78	0.17*
	8-digit	7054	233	81	63	79	0.10

<sup>†</sup> zeros predicted by Armenter and Koren's model

<sup>‡</sup> zeros predicted by our own model

\* computed by using equation Eq. 9

data for the US taken from Armenter and Koren (2010, Tab. 2)

data for China taken from Shi (2010, Tab. 11)

model does not replicate this property of trade systems. All in all, our probabilistic setup provides a parsimonious explanation for distribution of zeros and the extensive margin of trade. By knowing the initial set of goods traded by a given country and assuming that the capability to innovate by capturing new trading opportunities and to trade new products is proportional to the volume of trade, we generate a network of trade in which zero trade relationships have the same sparsity structure of the actual trade network.

Table 2: Number of shipments across product-destination categories

number of shipments	frequency (%)	
	data	RS model
0	82.0	82.0
1	5.2	5.2
2	2.3	2.7
3	1.4	2.0
4	1.0	1.5
5	0.7	1.2
6-9	1.8	2.6
> 10	5.7	2.9
TOT	100	100

data from Armenter and Koren (2010, Tab. 2)

## 4.2 Intensive margin

The assumption of proportional growth governing the dynamic of bilateral trade flows implies a lognormal distribution for the intensive margin of trade. This appears consistent with previous findings (Bhattacharya *et al.*, 2008; Fagiolo *et al.*, 2009),

and with the observation that trade flows are highly concentrated (Easterly *et al.*, 2009).

Here we estimate the parameters of the lognormal distribution by both maximum likelihood (as in Bee, 2006) and weighted least-squares (this estimator minimizes the distance between the empirical and the theoretical cumulative distribution; see Swain *et al.*, 1988). The latter method gives more weight to the tails of the distribution and yields our preferred estimates for the parameters of lognormal distribution ( $\mu = 6.326$  and  $\sigma = 2.213$ ).<sup>15</sup>

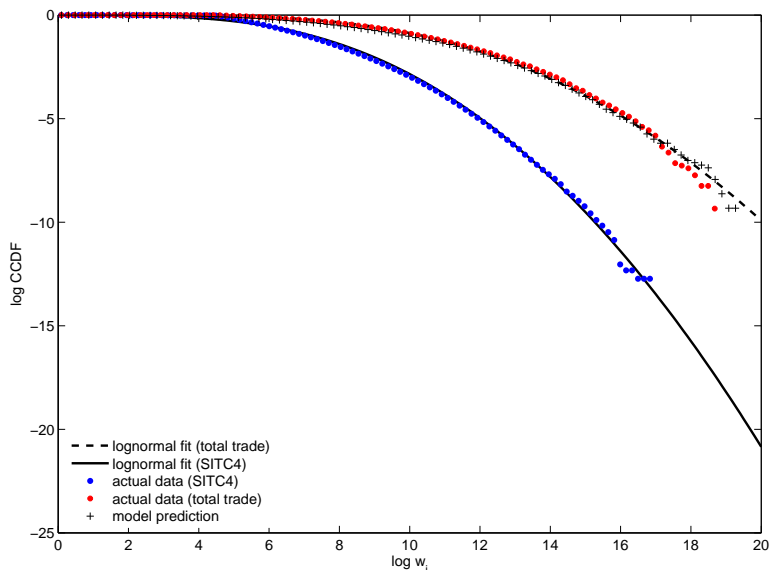


Figure 2: Complementary cumulative distribution of trade values (double log scale, data for 1997). Model predictions obtained by summing a number of values equal to the predicted extensive margin ( $K_i$ ) drawn from the lognormal distribution describing 4-digit trade flows.

Figure 2 displays the complementary cumulative distribution of bilateral trade flows at the 4-digit level, together with a lognormal fit.<sup>16</sup> In addition we plot the implied distribution of total trade flows that is obtained by summing up, for each country  $i$ ,  $K_i$  random draws from the underlying lognormal, where  $K_i$  is the number of trade relations maintained by country  $i$  according to the stochastic model (i.e. the number of nonempty bins associated with country  $i$ ). First of all the figure shows

<sup>15</sup>A Kolmogorov-Smirnov test rejects the hypothesis of normality for the log of bilateral trade flows, whereas an Anderson-Darling test cannot reject the null hypothesis of normality. As the latter statistic is known to be more sensitive to discrepancies in the tails of the distribution, and as we are mainly concerned about the tail behavior of the data that determines the high concentration of trade, we give more credit to this latter test.

<sup>16</sup>The CCDF is the probability that an observation is larger than a given value  $P(X > x)$  and is the complement to one of the cumulative distribution CDF.



that the lognormal provides a good fit of the disaggregated data and, moreover, the aggregate results obtained by combining the intensive and extensive margins implied of the model brings us very close to actual trade flows.

The data display some departure from the lognormal fit in the right tail of the distribution: in particular, the model appears to overpredict the number of very large bilateral export flows. This effect persists over different samples, datasets, and levels of aggregation and hence it seems not to depend on specific feature of the data such as truncations or measurement errors.<sup>17</sup> One simple explanation for the departure from lognormality is that this represents the underlying process at the microeconomic level, but what is actually observed, even at very fine levels of aggregation, is an aggregation whose distribution comes from a sum of lognormal, which is therefore not lognormal itself (and for which there exists no closed form solution).<sup>18</sup> Complementary explanations are possible, in particular the existence of market frictions that prevent firms and countries from fully exploiting their potential and therefore reduce the value of trade flows.

A further interesting feature of lognormality is that it implies that most measures of concentration —such as entropy, herfindhal, C5— can be written as simple functions of just two parameters, namely the number of elements of the distribution ( $n$ ) and the variance of the underlying normal distribution ( $\sigma^2$ , see Hart, 1971; Davies, 1979). So, for instance, the C5 concentration ratio, i.e. the cumulative share of total export of a country commanded by the 5 most popular export categories, can be expressed under lognormality as  $N(z_1 | 0, 1)$  where  $N(0, 1)$  is the standard normal distribution,  $z_1 = \sigma + z_2$ , and  $z_2$  is computed such that  $prob(x < z_2) = \frac{5}{n}$  from the standard normal tables.<sup>19</sup> Similarly, the Herfindahl index ( $\sum_{n=1}^N s_n^2$ ) takes the form  $e^{\sigma^2}/n$  under lognormality, whereas entropy ( $-\sum_{i=1}^n s_n \log s_n$ ) boils down to  $\log n - \frac{1}{2}\sigma^2$ .

Since in our setup the number of zero and nonzero elements are strictly related and depends on the entry rate, as discussed above, lognormality implies that the degree of concentration is fully determined by the dispersion (variance) of trade values and the entry rate in the stochastic process characterizing the extensive margin.

---

<sup>17</sup>In addition to various years of the NBER-UN dataset we have used Comtrade data (8-digit HS) for 2007 and we even experimented with the disaggregated data for US exports maintained by the US Census. Additional information about these results are available upon request.

<sup>18</sup>Recent evidence shows that the thickness of the right tail of a distribution increases upon aggregation (Bee *et al.*, 2012), and this holds across different domains. This means that upon aggregation the probability to observe larger than expected outcomes tends to increase.

<sup>19</sup>This formula, as well as the following ones concerning Herfindahl and Entropy are taken from Davies (1979).

Table 3 reports actual concentration measures computed from 4-digit SITC trade flows (NBER-UN data), alongside with the values the different indexes (C5, Herfindahl, Entropy) take when we use the formulas that apply under lognormality and the corresponding figures for  $n$  and  $\sigma$ .

Table 3: Concentration measures: actual and predicted

index	actual	predicted
C5	1.707%	1.520%
Herfindahl	0.020%	0.018%
Entropy	10.975	10.775

From the table we see that albeit high, actual concentration is always lower than its predicted counterpart, irrespective of the index of choice. From a mechanical point of view, this result depends on the underrepresentation of very large flows in the data relative to the predictions of the model that we have observed above. At the same time, the difference between actual and predicted concentration represents a good example of a significant departure from the stochastic benchmark that calls for an economic interpretation. We conjecture that remaining barriers to trade (distance, transport costs, nontariff barriers) can provide an explanation to this departure, as well as diseconomies of scale whereby the value of a single bilateral flow is constrained by firm size, input availability or other technical considerations.<sup>20</sup>

### 4.3 Gravity model

When it comes to analyzing trade flows the standard empirical tool is the gravity model, whereby (nonzero) bilateral flows are positively correlated with the size (mass) of the countries involved in the exchange, normally captured by GDP and/or population, and negatively to their distance (which proxies for trade costs).<sup>21</sup> Hence, given a sample of countries, using information on their GDP and their bilateral distance it is possible to estimate a benchmark level of bilateral trade among each country pair and, from this, replicate the global structure of international trade flows.

Over the years the gravity specification has acquired solid theoretical foundations (see for instance Anderson, 1979; Anderson and van Wincoop, 2003; Chaney, 2008;

---

<sup>20</sup>In a similar vein, the literature finds that many economic networks are characterized by a hub-and-spoke structure whereby peripheral nodes tend to link themselves with hubs, and vice-versa. Schiavo *et al.* (2010) reports this holds true for the international trade network as well. Given the relationship between extensive and intensive margin discussed below, such feature provides another explanation for lower-than-predicted concentration.

<sup>21</sup>Empirical analysis usually adds a host of other control variables such as common borders, cultural ties, WTO membership, trade agreements, common currencies and the like.

Helpman *et al.*, 2008), whereas our own endeavor is limited to presenting a stochastic setup, not a fully-fledged explanation of the determinants of bilateral trade flows. A further difference stems from the kind of information required to feed the two models and thus generate the full structure of world trade. In the case of the gravity model, in fact, distance and country size (GDP) both represent a required input. On the contrary, our setup does not need any information of the sort: our ‘bins’ (countries) are weightless or, better, their size is an outcome of the model rather than an input to it. In this respect we can claim that our stochastic setup represents a more parsimonious benchmark. This section aims at comparing the two models and, in particular, to see whether the structure of bilateral flows generated by our setup displays gravity-like features despite the fact that we make no reference to either economic mass or distance among trading partners.<sup>22</sup>

Results are summarized in Table 4. Column (A) displays estimated coefficients from a plain-vanilla OLS regression of aggregate bilateral trade on exporting and importing country GDP and distance (all in logs) using actual data. Aggregate trade flows are taken from the NBER-UN database documented in Feenstra *et al.* (2005), while distance comes from the CEPII database.<sup>23</sup> All coefficients have the expected sign and their magnitude is in line with the literature; this simple specification explains more than 60% of the variation in the data and represents our reference point.

In column (B) the dependent variable is no longer actual trade; rather, for each country pair  $i, j$ , we combine the actual number of products exported with an equal number of random draws from the estimated lognormal distribution. Hence, if, say, Italy ships 10 (4-digit) products to France, total trade in column (B) is computed as the sum of 10 values sampled from a lognormal distribution with mean  $\mu = 6.326$  and standard deviation  $\sigma = 2.213$  (see Section 4.2) Given the double-log specification of the gravity model, sampling the intensive margin from a lognormal distribution is tantamount to adding white noise to the extensive margin  $K_{ij}$ . The lower  $R^2$  in Column (B) comes therefore as no surprise; apart from this, estimated coefficients are significant and have the expected sign.

To better see the effect of associating the actual extensive margin with random values for the intensive one, Column (C) reports results of a gravity model run on the number of products shipped by country  $i$  to country  $j$ , i.e. using the sole extensive margin as the dependent variable. The difference between specifications

---

<sup>22</sup>This is a major departure from the bins-and-balls model developed by Armenter and Koren (2010), who impose a distribution of bin size and thus end up with a built-in relationship between trade and economic mass that remembers a gravity specification.

<sup>23</sup>Available at <http://www.cepii.fr/anglaisgraph/bdd/distances.htm>

Table 4: Gravity-type regression results

	(A)	(B)	(C)	(D)	(E)	(F)
dependent variable	actual trade	actual $K_{ij}$ random $w_{ij}$	actual $K_{ij}$ –	actual average $w_{ij}$	model $K_{ij}$ random $w_{ij}$	actual trade
const	-1.95***	2.09***	-4.21***	2.26***	1.80***	-2.33***
GDP exp	0.93***	0.71***	0.70***	0.22***	0.76***	0.77***
GDP imp	0.86***	0.50***	0.49***	0.37***	0.51***	0.75***
distance	-0.93***	-0.74***	-0.73***	-0.20***	-0.78***	-0.76***
balls-and-bins						0.21***
adjusted $R^2$	0.62	0.28	0.66	0.29	0.29	0.66
observations	11192	11192	11192	11192	11192	11192

\*\*\* indicates significance at 1% level

(B) and (C) lies therefore only in the random component  $w_{ij}$ , that is absent in the latter case. Estimated coefficients are basically the same, but removing the random noise improves the fit of the regression, bringing it back to the values seen in the baseline case (A).

On the other hand, when we only use the (actual) intensive margin (average  $w_{ij}$ ) as in Column (D), the fit is much worse, with GDP and distance explaining less than 30% of total variation in the data. Also, estimated coefficients are lower, though they remain significant and consistent with theoretical expectations. This result is in line with the finding that gravity effects are more pronounced for the extensive margin of trade. In particular, the fact that the distance elasticity is larger for the extensive margin is consistent with Chaney (2008) and Lawless (2010).

Column (E) displays results obtained using the trade flows generated by the stochastic model as the dependent variable. Note that here we need a further assumption in order to be able to label our bins, i.e. to assign a value for GDP to each location (and a value for distance to each pair). We assign country labels on the basis of the ranking of countries in terms of the number of export contacts, i.e. on the basis of  $K_i$ . This means that the country with the largest number of nonempty bins is assigned the largest GDP value, and so on following the ranking. In so doing we are therefore imposing a sort of proportionality between the number of trade links of a country and its GDP.<sup>24</sup> Estimated coefficients are significant, close to those obtained in Column (B), and not dissimilar from those of the baseline case.

The driving force behind this results is the ability of the stochastic model to closely match the data in terms of number of product traded, as discussed in Section 4.1 above. As before, sampling the values of bilateral flows from a lognormal implies

<sup>24</sup>The relationship between GDP and the number of trade partners is discussed at length by Garlaschelli and Loffredo (2004) in the context of network analysis. At industry- or product-level De Benedictis *et al.* (2009) find that export diversification increases monotonically with income, whereas Cadot *et al.* (2011) find evidence of re-concentration at very high levels of income.

adding white noise to the gravity model and therefore reduces the explanatory power of the equation. A further explanation for the lower value of  $R^2$  with respect to the baseline case comes from the fact that in the model we assume a single lognormal distribution for all exporting countries, while one can reasonably expect high-income countries to export higher value-added products and thus to draw from distributions characterized by higher mean values. Hence, although large exporters draw many values from the lognormal, and this bids up both total and average trade, this effect is not large enough to capture the relationships between income level and trade volumes implied by the gravity model. This is a good example of an ‘economically relevant’ empirical finding that cannot be fully explained by our simple stochastic model and requires further economic modeling.

Last, in column (F) we revert to actual data on the left-hand side of the gravity equation, as in the baseline case (A), but add the value of bilateral trade implied by the model among the control variables. In so doing we are controlling for the stochastic component of the gravity model. The effect of this additional control (dubbed *balls – and – bins* in Table 4) is first of all to improve the fit, by raising the adjusted- $R^2$  from 0.62 to 0.66, and second to reduce the values of the estimated coefficients.

All in all our simple stochastic setup does a good job in replicating a gravity-type structure, especially when we consider that both country GDP and distance are remarkably absent from the model.

## 5 Discussion and conclusions

The paper proposes a simple stochastic model based on proportionate growth to describe the network structure of bilateral trade flows. The main contribution of our framework is to provide a parsimonious representation of economic interaction in a multi-country world, achieved by extending the proportionality law long used to describe the intensive margin of trade (e.g. the gravity model), to apply to the extensive margin as well. To do so, we employ a discrete version of proportionate growth that builds on urn models à la Simon (1955) and, more recently, on network models such as Barabási and Albert (1999).

We show that this setup is able to adequately replicate many of the relevant stylized facts described in the literature, that are not well explained by existing models of international trade. Even more relevant is the ability of the model to match several regularities simultaneously. In particular, the paper replicates the large number of zeros found in trade data at various levels of disaggregation, generates a

very skewed distributions for both the intensive and the extensive margin of trade, and confirms that the latter accounts for a large fraction of greater exports from large economies. Indeed, the preferential attachment implies a degree of persistence in the relative connectivity of countries, whereby “large” countries are those with a large number of links. The very skewed distribution of trade values reinforces this feature as it is very unlikely (albeit possible) that countries with few links feature very large export flows (the so-called “big hits”).

The ability of the model to accurately describe the data depends crucially on two parameters: the rate of entry of new firms into trade, and the variance of the lognormal distribution from which the value of bilateral flows (the intensive margin) is drawn at the microeconomic level.

A further implication of the model is that the various stylized facts are strongly interconnected. So, for instance, the number of zeros in trade data is related to concentration in trade flows, whose skewed distribution in turn determines the relationship between the intensive and the extensive margin. Hence, international trade theories should not consider the different stylized facts in isolation, but rather tackle them together.

The paper shows that the same underlying process is capable of matching data at different levels of aggregation. Our results can be interpreted also as a stochastic benchmark against which to assess the ability of competing theories to explain reality: departures from this benchmark signal the presence of interesting economic phenomena that are worth further theoretical exploration. The most relevant departures from the model predictions that we observe in the data concern the right tail of the distribution of trade values, that is underrepresented in the data, and trade concentration, which is lower than predicted. They signal the presence of trade frictions that prevent firms and countries to fully exploit their trade potential and the possibility that large traders draw their trade values from distributions characterized by different parameters.

We consider the paper as a first step toward the integration of stochastic mechanisms into standard economic modeling (similar to what has been done in other fields for instance by Sutton, 1998; Klette and Kortum, 2004). Having shown that a simple stochastic setup can effectively match many of the relevant stylized facts, the paper offers a promising first step in this direction. To take this route, the two key parameters of the model, i.e. the entry rate and the variance of the lognormal distribution, need to be more firmly grounded into economic theory. From an empirical point of view, access to firm-level data would allow one to more closely test the model predictions, moving their validation one step closer to the theoretical setup.

## References

- Anderson, J., 1979. A theoretical foundation for the gravity equation, *American Economic Review*, 69, 106–116.
- Anderson, J. and van Wincoop, E., 2003. Gravity with gravitas: A solution to the border puzzle, *American Economic Review*, 93, 170–192.
- Anderson, J.E., 2011. The gravity model, *Annual Review of Economics*, 3 (1), 133–160.
- Arkolakis, C., Costinot, A., and Rodríguez-Clare, A., 2013. New trade model, same old gains?, *American Economic Review*.
- Armenter, R. and Koren, M., 2010. A balls-and-bins model of trade, Discussion Papers 7783, CEPR.
- Baldwin, R. and Harrigan, J., 2011. Zeros, quality and space: Trade theory and trade evidence, *American Economic Journal: Microeconomics*, 3 (2), 60–88.
- Barabási, A.L. and Albert, R., 1999. Emergence of scaling in random networks, *Science*, 286 (5439), 509–512.
- Bee, M., 2006. Estimating the parameters in the loss distribution approach: how can we deal with truncated data?, *in*: E. Davis, ed., *The advanced measurement approach to operational risk*, London: Riskbooks, 123–144.
- Bee, M., Riccaboni, M., and Schiavo, S., 2012. A trick of the (pareto) tail, Working Papers 1206, Department of Economics, University of Trento.
- Bergstrand, J.H. and Egger, P., 2011. Gravity equations and economic frictions in the world economy, *in*: D. Bernhofen, R. Falvey, D. Greenaway, and U. Krieckemeier, eds., *Palgrave Handbook of International Trade*, Palgrave-Macmillan Press.
- Bernard, A., Eaton, J., Jensen, J.B., and Kortum, S., 2003. Plants and productivity in international trade, *American Economic Review*, 93 (4), 1268–1290.
- Bernard, A., Jensen, B., Redding, S., and Schott, P., 2007. Firms in international trade, *Journal of Economic Perspectives*, 21 (3), 105–130.
- Bernard, A., Jensen, J.B., Redding, S., and Schott, P., 2009. The margins of us trade, *American Economic Review*, 99 (2), 487–93.

- Besedes, T. and Prusa, T.J., 2010. The role of extensive and intensive margins and export growth, *Journal of Development Economics*, In Press, Accepted Manuscript, –.
- Bhattacharya, K., Mukherjee, G., Saramaki, J., Kaski, K., and Manna, S.S., 2008. The international trade network: weighted network analysis and modelling, *Journal of Statistical Mechanics: Theory and Experiment*, 2008 (02), P02002.
- Cadot, O., Carrière, C., and Strauss-Kahn, V., 2011. Export diversification: What’s behind the hump?, *The Review of Economics and Statistics*, 93 (2), 590–605.
- Chaney, T., 2008. Distorted gravity: The intensive and extensive margins of international trade, *American Economic Review*, 98 (4), 1707–21.
- Chaney, T., 2011. The network structure of international trade, Working Papers 16753, NBER.
- Davies, S., 1979. Choosing between concentration indices: The iso-concentration curve, *Economica*, 46 (181), 67–75.
- De Benedictis, L., Gallegati, M., and Tamberi, M., 2009. Overall trade specialization and economic development: countries diversify, *Review of World Economics*, 145, 37–55.
- Easterly, W., Reshef, A., and Schwenkenberg, J., 2009. The power of exports, Policy Research Working Paper Series 5081, World Bank.
- Eaton, J., Kortum, S., and Kramarz, F., 2004. Dissecting trade: Firms, industries, and export destinations, *American Economic Review*, 94 (2), 150–154.
- Eaton, J., Kortum, S., and Kramarz, F., 2011. An anatomy of international trade: Evidence from french firms, *Econometrica*, 79 (5), 1453–1498.
- Eaton, J., Kortum, S., and Sotelo, S., 2012. International trade: Linking micro and macro, Tech. rep., National Bureau of Economic Research.
- Fagiolo, G., Reyes, J., and Schiavo, S., 2009. World-trade web: Topological properties, dynamics, and evolution, *Physical Review E*, 79 (3), 036115.
- Fagiolo, G., Schiavo, S., and Reyes, J., 2008. On the topological properties of the world trade web: A weighted network analysis, *Physica A*, 387 (15), 3868–3873.



- Feenstra, R.C., Lipsey, R.E., Deng, H., Ma, A.C., and Mo, H., 2005. World trade flows: 1962–2000, Working Papers 11040, NBER.
- Gabaix, X., 1999. Zipf’s law for cities: An explanation, *Quarterly Journal of Economics*, 114, 739–67.
- Garlaschelli, D. and Loffredo, M., 2004. Fitness-dependent topological properties of the world trade web, *Physical Review Letters*, 93, 188701.
- Growiec, J., Pammolli, F., Riccaboni, M., and Stanley, H.E., 2008. On the size distribution of business firms, *Economics Letters*, 98 (2), 207–212.
- Hart, P., 1971. Entropy and other measures of concentration, *Journal of the Royal Statistical Society*, 134 (1), 73–85.
- Helpman, E., Melitz, M., and Rubinstein, Y., 2008. Estimating trade flows: Trading partners and trading volumes, *Quarterly Journal of Economics*, 123 (2), 441–487.
- Hummels, D. and Klenow, P.J., 2005. The variety and quality of a nation’s exports, *American Economic Review*, 95 (3), 704–723.
- Ijiri, Y. and Simon, H., 1977. *Skew Distributions and the Sizes of Business Firms*, Amsterdam: North Holland.
- Klette, T.J. and Kortum, S., 2004. Innovating firms and aggregate innovation, *Journal of Political Economy*, 112 (5), 986–1018.
- Lawless, M., 2010. Deconstructing gravity: Trade costs and extensive and intensive margins, *Canadian Journal of Economics*, 43 (4), 1149–1172.
- Luttmer, E., 2007. Selection, growth, and the size distribution of firms, *The Quarterly Journal of Economics*, 122 (3), 1103–1144.
- Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity, *Econometrica*, 71 (6), 1695–1725.
- Riccaboni, M. and Schiavo, S., 2010. The structure and growth of weighted networks, *New Journal of Physics*, 12 (023003), 1–14.
- Savage, I.R. and Deutsch, K.W., 1960. A statistical model of the gross analysis of transaction flows, *Econometrica*, 28, 551–572.
- Schiavo, S., Reyes, J., and Fagiolo, G., 2010. International trade and financial integration: a weighted network analysis, *Quantitative Finance*, 10 (4), 389–399.

- Sen, A. and Smith, T.E., 1995. *Gravity Models of Spatial Interaction Behavior*, Berlin: Springer.
- Serrano, A. and Boguñá, M., 2003. Topology of the world trade web, *Physical Review E*, 68, 015101(R).
- Shi, H., 2010. The missing trade of China: balls and bins model, mimeo, Department of Economics, The Ohio State University,.
- Simon, H.A., 1955. On a class of skew distribution functions, *Biometrika*, 42 (3/4), 425–440.
- Sutton, J., 1998. *Technology and Market Structure*, Cambridge, MA: MIT Press.
- Swain, J., Venkatraman, S., and Wilson, J., 1988. Least-squares estimation of distribution functions in Johnson’s translation system, *Journal of Statistical Computation and Simulation*, 29 (4), 271–297.
- Yamasaki, K., Matia, K., Buldyrev, S., Fu, D., Pammolli, F., Riccaboni, M., and Stanley, H., 2006. Preferential attachment and growth dynamics in complex systems, *Physical Review E*, 74 (3), 035103R.

# Appendices

## A Discussion of the stylized facts

### Extensive margin

**Stylized fact 1.** Helpman *et al.* (2008) use the NBER-UN trade database (Feenstra *et al.*, 2005) to show that country-pairs not trading at all among themselves represent around 50% of the data, so that on average half of the potential trade links are never activated. Second, trade in both directions account for just around 30 to 40% of exchanges, with the remaining fraction of international trade due to transactions going in one direction only (country A exporting to country B but not vice-versa). The pervasiveness of zeros increases the higher the degree of disaggregation: Baldwin and Harrigan (2011) and Armenter and Koren (2010) look at US exports flows at the HS-10 level and report a share of zeros higher than 80% of all the potential trade flows for 2005. Similarly, Easterly *et al.* (2009) focus on product-destinations pairs and, using COMTRADE data at the HS-6 level for 151 countries, find a share of zeros relative to the number of potential flows that range between 69% and 99.5%, with a mean value of 96% and a median of 98.5%.

**Stylized fact 2.** Trade is highly concentrated. This can be appreciated by looking at the distribution of the number of destinations served by each exporting country, the number of products shipped, or the number of product-destination pairs. Moreover, this feature cuts across different levels of aggregation. The distribution of the number of trade linkages maintained by each country is found to be very skewed in many studies using different datasets, and such a feature appears to be rather stable over time (Serrano and Boguñá, 2003; Garlaschelli and Loffredo, 2004; Fagiolo *et al.*, 2008; Easterly *et al.*, 2009). Riccaboni and Schiavo (2010) provide similar evidence analyzing the number of products and product-destination pairs. Eaton *et al.* (2004) report similar evidence for data on French exporting firms.

### Intensive margin

**Stylized fact 3.** Easterly *et al.* (2009) investigate the concentration of export flows and find that the most popular export products account for a disproportionate share of a country's total export. As an example, they find that for the median country in their sample, the most popular 1% of products (at the HS-6

level) account for 47% of total export, with the number growing as high as 86% when one looks at the most exported 10% of products.<sup>25</sup> Similarly, when looking at product-destination pairs, the top 1% accounts for 52% of exports and the top 10% for 89% of exports in the median country.<sup>26</sup>

**Stylized fact 4.** Highly concentrated trade results in a very skewed distribution of export flows, with a small fraction of very large trade relationships existing alongside with a large number of very small links. A large body of work has documented this feature of trade data, and although there is no consensus on the actual shape of the distribution (candidates being the Pareto, the log-normal, or mixtures of the two), different authors agree on its highly skewed shape (Bhattacharya *et al.*, 2008; Fagiolo *et al.*, 2009; Easterly *et al.*, 2009).<sup>27</sup>

### Relationship between intensive and extensive margin

**Stylized fact 5.** Hummels and Klenow (2005) start from the observation that large countries export more, and conclude their analysis by claiming that 60% of this is due to the extensive margin, i.e. to the fact that large countries export a wider set of goods. A similar conclusion is reached by Bernard *et al.* (2009) who conclude that variations in trade flows across countries are primarily due to extensive margins.

**Stylized fact 6.** A complementary finding to the previous stylized fact is that the variation of trade flows over time is mainly driven by the intensive margin, i.e. the growth in the intensity of existing relationships. Helpman *et al.* (2008) report that since the 1970s, the rapid growth of world trade was predominantly due to the intensification of trade among countries that already traded with each other, rather than to the development of new trade relationships. Similarly, Besedes and Prusa (2010) look at the relative contribution of extensive and intensive margins to export growth and find that the former has only short-term impact, and almost no impact at all in the long term.

---

<sup>25</sup>The top 1% ratio ranges from a low of 18% to a high of 92%, whereas the top 10% ratio range from 43% to 99%.

<sup>26</sup>Minimum and maximum values are 20% and 85% for top 1%, 53% and 99% for top 10%.

<sup>27</sup>The debate is on whether the value of bilateral export flows is distributed as a power law or as a lognormal, which are both very skewed distributions.



INSTITUTE FOR ADVANCED STUDIES LUCCA