UNIVERSITY
OF TRENTO - Italy

Department of Economics

The Structure and Growth of International Trade

Massimo Riccaboni, Stefano Schiavo

n.
5/2010

The Structure and Growth of International Trade

Massimo Riccaboni[*]   Stefano Schiavo[†]

This version — March 2010

Abstract

The paper develops a model of proportionate growth to describe the dynamics of international trade flows. We show that a large number of the empirical regularities characterizing international trade —such as the fraction of zero trade flows across pairs of countries, the positive relationship between intensive and extensive margins, the high concentration of trade with respect to both products and destinations, the core-periphery structure of exchanges— are well explained by this simple stochastic setup. This helps us to distinguish among economically relevant regularities and those simply resulting from the mechanical interactions among agents. Furthermore, our model can be used to describe the process of 'self-discovery' that lie at the foundations of successful export-led growth and is thought to play a crucial role in the process of economic development. Our model correctly predicts that large export flows are rare events, as pointed out in the empirical literature: yet, countries characterized by large 'discovery' efforts are much more likely to draw a 'big hit' due to the (very skewed) shape of the distribution of bilateral export flows.

JEL Codes: F14, F43, O25

Keywords: international trade, development, weighted networks, proportionate growth, industrial policy

[*]Department of Computer and Management Sciences, University di Trento (Italy), IMT – Institute for Advanced Studies, Lucca (Italy) and Department of Physics, Boston University (United States). E-mail: massimo.riccaboni@unitn.it

[†]Department of Economics, University of Trento (Italy) and OFCE-DRIC (France). E-mail: stefano.schiavo@unitn.it

1 Introduction

We present a simple stochastic model of proportionate growth to describe international trade flows as a set of transactions of different magnitude occurring among countries, and we test it using both simulations and real data. With this simple setup we combine elements coming from different streams of the literature that, albeit related, have so far progressed in a parallel way and seldom interacted.

The use of stochastic models to assess the economic relevance of a given phenomenon and establish a benchmark against which measure its magnitude enjoys a long tradition in the industrial organization literature (Simon, 1955; Ijiri and Simon, 1977; Sutton, 2007), and has more recently been successfully applied to measure the degree of geographic concentration of economic activities by Ellison and Glaeser (1997) and Guimarães et al. (2009). The paper that is closer to ours in spirit is Armenter and Koren (2008) who develop a simple stochastic model that accounts for the large number of zeros appearing in any matrix of disaggregated bilateral trade flows. They describe US exports as a series of balls falling into bins of different size each representing a product-destination pair. In spite of its simplicity Armenter and Koren (2008) show that their model has a rich set of predictions that match a large number of stylized facts concerning US trade. They claim that, paradoxically, the best way such a setup can inform economic theory is by *missing* an empirical fact, as it signals that the latter is not the mere result of mechanic interactions, but rather the outcome of choices and decisions fed by economic principles.

In what follows 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, which is represented as a network of bilateral links of different weights among countries. We show that our simple setup is capable of matching many of the empirical regularities characterizing world trade, so that not all of them appear to be economically meaningful: this is to say that competing economic theories should not be judged on the basis of their ability to match the facts that are well explained by stochastic interactions of agents.

Our work is part of a larger trend involving the study of the empirical regularities characterizing international trade flows, and the development of theoretical models capable of explaining the stylized facts that are puzzling for the existing literature. So, for instance, the sparse nature of trade data, i.e. the large fraction of zero product-level trade flows has been receiving a good deal of attention in recent years. Baldwin and Harrigan (2007) look at 10-digit Harmonized System (HS) US trade data and conclude that 82% of potential product-partner trade flows are actually zero (the share goes up to 92% for imports). Similarly, Helpman et al. (2008) use

data on trade among 158 countries over the years 1970–1997 to show that just around 50% of all possible country-pairs engage in trade of any sort (either one country ships goods to the other or both do it), whereas bilateral trade is even rarer. Both papers start from the heterogeneous-firm trade model first proposed by Melitz (2003) and accommodate zero trade flows by relaxing the hypothesis of symmetric countries therefore generating patterns of export flows that are consistent with the empirical evidence, both in terms of zeros and with respect to the role played by intensive and extensive margins of trade.[1]

The distinction among the two dimensions along which it is possible to decompose total trade, the number of flows and their average value, has led researchers to dig deeper into the theoretical foundations of the gravity relationship between trade flows, distance and size, and come up with refined versions of the model capable of accommodating the growing stock of empirical evidence coming primarily from firm-level data (see for instance Chaney, 2008; Helpman et al., 2008). Empirically, Hummels and Klenow (2005) report that the extensive margin accounts for about 60% of the greater exports of larger economies, while Bernard et al. (2009) —who focus on data for the US— find that variation in trade flows across partner countries is mainly due to the extensive margin, with the intensive margin determining most of the variation in trade over short (one-year) time spans instead. These findings can be rationalized by means of yet another empirical regularity, namely the high concentration of trade. Indeed it appears that at different levels of aggregation (country, product, or firm-level data) total export flows are dominated by a small number of players making up the bulk of export. Thus, while most firms export (very) few products to (very) few destinations, a small 'club' of multi-product firms export almost everywhere and represent a disproportionate share of total export. This appears to be true for the US as well as for other countries (see for instance Bernard et al., 2007; Mayer and Ottaviano, 2008).

Empirical evidence consistent with such results has been gathered also by scholars using complex network analysis to describe real-world phenomena such as Internet traffic, airport connections, and international trade (Serrano and Boguñá, 2003; Garlaschelli and Loffredo, 2004, 2005; Bhattacharya et al., 2008; Fagiolo, Schiavo

---

[1]The definition of intensive and extensive margins of trade is not unequivocal in the literature, and crucially depends on the level of aggregation of the various studies. Thus, using microdata, Chaney (2008) and Crozet and Koenig (2010) look at the number of exporters (extensive margin) and the average volume of firm-level trade (intensive margin), whereas Bernard et al. (2009) decompose the former into the number of firms trading with country i and the number of distinct products exported. Hummels and Klenow (2005), who use country data at the six-digit HS level, distinguish among the number of six-digit categories exported (extensive margin) and the average export per category (intensive margin).

and Reyes, 2008; Fagiolo et al., 2009; Riccaboni and Schiavo, 2010). Our modeling strategy is closely related to the network literature and the model laid down in Section 3 can be used to describe the evolution of weighted networks in general (Riccaboni and Schiavo, 2010).

According to this approach, countries are described as nodes that establish (trade) links among themselves: these links are given different weight depending on the value of trade they carry. Although these contributions are mainly rooted in physics and therefore not particularly interested in the economics behind the phenomena they study, the stylized facts they uncover are not only broadly consistent with those addressed by economists, but can also shed new lights on them, as we will show in the paper. From this literature we learn that the distribution of trade flows assumes a log-normal form (Bhattacharya et al., 2008; Fagiolo et al., 2009), whereas their growth rate display fat tails (Fagiolo et al., 2009). Further features concerns the hierarchical structure of trade (consistent with the high concentration of trade mentioned above), and the presence of a 'rich-club' whereby a handful of countries command a disproportionately large share of world trade (Fagiolo et al., 2009). Finally, Barrat et al. (2004) and Eom et al. (2008) find a power-law relation linking the number of partners of each node (node strength) and the total weight of its links (node strength).[2] By adapting this finding to trade data, we end up with a relationship between total export and the number of destinations served that is not far from the correlation between the intensive and the extensive margins of trade emphasized by Hummels and Klenow (2005).

The last stream of the literature we come across in our journey is rooted in development economics and has to do with the relative merits of industrial policy in facilitating economic growth, and especially export-led growth. After years of neglect and skepticism at the very notion of industrial policy as an effective tool for economic development —fueled by the poor performance of import substitution policies in many countries, most notably in Latin America— the concept has been rehabilitated in a series of papers starting with Hausmann and Rodrik (2003) and Rodrik (2004). There, development is described as a process of 'self-discovery' about what a country is good at producing. Such learning process that occurs through trial-and-error generates important spillovers as success or failure send signals to other agents and is therefore beneficial to the economy as a whole. The problem is thus similar to that faced by innovators: sunk (entry) costs plus imperfectly appropriable returns are likely to result in too little investment (too little 'search' activity

---

[2]These papers do not use trade data. Rather, Barrat et al. (2004) look at the scientific collaboration network and the world-wide air-transportation network, whereas Eom et al. (2008) investigate several online bulletin board systems and a movie actor network.

and 'discovery' in the present context). This calls for a new kind of (industrial) policy intervention aimed at eliciting information from private sector activities, quite different from the traditional ideas of protecting domestic firms or 'picking winners'.

The theme of industrial policy has been recently addressed by Easterly and Reshef (2009) in the context of the relationship between export and development, in a way that can fruitfully interact with our own approach. The authors set off by observing that, for virtually all countries, manufacturing export is extremely concentrated both in terms of products and destination markets so that export value is made up of few 'big hits'. From a policy perspective Easterly and Reshef (2009) then warn against the ability of industrial policy to 'pick winners' since the probability of drawing a 'big hit' from such a skew distribution (as the one characterizing export flows) is very low.

We will show that our modeling strategy nicely accommodate both the Hausmann and Rodrik (2003) and the Easterly and Reshef (2009) views: large export flows are indeed rare events and our model correctly predicts that, yet countries characterized by large 'discovery' efforts are much more likely to draw a 'big hit', due to the (very skewed) shape of the distribution of bilateral export flows.

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

2   Empirical regularities

We use the NBER-United Nations Trade Data documented in Feenstra et al. (2005) and available through the Center for International Data at UC Davis. This source provides bilateral trade flows among a large number of countries over 1962–2000, both aggregate and at 4-digit SITC level (which is the finest available level of aggregation). Data are in thousands US dollars and, for product-level flows, there is a lower threshold at $ 100,000 below which transactions are not recorded. One point to note is that disaggregated data are not always consistent with country trade flows: in a number of cases we do not observe any 4-digit transaction recorded between two countries, but nevertheless find a positive total trade, and vice-versa. To avoid inconsistency we compute the total trade by aggregating commodity-level data.

In what follow we only consider data for the period 1992–2000, in order to

minimize the effects induced by the variation in the number of countries due to geopolitical events such as the breaking up of Yugoslavia and the Soviet Union. Moreover, we drop a number of small economies (e.g. Gibraltar) for which trade data exists but are not exhaustive; we also aggregate information for some countries (e.g. the Czech Republic and Slovakia) to keep the number of economies constant over time.[3] In this way we end up with a balanced panel of 166 countries.

Let us start this gallery of relevant stylized facts from the issue of zero trade flows, that has received a great deal of attention as of late (Baldwin and Harrigan, 2007; Helpman et al., 2008; Armenter and Koren, 2008). Looking at aggregate trade flows among 166 countries, Helpman et al. (2008) report 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). Our dataset displays a very similar pattern, with zeros making up 58.5% of the dataset (in terms of aggregate trade flows) in 1992, a figure that goes down to 52% and 51.7% in 1997 and 2000. Bilateral trade accounts for roughly 30–35% of the data, while country-pairs for which trade only flow in one direction represent 12–14% of the total.

When one looks at commodity-level data the number of zeros booms, consistently with previous findings. In particular, we find that between 98.1 and 98.5% of all possible commodity-destination pairs is void, a much higher figure than reported by Armenter and Koren (2008) for the US (a large country likely to export many products to many destinations), but in line with results discussed by Easterly and Reshef (2009).[4]

A second aspect that is particularly relevant for our work is the extreme concentration of export flows both in terms of commodities exported and destinations served. Following Easterly and Reshef (2009) we compute export shares relative to the top 1, top 3 and top 10 exported categories for each country, and display summary statistics in the left panel of Table 1.[5] Data show that for the median country the single most important export category represents roughly one quarter

---

[3]Detailed information on the issue are available upon request.

[4]Since the aim here is to give a feeling of the phenomenon rather than pinning down exact numbers, we define the number of potential commodity-destination pairs for each exporting country simply as the product of the number of 4-digit SITC subgroups (1320) by the number of destination countries (165). In doing this we disregard the fact that countries export a limited number of products to a limited number of destinations: see Easterly and Reshef (2009) for an alternative approach.

[5]Only nonzero export flows are considered in the computations.

of its total export with a low of 2–3% and a high of 95–99%. Similarly, the top 3 export commodities enjoy a share close to 50% of total export for the median country.

[Table 1 about here.]

The right panel of Table 1 repeats the exercise but looks at the combination of commodity and destination market as the unit of analysis, so that the same good exported to two different places represent two distinct export categories. The degree of concentration goes down as expected, but remains nonetheless strikingly high.

This issue of the concentration of exports can be further investigated along two different dimensions: the number of commodity exported and the size of bilateral trade flows. Figure 1 shows that the number of 4-digit SITC goods traded is Pareto with an exponential cutoff. 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 phenomenon in semi-log scale: this time it is the exponential part of the distribution that becomes a straight line, so that we can magnify what happens to the probability distribution as the number of goods exported grows large.

[Figure 1 about here.]

Moving to the distribution of bilateral trade flows, Figure 2 plots the complementary cumulative probability distribution of trade flows in log-log scale, both for commodity-level transactions and for aggregate bilateral flows.[6] We observe that both distributions display the parabolic shape typical of the log-normal distribution, thus conforming to previous findings by Bhattacharya et al. (2008) and Fagiolo et al. (2009).[7]

[Figure 2 about here.]

Figures 1 and 2 indicate that both the number of commodities exported by each country and the value of bilateral trade flows are characterized by very skew distributions. Thus, while the vast majority of countries export only a few goods, a small number of them trade in most export categories. Something similar occurs with respect to trade values: while the bulk of transactions has small size, a few 'big hits' make up a disproportionate share of world trade.

---

[6]Figure 2 refers to 1997 data, but other years display exactly the same behavior.

[7]Easterly and Reshef (2009) find a similar pattern but they stress the power-law behavior of the right tail of the distribution.

3 The model

We model international trade as a set of (stochastic) transactions of different magnitude occurring among countries, and can be though of as an extension of the preferential attachment model put forward by Barabási and Albert (1999) to describe the properties of many real-world networks (Internet traffic, air-transportation, scientific collaborations, to quote just a few). Our extension builds on a fairly old idea that goes back to Herbert Simon, and has been extensively used to model the dynamic of socio-economic systems (Simon, 1955; Ijiri and Simon, 1977), and is capable of accommodating the large degree of heterogeneity across trade flows (something that is not possible in the original Barabási and Albert model). The easiest (and less demanding in terms of assumptions) way to account for the heterogeneity in trade flows is to assume that their magnitude grows according to the so-called Gibrat's *law of proportionate effects.*[8] In recent years, generalization of this idea have been used to rationalize the stylized fact that the distribution of the growth rates of economic organizations ranging from company divisions up to country GDPs is very skewed (Growiec et al., 2008; Buldyrev et al., 2007).

Thus we end up with a simple stochastic model where (trade) link formation is governed by preferential attachment (for each country the probability of exporting a new product/destination increases in the number of existing relationships), whereas export volumes grow according to a geometric Brownian motion. Moreover, the two processes governing link formation and weight growth are assumed to be independent.

The model follows Riccaboni and Schiavo (2010) and the key assumptions are the following:

1. at time $t = 0$ there are N_0 countries each characterized by a self loop (this only serves for initialization purpose: self loops are never considered in the analysis). At each time step $t = \{1, \ldots, M\}$, a new link among two countries arises: thus the number of links existing at time t is $m_t = t$. A trade link represents the possibility to export a given product to a given destination and is therefore identified by a product-destination pair. We write $K_i(t)$ for the number of links of country i at time t (node *degree* in network jargon). To identify the countries connected by the newly formed link at time t we adopt the following procedure: with probability a the new link is assigned to a new country, whereas with probability $1 - a$ it is allocated to an existing country i. In the latter case, the probability of choosing country i is given by: $p_i(t) = K_i(t-1)/2t$. The two

---

[8]Gibrat's law postulates that the expected value of the growth rate of a business firm is independent of its current size; see Gibrat (1931) for the original formulation.

countries i and j connected by each new link are chosen symmetrically with $i \neq j$. Thus with probability a the new link is assigned to a new destination country, while with probability $1 - a$ it is allocated to an existing destination with probability $p_j(t) = K_j(t-1)/(2t - K_i(t-1))$ if $j \neq i$ and $p_j(t) = 0$ otherwise. Hence, at each time t this rule identifies the pair of (distinct) countries to be linked;

2. at time t each (existing) trade flow between countries i and j has weight $w_{ij}(t) > 0$, where K_i, K_j and w_{ij} are independent random variables. At time $t+1$ the weight of each link is increased or decreased by a random factor $x_{ij}(t)$, so that $w_{ij}(t+1) = w_{ij}(t)x_{ij}(t)$. The shocks and initial link values are taken from a distribution with finite mean and standard deviation.

Thus we assume that the value of each trade flow grows in time according to a random process. Moreover the two processes governing link formation (the *extensive* margin) and the growth of existing (bilateral) trade flows (the *intensive* margin) are assumed to be independent. We therefore combine a preferential attachment mechanism (Assumption 1), with an independent geometric Brownian motion characterizing the magnitude of bilateral trade flows (Assumption 2).

Based on the first assumption we derive the degree distribution $P(K)$ (Barabási and Albert, 1999; Buldyrev et al., 2007). In the absence of the entry of new countries ($a = 0$) the probability distribution of the number of links at large t, i.e. the distribution $P(K)$, is exponential:

$$P(K) \approx \frac{1}{\bar{K}} \exp(-K/\bar{K}), \tag{1}$$

where $\bar{K} = 2t/N_0$ is the average number of links per country, which linearly grows with time.[9]

If $a > 0$, $P(K)$ becomes a Yule distribution which behaves as a power law for small K:

$$P(K) \sim K^{-\varphi}, \tag{2}$$

where $\varphi = 2 + a/(1-a) \geq 2$, followed by the exponential decay of Eq.(1) for large K with $\bar{K} = (1 + 2t/N_0)^{1-a} - 1$ (Yamasaki et al., 2006).

Hence, in the limit of large t when $a = 0$ (no entry), the distribution of $P(K)$ converges to an exponential; on the contrary when $a > 0$ and small the connectivity distribution at large t converges to a power-law with an exponential cutoff (Yamasaki et al., 2006).

---

[9]$\bar{K}$ does not include initial self loops.

Using the second assumption we can compute the growth rate of total export for each country (node *strength* in network jargon). The strength of node i is given by $W_i = \sum_{K_i} w_{ij}$. The growth rate is measured as $g = \ln(W(t+1)/W(t))$. Thus, the resulting distribution of the growth rates of node strength $P(g)$ is determined by

$$P(g) \equiv \sum_{K=1}^{\infty} P(K)P(g|K), \tag{3}$$

where $P(K)$ is the connectivity distribution computed in the previous stage of the model and $P(g|K)$ is the conditional distribution of growth rates of nodes with given number of links determined by the distribution $P(w)$ and $P(x)$.

Fu et al. (2005) find an analytical solution for the distribution of the growth rates of trade flows $P(g)$ for the case when $a \to 0$ and $t \to \infty$,

$$P(g) \approx \frac{2V_g}{\sqrt{g^2 + 2V_g}\,(\bar{g} + \sqrt{g^2 + 2V_g})^2} \tag{4}$$

$P(g)$ has similar behavior to the Laplace distribution for small g, whereas for large g, $P(g)$ has power law tails.

A further implication of the model that can be derived from the second assumption concerns the distribution of the size of bilateral trade flows $P(w)$. The proportional growth process (Assumption 2) implies that the distribution of the weights $P(w)$ converges to a log-normal. Thus total export for each country W is given by the sum of K log-normally distributed stochastic values. Growiec et al. (2008) show that since the log-normal distribution is not stable upon aggregation, the distribution of total export $P(W)$ is multiplied by a stretching factor that, depending on the distribution of the number of links $P(K)$ could lead to a Pareto upper tail.[10]

Moreover, a negative relationship exits among the weight of links and the variance of their growth rate. Our model implies an approximate power-law behavior for the variance of growth rates of the form $\sigma(g) = W^{-\beta(W)}$ where $\beta(W)$ is an exponent that weakly depends on the strength W. In particular, $\beta = 0$ for small values of W, $\beta = 1/2$ for $W \to \infty$, and it is well approximated by $\beta \approx 0.2$ for a wide range of intermediate values of W (Riccaboni et al., 2008).

Finally, the model yields a prediction also on the relation between the number of product-destinations exported (the extensive margin K) and each country's total export flows (the intensive margin W). In Section 4 we show that since the weight of

---

[10]This result is consistent with the discussion in Easterly and Reshef (2009) and provides a theoretical foundation to it.

each link is sampled from a log-normal distribution (w are log-normally distributed), and given the skewness of such a density function, the law of large numbers does not work effectively. In other words, the probability to draw a large value for a link weight increases with the number of draws, thus generating a positive (power law) relationship between W and K, for small K.[11] From an economic point of view we can interpret this relationship as one between the extensive and the intensive margins of trade. Hence, since total export is just the product of the number of transactions by their average size, we end up with a relationship echoing the main finding in Hummels and Klenow (2005), namely that the extensive margin accounts for a large share of the greater exports of large economies.

4 Testing the model predictions

In this Section we first discuss in more details the correspondence between the main predictions of the model and the data. Then, we simulate the model and compare the results with trade data in order to verify the predictive capability of our theoretical framework and test alternative hypotheses about the evolution of the world trade.

4.1 A further look at trade data

Let us start from a closer examination of the main properties of international trade data sketched in Section 2.

Figure 1 above shows that the distribution of the number of commodities exported (a commodity here is identifies by a good-destination pair) is Pareto with an exponential cutoff, thus conforming to the predictions of the model. The cutoff suggests the existence of moderate entry of new players: empirically this is represented by the countries emerging from the collapse of the Soviet Union and former Yugoslavia that, though not starting from scratch, had nonetheless to rebuild their network of trade relationships from low levels of connectivity.[12]

Similarly, when looking at the values of bilateral trade flows as in Figure 2 we find they are log-normally distributed, as implied by the proportional growth

---

[11]Another way to think about this issue is in terms of convergence to the central limit theorem. Skewness of the underlying distribution causes convergence to normality to be slow: hence, repeated draws from a lognormal distribution will not converge to normality unless the number of draws is very large. Normality would imply no relation between the number of links and their average value, whereas departures from it (i.e. a slow convergence) determine a positive correlation between the two variables since a vast majority of trade relationships will has very small size due to the concentration of probability on the lower tail.

[12]In our dataset there are 17 countries that were formed after 1991 and represent therefore new entrants.

process (geometric Brownian motion) governing the dynamic of trade flows. Upon aggregation the power-law behavior of the upper tail become more apparent, as predicted by Growiec et al. (2008), but this departure from log-normality concerns a very small number of observations (0.16% in the case of commodities flows, 2.21% for aggregate ones).[13]

[Figure 3 about here.]

Figure 3 shows that the growth rates of aggregate trade flows display a distribution that fits the model's prediction. Goodness of fit tests, reported in Table 2, lead us to reject the hypotheses of a Gaussian or a Laplace distribution, whereas both the distribution described by equation (4) and a Generalized Exponential (GED) perform much better in terms of Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests, making it difficult to discriminate among them. Hence, trade flows appear to follow a growth path similar to the one characterizing products, firms, industries, and country GDPs (Fu et al., 2005; Fagiolo, Napoletano and Roventini, 2008).

[Table 2 about here.]

As discussed in Section 3, a simple model like the one presented here implies a negative relationship between the size and the variance of trade growth rates. Figure 4 reports the standard deviation of the annual growth rates of total bilateral trade flows (g), and their initial magnitude (W). The standard deviation of the growth rate of link weights exhibits a power law relationship $\sigma(g) = W^{-\beta}$ with $\beta \approx .2$, as predicted by the model (Riccaboni et al., 2008). This implies that the fluctuations of the most intense trade relationships are more volatile than expected based on the central limit theorem.

Hence, countries relying disproportionately on a small number of large export flows will be subject to substantial volatility in their export revenues. Developing countries exporting raw materials or primary commodities are textbook examples of this phenomenon, but the case can be easily extended to non-diversified manufacturing export.

[Figure 4 about here.]

All in all, the main predictions of the model in terms of growth and size distribution of trade flows, number of commodities traded and size-variance relationship of

---

[13]Estimations of the power-law fit have been obtained applying the methodology described in Clauset et al. (2009).

trade flows are verified empirically. Thus we can conclude that a stochastic model that assumes a proportional growth of transactions as well as a multiplicative random growth of the value of each transaction can reproduce most of the observed structural features of international trade data. As long as stylized facts can be matched by the mechanic interaction of agents without any particular economic rule, their usefulness as a testbed for discriminating among competing trade models is questioned. With respect to this issue we therefore agree with Armenter and Koren (2008) in saying that from the point of view of economic theory our model is mainly useful when it misses an empirical regularity rather than when it matches it, as in that way it signals that something else is at work beyond the mere stochastic interaction of agents. We now turn to simulations to further investigate the ability of the model to match empirical facts.

4.2 Simulations

Simulations proceed in two steps. In the first stage, we generate the basic structure characterizing the network of international trade flows by determining the number of commodities K exported by each country. In the second stage, we assign the value of the transactions based on a random sampling of K values from a log-normal distribution whose parameters are obtained through a maximum likelihood fit of the real world distribution.

In the present context we model trade as a system where at every instant t a new trading opportunity arises, which represents the possibility to export one commodity to a destination country. We need to slightly modify the original setting in order to account for the possibility that these new links could be assigned randomly rather than proportionally to the number of existing trade relationships. In our simulations the parameter a governs the entry of new nodes according to Assumption 1, whereas parameter b is the probability that a new link is assigned randomly. Thus, with probability a the new link is assigned to a new country, whereas with probability $1 - a$ it is allocated to an existing country i. In this latter case, the probability of choosing country i is now given by $p_i(t) = (1 - b)K_i(t - 1)/2t + b/N_{t-1}$ where N_{t-1} is the number of countries at time $t - 1$. The destination market served by the new trade link is chosen in the same way with $i \neq j$.

Tuning the two model parameters a and b we generate different structures of world trade in terms of (the distribution of) the number of products exchanged by each country pair. In particular, without entry ($a = 0$) and completely random allocation of opportunities ($b = 1$) one obtains a random network characterized by a Poisson connectivity distribution (Erdös and Rényi, 1959), whereas allowing entry

$(a > 0)$ one moves towards an exponential distribution. Keeping a positive entry rate, but assigning trade links according to a pure preferential attachment model $(b = 0)$, one obtains a Pareto distribution for the number of commodities traded by each country (this is the original formulation by Barabási and Albert, 1999).[14] In the limit case in which entry of new players is ruled out $(a = 0)$, the connectivity distribution tends toward a Bose-Einstein geometric distribution.

We compare the structure of random scale-free model networks with the real world trade network in 1997. Since the structure of the network is highly stable over time, results do not change substantially by comparing simulations with the structure of trade for different years. In the first stage, we generate one million networks with a and b ranging from 0 to 1. We simulate a system with 166 countries and 1,079,398 trade links (number of different commodities traded). Next we select the random networks that better fit the real world pattern in terms of correlation, as measured by the Mantel r test, and connectivity distribution.[15] This test allows us to assess not only whether the model correctly predicts the number of zeros in the export matrix that represents world trade (as in Armenter and Koren, 2008), but also how close it gets to matching the number of commodities exchanged by each country pair.

[Figure 5 about here.]

Figure 5 reports the value of the Mantel test for networks with $0 \leq b \leq 1$ and an entry rate a which implies the entry of 0 to 66 countries. The Mantel correlation statistics reach a peak of 0.88 (p-value< 0.01) for purely preferential attachment regimes $(b = 0)$. However, the Mantel test does not discriminate among different entry regimes. Hence, we now compare the connectivity distribution of simulated networks with the real world distribution of the number of traded commodities K by means of the Kolmogorov-Smirnov (KS) goodness of fit test. Figure 6 confirms that the best fit is obtained in the case of a purely preferential attachment regime $(b = 0)$. However the KS tests provides additional information on the most likely value of a (entry rate of new countries).

[Figure 6 about here.]

---

[14]Many empirical studies have found that such kind of connectivity distribution characterizes a large number of (real-world) social, economic, or technological networks. This results explain the popularity of the Barabási and Albert (1999) model in network analysis .

[15]The Mantel test is a non-parametric statistical test of the correlation between two matrices Mantel (1967). The test is based on the distance or dissimilarity matrices which, in the present case, summarize the number of links between two nodes in the simulated and real networks. A typical use of the test entails comparing an observed connectivity matrix with one posed by a model. The significance of a correlation is evaluated via permutations, whereby the rows and columns of the matrices are randomly rearranged.

Figure 5 shows that the our model can better reproduce the distribution of the number of traded commodities $P(K)$ with and entry rate $a > 0$, that implies the entry of 14–18 countries. This closely corresponds to the empirically observed number of new countries. Thus we can conclude that a simple proportional growth model with mild entry can account for the distribution of the number of commodities traded by each pair of countries.

[Figure 7 about here.]

Introducing the value of the transactions we can show that the model generates the observed relationship between intensive and extensive margins of trade. Figure 8 depicts the relationship between total bilateral trade flows (W) and the number of commodities exported by each country (K). The Figure displays the relationship emerging from 1997 trade data, and confirms that there exists a positive correlation between the two variables. The slope of the interpolating line in double logarithmic scale reveals a positive relationship between the number of commodities exported and the average value of trade flows of the kind $W = K^{\theta}$, with $\theta \approx 1.33$.

[Figure 8 about here.]

The curve displays and upward departure in the upper tail. This can be explained by noting that the 4-digit SITC classification that we use imposes a ceiling to the number of goods a country can trade since there are only around 1,300 4-digit categories (vertical dotted line).

Apart from the upper decile of the distribution, the simulated version of the network shows exactly the same dependence among the size and the number of the transactions. This seems surprising, by considering that the model assumes two independent growth processes for the number of transactions (K) and their values (w). However, it should be noted that the law of large numbers does not work properly in case of skew distributions such as the log-normal. Given a random number of transactions with a finite expected value, if its values are repeatedly sampled from a log-normal, as the number of transactions increases, the average value of the transactions will tend to approach and stay close to the expected value (the average for the population). However this is true only for a sufficiently large K, whereas we know from the distribution $P(K)$ that the vast majority of countries are characterized by small K, i.e. they export a limited number of commodities. The higher is the variance of the growth process of link weights, the larger has to be K to start observing convergence toward $W = wK^{\theta}$ with $\theta = 1$ predicted by the law of large numbers. Thus only the largest countries approach the critical threshold. In

14

sum, our simulations show that the model can account for the relationship between the number of commodities exported (K) and the magnitude of trade flows (W), in a way consistent with previous findings on the relation between extensive and intensive margins of trade (Hummels and Klenow, 2005; Bernard et al., 2009).

5 Discussion and conclusions

Using a simple model of proportionate growth and preferential attachment we are able to replicate the main structural properties of international trade data. In particular, our setup is capable of generating the power-law distribution that characterizes the number of commodities traded by each country, as well as the log-normal distribution of bilateral trade flows. These features of the data testify for the high concentration of trade flows whereby a small number of products/destinations account for a large share of export revenues.

Additionally, the model matches the fat tails displayed by the distribution of the growth rates of trade flows, and the negative relationship between the size of trade flows and the variance of their growth rates. Through this channel the model is thus able to provide an explanation to the observation that developing countries specialized in the export of a small number of goods tend to suffer high volatility.

Last, the model confirms the fact that the extensive margin of trade (here defined as the number of commodities exported) accounts for a large fraction of the greater exports of large economies.

In the spirit of Armenter and Koren (2008) we claim that the empirical regularities well matched by stochastic models are not very informative for economic theory and should probably not be used as the main testbed for discriminating among competing international trade models. Economic forces should rather account for departures from a stochastic benchmark as the one proposed here.

Beside matching many of the stylized facts about international trade, what does the model tell us? As hinted at in the Introduction, the paper can successfully contribute to the debate on the relevance (and ability) of industrial policy to spur export-led economic growth and development. Indeed, the preferential attachment mechanism the lies at the core of our model can be see as a simple formalization of the idea originally put forward by Hausmann and Rodrik (2003) and Rodrik (2004): development as a process of 'self-discovery' through which countries need to find out which goods (or services) they are good at producing and exporting. Entrepreneurial activity serves a public as well as a private role since it provides useful information to all economic agents in the form of a knowledge spillover resulting from business

success or failure. Then it is reasonable to assume that countries that have already performed successful discoveries in the past will find easier to discover again. In the context of our mode this simply means that the probability of capturing a new trading opportunity (i.e. a new link) is positively related to the number of links already established.

Krautheim (2007) uses a simple network formation game to microfound the (positive) spillover effect coming from having many firms exporting to a single destination. Exchanging relevant information lowers the fixed costs of entering a foreign market and therefore makes serving that particular destination more profitable for everyone. Similarly, we can imagine that exporting many commodities (having a high K) reduces the (fixed) costs of 'discovery' and therefore increases the amount of investment in such an activity, thus improving the odds of appropriating a new trading link.

The role of industrial policy is then simply one of creating the conditions for having the socially desirable level of searching (discovering) activity. Since private entrepreneurs do not fully appropriate the benefits of the information they produce, the market will generate too little investment in this kind of activity.[16]

Our model thus provides yet another way to rationalize public intervention in the form of industrial policy, and one that is consistent with the warning put forward by Easterly and Reshef (2009) who are concerned about the small probability of 'picking winners' given the high concentration of exports. We show that large export flows are indeed rare events (as our model correctly predicts): yet, countries characterized by large 'discovery' efforts are much more likely to draw a 'big hit' due to the (very skewed) shape of the distribution of bilateral export flows.

References

Armenter, R. and Koren, M. (2008), A balls-and-bins model of trade, CeFiG Working Papers 3, Center for Firms in the Global Economy.

Baldwin, R. and Harrigan, J. (2007), Zeros, quality and space: Trade theory and trade evidence, NBER Working Papers 13214, National Bureau of Economic Research.

---

[16]The higher than expected volatility of large export flows discussed in Section 4.1 above provides another reason for government intervention in the form of both a strategy to spur diversification in exports, and in the design of appropriate institutional mechanisms capable of dampening the effects of such volatility on the real economy. The experience of Chile with respect to its mining sector is one of the most successful recent examples.

Barabási, A.-L. and Albert, R. (1999), 'Emergence of scaling in random networks', *Science* **286**(5439), 509–512.

Barrat, A., Barthélemy, M., Pastor-Satorras, R. and Vespignani, A. (2004), 'The architecture of complex weighted networks', *Proceedings of the National Academy of Sciences* **101**(11), 3747–3752.

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.

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.

Buldyrev, S. V., Growiec, J., Pammolli, F., Riccaboni, M. and Stanley, H. E. (2007), 'The growth of business firms: Facts and theory', *Journal of the European Economic Association* **5**(2-3), 574–584.

Chaney, T. (2008), 'Distorted gravity: The intensive and extensive margins of international trade', *American Economic Review* **98**(4), 1707–21.

Clauset, A., Shalizi, C. and Newman, M. (2009), 'Power-law distributions in empirical data', *SIAM Review* **51**, 661–703.

Crozet, M. and Koenig, P. (2010), 'Structural gravity equations with intensive and extensive margins', *Canadian Journal of Economics* **43**(1), 41–62. forthcoming Canadian Journal of Economics.

Easterly, W. and Reshef, A. (2009), Big hits in manufacturing exports and development, mimeo.

Ellison, G. and Glaeser, E. (1997), 'Geographic concentration in U.S. manufacturing industries: A dartboard approach', *Journal of Political Economy* **105**(5), 889–927.

Eom, Y.-H., Jeon, C., Jeong, H. and Kahng, B. (2008), 'Evolution of weighted scale-free networks in empirical data', *Physical Review E* **77**(5), 56105.

Erdös, P. and Rényi, A. (1959), 'On random graphs', *Publicationes Mathematicae* **6**, 290–297.

Fagiolo, G., Napoletano, M. and Roventini, A. (2008), 'Are output growth-rate distributions fat-tailed? some evidence from OECD countries', *Journal of Applied Econometrics* **23**, 639–669.

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.

Fu, D., Pammolli, F., Buldyrev, S. V., Riccaboni, M., Matia, K., Yamasaki, K. and Stanley, H. E. (2005), 'The growth of business firms: Theoretical framework and empirical evidence', *Proceedings of the National Academy of Sciences* **102**(52), 18801–18806.

Garlaschelli, D. and Loffredo, M. (2004), 'Fitness-dependent topological properties of the world trade web', *Physical Review Letters* **93**, 188701.

Garlaschelli, D. and Loffredo, M. (2005), 'Structure and evolution of the world trade network', *Physica A* **355**, 138–44.

Gibrat, R. (1931), *Les Inegalites Economiques*, Sirey, Paris.

Growiec, J., Pammolli, F., Riccaboni, M. and Stanley, H. E. (2008), 'On the size distribution of business firms', *Economics Letters* **98**(2), 207–212.

Guimarães, P., Figueiredo, O. and Woodward, D. (2009), 'Dartboard tests for the location quotient', *Regional Science and Urban Economics* **39**(3), 360–364.

Hausmann, R. and Rodrik, D. (2003), 'Economic development as self-discovery', *Journal of Development Economics* **72**(2), 603–633.

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*, North Holland, Amsterdam.

Krautheim, S. (2007), Gravity and information: Heterogeneous firms, exporter networks and the 'distance puzzle', Economics Working Papers ECO2007/51, European University Institute.

Mantel, N. (1967), 'The detection of disease clustering and a generalized regression approach', *Cancer Research* **27**, 209–220.

Mayer, T. and Ottaviano, G. (2008), 'The happy few: The internationalisation of european firms', *Intereconomics: Review of European Economic Policy* **43**(3), 135–148.

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.

Riccaboni, M., Pammolli, F., Buldyrev, S. V., Ponta, L. and Stanley, H. E. (2008), 'The size variance relationship of business firm growth rates', *Proceedings of the National Academy of Sciences* **105**(50), 19595–19600.

Rodrik, D. (2004), Industrial policy for the twenty-first century, CEPR Discussion Papers 4767, C.E.P.R. Discussion Papers.

Serrano, A. and Boguñá, M. (2003), 'Topology of the world trade web', *Physical Review E* **68**, 015101(R).

Simon, H. A. (1955), 'On a class of skew distribution functions', *Biometrika* **42**(3/4), 425–440.

Sutton, J. (2007), 'Market share dynamics and the 'Persistence of leadership' debate', *American Economic Review* **97**(1), 222–241.

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.
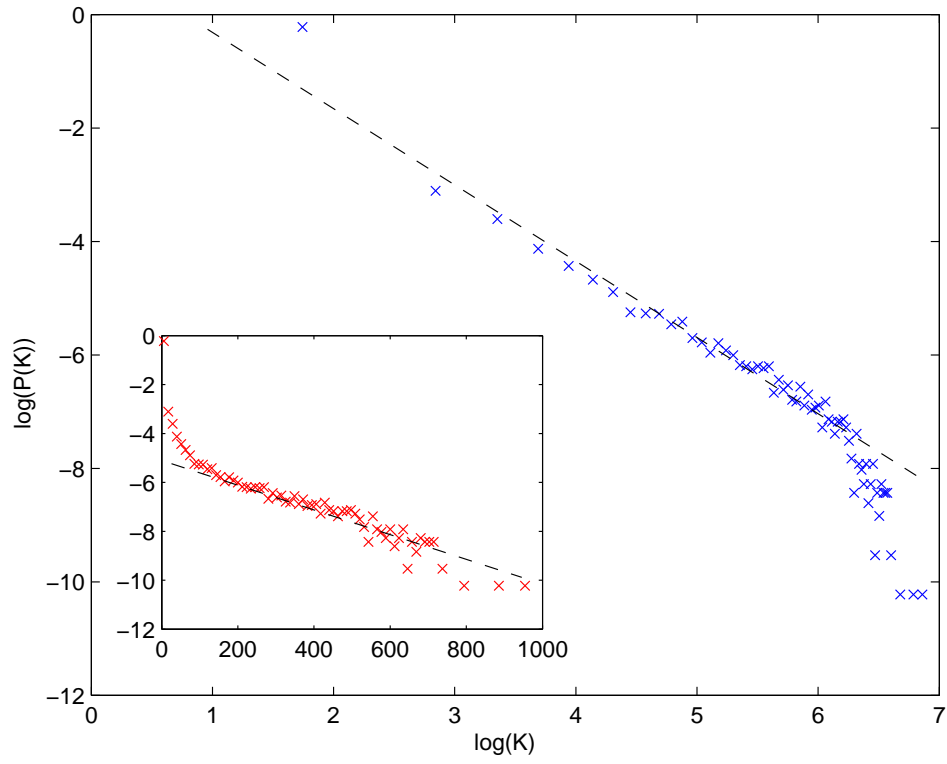
Figure 1: Distribution of the number of products traded, 1997. Double logarithmic scale (main plot) and semi logarithmic scale (inset).
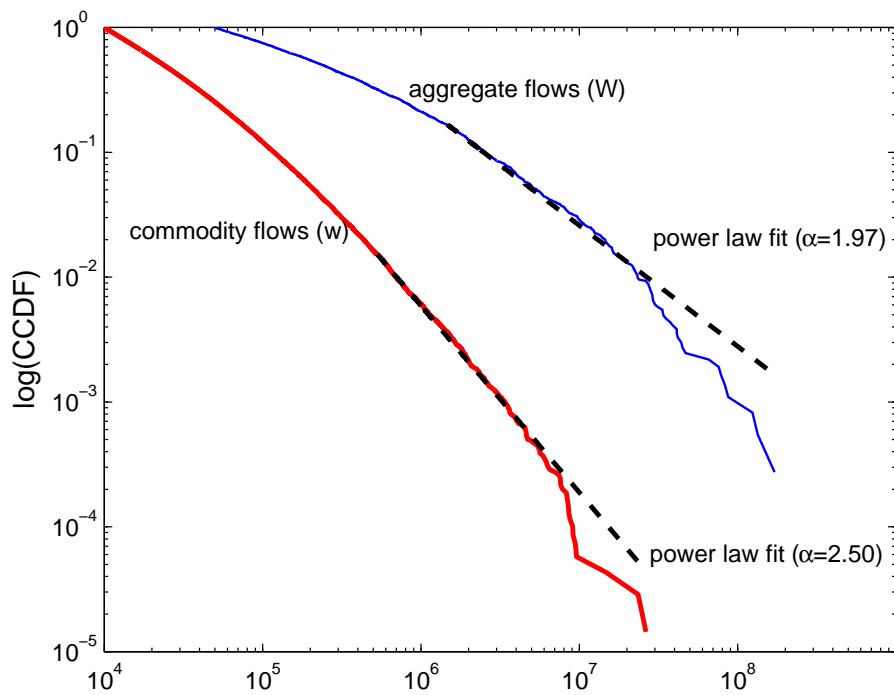
Figure 2: Distribution of nonzero trade flows, 1997. Complementary cumulative distribution of aggregate (blue) and commodity (red) flows and power-law fits (dashed lines).
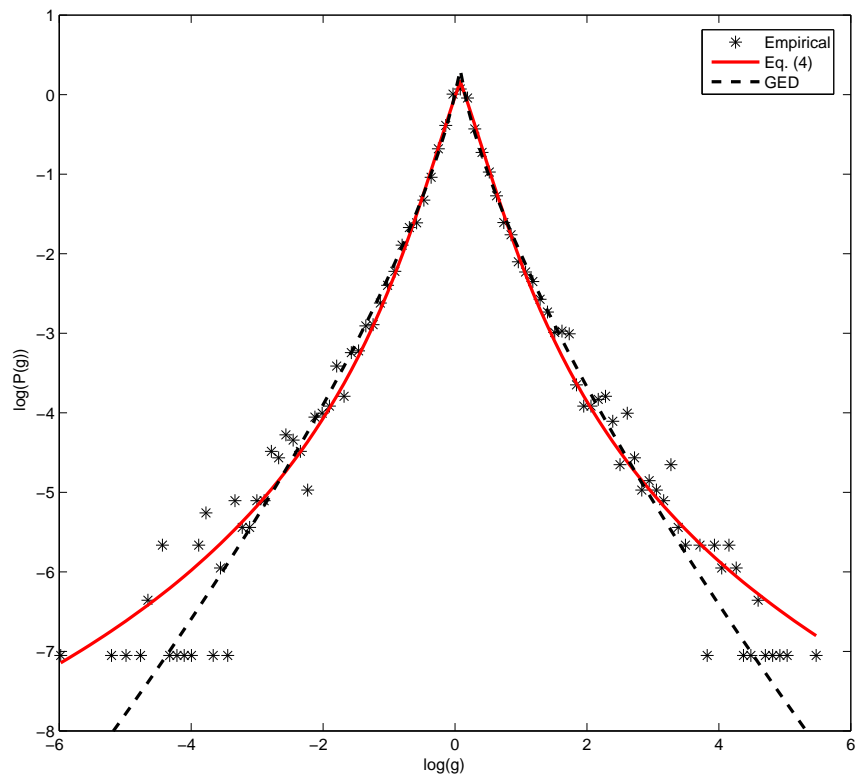
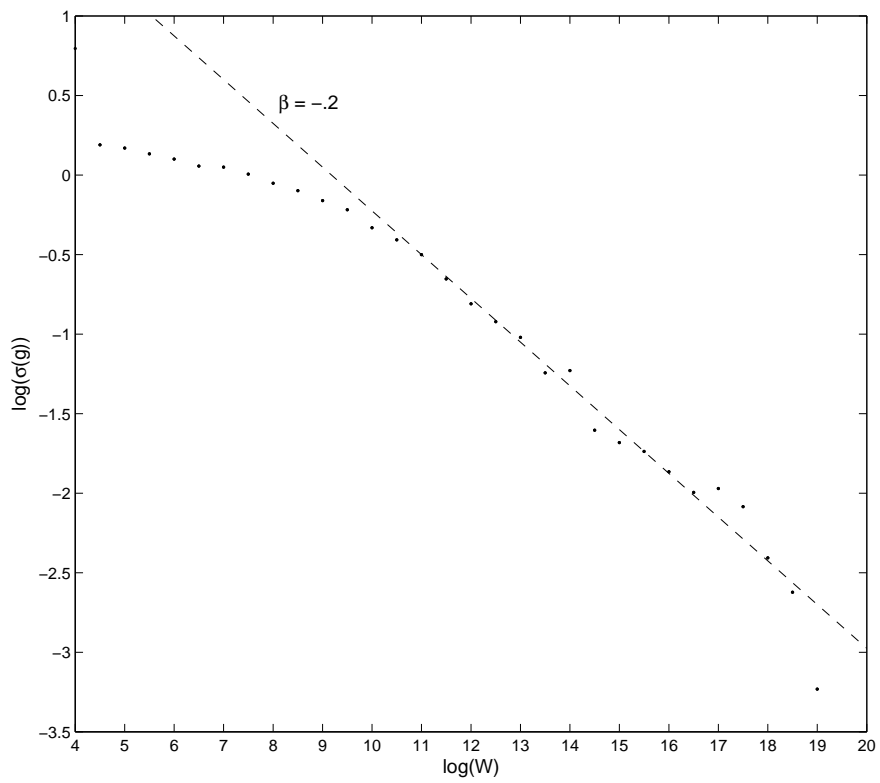Figure 3: Distribution of the growth rates of aggregate trade flows $P(g)$.

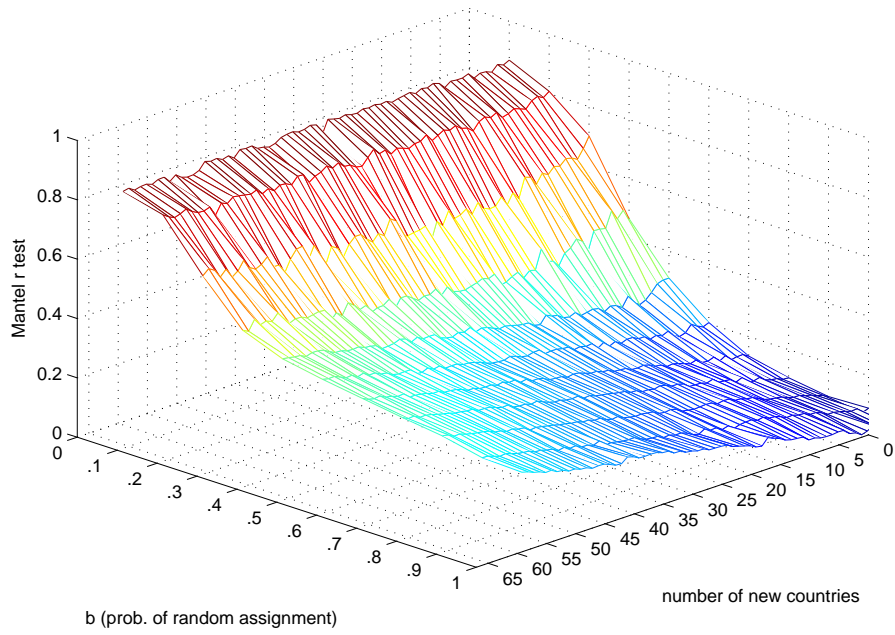Figure 4: Relationship between aggregate trade flows W and the standard deviation of its growth rate.

Figure 5: Mantel test comparing simulated and real K_ij
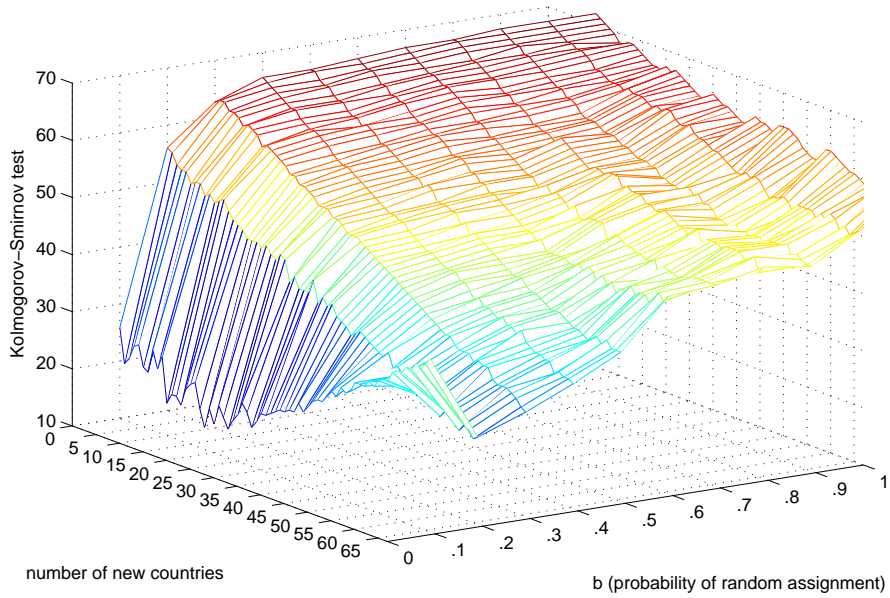
Figure 6: Kolmogorov-Smirnov goodness-of-fit test for different entry rates and probabilities of random assignment.
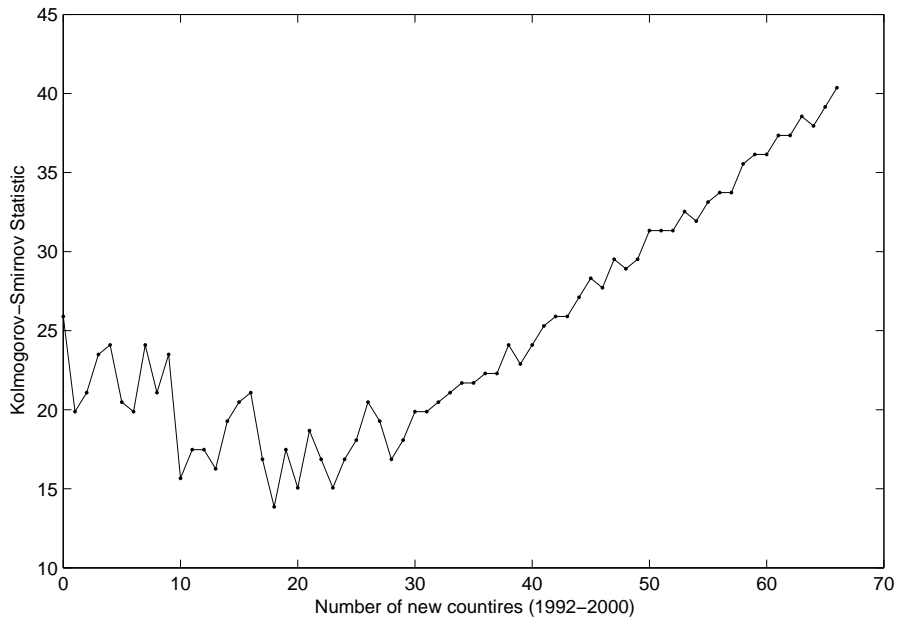
Figure 7: Kolmogorov-Smirnov goodness-of-fit test for different entry rates in a pure preferential attachment regime ($b = 0$).
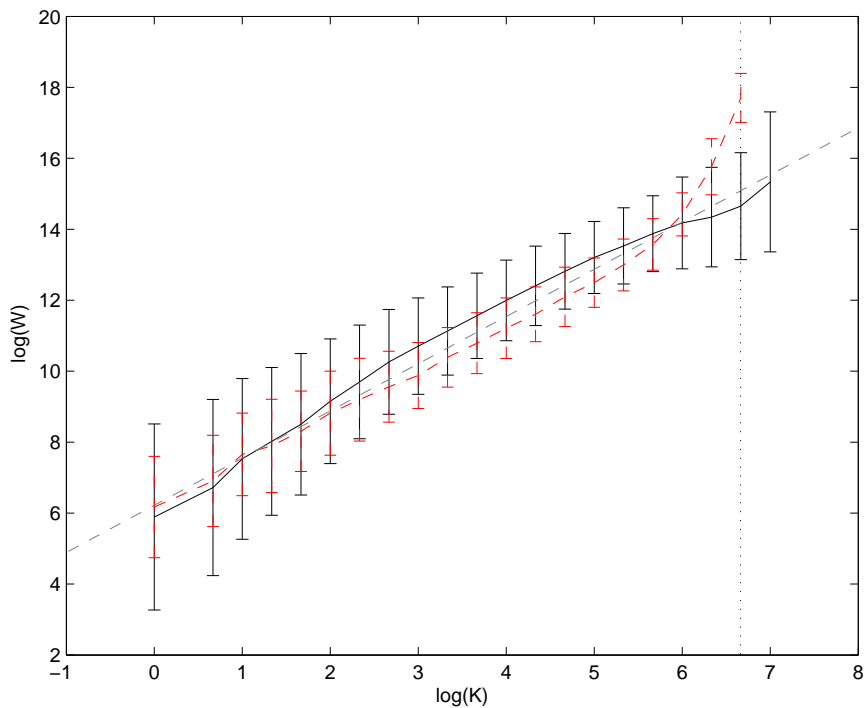

Figure 8: Relationship between the number of commodities exported K and trade flows (W). Simulated (black) and real-world (red) data, mean and one standard deviation in each direction. The dashed line represents the reference line $W = K^{\theta}$ with $\theta \approx 1.33$.

Table 1: Concentration of Exports

	by commodity				by commodity-destination			
1992	median	mean	min	max	median	mean	min	max
top 1	27.5%	33.8%	2.2%	94.7%	14.1%	18.5%	1.0%	91.4%
top 3	51.0%	51.3%	6.0%	97.9%	29.0%	32.6%	2.6%	95.5%
top 10	76.7%	69.6%	16.0%	100.0%	52.5%	51.7%	6.4%	98.1%
1997	median	mean	min	max	median	mean	min	max
top 1	24.6%	30.9%	2.6%	98.9%	10.8%	15.5%	0.7%	79.4%
top 3	45.1%	48.7%	6.9%	99.7%	24.7%	29.2%	1.9%	91.2%
top 10	72.5%	67.4%	18.6%	100.0%	46.8%	48.1%	4.9%	98.3%
2000	median	mean	min	max	median	mean	min	max
top 1	24.2%	32.4%	3.2%	99.0%	12.5%	17.0%	6.7%	72.7%
top 3	47.8%	50.0%	8.5%	99.3%	26.2%	31.1%	1.9%	95.0%
top 10	76.7%	68.7%	20.6%	99.8%	49.9%	50.0%	5.4%	97.9%

Only nonzero export flows considered

Table 2: Goodness of fit tests for the distribution of growth rates of aggregate trade flows.

1992–1993						
distribution	mode	par. 1	par. 2	par. 3	KS stat	AD stat
Gauss	0.0331	0.0014	0.9127		11.2891	2726.682
Laplace	0.0208	0.0035	0.5754		3.3476	0.3374
GED	0.0834	-0.0494	0.3083	0.7185	1.5507	0.0453
Equation (4)	0.0219	0.4549			1.2391	0.0389

1999–2000						
distribution	mode	par. 1	par. 2	par. 3	KS stat	AD stat
Gauss	-0.0006	0.0826	0.8032		10.9902	298.8943
Laplace	0.0736	0.0014	0.5086		3.3075	0.1146
GED	0.0273	0.04421	0.2752	0.7214	1.2250	0.0297
Equation (4)	0.0798	0.3576			0.9325	0.0356

Pooled						
distribution	mode	par. 1	par. 2	par. 3	KS stat	AD stat
Gauss	0.0385	0.0333	0.8417		10.8305	117329
Laplace	0.0605	0.0040	0.5338		2.8414	1.4107
GED	0.0191	0.0444	0.2899	0.7224	1.0915	0.0314
Equation (4)	0.0651	0.3658			0.8214	0.0477

Elenco dei papers del Dipartimento di Economia
pubblicati negli ultimi due anni

2008.1 *A Monte Carlo EM Algorithm for the Estimation of a Logistic Auto-logistic Model with Missing Data,* by Marco Bee and Giuseppe Espa.

2008.2 *Adaptive microfoundations for emergent macroeconomics*, Edoardo Gaffeo, Domenico Delli Gatti, Saul Desiderio, Mauro Gallegati.

2008.3 *A look at the relationship between industrial dynamics and aggregate fluctuations*, Domenico Delli Gatti, Edoardo Gaffeo, Mauro Gallegati.

2008.4 *Demand Distribution Dynamics in Creative Industries: the Market for Books in Italy*, Edoardo Gaffeo, Antonello E. Scorcu, Laura Vici.

2008.5 *On the mean/variance relationship of the firm size distribution: evidence and some theory*, Edoardo Gaffeo, Corrado di Guilmi, Mauro Gallegati, Alberto Russo.

2008.6 *Uncomputability and Undecidability in Economic Theory*, K. Vela Velupillai.

2008.7 *The Mathematization of Macroeconomics*: *A Recursive Revolution*, K. Vela Velupillai.

2008.8 *Natural disturbances and natural hazards in mountain forests: a framework for the economic valuation*, Sandra Notaro, Alessandro Paletto

2008.9 *Does forest damage have an economic impact? A case study from the Italian Alps,* Sandra Notaro, Alessandro Paletto, Roberta Raffaelli.

2008.10 *Compliance by believing: an experimental exploration on social norms and impartial agreements,* Marco Faillo, Stefania Ottone, Lorenzo Sacconi.

2008.11 *You Won the Battle. What about the War? A Model of Competition between Proprietary and Open Source Software*, Riccardo Leoncini, Francesco Rentocchini, Giuseppe Vittucci Marzetti.

2008.12 *Minsky's Upward Instability: the Not-Too-Keynesian Optimism of a Financial Cassandra,* Elisabetta De Antoni.

2008.13 *A theoretical analysis of the relationship between social capital and corporate social responsibility: concepts and definitions,* Lorenzo Sacconi, Giacomo Degli Antoni.

2008.14 *Conformity, Reciprocity and the Sense of Justice. How Social Contract-based Preferences and Beliefs Explain Norm Compliance: the Experimental Evidence,* Lorenzo Sacconi, Marco Faillo.

2008.15 *The macroeconomics of imperfect capital markets. Whither saving-investment imbalances?* Roberto Tamborini

2008.16 *Financial Constraints and Firm Export Behavior*, Flora Bellone, Patrick Musso, Lionel Nesta and Stefano Schiavo

2008.17 *Why do frms invest abroad? An analysis of the motives underlying Foreign Direct Investments Financial Constraints and Firm Export Behavior*, Chiara Franco, Francesco Rentocchini and Giuseppe Vittucci Marzetti

2008.18 *CSR as Contractarian Model of Multi-Stakeholder Corporate Governance and the Game-Theory of its Implementation*, Lorenzo Sacconi

2008.19 *Managing Agricultural Price Risk in Developing Countries*, Julie Dana and Christopher L. Gilbert

2008.20 *Commodity Speculation and Commodity Investment*, Christopher L. Gilbert

2008.21 *Inspection games with long-run inspectors*, Luciano Andreozzi

2008.22 *Property Rights and Investments: An Evolutionary Approach*, Luciano Andreozzi

2008.23 *How to Understand High Food Price*, Christopher L. Gilbert

2008.24 *Trade-imbalances networks and exchange rate adjustments: The paradox of a new Plaza*, Andrea Fracasso and Stefano Schiavo

2008.25 *The process of Convergence Towards the Euro for the VISEGRAD – 4 Countries*, Giuliana Passamani

2009.1 *Income Shocks, Coping Strategies, and Consumption Smoothing. An Application to Indonesian Data*, Gabriella Berloffa and Francesca Modena

2009.2 *Clusters of firms in space and time*, Giuseppe Arbia, Giuseppe Espa, Diego Giuliani e Andrea Mazzitelli

2009.3 *A note on maximum likelihood estimation of a Pareto mixture*, Marco Bee, Roberto Benedetti e Giuseppe Espa

2009.4 *Job performance and job satisfaction: an integrated survey*, Maurizio Pugno e Sara Depedri

2009.5 *The evolution of the Sino-American co-dependency: modeling a regime switch in a growth setting*, Luigi Bonatti e Andrea Fracasso

2009.6 *The Two Triangles: What did Wicksell and Keynes know about macroeconomics that modern economists do not (consider)?* Ronny Mazzocchi, Roberto Tamborini e Hans-Michael Trautwein