The Structure and Growth of World Trade, and the Role of Europe in the Global Economy

SIS Research Paper N^o 2014–1

January 2014

Shamnaaz B. SUFRAUJ
University of Trento

Stefano SCHIAVO
University of Trento and OFCE

Massimo RICCABONI
IMT Lucca

School of International Studies
via Tommaso Gar, 14
38122 Trento – ITALY
http://www.unitn.it/ssi

The Structure and Growth of World Trade, and the Role of Europe in the Global Economy*

Shamnaaz B. Sufrauj
School of International Studies, University of Trento

Stefano Schiavo
School of International Studies & Dept. of Economics and Management, University of Trento and OFCE-DRIC

Massimo Riccaboni
IMT Lucca

Abstract

This paper presents a simple stochastic model of proportionate growth to describe international trade and it applies this set-up to the relationship between export dynamics and economic development. Trade flows are assumed to grow as a geometric Brownian motion while new trade links follow a preferential attachment mechanism, and these two processes are assumed to be independent. This simple set-up accurately describes many of the empirical features that characterize the structure and growth of the international trade network. Furthermore, it reconciles diverging views of industrial policy in the economic development literature: although export is very concentrated so that large bilateral flows are rare, countries characterized by a large number of export relations are more likely to capture such "big hits". The stochastic model provides a simple benchmark against which we can assess countries' export performance. We then investigate the determinants of deviation of empirical data from the predictions of the model in terms of the number of "big hits".

*This research has been supported by Banca Intesa San Paolo with a grant in memory of the late Prof. Ferdinando Targetti, first director of the School of International Studies at the University of Trento. Massimo Riccaboni and Stefano Schiavo acknowledge financial support received through the project 'The international trade network: empirical analyses and theoretical models' funded by the Italian Ministry of Education, University and Research (PRIN 2009).

1 Introduction

The use of industrial policy as an instrument of export growth has more than often solicited scepticism or even strong opposition in mainstream economics (Stiglitz et al., 2013). The standard argument is based on the view that any government intervention is likely to breed inefficiency and encourage rent-seeking behaviour. Growth models formalising the supremacy of a liberal over a closed-economy have further led economists and policy makers to have greater recourse to markets to redress the economy. However, the free play of markets has not always produced the desired results; for instance, ardent reform in Latin America did not improve their performance by the 1990s. Contrarily, increased state-involvement in other economies has sometimes led to greater economic success, such as, in Japan. As a result of these discrepancies, development economists have recently revived the debate regarding the actual and potential economic merits of industrial policy. We particularly refer to the debate confronting Hausmann and Rodrik (2003) to Easterly et al. (2009).

Hausmann and Rodrik (2003) and subsequent literature (Rodrik, 2004; Hausmann and Rodrik, 2005) provide a formalism of the case for industrial policy through a process they call "self-discovery", where investors learn what they are good at producing among the wide set of investment possibilities. Once an investor discovers what products would bring the highest returns, this knowledge generates positive spillovers in the economy as it signals other investors where to direct their entrepreneurial efforts. While these informational externality is good for the economy, it discourages initial investment. Consequently, investment is under-supplied or, in other words, there is too little "search" and "discovery". Hence, these authors argue that there is a need for some optimal industrial policy to counteract this distortion. It is made clear in Rodrik (2004) that industrial policy is not about imposing taxes or subsidies but it refers to a greater collaboration between the state and the private sector so as to discover what investment decision would benefit the economy and what kind of intervention is needed or not. In sum, industrial policy is simply the "provision of public goods for the productive sector", such as R&D, infrastructure and training (p.39). In Hausmann and Rodrik (2005), they emphasise that the focus is on process rather than on specific policies, in their words "[i]n view of the inherent uncertainty about what is likely to work, it is more important to design robust institutional arrangements than to adopt an agenda of specific policy actions. The process of self-discovery is as much about policy learning–which types of policies work and which do not under existing realities–as it is about entrepreneurial learning" (p.77).

On the other side of the debate, Easterly et al. (2009) contend that policy aimed at "picking winners"–selective industrial policy–is bound to be unsuccessful.[1] They build their argument on the fact that manufacturing exports are highly concentrated where a few product-destinations account for a disproportionately large share of export values. These few exports that make up most of total export value are termed "big hits"; they capture both the right product and the right market. With such specialisation, the shape of the distribution of exports is highly skewed (close to being a power law) which implies that the probability of identifying a "big hit" is very low, as this probability decreases exponentially with the size of the hit. Hence, these authors claim that it is better to leave markets unhindered by policy. Nevertheless, they

[1] The authors document that manufacturing export success is intricately related to development success.

do mention that in the presence of externalities, "discovery efforts" could be subsidised while letting the market "pick winners" and, hence, getting the best of both worlds.

In this paper, we show that our modelling strategy nicely accommodates the "two worlds"– Hausmann and Rodrik (2003) versus Easterly et al. (2009) as it is able to account for several stylised facts simultaneously. While the distribution of exports is fat-tailed, i.e. "big hits" are rare events, the probability of drawing a large export flow increases with the number of trade relationships. This follows that countries characterised by large "discovery efforts", stimulated or not through industrial policy, are much more likely to draw "big hits".

We describe international trade as a set of transactions of different magnitude occurring between countries using a simple stochastic model of proportionate growth and we apply this set-up to describe the relationship between export dynamics and economic development. Stochastic models have long been used in the literature to assess the economic relevance of a given phenomenon and to establish a benchmark for measuring its extent (Simon, 1955; Simon and Ijiri, 1977). More recently, Ellison and Glaeser (1997), propose a stochastic process whereby firms choose their location by throwing darts on a map; they compare this measure with the observed level of geographic concentration of economic activities. The balls-and-bins model of Armenter and Koren (2010) probably represents the closest approach to our own. These authors describe US exports as balls falling into bins of different sizes each representing a product-destination pair and their simple set-up is able to predict the pattern of zero-trade flows at the extensive margin and to match some stylised facts of US trade flows. They argue that such a model can be more informative to theory when it *misses* an empirical fact rather than when it succeeds to match it. The present paper departs from Armenter and Koren (2010) by focusing on global trade rather than just US trade with the rest of the world. Our model is also able to match many of the stylised facts regarding international trade both at the intensive and extensive margins.

One of the facts that is recurrent in many trade studies and relevant to this paper is the large share of zeros in trade. Such sparsity has been documented by Helpman et al. (2008) who using data for 158 countries for the period 1970-1997 report that about only 50% of country-pairs engage in trade. Similarly, Baldwin and Harrigan (2011), describing US trade at the 10-digit Harmonised System, find that zeros are as present in import as in export data, making up about 82% of the latter. Both papers build on the heterogeneous firms trade model of Melitz (2003) which they improve to account for the share of zeros and other trade patterns. Second, we refer to the literature which report the high concentration of trade both at the intensive (the value of trade) and extensive margin (the number of trade relationships). These facts are well-documented in Easterly et al. (2009). They find that the top 1% export products accounts for 49% of the median export share. At the intensive margin, they report that the top 1% product-destination makes up about 53% of the median export value. Trade concentration seems to have been here for long; investigating the role of distance and the patterns of intra-European trade, Beckerman (1956) stresses the non-normal distribution of trade where trade accumulates at the extremities instead of around the mean. He finds that on average each country exported about 56% of its total export to its 3 major markets in 1938 and about 50% in 1953.

Another pattern closely related to trade concentration is the observation that most countries export only a few products to a few destinations while a small club of countries export most products to many destinations (Riccaboni and Schiavo, 2010). Similarly, the distribution of bilateral trade values is highly skewed assuming a log-normal form (Bhattacharya et al., 2008; Fagiolo et al., 2009) and this shape holds for different level of aggregation of the data. Here too, a small fraction of very large trade relationships exists alongside a large number of small trade flows.

Hence, skewness is evident both at the intensive and at the extensive margin. There is extensive debate and disagreement in recent trade literature as to what role do the intensive and extensive margins play in trade growth. While some papers find that the extensive margin drives trade growth (Hummels and Klenow, 2005), other authors report the dominant role of the intensive margin (Besedeš and Prusa, 2011). In fact, the extensive margin seems to exert a greater influence on trade at the cross-sectional level, for instance, Hummels and Klenow (2005) find that it accounts for 60% of export. On the contrary, the intensive margin explains most of the dynamics of trade over time. Helpman et al. (2008) report that the rapid growth of world trade since the 1970s was mostly trade between existing countries rather than with new partners. Similarly, Besedeš and Prusa (2011) emphasise the temporal nature of extensive trade in export growth so that extensive trade has no effect in the long-term.

While there is a largely held view and empirical evidence that diversification of exports is positively related to economic development, Cadot et al. (2009) and previously Imbs and Wacziarg (2003) find that there is an increase in concentration at higher level of development. They show that most of the action happen at the extensive margin by using the Theil Index of concentration. The latter is an interesting measure as it can be decomposed additively into a between-group and a within-group components. Changes in the between-group component can be matched into changes in the extensive margin of trade while changes in the within-group component can be mapped into changes in the intensive margin of trade. This property is explained and proved by Cadot et al. (2009).

We also draw from studies that explore the dependency of growth rates on size of firms. Stanley et al. (1996) find that the variances of growth rates differ with firms of different sizes and decreases with larger sizes. Indeed, the distribution of growth rates is not Gaussian but rather exponential. The same holds for the size of an economy as reported in Canning et al. (1998) and for many other economic phenomena. Plotting the log of the standard deviation of firms' growth rate against the log of their initial sales value, a power-law relationship emerges. This result challenges beliefs that volatility in growth rates decreases quickly with size.

Among other branches of research, we refer much to development literature that attempt to explain export success, or in our case, the identification of "big hits". We look into the literature on country capabilities which contends that economic development rests on both tradable and non-tradable capabilities present in a country (Hidalgo and Hausmann, 2009; Hausmann and Hidalgo, 2011). Further, we draw from literature that stresses the importance of financial development, institutions and geography in economic performance.

2 The model

We describe international trade as a set of links (bilateral transactions) of different magnitude occurring among pair of nodes (countries). Consistently with the empirical analysis, we define a trade link as a product-destination pair. We assume the binary structure of trade (i.e. the presence/absence of a link between two specific countries and the number of links maintained by each country) is governed by a process of "preferential attachment": countries establish new trade links based on the number of connection they already have. Hence, more active exporters are more likely to export new products and/or reach new markets.[2]

Apart from being remarkably consistent with a large number of empirical networks, this mechanism of network formation and growth is consistent with the notion that exporting represents a key engine of economic growth thanks to the endogenous forces set in motion by learning effects in the manufacturing sector (Kaldor, 1957). More recent framings of this view of cumulative causations are the "self-discovery" story presented by Hausmann and Rodrik (2003) and the "complexity" approach proposed by (Hidalgo and Hausmann, 2009). The former postulates that attempts to set up new businesses and export new products to new destinations generate valuable public information as they signal profitable opportunities or dead ends. The latter builds on the notion that by producing a given set of goods each country cumulates a number of capabilities: the more capabilities are present, the easier it is to recombine them and put them to a novel use. Microfounded accounts of preferential attachment are offered by Chaney (2011) and Krautheim (2012). The first assumes that firms can establish links with suppliers either at random or via existing connections (meeting friends of friends); the second models the fixed costs associated with penetrating a foreign market as a decreasing function of the number of firms already exporting there (from a given source country) due to the presence of spillover effects.

Building on the extension to weighted networks proposed by Riccaboni and Schiavo (2010) we assume that, while the binary structure of international trade follows a purely preferential attachment mechanism, the dollar value of each trade flow grows according to a geometric Brownian motion. Moreover, the two processes governing the formation of links and their growth are assumed to be independent.

More formally, we can describe the stochastic model as follows:

- at time $t = 0$ there are N_0 countries each characterized by a self loop;[3]
- at each time step $t = \{1, \ldots, T\}$, a new link among two countries arises: thus the number of links existing at time t is $m_t = t$;

We identify the number of links of country i at time t with $K_i(t)$, whereas $K_{ij}(t)$ represents the number of products traded between countries i and j. To identify the countries connected to each link we adopt the following procedure:

[2] Preferential attachment is typically associated with the pioneering work by Barabási and Albert (1999) and gives rise to a very skewed connectivity distributions (few well-connected nodes coexist alongside a large number of peripheral actors), which is found to characterize many real-world applications beside trade (the internet, world-wide air trasportation, mobile communication, interbank payments to quote just a few; see Faloutsos et al., 1999; Guimera et al., 2005; Onnela et al., 2007; Soram aki et al., 2007).

[3] This only serves for initialization purpose: self loops are never considered in the analysis.

- with probability a the new link is assigned to a new country, whereas with probability $1-a$ it is allocated to an existing country;
- in the latter case, the probability of choosing country i is given by $p_i(t) = K_i(t-1)/2t$; i.e. the probability depends on the number of links already secured by country i;
- the same process governs both side of the trade link, meaning that the two partner countries are chosen symmetrically with $i \neq j$;

Hence, at each time t this set of rules identifies the pair of (distinct) countries to be linked and the process generates the binary structure of the network, meaning the number of active bilateral links, the number of zero trade flows, the number of links associated with each country.

For what concerns the value of each trade flow, we assume that it grows in time according to a simple random process:

- at time t each (existing) trade flow between countries i and j has weight $w_{ij}(t) > 0$;[4]
- at time $t+1$ the weight of each link is increased or decreased by a random shock $x_{ij}(t)$, so that $w_{ij}(t+1) = w_{ij}(t)x_{ij}(t)$. All we need to assume is that the shocks and initial link values are taken from a distribution with finite mean and standard deviation.[5]

We therefore combine a preferential attachment mechanism, with an independent geometric Brownian motion characterizing the magnitude of bilateral trade flows. As detailed in Riccaboni and Schiavo (2010), this simple setup gives rise to very skewed distributions that are consistent with the empirical evidence. In particular, in the limit of large t, with $a > 0$ and small, the connectivity distribution converges to a power-law with an exponential cutoff (Yamasaki et al., 2006). For what concerns the distribution of bilateral trade flows, the proportional growth process described above implies that the distribution of the weights $P(w)$ converges to a lognormal.

We do not regard this setup as a full-fledged economic model that *explains* bilateral trade flows, but rather as a stochastic benchmark against which to compare actual data.

3 Data and Empirics

3.1 Data

We use data on bilateral trade flows contained in the BACI dataset that reconciles data reported by exporting and importing countries to the United Nations Statistics Division.[6] The bilateral trade flows we use in the present paper cover a maximum of 194 countries and 5039 product categories at the 6-digit Harmonised System(HS) level. We clean the data for very small countries (such as the Cook Islands) for which data are sparse and not reliable, and end up with a dataset covering 186 countries over a 15-year period between 1995 and 2009.

[4] We further assume that K_i, K_j and w_{ij} are independent random variables, i.e. the value of bilateral trade is independent of the connectivity of the two partner countries.

[5] So in principle the initial value of trade could be 1, or can be defined in terms of some exogenous determinant such as size and distance, as in a standard gravity model.

[6] For more information on the construction of BACI, refer to Gaulier and Zignago (2010).

3.2 Regularities in empirical data

As mentioned previously, sparsity of trade is one of the most documented stylised facts regarding bilateral trade flows. We also find in our data that the number of active trade links is far below the potential one. Over the period 1995-2009, in terms of aggregate trade flows to a destination, the average share of zeros is about 47%, and the highest share of 57% is in 1995. Since many products are exported to just one or few destinations, we also look at the share of zeros at the product-destination level and the average is as high as 97%. As shown in figure 1, both shares exhibit a declining trend over the years, except in 2009, but remain high.

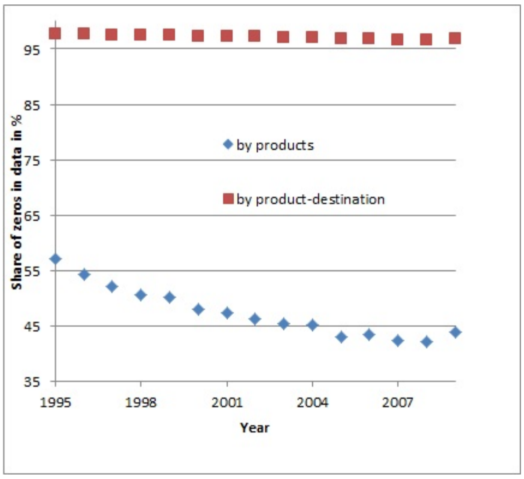


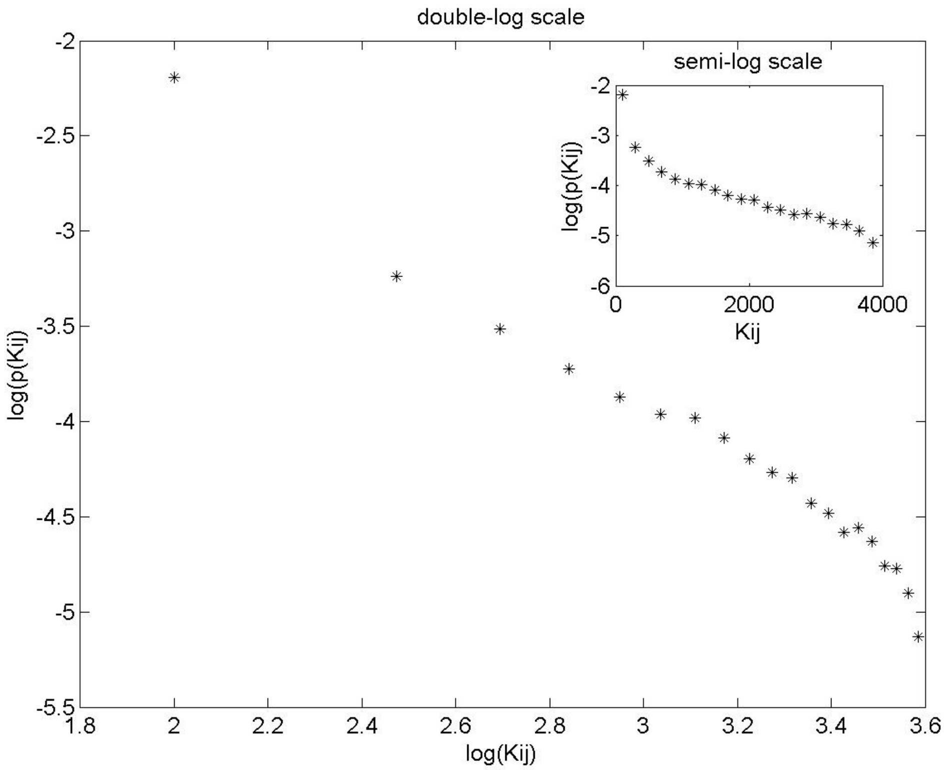
Figure 1 Share of zero trade flows at the products and product-destinations level

Focusing now on active trade links, we look at the distribution of the number of trade relationships (at the product-destination level) maintained by each country. Figure 2 shows such distributions for the years 2000 (panel a) and 2005 (panel b). These distributions are well approximated by a Pareto in the body (the straight line behaviour is apparent in the plot), with an an exponential cut-off in the upper tail, where the number of product-destinations is very large. This departure is magnified in the inset (in semi-log-scale) of the figures where the exponential part now appears as a straight line. This behaviour conforms to the prediction of our model.

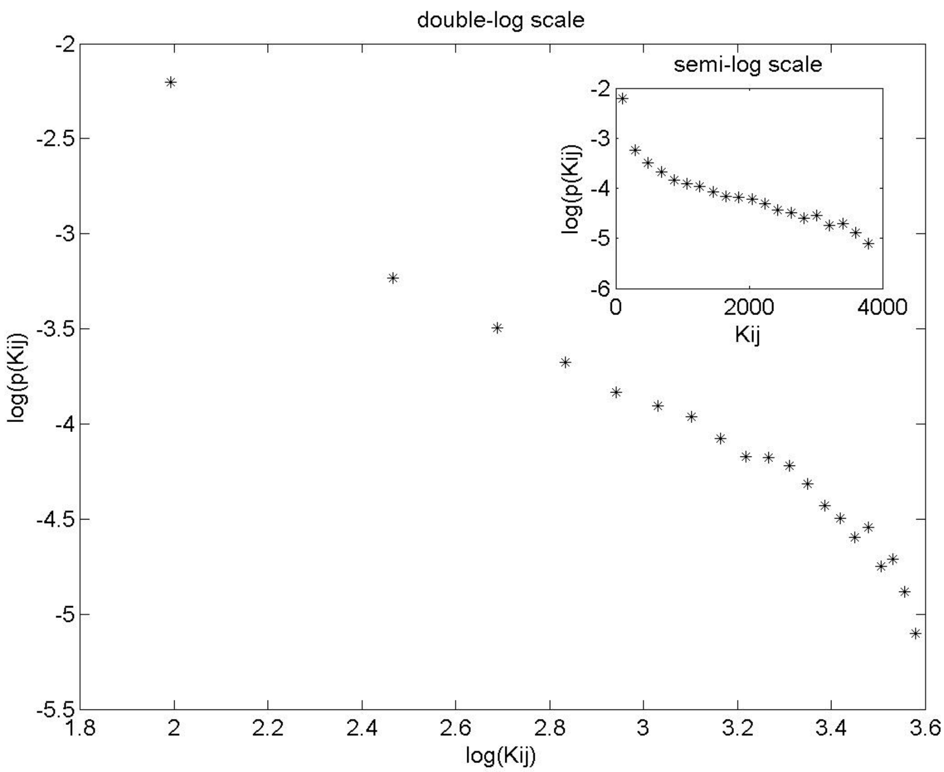
We also expect the distribution of trade values to be log-normally distributed as implied by the Gibrat process of proportionate growth driving the intensive margin. Figure 3 a) shows the complementary cumulative distribution function of manufacturing trade for both bilateral trade values and total trade values, i.e., total trade values of country i, together with the log-normal estimates for the year 2005. The log-normal estimates are calculated using maximum likelihood for truncated data.[7] Figure b) shows the probability-probability plot using least-square estimation method. The distributions of trade values is fat-tailed and in the aggregate the lognormal estimates provide a good fit. Using the method of least-square a log-normal distribution fits well even the disaggregated data as shown in figure 3 b).

The data also show the striking degree of specialisation in exports. For example, in 2005, the median export shares for the top 20%, 10% and 1% products are 97%, 93% and 66% re-

[7] See Bee (2006).

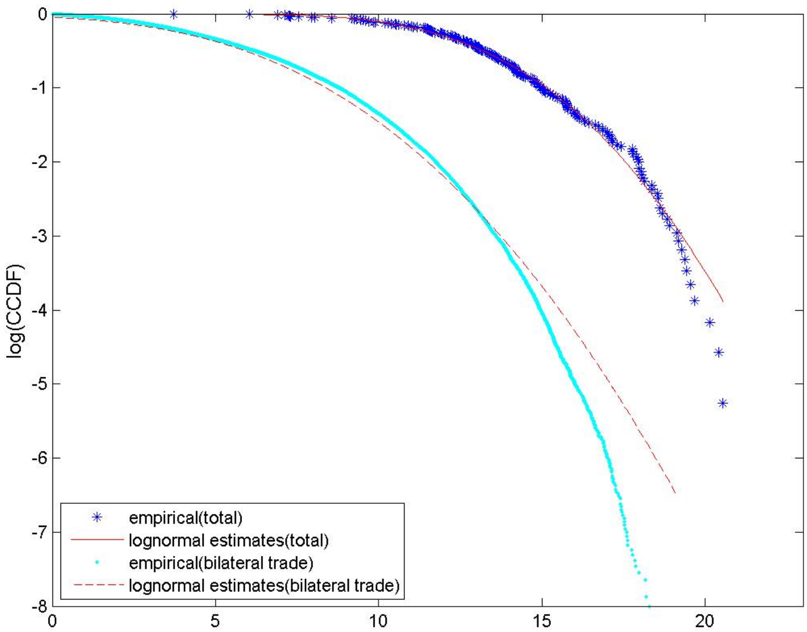


(a) Year 2000

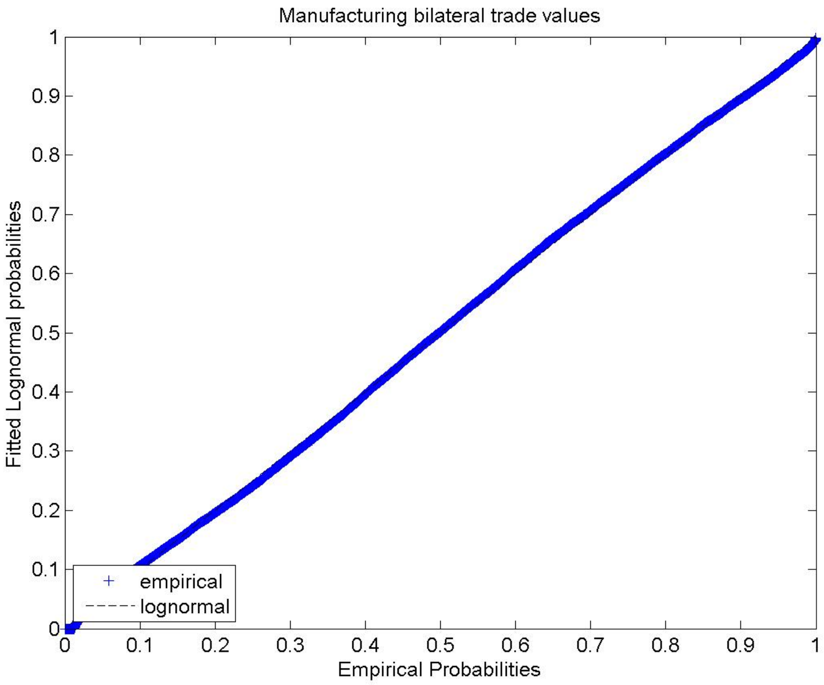


(b) Year 2005

Figure 2 Connectivity distributions at HS6: Main plot in double log and inset in semi-log



(a)



(b)

Figure 3 Distributions of bilateral and total manufacturing trade flows and lognormal fits spectively as shown in table 1. The median share for just the top 3 products is as high as 42%. A comparison of the 2005 figures with those of 1995 show that export concentration has increased slightly over time. Concentration at the product-destination level is very high and increases over time. At low levels, such as, for the top 10 or top 3 products the figures are lower than those at the product level since there are many more product-destinations than products.[8]

Table 1 Export share for the year 1995, 2000, 2005 and 2009 (top 20% stands for top 20 percent and top 10 stands for top 10 products)

Share of exports by products											
Year 1995	***mean***	***p50***	***sd***	***min***	***max***	***Year 2000***	***mean***	***p50***	***sd***	***min***	***max***
top 20%	94.7	96.9	6.7	54.8	99.9	**top 20%**	95.1	96.9	5.7	58.5	100
top 10%	89	92.3	10.7	35.5	99.8	**top 10%**	89.8	92.6	9.9	34.3	100
top 1%	59.9	59.4	20.3	9.2	97.7	**top 1%**	62.2	61.3	19.7	14.1	100
top 10	61.9	65.6	28.1	10.4	100	**top 10**	62.3	63.7	26.1	12.1	100
top 3	45.5	40.1	28.5	4.2	100	**top 3**	45.1	40.2	26.8	5	100
top 1	28.3	20.4	23.6	1.7	97.4	**top 1**	28.7	20.5	23.7	1.8	100
Year 2005	**mean**	**p50**	**sd**	**min**	**max**	**Year 2009**	**mean**	**p50**	**sd**	**min**	**max**
top 20%	95.8	97.2	5.2	60.1	100	**top 20%**	96.8	97.6	3.1	76	100
top 10%	91.2	93.4	8.2	49.6	99.9	**top 10%**	92.9	93.9	5.6	59.6	100
top 1%	65.5	65.5	18.9	16.4	99.5	**top 1%**	66	65.5	14.5	13.6	95.6
top 10	62	62.8	26.2	12.3	100	**top 10**	46.9	43.9	28	3.4	97.9
top 3	45.7	41.9	27.5	6.5	100	**top 3**	30	25	22.4	1.4	94.9
top 1	28.9	20.9	23.8	2.9	98.2	**top 1**	16.8	13.2	15.8	0.5	86.7
Share of exports by product-destinations											
Year 1995	***mean***	***p50***	***sd***	***min***	***max***	***Year 2000***	***mean***	***p50***	***sd***	***min***	***max***
top 20%	95	96.1	6.2	54.8	99.8	**top 20%**	95.6	96.8	5.2	58.4	100
top 10%	89.4	91.2	9.4	35.5	99.4	**top 10%**	90.7	92.3	8.9	34.3	100
top 1%	58.6	58.8	16.3	9.2	97.4	**top 1%**	62	60.4	16.5	14.1	100
top 10	47.5	43.5	29	3	100	**top 10**	47.4	44.5	28	3.3	100
top 3	31.4	22.9	24.8	1.3	100	**top 3**	30.8	23.6	23.8	1.3	100
top 1	17.6	11	17.9	0.5	97.4	**top 1**	17.2	11.9	17.7	0.6	100
Year 2005	***mean***	***p50***	***sd***	***min***	***max***	***Year 2009***	***mean***	***p50***	***sd***	***min***	***max***
top 20%	96.5	97.3	4.4	60.1	100	**top 20%**	96.8	97.6	3.1	76	100
top 10%	92.6	93.7	6.6	49.6	99.9	**top 10%**	92.9	93.9	5.6	59.6	100
top 1%	66.5	65.1	14.8	16.4	96.8	**top 1%**	66	65.5	14.5	13.6	95.6
top 10	46.6	42.9	27.9	2.9	100	**top 10**	46.9	43.9	28	3.4	97.8
top 3	30.6	24.6	23.6	1.2	100	**top 3**	30	25	22.4	1.4	94.9
top 1	17.2	11.1	16.7	0.5	90.3	**top 1**	16.8	13.2	15.8	0.5	86.7
Share of exports by destinations											
Year 1995	***mean***	***p50***	***sd***	***min***	***max***	***Year 2000***	***mean***	***p50***	***sd***	***min***	***max***
top 20%	92.3	93.3	5.3	71.6	100	**top 20%**	92.9	94	5.5	63.5	100
top 10%	80.8	81.9	9.9	52.4	100	**top 10%**	82.9	83.9	9	50.8	100
top 1%	39.9	35.8	20	11.4	100	**top 1%**	41.2	36.1	18.8	13.1	100
top 10	84.3	83.7	11.3	59.6	100	**top 10**	82.4	82.7	11.8	53.6	100
top 3	59.7	57	19.2	27.2	100	**top 3**	57.6	53.5	19.5	24.1	100
top 1	36	30.5	21.2	11.4	100	**top 1**	35.1	28.6	20.9	10.3	100
Year 2005	mean	p50	sd	min	max	**Year 2009**	mean	p50	sd	min	max
top 20%	93.7	94.5	4.2	71.6	99.6	**top 20%**	93.5	94.1	4.3	63.7	100
top 10%	83.5	84.1	7.9	55.1	97.9	**top 10%**	83	83.2	8.6	39.3	100
top 1%	39.4	35.2	17.3	14.3	97.6	**top 1%**	39.1	34.4	18	15.1	100
top 10	80.6	81.4	11.9	51.4	100	**top 10**	80.2	81.2	13.1	49.4	100
top 3	54.8	50.9	18.3	21.8	100	**top 3**	54.6	50.8	19.4	22.5	100
top 1	32.2	25.2	19.1	8	97.6	**top 1**	32.1	24.5	19.9	8.6	100

[8]Our figures are higher than those reported by Easterly et al. (2009). They restrict their analysis to manufacturing exports, thus dropping export of natural resources which is dependent on a country's endowments and therefore inflate concentration. For comparative purposes, we also consider the restricted sample of manufacturing exports for the year 2000 (the same used by Easterly et al., 2009): as we expected, we obtain lower concentration figures than using the whole sample, and figures that are even lower than those reported by Easterly et al. (2009). This may be due to the fact that their dataset contains fewer categories of products (2950) and fewer countries (151) than ours.

Table 2 further investigates the issue of concentration by reporting the standard concentration indices.[9] Except for the Herfindahl index, the Gini, relative entropy and Theil index, all show that export is highly concentrated and confirm the previous findings. The Theil can be decomposed additively into a between- and a within-group component: the former reflects changes in the extensive margin of trade while the latter mirrors the behaviour of the intensive margin (Cadot et al., 2009). Results (not displayed) confirm the findings obtained by Cadot et al. (2009): diversification appears to be mainly driven by the extensive margin. Furthermore, while at the intensive margin there is a positive relation between the Theil index and GDP per capita (concentration increases with income), at the extensive margin this relationship is U-shaped.

Table 2 Concentration indices and descriptive statistics

	mean	**st. dev.**	**min**	**max**	**obs.**
Gini	0.998	0.004	0.977	1.000	194
Entropy	0.639	0.150	0.291	0.893	194
Theil	8.823	2.074	4.022	12.327	194
Herfindahl	0.079	0.098	0.000	0.470	194

Finally, figure 4 shows the relationship between connectivity (the extensive margin) and the value of bilateral trade flows (intensive margin) for manufacturing exports on a log-log scale. A positive relationship emerges with the coefficient of the interpolating line being $\theta \approx 1.32$ for $W = K^{\theta}$ for the year 2005.[10] In economic terms, a $\theta > 1$ implies a positive relationship between the extensive and the intensive margin so that countries entertaining a larger number of trade links feature higher average and total trade. This is consistent with the empirical evidence that the extensive margin plays a prominent role in explaining cross-country differences in total exports (Hummels and Klenow, 2005) and with the findings reported in Easterly et al. (2009).

4 Big hits

In this section, we lay down our contribution to the debate regarding the role of industrial policy in export success. To re-frame, Hausmann and Rodrik (2003) regard development as a self-discovery process wherein policies are designed to discover the productive potential of a country. Without public support, entrepreneurial activity (discovery efforts) will be underprovided because of the non-appropriability of the returns from the investment. In fact, since "search and discovery" provides more than just private returns, so that the knowledge that is produced tends to spillover to the whole economy, other entrepreneurs can exploit the information and the returns to the society are larger than private returns. This entails a market failure and call for public intervention.

[9] We calculated these indices for both the empirical and theoretical maximum number of product-destinations. The theoretical statistics do not differ much from the empirical ones because the theoretical maximum number of product-destinations is not very different from its empirical counterpart (5039 and 5014 respectively) The table only shows the empirical results.

[10] This coefficient increases over time (results are not reported here).

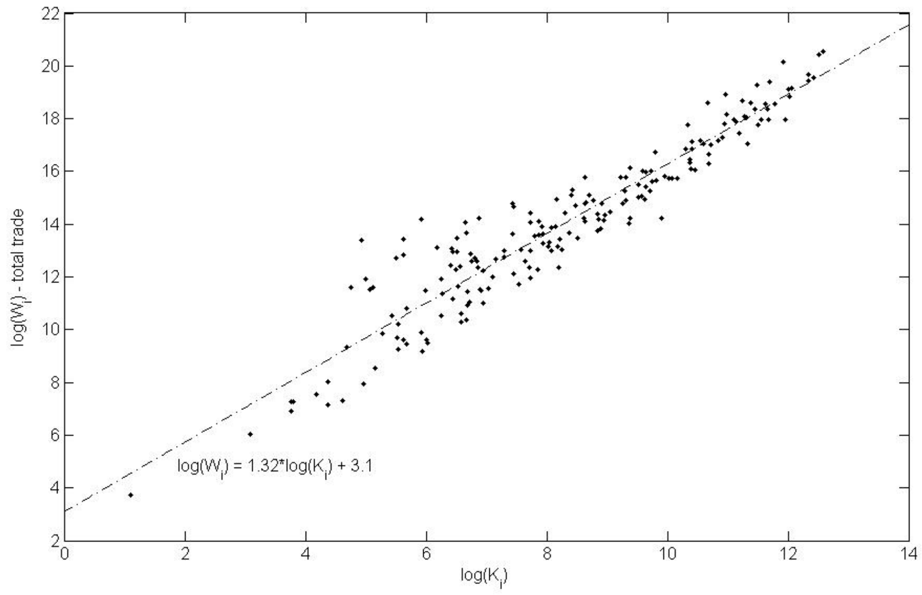


Figure 4 Relationship between the intensive and extensive margin

This idea is well captured by our preferential attachment assumption which can be interpreted in the following way: countries that make a successful discovery are more likely to discover in the future. That is to say, the probability of capturing a new export link is increasing in the number of links already established. This also means that discovery costs will be spread over more products or links, hence, further encouraging initial investment.

Although Easterly et al. (2009) are of the view that a strategy of "picking winners" is unlikely to be successful, we argue that the role of industrial policy is to stimulate as much search as is optimally desirable. As Easterly et al. (2009) contends, we agree that "big hits" are rare events. This follows from the skewness of the distribution of export values. However, we focus on the fact there exists a more than proportional relationship between the extensive (the number of export links) and the intensive margin of trade (export values) as shown in figure 4. It follows that the probability of drawing a large flow increases with the number of trade relationships. The more a country searches, the more likely it is to discover a successful export link. As such, policies that stimulate "search" (the establishment of as many links as possible) and "discovery" efforts (the identification of a big hit) would be beneficial for development.

To prove our argument, we plot the number of big hits against the number of product-destinations as shown in figure 5 and we note the positive relationship. An export value, w_{ij} is defined as a big hit if w_{ij} > three standard deviations from the mean export value.[11] The figure also shows the simulated big hits which we obtained by counting the number of big hits from drawing a hundred simulations from a log-normal distribution with the empirical parameters. Thus, we can compare how the observed data deviate from the simulated benchmark. Table 3 reports the descriptive statistics which show that on average the actual data have less big hits than the simulated benchmark and the actual data have 65 more countries with zero big hits.

[11] While Easterly et al. (2009) imply that a big hit is a large export value, they do not provide a definition of big hits.

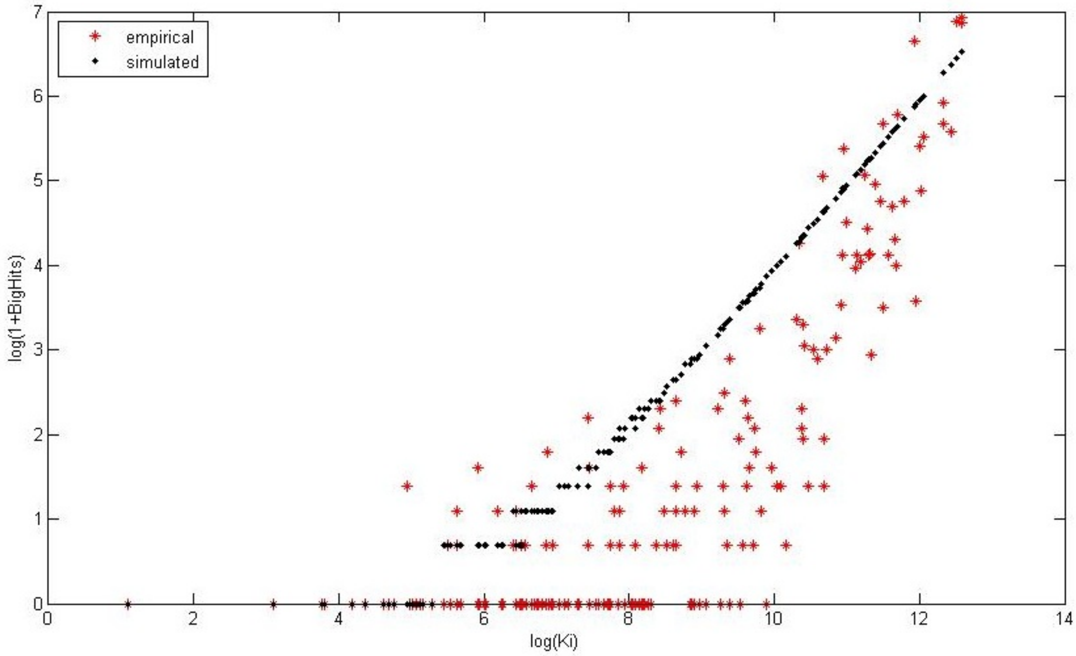


Figure 5 "Big hits" versus K (Number of product-destinations–empirical and simulated data, year 2005

Table 3 Descriptive statistics: simulated and empirical big hits

	Simulated	**Empirical**
No. of Countries	188	188
Mean	68	44
Std. Dev.	131	146
Max.	679	1027
Quantile 0.25	2	0
Median	9	1
Quantile 0.75	70	9
No. of big hits = 0	14	79
No. of big hits = 1	25	20
No. of big hits = 20	1	1
No. of big hits > 200	21	12
No. of big hits > 500	6	4

Note:
95 % confidence interval of simulated data is 5.5 to 185.6

4.1 Big hits in the world and in Europe

Table 4 lists the 14 countries that are doing better than what the model predicts. Of those countries that have more than a 100 "big hits", Japan and Mexico exceed the model prediction by a remarkable 50% and 33% respectively. Germany and Ireland are the only European countries that fare better. On the other side, there are many more countries (156) that fall short of the benchmark (see list of all countries in appendix A.2, these include countries such as Italy, Spain, Turkey and Singapore. Table 5 shows the European countries that perform worse than the benchmark.

In terms of percentage deviation from the benchmark, the Baltic countries perform poorly with both Lithuania and Latvia having a single empirical big hit much less than their simulated benchmark. They are small economies, hence, they have a limited market size. Despite their recent integration to the EU, they are technologically less-advanced than most developed countries. Their economic structure is mainly agricultural and light industry, hence further reducing their probability of having big hits in manufacturing. Most eastern European countries also lag behind their benchmark by a great deal. However, the actual big hits of the Czech Republic, Hungary and Poland ranks fairly well, doing even better than the advanced nation Denmark which only has 31 "big hits". Other advanced European countries, such as, France, Netherlands, Italy or the UK, show good performances. Greece shows a very poor performance given its size and its high simulated benchmark.

Table 4 List of countries doing better than the benchmark, year 2005

Country	Empirical	Simulated	% Deviation
Japan	705	354	50
China	939	678	28
USA	887	627	29
Germany	865	679	22
Mexico	199	133	33
Ireland	148	102	31
Canada	265	230	13
Rep. of Korea	296	280	5
Marshall Isds	3	0	100
Liberia	3	1	67
Mozambique	4	2	50
Zambia	6	4	33
Bahamas	3	2	33
Suriname	2	1	50

Note: Except for Marshall Islands, Liberia and Suriname, all empirical values fall within the 95% CI of the simulated big hits

4.2 Potential determinants of big hits deviation

To investigate why some countries fare better while others lag behind as compared to the benchmark, we regress the deviations in "big hits" from the simulated benchmark on a number

Table 5 The European picture (2005)

Countries	Empirical	Simulated	%deviation
France	328	529	-61
United Kingdom	257	530	-106
Netherlands	229	405	-77
Belgium-Luxembourg	186	383	-106
Spain	113	388	-243
Switzerland	106	309	-192
Sweden	99	262	-165
Austria	59	273	-363
Hungary	57	134	-135
Poland	55	190	-245
Czech Rep.	52	190	-265
Finland	46	157	-241
Denmark	31	231	-645
Portugal	29	129	-345
Romania	12	93	-675
Slovenia	5	103	-1960
Iceland	3	13	-333
Croatia	3	56	-1767
Greece	3	102	-3300
Estonia	2	43	-2050
Bulgaria	2	84	-4100
Latvia	1	38	-3700
Lithuania	1	60	-5900
Cyprus	0	20	-100

Note: Only Germany's and Ireland's empirical values fall within the 95% CI of the simulated big hits

of potential determinants of these deviations. We get our inspiration from the industrial policy debate and the larger development literature that tries to assess the determinants of economic performance.

Industrial policy is a tricky term; what it implies depends much on the context. In his work, Rodrik stresses that there is no one recipe that fits all when it comes to policy formulation. What is understood is that the main role of industrial policy is both to inform the private sector of externalities and to provide the optimal environment for growth through implementation of policies. He made it clear that it cannot also be reduced to the application of taxes or subsidies. Thus, it becomes very difficult to find a measure of industrial policy and even more challenging to assess the impact of industrial policy on economic performance.

Even Rodrik himself warns that nothing can be learnt from regressing economic growth or any other performance indicator on policies (Rodrik, 2012). This is because while policy endogeneity can be dealt with, the problem lies beyond econometrics and is conceptual since the endogeneity is an integral part of the null hypothesis. We operationalise the concept of industrial policy through crude measures, such as, tariffs and other institutional variables, such as, rule of law and ease-of-doing business variables.

The next set of potential determinants of big hits deviation comes from closely related literature, namely the literature on country capabilities as recently developed by Hidalgo et al. (2007) and further explored in Hidalgo and Hausmann (2009). The classical view is that the performance of a country is dependent upon factor accumulation. Increases in total factor productivity results in increases in income per capita. Hence, policy implies human and physical capital accumulation. However, total factor productivity explains only a fraction of economic performance.

According to the capability discourse the productive structure of a country depends on the availability of a specific set of capabilities, including tradables and non tradables, such as, norms, institutions, and social networks. Countries differ in their capabilities; high-income ones tend to produce more varieties and more complex products, which require many capabilities whereas poor countries make few and less complex products that require fewer capabilities. Countries have capabilities and products require capabilities. More sophisticated products require a larger range of capabilities and will tend to be produced by fewer countries, i.e. these products are less ubiquitous. They show a combination of the diversification of a country and product ubiquity are indirect measures of a country's capability.

Since the work of Schumpeter (1934), the role of financial development in economic growth has received considerable scholarly attention. The latter theorises that a well-developed financial system nurtures innovation and economic growth by providing entrepreneurs with the necessary conditions for success including better allocation of resources, lesser information asymmetries and so on. Using cross-country regressions for 77 countries over the period 1960-1989, King and Levine (1993) observe a significant positive impact of financial depth on growth. Their results are consistent with different definitions of financial depth and, more interestingly, with different growth measures, namely, real GDP per capita growth, real per capita capital growth and productivity growth. We, thus, include financial development variables which could potentially influence big hits.

We also include a set of network measures to assess the importance of connectivities on deviations in big hits (Miura, 2011). Degree centrality refers to the importance of a node in a network based on its number of connections. Closeness centrality gives higher centrality scores to nodes that are situated closer to other nodes of their component where closeness refers to the inverse of the average shortest paths. Betweenness centrality gives larger centrality scores to nodes that lie on a higher proportion of shortest paths linking nodes other than itself; in simpler terms, a node that lies on a communication path between two other nodes, is an important node. The extent to which nodes in a network are concentrated is measured using the clustering coefficient[12].

4.3 Quantile and WALS regression

We estimate the following regression model, using quantile regression and model averaging techniques:

$$DeviationBH_i = \alpha + \beta_1 InitialY_i + \beta_2 Size_i + \beta_\theta \mathbf{X}_i + \epsilon_i$$

where $InitialY_i$ is initial GDP per capita (year 1995), $Size_i$ is GDP in constant US$ to control

[12] See the appendix for mathematical definitions

for size, and X_i is a set of controls. A full description of the variables and their sources is reported in the appendix.

The use of quantile techniques is motivated by the shape of the distribution of the deviation of big hits which is fat-tailed[13]. Quantile regression (Koenker and Bassett Jr, 1978) differs from standard least-square regression techniques as it allows the estimation of different quantiles of the dependent variable, whereas standard least-square provides summary estimates that approximate the mean effect of the dependent variable given a set of values of the independent variables. Hence, quantile regressions provide a more complete picture of the relationship between the response and predicted variables.

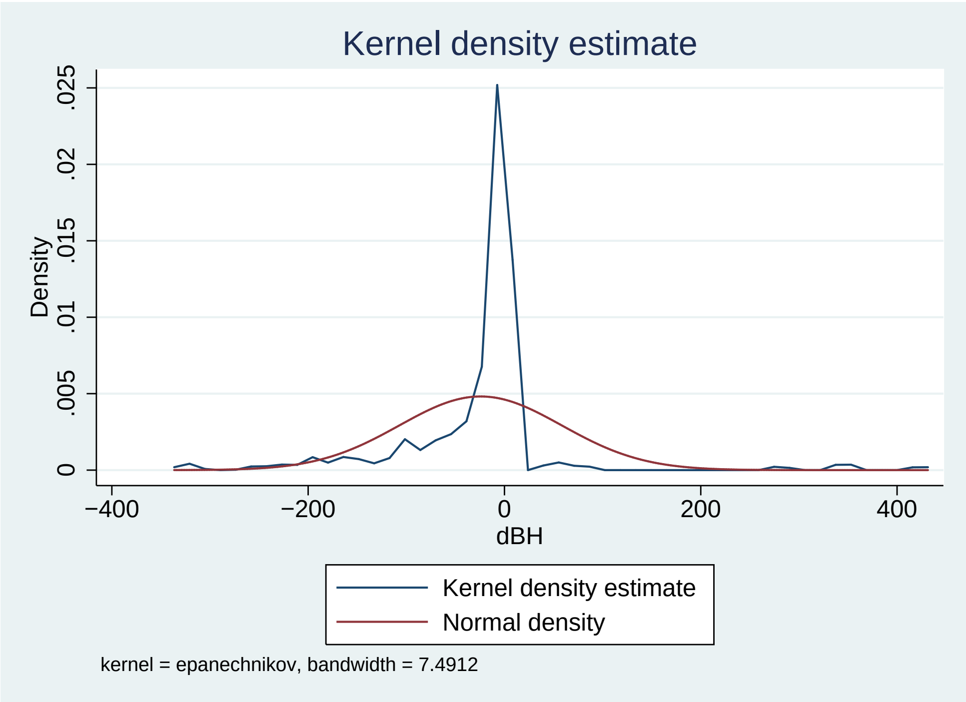


Figure 6 Kernel density of deviations in big hits

The fact that there are only 16 countries out of 188 doing better than the benchmark gives a non-Gaussian distribution of deviations in big hits as shown in figure 6. As such, we cannot assume normally distributed errors. Hence, we suspect there might be different factors working at different portions of the distribution. It is as informative to understand what factors cause some countries to perform badly compared to the benchmark as it is to uncover why others excel as regards the reference point. Quantile techniques allow the estimation of coefficients at different quantiles of the dependent variable. In our case, it allows us to acknowledge the heterogeneity of countries in their export performance.

Furthermore, our largest deviation 424 is much larger than the mean deviation of -24. In our case, it would be major loss of information to remove such extreme observations because they form the basis of successes or failures in big hits. Since the objective function of a quantile regression is a weighted sum of absolute deviations, it provides a robust measure of location (Buchinsky, 1998). As quantile regressions are more robust to outliers and skewed distributions, it remains one of our preferred estimation technique.

[13] with a positive kurtosis of more than 13

Koenker and Bassett Jr (1978) introduced the quantile regression model which can be written as

$$y_i = x_i^{'}\beta_\tau + u_{\tau i} \text{ with } Q_\tau(y_i|x_i) = x_i^{'}\beta_\tau \tag{1}$$

where $Q_\tau(y_i|x_i)$ is the conditional quantile of y_i given the regressor x_i, β is the vector of parameters to be estimated and u is a vector of residuals. The τ^{th} regression quantile estimator is found by minimising:

$$\min_{\beta} \frac{1}{n}\left[\sum_{i\in y_i\geq x_i^{'}\beta} \tau|y_i - x_i^{'}\beta| + \sum_{i\in y_i< x_i^{'}\beta} (1-\tau)|(y_i - x_i^{'}\beta|\right] \tag{2}$$

Linear programming is used to solve Eq. 4.3 (Buchinsky, 1994; Koenker, 2006).

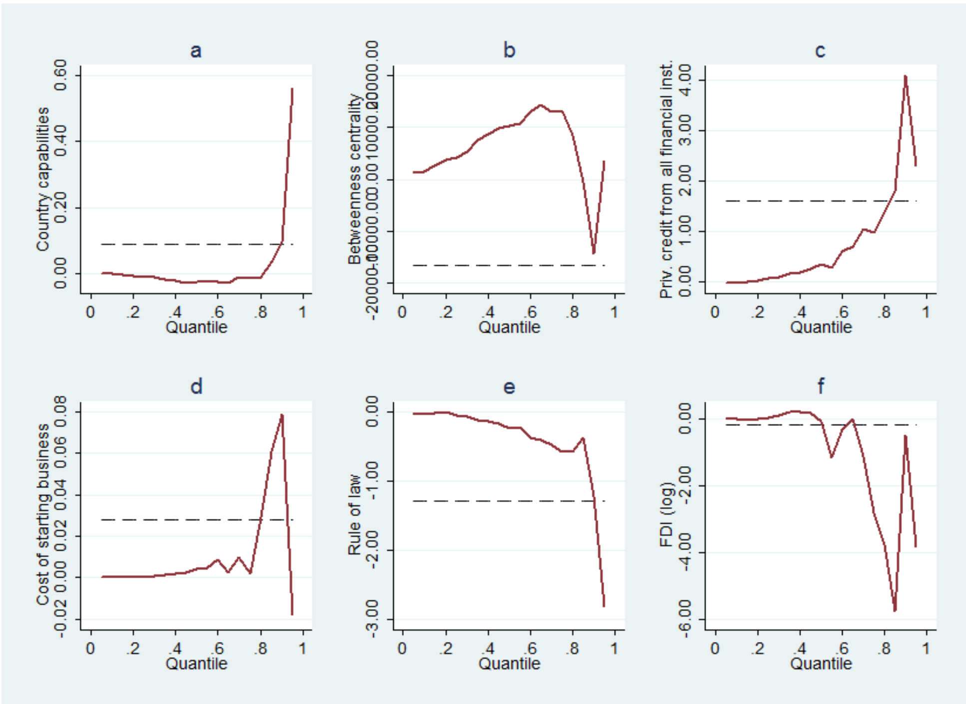


Figure 7 Quantile results for selected variables

Figure 7 shows the graphical results of the quantile regression for selected variables of interest over the conditional deviation in "big hits" distribution. The horizontal lines represent the OLS estimates. The graphs clearly shows how the effect of the different variables vary over quantiles, for instance, as shown in figure 8a country capabilities have a negative effect at low quantiles but a positive one for well-performing "big hits" at the upper quantile. The OLS estimate fail to capture this effect. As for betweenness centrality, figure 8b demonstrates that the OLS regression provides a misrepresentation of the variation in the variable.

While we are guided by related theory, there is no established theoretical background for finding the best model specification for the deviations in big hits, let alone for economic performance. Model averaging techniques are useful to overcome the uncertainty linked in finding the "true" modelling strategy when there exists a large number of potential regressors. They are best used as robustness check to confirm that inferences of the estimated coefficients do not vary widely between model specifications. They aim at finding the best possible estimates as opposed to the best possible model. As such, we also check the robustness of our results by making use of the recent model averaging technique developed by Magnus et al. (2010), the weighted average least square (WALS).

Model averaging, such as WALS, consists of making inferences on the estimates from all models in a model space and the estimates are weighed according to their statistical strengths, i.e. their posterior model probabilities. In fact, the WALS estimator is a Bayesian combination of frequentist estimators. WALS is considered theoretically and practically superior to standard techniques such as Bayesian model averaging (BMA) as it explicitly takes into account model uncertainty and the amount of computing time is linear in the number of regressors rather than exponential as in standard BMA[14]. In fact, the WALS estimator is a Bayesian combination of frequentist estimators. Explanatory variables can be distinguished between focus and auxiliary. Focus regressors are those included on theoretical or other basis while auxiliary regressors are those over which uncertainty exists and over which model selection takes place.

The numerical results of the different regressions are shown in table 6. Column OLS is a robust ordinary least square regression, columns Q(0.20), Q(0.50), Q(0.70), Q(0.90) are the 20%, 50%, 70% and 90% quantiles respectively. Column WALS_0 is a WALS regression with no focus regressors, that is, we perform model selection and assume uncertainty over all the regressors. WALS_f includes 10 focus regressors included as such based on their significance level in previous regressions;they are country capabilities, betweenness centrality, start-up costs and the regional dummies. They are marked by (f) in the regression table 6. WALS_33 is a model with a restricted sample of 102 observations but boasting 33 auxiliary regressors.

The impact of the variables on deviations in “big hits” differ between quantiles and also differ considerably from the OLS results. While the OLS result implies that initial income has a negative impact on deviations, looking only at the OLS result would be misleading since initial income has a significant impact only at the 70% quantile regressions. Model averaging results confirm the negative impact of initial income level. Size of an economy or GDP has a positive impact on deviations but quantile results show that size has a negative impact at low quantiles but a positive and significant impact at higher quantiles, where deviations are large and positive. Size, however, is not identified as robust as shown by WALS results.

Although country capabilities have positive and significant impact on deviations in “big hits” at the 90% quantile, they have a negative significant effects at lower quantiles. Model averaging also found this negative effect and it is significant as a focus regressor. Since this measure of country capabilities is synonymous with product diversification, it follows that countries that have very large positive deviations in “big hits” benefit from an increase in diversification. Contrarily, where deviations are negative, diversification does not encourage “big hits”; this result is plausible since “big hits” are by definition large export flows, hence, diversifying too much would not facilitate the formation of large exports. Product ubiquity impacts deviations negatively but is significant at higher quantiles and is robust to model averaging estimations with and without focus regressors. In simple words, the more a certain product is popular, the worse will it perform in terms of deviations in “big hits”.

[14]See Magnus et al. (2010) and Magnus et al. (2011) for a description of the theoretical and practical superiority of WALS over BMA.

Table 6 Regression results, dependent variable: deviation in big bits

	OLS	Q(0.20)	Q(0.50)	Q(0.70)	Q(0.90)	WALS_0	WALS_f	WALS_33
Initial Income	-34.44+	-6.954	-6.760	-9.850*	-20.63	-24.75+	-26.03+	-31.31
GDP	11.58+	-1.228	-0.313	0.523	19.35*	8.722	7.553	-0.323
Country capabilities	-0.074	-0.077*	-0.072	-0.049*	0.099*	-0.054	(f) -0.059+	-0.112*
Product ubiquity	-7.036*	-0.705	-1.339	-1.098+	-6.685*	-5.542*	-5.299*	-2.942
Betweenness centr.	-24578*	-13197*	-7531	-7958*	-26766*	-19349*	(f) -197920*	-12920+
Clustering coeff.	45.68	14.78	16.85	20.05	37.56	34.86	38.87	66.63
Private credit	0.521	-0.074	0.030	0.120	0.478+	0.448+	0.384	0.196
Tariffs	-1.427	-0.231	-0.504	-0.374	3.066*	-1.571	-1.050	-2.516
Start-up costs	-0.037+	-0.026	-0.010	-0.012*	-0.031	-0.025	(f) -0.029	-0.097
No. export docs	-6.955+	-1.999	-0.652	-0.554	-2.847	-5.027	-5.524	-7.083
Inflation	-1.749	0.337	0.341	-0.390	-0.457	-1.201	-1.202	-3.144
Share export:GDP	-0.625	0.0547	-0.124	0.102	-0.737*	-0.466*	(f) -0.576*	-0.128
Rule of law	-1.025*	-0.0788	-0.161	-0.118	-0.650	-0.764+	-0.694	-1.211+
Telephones	1.499	0.164	0.133	0.0973	1.017	0.908	0.898	1.350
FDI	0.728	1.554	0.726	0.172	-4.497*	0.898	0.432	-0.140
Oil-producer dummy	42.56	4.287	9.741	1.560	38.81	25.54	33.92	28.42
Sub-Saharan Africa	48.74+	3.802	12.02	14.10*	26.36	40.92	(f) 42.43	113.2*
South Asia	27.70	14.62	22.19	13.05	-19.76	26.66	(f) 12.77	71.19
North America	224.2+	218.7*	159.5*	440.3*	117.1*	190.1*	(f) 247.4*	52.81
M. East & N. Africa	57.15*	31.23	27.01	17.12*	-20.83	51.01+	(f) 49.55+	84.36*
Latin Ame. & Carib.	44.16*	7.776	17.20	13.60*	49.62+	34.08	(f) 40.46	73.33*
East Asia Pacific	110.5*	4.498	18.29	32.35*	125.2*	84.81*	(f) 101.2*	92.39*
Capital Stock								0.0179*
Investment								1.555
Savings								0.322
Population density								-0.00208
No. prod-dest.								5.500
Herfindahl Index								-60.17
Employment agric.								0.843
Employment indus.								3.258+
Primary School								3.923
Secondary School								15.28*
Export volatility								3.888
Observations	141	141	141	141	141	141	141	102

Significance level: + p<0.1 * p<0.05

Centrality in the network (betweenness) has a consistent negative and significant impact on deviations. However, the extent of the impact differs between low and high quantiles. The negative impact is strongly robust to the OLS and model averaging estimation techniques. The betweenness centrality measure takes the number of connections, which reflects diversification, as a core element in its computation, as such, the negative impact is thus consistent with the diversification argument explained above.[15]

Financial development (private credit) has a positive impact on deviations at the highest quantiles where there are positive deviations while a negative impact at the 20% quantile. This implies that more developed manufacturers or "big hits" are more likely to benefit from well-developed financial structure than the less developed ones or "smaller hits". Beck et al. (2008), using a firm level survey covering about 3000 firms in 48 countries, find that small firms are less able to take advantage of external finance than larger firms, hence if we assume that small firms are less likely to be "big hits", Beck's et al. reasoning can explain our result. Moreover, they also find that less-developed institutions negatively impact the use of external finance so

[15]See appendix for a formal definition of this measure.

that small firms resort to informal finance. In our data, we find a positive correlation between rule of law and financial development for lower quantiles to support such an argument. Model averaging supports the thesis of a positive impact.

Tariffs on manufactured goods have negative but insignificant impact on deviations at lower quantiles. However, at the highest quantile, tariff protection relates positively with deviations. The result is not robust and points towards the ineffectiveness of direct and obsolete policy measures as warned by Rodrik (2004). We expect cost of starting a business to negatively affect deviations. We only observe a significant effect at the 70% quantile and the OLS. Being a "policy" measure reflecting ease of doing business, we include it as a focus regressor in WALS_f; while it has the expected sign, it has low probability of inclusion. Openness to trade, as measured by share of exports to GDP, has mixed effects but a significant negative impact on the upper-most quantile where deviations are positive. It is robust to model averaging. Contrary to common and academic belief (Rigobon and Rodrik, 2005; Barro, 2001), rule of law which reflects a sound institutional and legal system negatively influences deviations.

The effect of FDI on deviations is also mixed: while it exerts positive but insignificant impact at low quantiles, it has a strong negative effect at the highest quantile but is not robust to model averaging. In other words, FDI positively affects countries that are lagging behind in “big hits” but is detrimental to countries that are already well-performing. The former result is well-explained by the FDI-led growth in developing countries literature (Blomstrom et al., 1994). Recent empirical studies may explain the latter result, for instance, using robust mixed fixed random estimator Nair-Reichert and Weinhold (2001) find that the effect of FDI on growth suffers from considerable heterogeneity.

The estimated coefficients of the regional dummies show that they are doing better than the reference region EU-Central Asia[16]. North America is the best performing region followed by East Asia Pacific region and Middle East and North Africa region; they are robust to model averaging estimation.

Clustering coefficient, inflation, number of documents to export, basic telecommunication development as measured by telephones per 100 persons appear not to matter. Being an oil-producing country is beneficial but not significant. The model averaging estimation with the restricted sample identifies capital stock, the percentage of employment in industry and years of secondary schooling as positive determinants of deviations.

In sum, the use of quantile regressions alongside model averaging provides a complete picture of the forces at work as regards the performance of countries in terms of deviations in “big hits” from the established benchmark.

5 Conclusion

In this paper, we show that a simple stochastic model of proportionate growth is not only able to match a number of stylised facts documented in the literature but it also reconciles diverging views in the economic development literature as regards the potential merits of industrial policy. On the one hand, Hausmann and Rodrik (2003) argue for industrial policy that would

[16]except for Middle East and North Africa at the 90% quantile.

lead to successful export-led growth. Growth is achieved by subsidising discovery efforts that would be under-supplied in an unregulated market. On the other hand, Easterly et al. (2009) argues against policies aimed at picking winners as they are likely to be unsuccessful given that the distribution of exports is fat-tailed, that is, “big hits” are rare. Our model suggests that the probability of drawing a big hit is larger for countries that engage in more discovery efforts, that is, trying to establish more trade links. It, thus, brings together the above mentioned diverging views regarding industrial policy.

We exploit the model as a benchmark against which we evaluate the export performance of different countries in terms of big hits. Given the number of export relationship of each country and the skewed distribution of export values, we can derive (by simulations) an expected number of big hits for each country. These represent the “normal” performance of countries according to our model. By comparing actual and expected “big hits” we can define whether countries perform better or worse than the benchmark. Interestingly, we find that most European countries (save Germany and Ireland) underperform, while the best performers are no surprise and include countries such as China, Germany, USA and Japan which have a solid manufacturing sector. At the bottom of the list appears many of the countries that have been hit by the recent financial crisis, or are plagued by structural problems such as Italy and Spain.

What are the reasons behind the good and bad performance of countries in terms of big hits? The use of quantile regressions provide some insights into the determinants of these deviations. It further shows that countries are heterogeneous so that the impact of many variables are different for countries that are at the bottom of the list from those that are at the top. We check the robustness of our results with the recently developed model averaging estimation WALS. Initial income, country capabilities, product ubiquity, betweenness centrality, financial development, openness of an economy and rule of law negatively affect deviations in “big hits”. The regional dummies, North America, Middle East and North Africa, and East Asia Pacific positively affect deviations. Further research is needed to identify the policy actions that can revamp competitiveness and help European countries to improve their performance.

References

Armenter, R. and Koren, M. (2010). A balls-and-bins model of trade. Technical Report 7783, Centre for Economic Policy Research, London, UK.

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.

Barro, R. J. (2001). Human capital and growth. *The American Economic Review*, 91(2):12–17. ArticleType: research-article / Issue Title: Papers and Proceedings of the Hundred Thirteenth Annual Meeting of the American Economic Association / Full publication date: May, 2001 / Copyright 2001 American Economic Association.

Barro, R. J. and Lee, J.-W. (2012). A new data set of educational attainment in the world, 1950?2010. *Journal of Development Economics*.

Beck, T., Demirgüç-Kunt, A., and Maksimovic, V. (2008). Financing patterns around the world: Are small firms different? *Journal of Financial Economics*, 89(3):467–487.

Beck, T. and Demirguc-Kunt, A. (2009). Financial institutions and markets across countries and over time - data and analysis. *World Bank Policy Research Working Paper Series*, 4943.

Beck, T., Demirguc-Kunt, A., and Levine, R. (2000). A new database on financial development and structure. *World Bank Economic Review*, 14:597–605.

Beckerman, W. (1956). Distance and the pattern of intra-european trade. *The Review of Economics and Statistics*, 38(1):31–40.

Bee, M. (2006). Estimating the parameters in the loss distribution approach: how can we deal with truncated data? In Davis, E., editor, *The advanced measurement approach to operational risk*, pages 123–144. Riskbooks, London.

Besedeš, T. and Prusa, T. J. (2011). The role of extensive and intensive margins and export growth. *Journal of Development Economics*, 96(2):371–379.

Bhattacharya, K., Mukherjee, G., Saramäki, 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.

Blomstrom, M., Lipsey, R. E., and Zejan, M. (1994). What explains developing country growth? In *Convergence of productivity: cross-national studies and historical evidence*. National Bureau of Economic Research.

Buchinsky, M. (1994). Changes in the U.S. wage structure 1963-1987: application of quantile regression. *Econometrica*, 62(2):405–458. ArticleType: research-article / Full publication date: Mar., 1994 / Copyright © 1994 The Econometric Society.

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. *Journal of applied econometrics*, 13(1):1–30.

Cadot, O., Carrère, C., and Strauss-Kahn, V. (2009). Export diversification: What's behind the hump? Working Paper 200934, CERDI.

Canning, D., Amaral, L., Lee, Y., Meyer, M., and Stanley, H. (1998). Scaling the volatility of GDP growth rates. *Economics Letters*, 60(3):335–341.

Chaney, T. (2011). The network structure of international trade. *CEPR Discussion Paper*, No. DP8240.

Easterly, W., Reshef, A., and Schwenkenberg, J. (2009). The power of exports. *World Bank Policy Research Working Paper Series*, 5081.

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

Fagiolo, G., Reyes, J., and Schiavo, S. (2009). World-trade web: Topological properties, dynamics, and evolution. *Physical Review E*, 79(3):36115.

Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationships of the internet topology. In *ACM SIGCOMM Computer Communication Review*, volume 29, page 251–262.

Gaulier, G. and Zignago, S. (2010). Baci: international trade database at the product-level (the 1994-2007 version). Working Paper 2010-23, CEPII.

Guimera, R., Mossa, S., Turtschi, A., and Amaral, L. N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences*, 102(22):7794–7799.

Hausmann, R. and Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth*, 16(4):309–342.

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

Hausmann, R. and Rodrik, D. (2005). Self-discovery in a development strategy for El Salvador. *Economia*, 6(1):43–101.

Helpman, E., Melitz, M., and Rubinstein, Y. (2008). Estimating trade flows: trading partners and trading volumes. *The Quarterly Journal of Economics*, 123(2):441–487.

Hidalgo, C. A. and Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26):10570–10575.

Hidalgo, C. A., Klinger, B., Barabási, A.-L., and Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837):482–487.

Hummels, D. and Klenow, P. J. (2005). The variety and quality of a nation's exports. *The American Economic Review*, 95(3):704–723.

Imbs, J. and Wacziarg, R. (2003). Stages of diversification. *American Economic Review*, page 63–86.

Jackson, M. O. (2008). *Social and economic networks*. Princeton University Press.

Kaldor, N. (1957). A model of economic growth. *The Economic Journal*, 67(268):591–624. ArticleType: research-article / Full publication date: Dec., 1957 / Copyright © 1957 Royal Economic Society.

King, R. G. and Levine, R. (1993). Finance and growth: Schumpeter might be right. *The Quarterly Journal of Economics*, 108(3):717–737.

Koenker, R. (2006). Quantile regresssion. In *Encyclopedia of Environmetrics*. John Wiley & Sons, Ltd. Cited by 0000.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, page 33–50.

Krautheim, S. (2012). Heterogeneous firms, exporter networks and the effect of distance on international trade. *Journal of International Economics*, 87(1):27–35. Cited by 0011.

Magnus, J. R., Powell, O., and Prüfer, P. (2010). A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics*, 154(2):139–153.

Magnus, J. R., Wan, A. T., and Zhang, X. (2011). Weighted average least squares estimation with nonspherical disturbances and an application to the hong kong housing market. *Computational Statistics & Data Analysis*, 55(3):1331?1341.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Miura, H. (2011). SGL: stata graph library for network analysis. CHI11 Stata Conference 7, Stata Users Group.

Nair-Reichert, U. and Weinhold, D. (2001). Causality tests for cross-country panels: a new look at fdi and economic growth in developing countries. *Oxford Bulletin of Economics and Statistics*, 63(2):153–171.

Onnela, J.-P., Saram aki, J., Hyv onen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., and Barabási, A.-L. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336.

Riccaboni, M. and Schiavo, S. (2010). Structure and growth of weighted networks. *New Journal of Physics*, 12(2):023003.

Rigobon, R. and Rodrik, D. (2005). Rule of law, democracy, openness, and income. *Economics of Transition*, 13(3):533?564.

Rodrik, D. (2004). Industrial policy for the twenty-first century. *CEPR Discussion Paper*, No. 4767.

Rodrik, D. (2012). Why we learn nothing from regressing economic growth on policies. *Seoul Journal of Economics*, 25(2):137–151.

Schumpeter, J. A. (1934). *The theory of economic development: an inquiry into profits, capital, credit, interest, and the business cycle*. Transaction Publishers.

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

Simon, H. A. and Ijiri, Y. (1977). *Skew distributions and the size of business firms*. Amsterdam, North Holland.

Soram aki, K., Bech, M. L., Arnold, J., Glass, R. J., and Beyeler, W. E. (2007). The topology of interbank payment flows. *Physica A: Statistical Mechanics and its Applications*, 379(1):317–333.

Stanley, M. H. R., Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M. A., and Stanley, H. E. (1996). Scaling behaviour in the growth of companies. *Nature*, 379(6568):804–806.

Stiglitz, J. E., Lin, J. Y., and Monga, C. (2013). The rejuvenation of industrial policy. Policy Research Working Paper 6628, The World Bank.

Yamasaki, K., Matia, K., Fu, D., Buldyrev, S. V., Pammolli, F., Riccaboni, M., and Stanley, H. E. (2006). Preferential attachment and growth dynamics in complex systems. *Physical Review E, Statistical, Nonlinear and Soft Matter Physics*, 74:035103.

A Appendix

A.1 List of variables and their description

- **Initial Income** is GDP per capita at PPP (constant 2005 international \$) in logarithm for the year 1995 which is the beginning of the period under consideration (source from World Development Indicators (WDI)).

- **GDP** represents the size of the economy. It is total GDP amount in logarithm for each country at constant US\$.

- **Country capabilities** is calculated as in Hidalgo and Hausmann (2009) ,

  $$k_{c,0} = \sum_p M_{cp}$$

  i.e. the diversification of country c as the sum of M_{cp} over all products p where

  $M_{cp} = 1$ if country c exports product p with a Revealed Comparative Advantage above 1, $M_{cp} = 0$ otherwise. We use our own data and average the figure for each country.

- **Product ubiquity** is also taken from Hidalgo and Hausmann (2009) and represents the pervasiveness of the products of a country,

  $$k_{p,0} = \sum_c M_{cp}$$

- **Betweenness centr.**, following Jackson (2008), is defined as

  $$\sum_{ij:i\neq j,k\notin ij} \frac{P_{ij}(k)}{P_{ij}}$$

  where P_{ij} denotes the number of shortest paths from node i to j and $P_{ij}(k)$ denotes the number of shortest paths from node i to j that node k lies on. It is a measure of the importance of a country based on its connections.

- **Clustering coef.** of a node is the probability that two randomly selected neighbours are connected to each other. The clustering coefficient C_c of node i:

  $$C_c(\mathbf{A}) = \frac{\sum\limits_{j\neq i;k\neq j;k\neq i} A_{ji}A_{ik}A_{jk}}{\sum\limits_{j\neq i;k\neq j;k\neq i} A_{ji}A_{ik}}$$

- **Private credit** is credit from all financial institutions and sourced from the Financial Development and Structure Dataset, World Bank developed by Beck et al. (2000). The April 2013 version has been used. The variable is Private Credit by Deposit Money Banks and Other Financial Institutions to GDP (%) which is defined as claims on the private sector by deposit money banks and other financial institutions divided by GDP. As described in

Beck and Demirguc-Kunt (2009), this asset side variable captures one of the most important function of financial intermediaries, that is, allocation of credit.

- **Tariffs** is average tariff rate on manufactured products (%) (source WDI).
- **Start-up costs** represents costs of business start-up procedures as a percentage of gross national income per capita (source WDI).
- **No. export docs** is simply the number of documents required per shipment to export goods such as documents required for clearance by government ministries, customs authorities, port and container terminal authorities, health and technical control agencies and banks are taken into account. It is an ease of doing business indicator.
- **Inflation** is used to capture changes in the price level. It reflects annual percentage changes in consumer prices (source WDI).
- **Share export:GDP** is a standard indicator of openness to trade (source WDI).
- **Rule of law**, percentile rank, is a World Governance Indicator, World Bank.
- **Telephones** captures technological development (source WDI).
- **FDI** is net inflows of Foreign Direct Investment in current US$ in logarithm (source WDI).
- **Oil-producer dummy** is a dummy variable capturing whether a country is a significant oil-producer, that is, its oil rents is more than 40% of its GDP (source WDI).
- **Sub-Saharan Africa, South Asia, North America, M. East & N. Africa, Latin Ame & Caribbeans, East Asia Pacific** are regional dummy variables representing regions as per the WDI 7 regions classification.
- **Capital Stock, Investment and Savings** are capital stocks in billions constant 2005 $US, investment rate as a percentage of GDP, Savings rate as a percentage of GDP respectively (source CEPII Econmap baseline database 2.1).
- **Population density** is population density (people per sq. km of land area) (source WDI).
- **No. prod-dest.** is population density (people per sq. km of land area) (source WDI).
- **Employment agric. and Employment indus.** represent employment in the agricultural sector and employment in industry (% of total employment) respectively (source WDI).
- **Primary school and Secondary school** are years of primary schooling and secondary schooling for population aged 15 and over respectively (source (Barro and Lee, 2012)).
- **Export volatility** is the standard deviation in export growth from 1995 to 2005 (calculated using data from BACI database).

  All the variables are for the year 2005 except where specified above. To maximise the number of observations, where data was missing for the year 2005, averages of previous and successive years were used.

A.2 List of all countries with simulated and empirical big hits, and the actual deviations between simulated and empirical big hits

Country	Empirical	Simulated	Dev.	Country	Empirical	Simulated	Dev.
Afghanistan	0*	2	-2	Kuwait	0	10	-10
Albania	0	7	-7	Kyrgyzstan	0	5	-5
Algeria	1	5	-4	Lao	0	4	-4
Andorra	0	4	-4	Lebanon	0	47	-47
Angola	0*	2	-2	Latvia	1	38	-37
Antigua & Barbuda	1*	2	-1	Liberia	3	1	2
Azerbaijan	1	6	-5	Libya	0*	2	-2
Argentina	7	75	-68	Lithuania	1	60	-59
Australia	45	168	-123	China, Macao SAR	2	17	-15
Austria	59	273	-214	Madagascar	0	9	-9
Bahamas	3*	2	1	Malawi	0*	2	-2
Bahrain	1	13	-12	Malaysia	144	178	-34
Bangladesh	16	28	-12	Maldives	0*	2	-2
Armenia	0	5	-5	Mali	0*	3	-3
Barbados	0	5	-5	Malta	4	14	-10
Belgium-Lux.	186	383	-197	South America, nes	0*	0	0
Bermuda	1*	1	0	Mauritania	0*	1	-1
Bhutan	0	0	0	Mauritius	1	17	-16
Bolivia	0	8	-8	Mexico	199*	133	66
Bosnia Herz.	0	25	-25	Mongolia	0*	3	-3
Bouvet Island	0*	0	0	Moldova	1	10	-9
Brazil	76	186	-110	Morocco	5	38	-33
Belize	0	2	-2	Mozambique	4*	2	2
Solomon Isds	0	0	0	Oman	1	13	-12
Brunei Darussalam	0	2	-2	Nauru	0	0	0
Bulgaria	2	84	-82	Nepal	0	8	-8
Myanmar	1	7	-6	Netherlands	229	405	-176
Burundi	0*	1	-1	Vanuatu	0	0	0
Belarus	8	34	-26	New Zealand	6	76	-70
Cambodia	4*	8	-4	Nicaragua	2*	6	-4
Cameroon	0	9	-9	Niger	1*	2	-1
Canada	265*	230	35	Nigeria	0	8	-8
Cape Verde	0*	2	-2	Norway	18	120	-102
Cayman Isds	1*	1	0	FS Micronesia	0*	0	0
C. African Rep.	0*	1	-1	Marshall Isds	3	0	3
Sri Lanka	2	37	-35	Palau	0*	0	0
Chad	0	0	0	Pakistan	7	74	-67
Chile	25	41	-16	Panama	3	35	-32
China	939*	678	261	Papua New Guinea	1*	2	-1
Colombia	4	50	-46	Paraguay	0	5	-5
Comoros	0*	1	-1	Peru	6	32	-26
Congo	0*	2	-2	Philippines	65*	71	-6
Dem. Rep.Congo	0*	1	-1	Pitcairn	0*	0	0
Costa Rica	11	26	-15	Poland	55	190	-135
Croatia	3	56	-53	Portugal	29	129	-100
Cuba	3*	5	-2	Guinea-Bissau	0	0	0
Cyprus	0	20	-20	Timor-Leste	0*	1	-1
Czech Rep.	52	190	-138	Qatar	1	13	-12
Benin	0*	2	-2	Romania	12	93	-81
Denmark	31	231	-200	Russian Fed.	84	140	-56
Dominica	0*	2	-2	Rwanda	0*	2	-2
Dominican Rep.	8*	10	-2	Sao Tome & Princ.	0	0	0
Ecuador	0	17	-17	Saudi Arabia	24	70	-46
El Salvador	0	17	-17	Senegal	1	12	-11
Equatorial Guinea	1*	1	0	Serbia	0	32	-32
Ethiopia	0*	3	-3	Seychelles	0*	1	-1

Country	Empirical	Simulated	Dev.	Country	Empirical	Simulated	Dev.
Eritrea	0	0	0	Sierra Leone	0*	2	-2
Estonia	2	43	-41	India	31	364	-333
Fiji	0	8	-8	Singapore	129	206	-77
Finland	46	157	-111	Slovakia	17	89	-72
France	328	529	-201	Viet Nam	17	77	-60
Fr. S. Antarc. Terr.	0*	0	0	Slovenia	5	103	-98
Djibouti	0*	1	-1	Somalia	0*	1	-1
Gabon	1*	3	-2	Zimbabwe	0	7	-7
Georgia	0	8	-8	Spain	113	388	-275
Gambia	0*	1	-1	Sudan	0*	2	-2
Palestinian Terr.	0*	1	-1	Suriname	2	1	1
Germany	865*	679	186	Sweden	99	262	-163
Ghana	1	7	-6	Switzerland	106	309	-203
Gibraltar	0*	1	-1	Syria	0	28	-28
Kiribati	0*	0	0	Tajikistan	1*	1	0
Greece	3	102	-99	Thailand	104	222	-118
Greenland	0*	1	-1	Togo	0	5	-5
Grenada	0*	1	-1	Tokelau	0*	1	-1
Guam	0*	1	-1	Tonga	0*	0	0
Guatemala	2	26	-24	Trinidad & Tobago	2	16	-14
Guinea	1*	2	-1	Utd Arab Emirates	16	193	-177
Guyana	0	3	-3	Tunisia	6	35	-29
Haiti	1*	1	0	Turkey	49	277	-228
Honduras	5*	10	-5	Turkmenistan	0*	2	-2
Hong Kong (China)	59	247	-188	Tuvalu	0*	0	0
Hungary	57	134	-77	Uganda	0	6	-6
Iceland	3	13	-10	Ukraine	25	77	-52
Indonesia	54	161	-107	TFYR of Macedonia	0	18	-18
Iran	3	40	-37	Egypt	2	53	-51
Iraq	0*	1	-1	United Kingdom	257	530	-273
Ireland	148*	102	46	Tanzania	0	9	-9
Israel	15	106	-91	USA	887*	627	260
Italy	225	585	-360	Burkina Faso	1*	2	-1
Côte d'Ivoire	2	11	-9	Uruguay	0	16	-16
Jamaica	2	6	-4	Uzbekistan	4*	4	0
Japan	705*	354	351	Venezuela	7	23	-16
Kazakhstan	9*	13	-4	Samoa	0*	1	-1
Jordan	3	25	-22	Yemen	0	4	-4
Kenya	1	27	-26	Serbia & Monten.	1	34	-33
Dem. Korea	0	16	-16	Zambia	6*	4	2
Rep. of Korea	296*	280	16				

Values marked with an * indicates that the empirical big hits fall within the 95% CI of the simulated big hits