The Birth and Growth of New Export Clusters:

Which Mechanisms Drive Diversification?

Dany Bahar, Ernesto Stein, Rodrigo Wagner,
and Samuel Rosenow

CID Research Fellow and Graduate Student
Working Paper No. 86
September 2017

Working Papers
Center for International Development
at Harvard University

The birth and growth of new export clusters: which mechanisms drive diversification?*

Dany Bahar† — *Brookings Institution*, *Harvard CID*

Ernesto Stein — *IADB*

Rodrigo Wagner — *Universidad de Chile*

Samuel Rosenow — *IADB*

Latest Update: September 11, 2017

Keywords: comparative advantage, exports, relatedness, spillovers, R&D, patents, labor, upstream, downstream

JEL Classification Numbers: O14, O33, F14

*Authors acknowledge helpful comments from seminars at the Brookings Institution, the Inter-American Development Bank, George Washington University, LACEA Annual Meetings 2015, University of Chile, Harvard Business School and Harvard's Center for International Development. All errors are our own. The views expressed by the authors in this paper do not reflect the opinions of the Inter-American Development Bank, its Board of Directors or the countries it represents.

†Corresponding Author: db21@post.harvard.edu

Abstract

Export diversification is associated with economic growth and development. Our paper explores competing mechanisms that mediate the emergence and growth of export products based on their economic relatedness to pre-existing exports. Our innovation is to simultaneously consider supply factors like labor, sourcing and technology; as well as demand factors like industry specific customer-linkages in a global setting. We find that, while technology and workforce similarity explain emergence and growth, pre-existing downstream industries remain a robust predictor of diversification, especially for jumpstarting new exports in developing countries. Our global stylized fact generalizes Javorcik's (2004) view that spillovers are more likely in backward linkages.

1 Introduction

Export diversification has been associated with economic development and macroeconomic stability.[1] Nonetheless, the literature reveals little about the relative relevance of the mechanisms mediating the process of export diversification. A large literature focuses on studying changes in export baskets as a result of changes in the relative abundance of factor endowments (e.g. Heckscher and Ohlin, 1991; Romalis, 2004; Bernhofen et al., 2016) or changes in (mostly exogenous) productivity parameters (e.g. Ricardo, 1821; Eaton and Kortum, 2002; Melitz, 2003; Costinot et al., 2012). This literature typically takes the evolution of the comparative advantage of a single country-industry as independent of other industries within the same country. Our paper explores how the new products a country exports relate to the composition of its current export basket.

A body of literature in international economics and economic geography has studied how the birth of a new product, or even future growth of already existing products, relates to incumbent economic activities in the same geographic unit (e.g. Ellison and Glaeser, 1997; Ellison et al., 2010; Hidalgo et al., 2007; Hausmann and Hidalgo, 2011). However, these studies define "relatedness" solely on geographic terms: two products are related if they coexist within the same geographic unit (i.e. a country, a county, a city, etc.). Thus, in essence, these measures are agnostic in terms of why two products are related, beyond co-location. Our paper contributes to this literature by studying particular channels, beyond co-location, through which inter-industry productivity spillovers shape the comparative advantage of nations.

We study cross-industry spillovers based on both supply and demand driven channels. From the supply side, we look at technology, labor and supply (of

[1] For references regarding the relationship between diversification and economic development see, for example, Imbs and Wacziarg (2003); Hausmann et al. (2007); Cadot et al. (2011). For effects on macroeconomic stability, see, for example, Krishna and Levchenko (2009); Koren and Tenreyro (2007); Caselli et al. (2015); Hausmann et al. (2006). Also, Rodrik (2016) argues that reductions in diversification, like those induced by premature deindustrialization, can jeopardize the process of economic development, for example because it prevents unconditional convergence in manufacturing (Rodrik, 2012).

intermediate goods) linkages. From the demand side, we look at customer linkages. There are several explanations on how the emergence of new sectors can be fueled by the existence of related industries. For example, technology generated by or for a specific sector could explain the emergence or growth of another sector if the latter utilizes the knowledge created by the former (Hall and Trajtenberg, 2006; Griffith et al., 2006). Regarding labor, the existence of competitive industries using a trained (and competitive) workforce similar to that required by a non-existent product might play a role in explaining the emergence of the latter (Neffke et al., 2016; Hausmann and Neffke, 2016; Bahar and Rapoport, 2017). A similar logic can be applied to the existence of competitive industries producers of goods that are intermediate to products of a yet inexistent industry, which could build on those to add value and become an exporter. Finally, on demand channels, the existence of a critical mass of firms in certain industries could "pull" the development of a new upstream export product (Pietrobelli and Saliola, 2008; Kee and Tang, 2016).[2] Our paper explores all these alternative hypotheses.

We start by confirming in our global dataset of exports across country that measures of relatedness based on co-location predict both the emergence of new products and future growth of existing ones. The first relatedness index, from Hausmann and Klinger (2006), measures the probability that two products are competitively co-exported by the same country. The second measure uses the co-location of export industries in the same country of origin, an adaptation of the index used by Ellison and Glaeser (1997). In particular, we find that –for a given country– the probability of exporting a new product within the following decade is, on average, 1.6 to 2.7 percentage points higher (a 80 to 140 percent increase) given an increase of one standard deviation in the (agnostic) relatedness of that product to the rest of the export basket. Consistently with this, we also show that existing export sectors tend to experience faster growth of about 6.5 to 8 percentage points annually when other industries present in

[2]Unlike specialized business-to-business demand, the channel of a critical mass in final demand (e.g. Krugman, 1991), cannot be identified with our empirical approach, because we extract all country-year variation with our fixed effects.

that country's export basket at the beginning of the period are related to the product under consideration by one standard deviation above average.

Our challenge in this paper, however, is to disentangle the mechanisms behind these agnostic measures. To this end, we build on the work by Greenstone et al. (2010), and add five other relatedness measures to our analysis (similar to Ellison et al. (2010)): two proxying technology relatedness (cross industry patent-citations and the share of research and development done in one industry that another industry uses), a third one measuring workforce similarity (based on labor flows) and two others measuring downstream and upstream linkages between industries (quantified through input/output tables).

Our results indicate that, as expected, supply factors matter for diversification, most notably technology. A one standard deviation of higher technology linkages makes the emergence of a new product up to three times more likely, and is associated with a subsequent annual export growth of 14.8 extra percentage points over the next decade. This result is consistent with the idea that the diffusion of knowledge across industries is key to sectorial productivity shifts that result in changes in comparative advantage (Bahar et al., 2014). Workforce similarity across industries is also associated with industry growth, although not with its emergence; suggesting that, once an export sector is already established (i.e. after fixed costs have been absorbed), it is skills embodied in labor –on top of technology– what can induce productivity and with it changes in comparative advantage. When it comes to supplier linkages, we do not find any evidence of their role in driving neither the emergence nor growth of export sectors, except in developed countries. On the demand side, however, we do find that the existence of competitive industries can explain the emergence of new sectors that are related via customer linkages. In particular, one standard deviation of higher downstream relatedness increases the odds of emergence of a new export by around 2.5 times; almost as strong as technology. This effect is mostly driven by developing countries where markets are less sophisticated, and the existence of downstream sectors can mitigate risks for investment and entrepreneurship in new sectors. All these findings survive the inclusion of a battery of fixed effects, including time invariant determinants of comparative advantage within a country-product pair. In addition, channel specific relatedness measures are explanatory of emergence and growth of sectors beyond agnostic measures and can explain part of them, though not entirely.

Our findings suggest a global stylized fact that generalizes the view by Javorcik (2004) that spillovers are more likely in backward linkages.[3] There are many other case studies complementing this view that customers as drive change in comparative advantage. Kee and Tang (2016), for example, show that China gained comparative advantage in those intermediate inputs that are used by pre-existing Chinese exporters. This implies that competitive customers for one's products can elicit the emergence of new industries. Pietrobelli and Saliola (2008) also find that selling to a competitive customer, in their case a multinational, enhances supplier productivity. Our work, while more aggregate but global in scope, naturally complements these previous micro-level findings on the relevance of specialized demand as a catalyst of a new industry.

Our paper contributes to a number of different strands in the economic literature. First, we contribute to the literature of co-agglomeration and economic geography, since we provide evidence of particular channels explaining co-location of export sectors (e.g., Ellison and Glaeser, 1997; Ellison et al., 2010; Greenstone et al., 2010). Second, we contribute to the literature in international economics studying the determinants of dynamic comparative advantage, and in particular, of the emergence of new export sectors. Third, by looking at the determinants of the emergence of new export sectors, we contribute to a line of research in economic development that studies the role of structural transformation and diversification in the process of growth and development (e.g. Imbs and Wacziarg (2003); Cadot et al. (2011)). Yet, our interest in exploring connections among sectors in the economy does not come from the transmission of aggregate fluctuations (e.g. Carvalho et al., 2012),

[3]Customer linkages are twice as powerful as the agnostic measures in both the extensive and intensive margin. The emergence of new sectors are better explained by local supply chains that would, to some extent, reduce risks related to self-discovery costs. For evidence on self discovery costs in export, see Hausmann and Rodrik (2003).

but rather from structural changes in comparative advantage. To the best of our knowledge, our study is the first at combining the literature of exports diversification (e.g. Hidalgo et al., 2007; Hausmann and Hidalgo, 2011) with the literature on cross-industry spillovers (Ellison and Glaeser, 1997; Ellison et al., 2010; Javorcik, 2004).

The rest of the paper is structured as follows. Section 2 explains data sources and variable definitions, notably the measures of relatedness between products and the implied measures of density around each product. Section 3 presents the basic “sanity check” in which various agnostic measures of relatedness predict the birth and growth of export products. Section 4 is the core of our paper and studies the various mechanisms through which existing related industries predict the emergence of new export sectors and the growth of existing ones. Finally, Section 5 concludes and discusses the implications of our results.[4]

2 Data

2.1 Data Sources

The main data for our exercise is bilateral trade data compiled from UN Comtrade by Feenstra et al. (2005) with extensions and corrections suggested by Hausmann et al. (2014). To compute the different variables we mainly use exports by product from each country to the rest of the world from 1984 to 2014. Products are defined using the 4-digit Standard Industry Trade Classification (SITC) revision 2.[5] This product classification provides a disaggregation level that enables a meaningful discussion about export diversification patterns and includes 786 products. Some examples of products in this level of disaggregation are, for example, "Knitted/Crocheted Fabrics Elastic or Rubberized” (SITC 6553), or "Electrical Measuring, Checking, Analyzing Instru-

[4]The appendix includes additional results with alternative measures and the role of quality of export products on diversification.

[5]The words product, good, sector and industry interchangeably refer to the same concept throughout the paper.

ments" (SITC 8748). Following Hausmann et al. (2014), we exclude countries below 1 million citizens and total trade below USD $1 billion in 2010. We also exclude former Soviet Union countries from the analysis since their data does not exist prior to 1990 and remain sparse until 1995, and countries with no reported exports in any products for a particular year. This leaves us with 114 countries to construct the total value of exports per product and country to the rest of the world for each year. The sample represents over 90% of world trade. Since we do not want our results to come from overlapping in time, our analysis uses data for two periods: 1990 to 2000 and 2000 to 2010. For the exercise, we also use relatedness measures computed by Ellison et al. (2010), which we describe in detail below.

2.2 Variables and Measures

2.2.1 Revealed Comparative Advantage

We compute Revealed Comparative Advantage, or RCA, as our indicator of relative export intensity (Balassa, 1965). RCA is the share of a given product on the country's total exports, divided by the share of the same product in world's exports:

$$RCA_{c,p,t} \equiv \frac{x_{c,p,t} / \sum\limits_{p} x_{c,p,t}}{\sum\limits_{c} x_{c,p,t} / \sum\limits_{c} \sum\limits_{p} x_{c,p,t}}$$

with $x_{c,p,t}$ is total export value of product p from country c to the world in year t. Thus, for instance, in the year 2000, soybeans represented 4% of Brazil's exports, but accounted only for 0.2% of total world trade. Hence, Brazil's RCA in soybeans for that year was $RCA_{Brazil,Soybeans} = 4/0.2 = 20$, indicating that soybeans are 20 times more prevalent in Brazil's export basket than in that of the world.

Using this definition, we create the dependent variable for most of our specifications, $Y_{c,p,t \to T}$, to measure the change in comparative advantage of country c and product p between periods t and T. $Y_{c,p,t \to T}$ alternates according to whether the specification is studying the intensive or the extensive margin of trade. When studying the *extensive* margin, $Y_{c,p,t\to T}$ is 1 if the good was exported with a RCA below 0.1 in year t, and with a RCA above 1 in year T, where $T > t$:

$$Y_{c,p,t\to T} = 1[RCA_{c,p,T} \geq 1 | RCA_{c,p,t} < 0.1\ ] \tag{1}$$

When studying the extensive margin, products in which the country already had comparative advantage above 0.1 at initial time t are excluded from the calculations, since we are focusing only on potentially new goods to be exported. That is, $Y_{c,p,t\to T} = 1$ if country c achieved an RCA of 1 or more in product p in the period of time between t and T (conditional on having an $RCA_{c,p,t} < 0.1$). To avoid noise on the dependent variable, we restrict $Y_{c,p,t\to T} = 1$ to two additional conditions: first, the country-product under consideration must keep an average RCA above 1 after the end of the period, year T, for two more years; and second, the country-product under consideration must have exhibited an average RCA below 0.1 for two years before the beginning of the period, year t. This way we avoid our results being driven by noise in the data.

When studying the *intensive margin*, $Y_{c,p,t\to T}$ measures the compound annual growth rate (CAGR) in the export value of product p by country c between years t and T, conditional on having $x_{c,p,t} > 0$. That is:

$$Y_{c,p,t\to T} = \left(\frac{x_{c,p,T}}{x_{c,p,t}}\right)^{1/T-t} - 1 \text{ if } x_{c,p,t} > 0 \tag{2}$$

We presents results throughout the paper for which $T - t = 10$.

2.2.2 Relatedness

To test how the emergence of a new industry can be predicted by the existence of related products, one has to use a measure of product relatedness. To do so we follow the work of both Hausmann and Klinger (2006) and Ellison and Glaeser (1997). Both studies suggest a formulation to compute a geography-based measure of relatedness between each pair of industries. Hausmann and Klinger (2006), HK here onwards, use what they denominate "proximity", and it measures the probability of co-export of two given products by the same country. The proximity measure posits that two products are more related to each other the higher the probability of being co-exported is. In particular, the relatedness (proximity) index between products i and j for a particular year is defined as:

$$\varphi_{i,j}^{HK} = \min\left\{ \Pr(RCA_i \geq 1 | RCA_j \geq 1) \quad ; \quad \Pr(RCA_j \geq 1 | RCA_i \geq 1) \right\}$$

where the probabilities are assessed from global world exports data. Following Klinger and Hausmann (2007) and Hidalgo et al. (2007), we use $RCA \geq 1$ as the threshold to define whether a product is being exported in a particular year and country.

As an additional measure of relatedness, we adapt the co-agglomeration index implemented by Ellison and Glaeser (1997), EG here onwards, to export data. The EG index measures the intensity with which two given products are co-located in the same area, and in our case, co-exported by the same country. To compute the EG relatedness index between products i and j for a particular year, we use the formulation of the EG co-agglomeration index suggested by qEllison and Glaeser (1997):

$$\varphi_{i,j}^{EG} = \frac{\sum_{c=1}^{C}(s_{c,i} - x_c)(s_{c,j} - x_c)}{1 - \sum_{c=1}^{C} x_c^2}$$

where $s_{c,i}$ and $s_{c,j}$ are, respectively, the country share of product i and j in world product exports and x_c represents the share of country c's exports in total world markets. The EG co-agglomeration index posits that two products are more related to each other the more similar their proportion in product markets is relative to that of their respective country in global exports.

Both relatedness measures are averaged over the previous three years (i.e. the value of $\varphi_{i,j}^{HK}$ in year 2010 is the average between the values for years 2008, 2009 and 2010), and normalized such that it will distribute between 0 and 1 by using the corresponding percentiles of the values in the distribution (i.e. when $\varphi_{i,j}^{HK} = 0.9$ it implies that the relatedness value between products i and j is in the 90th percentile). Both the HK and EG indices are, in essence, measuring the intensity with which two products are exported in the same location above the average. It is important to note that the EG index uses continuous export data values, as opposed to the HK index which relies on a threshold of RCA above 1 to compute the probabilities.[6]

Both measures of relatedness are geographic-based. Yet, they are agnostic in the sense that they measure the intensity with which two industries coexist but provide no understanding on *how or why* they are related. As this paper aims to explore what channels explain dynamic comparative advantage and the emergence of related sectors, we employ five other measures of relatedness between industries taken from Greenstone et al. (2010) (similar measures are used by Ellison et al., 2010), which they refer to as measures of economic distance.[7]

First, we use two different measures to measure technological relatedness: the fraction of patents manufactured in a industry i that cite patents manufactured in each industry j; and the amount of R&D expenditure in a 3-digit industry that is used in other 3-digit industries.[8] Both technological measures

[6] See tables A6 and A7 in Appendix D for the fifteen most related product pairs based on the two relatedness measures HK and EG, respectively.

[7] These measures are based on US-manufacturing data from 1973-1998. We use these to extrapolate the technological relationship to other countries and years. We follow the methodology suggested by Cuñat and Melitz (2012) in their footnote 24 to create a concordance table from SIC classification to SITC using data from 2000-2006. We use that concordance table to define these relatedness measures in terms of SITC 4 digit industries. We thank Muhammed Yildirim for guidance in this task.

[8] The R&D measures were originally constructed using the technology matrix by Scherer (1984). In particular, Scherer matched R&D expenditure data organized by industry of origin (based on survey data from 443 large US companies in 1974) with a representative 1974 patent sample that contained information on industry of origin and industry of anticipated use. As a result of this matching, he obtained R&D dollars associated with each patent. To distribute the R&D dollars of patents to destination industries, Scherer then assigned industries in proportion to the industries' purchases from the origin industry, based on a 1972 input-output table for the US economy. Ellison et al. (2010) explain that this matrix "*captures how R&D activity in one industry flows out to benefit another industry. This*

use the average between both i to j and j to i directional values. Second, we use labor relatedness based on the fraction of separating workers from each industry i that move to firms in each industry j.[9] Finally, we employ supplier and customer linkages as two additional measures of relatedness; that is, the fraction of each industry's manufactured inputs purchased from other industry (e.g., supplier linkages) and the fraction of outputs that are sold to each other industry (e.g., customer linkages).[10]

We normalize all these channel-specific measures to represent percentiles in the distribution of relatedness based on the original values (e.g. a relatedness of 0.3 implies that those two products are in the 30th percentile of relatedness), such that they always distribute between 0 and 1. This normalization ensures that the channel-specific measures have the same range as the agnostic ones, enabling a proper comparison and more importantly, as discussed in greater detail below, to incorporate them into a relatedness measure at the country, product and year level. Table 1 presents a correlation matrix between all of the relatedness measures described above. As can be seen, correlations are all positive and statistically different from zero. The correlation between the agnostic measures, φ^{EG} and φ^{HK} is 0.54. The correlations of the agnostic relatedness measures with the channel-defined relatedness measures range between 0.06 to 0.19.

[Table 1 about here.]

2.2.3 Density based on relatedness

In order to study the emergence of new sectors as they relate to other existing related sectors, we still need a measure that can be expressed at the country,

technology transfer occurs either through supplier-customer relationships between these two industries or through the likelihood that patented inventions obtained in one industry will find applications in the other industry".

[9]It originally comes from the Current Population Survey (CPS) outgoing rotation file published by the Bureau of Labor Statistics (BLS).

[10]See tables A8, A9, A10, A11 and A12 in Appendix D for the top fifteen product pairs in terms of relatedness for each one of these measures: R&D, labor, patent, customer and supplier linkages,

product and year level. For this purpose we follow the work of Hausmann and Klinger (2006) and build a measure of "density" around each product, which captures the intensity with which the product under consideration is related to the current export basket of the same country, using as an input the product-product relatedness measures discussed above. This measure is defined for each country c, industry i in time t as:

$$\Phi_{c,p,t} = \frac{\sum_{j \neq p} \varphi_{p,j} \times R_{c,j,t}}{\sum_{j \neq p} \varphi_{p,j}} \tag{3}$$

where $R_{c,j,t} = 1$ is a dummy if $RCA_{c,j,t} \geq 1$ and 0 otherwise. A series of studies (e.g. Hausmann and Klinger, 2006; Hidalgo et al., 2007) show that $\Phi_{c,p,t}$ has explanatory power for the future export value of industry i in country c. That is, countries are more likely to start exporting new goods related to other goods in their current export basket, suggesting the dynamics of a country's export basket are path dependent.

Note that $\Phi_{c,p,t}$ can be computed using different measures of relatedness $\varphi_{i,j}$. In fact, throughout the paper we compute $\Phi_{c,p,t}$ using as inputs all the above described relatedness indices. This allows us to define $\Phi_{c,p,t}$ for the same c, p and t combination using different measures of $\varphi_{i,j}$ (agnostic or channel specific) as our main right hand side variable in most specifications.

2.3 Descriptive statistics of our sample

Table 2 presents the summary statistics of the dependent variables, agnostic and channel-specific density measures and control variables, based on information for ten years. Since we analyze both the emergence of new sectors (the extensive margin) and the growth of already existing sectors (the intensive margin), the panel is split up in two sections. Panel A presents the statistics for the extensive margin sample (i.e., for all observations in a country, product and year combination for which $RCA_{c,p,t} < 0.1$), while Panel B does so for the intensive margin sample (i.e., for all observations in a country, product and year combination for which $Exports_{c,p,t} > 0$).

Panel A shows that a country, on average, gains export competitiveness for a given product with a RCA above 1 (starting with a RCA equal or below 0.1 at the beginning of the period) in a 10 year period in 1.9% of all cases. That is, the probability that a new export product will "emerge" within those ten years is 1.9 percent. Similarly, Panel B indicates that the average country experiences annual export growth of about 4.6% for a given existing product in a 10 year period. The tables also include summary statistics for the two agnostic density measures ($\Phi_{c,p,t}$, using $\varphi_{i,j}$ for both HK and EG relatedness measures) as well as the five channel-specific ones. The density of a product, which distributes between 0 and 1, proxies for the existence of other products that share similar technologies or inputs. For example, values of $\Phi^{HK}_{c,p,t}$ closer to 1 indicate that a given product is highly related to the composition of its country's export basket. Conversely, values closer to 0 mean that there is little relatedness between the product under consideration and the rest of the country's export basket. The same logic applies to the channel-specific density measures, where relatedness is defined through characteristics common to industry pairs. For example, values of $\Phi^{Labor}_{c,p,t}$ close to 1 reveal that a given product employs labor that gets used intensively in other products of the country's export basket. Note that data for channel-specific density measures at both the extensive and intensive margin is limited to fewer products since the original relatedness measures in the SIC classification exhibited less industries.

[Table 2 about here.]

3 Birth and growth of exports using agnostic measures of colocation

Our empirical analysis in this section aims to be a starting point, showing that the emergence of new export products depends on incumbent exports as measured by co-location patterns. After this first step, in the next section we unpack this effect into the various channels, which are the center of our paper.

We analyze both the emergence of new sectors at the extensive margin and the growth of already existing sectors at the intensive margin. For the extensive margin, we exploit whether a product p that did not exist in the basket at the beginning of the period is exported by the end of such period. For the intensive margin, we study the growth rate of industries that already existed at the beginning of the period. We start by estimating the following specification:

$$Y_{c,p,t\rightarrow T} = \beta_d \Phi_{c,p,t} + Controls_{p,c,t} + \eta_{c,t} + \delta_{p,t} + \alpha_{c,p} + \varepsilon_{c,p,t} \quad (4)$$

The left hand side $Y_{c,p,t\rightarrow T}$, as explained above in section 2.2, alternates between the intensive and extensive margin. The extensive margin uses a dummy as the dependent variable, whereas the intensive margin uses a continuous variable.

$\Phi_{c,p,t}$ measures the intensity with which product p is related to the current export basket of the same country. A positive and significant β_d in the extensive margin regression would imply that the likelihood of product p appearing in the export basket of country c at time T is positively correlated with how much product p is related to other exported products in the export basket in the baseline period t (our sample is based on $T - t = 10$ years).

Note that our specification controls for a battery of fixed effects, strongly reducing concerns of misidentification. First, $\eta_{c,t}$ represents country-by-year fixed effects, which capture all time-variant country level variables, such as income, institutions, population, etc. Second, $\delta_{p,t}$ represent product-by-year fixed effects, which capture all time-variant product level variables, such as global demand for product p, common technological changes in the production of such product, among others. We also include $\alpha_{c,p}$ which represents country-by-product fixed effects, controlling for all possible country-product interactions that might explain intrinsic comparative advantage driven by initial observable and unobservable country-product productivity parameters or time invariant factor endowments and factor intensity interactions.

In addition, we include a vector of controls, which varies depending on whether we estimate the extensive or the intensive margin. For the extensive margin, we control for the baseline RCA level (e.g., $RCA_{c,p,t}$), and when estimating the intensive margin equations it also includes the baseline level of exports for that same product; as well as the compound average growth rate (CAGR) of the export value in the previous period, in order to control for the previous growth trend.[11] Finally, we include a a dummy variable indicating whether $x_{p,c,t} = 0$ in the baseline year of the previous period for which the growth rate is calculated (see footnote 11).

Table 3 shows the estimates of (4), using a 10 year period to define the change $Y_{c,p,t \to T}$. The upper panel shows results for the estimation of the extensive margin while the bottom does so for the intensive margin.

[Table 3 about here.]

The main finding from this table 3 is that the emergence of new sectors (upper panel) and the future growth of already existing sectors (lower panel) tend to be positively correlated with the existence of highly related products in the same country's export basket ten years earlier. These results are not new to the literature, as it has been extensively documented (e.g., Klinger and Hausmann, 2007; Hidalgo et al., 2007; Hausmann et al., 2014). In particular, the results in Columns 1 and 2, which use HK proximities, imply that a product is 2.7 percentage points more likely to emerge – a roughly 140 percent increase in the unconditional probability of emerging (of 1.9 percent)– if the baseline density is larger by one standard deviation, on average. Using EG proximities the corresponding numbers represent an increase of 1.6 to 2.7 percentage points, or an increase of 80 to 140 percent compared to the baseline scenario. The estimation using both different measures are strikingly similar.

The lower panel reveals that an increase of one standard deviation of a product's density is associated with an export annual growth larger by 6.5

[11] CAGR during 1985-1990 for the 1990-2000 period, and 1990-2000 for the 2000-2010 period. In order to correct for undefined growth rates caused by zeros in the denominator, we compute the CAGR following the above equation using $exports_{c,p,t} + 1$ for all observations. Note that when studying the intensive margin the CAGR of export value in the dependent variable will always be defined, given that we limit the sample only to products which are being exported at the beginning of the period (that is, $exports_{c,p,t} > 0$). However, the CAGR in the previous period included as a control may have an undefined growth rate; therefore, to control for our own correction, we also add as an additional control a binary variable indicating whether $exports_{c,p,t-1} = 0$ (at the beginning of the previous period, i.e. 1985 or 1990), which correspond to the observations most likely to be distorted.

to 8 percentage points, using the HK relatedness measure. Similarly, using the EG relatedness measure, an equivalent increase in a product's density is associated with an increase in export growth of 2.5 to 3.7 percentage points annually.[12] Note that the results in both panels are robust to using the very conservative specification that includes country-product fixed effects ($\alpha_{c,p}$).

After having re-established the empirical fact that the existence of related products in a country's export basket predicts new exports birth and expansion, we next aim to understand whether specific channels of relatedness, beyond the agnostic measures used in this section, can also explain dynamics of a country's export basket in systematic ways. Do supply side factors like labor, suppliers or technology play a role? Or is the emergence of new exports more related to pre-existing demand? The next section explore this.

4 Unpacking the channels behind birth and growth of exports

The goal of this section is to understand to what extend all, some or none of the specific channels with which we measure relatedness between products can explain the relationship between emergence of new sectors (and growth of existing sectors) and the baseline composition of a country's export basket.

4.1 The definition of channel-specific densities

We start by computing the same "density" measure to create five new variables of export intensity of related industries according to each dimension: patents, R&D, labor, as well as customer and supplier links (see Section 2.2.2 for a description of each one of these dimensions). For example, we define the density of product p in country c in year t based on labor relatedness between products as follows:

[12] See Online Appendix Section B for results that use the log-growth rate as the dependent variable. The results are maintained.

$$\Phi^{Labor}_{c,p,t} = \frac{\sum_{j \neq p} \varphi^{Labor}_{p,j} \times R_{c,j,t}}{\sum_{j \neq p} \varphi^{Labor}_{p,j}} \tag{5}$$

where $\varphi^{Labor}_{p\,j}$is the relatedness measure using relatedness based on labor flows.

Table 4 estimates equation (4) using each one of the different constructions of $\Phi_{c,p,t}$ for each one of the five defined channels of relatedness plus an additional column which adds them all simultaneously. For comparison purposes we report standardized coefficients by normalizing the regressors to have mean zero and unit standard deviation. As a reference point, we include in the bottom rows of the table the standardized coefficients for the same regression using instead both agnostic density measures (with the corresponding sample size).

[Table 4 about here.]

The results from the extensive margin, seen in panel A, show that density measured across patents, customer linkages and, –to a lesser extent– labor can explain export emergence. In particular, the coefficients for patents and customer linkages are almost twice as large as the coefficients for the agnostic measures (bottom rows), whose standardized coefficients are around 0.29, suggesting that these two have stronger explanatory power than the agnostic ones. In Column 6 and 7 we put on a "horse race" between all these measures, distinguishing between R&D and patent channels because both capture technological linkages. When adding all measures simultaneously we find that patent linkages and customer linkages (i.e. the existence of products that are customers of product p) maintain statistical significance and a big portion of the size of the coefficient. The estimators of patent and customer linkages in related industries are about twice as large as those of both agnostic channels.

The interpretation of this results is as follows. A product with stronger customer linkages to the current export basket by one standard deviation is 5 percentage points more likely to emerge in the next decade. This represents an increase of 263 percent compared to the baseline scenario, where the unconditional probability of a new product emerging is 1.9 percent (see Table 2). The corresponding effect for patents representing an increase of 310 percent compared to the baseline scenario.

The findings from the intensive margin, seen in panel B, show that density measured across all channels can explain future growth of existing products better than the agnostic measures (the coefficients are typically 2 to 3 times larger, excluding supplier linkages). However, when putting all of them together under the same regression model, only the presence of existing products related in terms of technology linkages, both through R&D and patents citations, as well as labor can explain future export growth. In particular, an increase of one standard deviation in a product's patent linkages to the current export basket is on average associated with an additional annual export growth of 14.8 percentage points. The corresponding effect for the R&D is on average associated with an annual increase in the export growth rate of 8 percentage points. A similar marginal effect corresponds to labor-based relatedness. Compared to the standardized effect of the HK and EK agnostic measures, which suggest increases in the annual growth rate of exports of 6.4 and 4.0 percentage points, respectively, both technology and labor channels have more up to two times the explanatory power.

All in all, technology (patent) and downstream linkages seems to explain the emergence of new sectors, much better than other channels. Conversely, density in the broader technological space and labor linkages dominate future growth of existing exports. The explanatory power of these channels is striking.

The fact that the construction of the density measures uses the same information except for the relatedness raises the question of whether adding density measures separately is misleading or jointly is subject to high multicollinearity problems. We argue that this is not the case. Even though the density measures correlate quite highly, the important question is how they correlate after controlling for the set of fixed effects which define our identification strategy, in particular the country-product fixed effect. We find in Table 5 that conditional on country-by-year, product-by-year and product-by-country fixed effects, the correlations are positive but not strikingly large, consistent with Table 1 which correlates the relatedness measures before including them in the calculation of the channel-specific density measures. Thus, we remain confident that multicollinearity does not affect the precision of our estimates in Table 4.[13]

[Table 5 about here.]

Discussion

When it comes to the emergence of new sectors, our results emphasize the importance of technology in the process of cross-industry spillovers, but –perhaps more surprisingly– of customer linkages: the existence of a competitive sector that could become a "buyer" to the yet-to-emerge industry. Others have looked at this phenomenon in the context of a particular country. For example, Javorcik (2004) claims that FDI productivity spillovers occur more frequently from customer to supplier, using firm-level data from Lithuania. Pietrobelli and Saliola (2008) also find that selling to a competitive customer, in their case a multinational, enhances supplier productivity. In our setting, productivity is measured by the emergence of a new export sector. More recently, Kee and Tang (2016) show that China gained comparative advantage in sectors that were upstream to Chinese exporters, implying again that the pre-existence of a downstream sector results in spillovers to the supplier. The existence of a competitive downstream industry might work both as a source of spillovers and also as a mechanism to mitigate risks for entrepreneurs and investors to start off a new sector upstream to it, that results in new exports. This is consistent with the idea that the emergence of a new sector is subject to a fixed cost that contains uncertainty of markets (e.g., Wagner and Zahler, 2015).

In terms of the intensive margin, our results point also to technology, but also to labor flows. In the case when the sector is already being exporter, a qualified labor force in other sectors might be the source of spillovers (measured

[13]In fact, the variance inflation factor for all density measures is 1.92, which is within the acceptable range.

by export growth) to other sectors. The availability of a qualified workforce (and thus, its abundance) is key in reducing production costs that could result in comparative advantage (Heckscher and Ohlin, 1991).

4.2 Industrialized vs. developing countries

We now turn to explore how our results differ for countries at different levels of development. In particular, we are interested in exploring the extent to which the economic channels exhibit differential impact for industrialized vis-a-vis developing countries. Thus we proceed to repeat the same exercise as in Table 4 separating the sample between OECD and non-OECD countries. The results are presented in Table 6.

The table has 8 columns. The first column is simply the number of observations used in each estimation. The other seven columns estimate specification (4) using each one of the density measures with different relatedness indices: HK, EG, R&D, patents, labor, supplier linkages and customer linkages. Presented coefficients are standardized. The upper panel presents results for the extensive margin, while the lower panel does so for the intensive margin.

[Table 6 about here.]

The results suggest that, indeed, some channels exhibit differential effects for developed and developing countries. Columns (2) and (3) show that the results discussed earlier with regards to the impact of the agnostic density measures are related mostly to the developing country experience. In fact, once we focus exclusively on developed countries, the HK measures completely lose their explanatory power, both in relation to the emergence of new export sectors, and the growth of existing ones. The EG agnostic densities, in contrast, retain their relevance for the extensive margin only.

Focusing on the extensive margin for the case of the different economic channels in columns (4) through (8), the two results that stand out are the ones concerning customer and supplier linkages. Both technology and labor channels are in one way or another present in both OECD and non-OECD countries, though it is important to notice that the point estimates for technological linkages (both R&D and patents) are estimated to be larger for non-OECD than for OECD countries.

With respect to customer linkages their impact is larger and significant (statistically and economically) in non-OECD countries, but not in OECD countries. Our interpretation is the following. As developing countries have less sophisticated and more uncertain markets, the existence of a competitive industry that is a "potential buyer" can mitigate risks and expedite the process of development of new upstream sectors. In the presence of frictions in credit markets, for example, the mere existence of a local market for a product can reduce uncertainty both for the creditor and the investor. In contrast, in OECD countries with more complete markets, customer linkages are less relevant for the development of new sectors.

On the other hand, our results suggest that supplier linkages are relevant only for OECD economies. We also provide a plausible interpretation for this: in these countries in which capital and technology is abundant, the ability to become more productive, plausibly, highly depends on access to better and less costly intermediate goods. Thus, the competitiveness of intermediate goods could explain changes in comparative advantage in particular for countries at higher levels of development that export more sophisticated goods intensive in intermediate inputs. According to our results, however, this link does not exist for developing countries which have, on average, less sophisticated export baskets: diversification does not seem to entail adding value to their natural resources, as new activities do not seem to depend on the existence of upstream ones.

4.3 Explaining agnostic measures: opening the "black box"

The previous exercise suggests that some channels are more important than others when explaining the emergence of new sectors or future growth of existing sectors. Moreover, we are interested in understanding to what extent these channels dominate over the explanatory power of the agnostic relatedness measures. One way to answer this question is by reestimating Table 4, but this time adding $\Phi^{HK}_{c,p,t}$ or $\Phi^{EG}_{c,p,t}$ to explore whether these two agnostic density measure lose explanatory power when added in conjunction with the channel-specific density measures. This approach answers the question: do channel-specific density measures have explanatory power beyond the agnostic measure (e.g., holding the agnostic measure constant)?[14] We, however, are interested in another question which is: how much of the agnostic measure can be explained by each channel-specific measure? That is, we put forward an two-step approach that starts by finding the component of the agnostic measure that is orthogonal to the channel-specific measure, say $\widetilde{\Phi^{channel}_{c,p,t}}$. By doing this we are extracting all the joint variance from the agnostic measure. In the second step we explore the explanatory power of $\widetilde{\Phi^{channel}_{c,p,t}}$.

Specifically, our goal is to estimate the following specification:

$$Y_{c,p,t\to T} = \beta_d \widetilde{\Phi^{channel}_{c,p,t}} + Controls_{p,c,t} + \eta_{c,t} + \delta_{p,t} + \alpha_{c,p} + \varepsilon_{c,p,t} \tag{6}$$

where $\widetilde{\Phi^{channel}_{c,p,t}}$ is the part of the agnostic measure of density (either HK or EG) that is orthogonal to a channel-defined density, which is achieved by computing the residual term of a linear regression of $\Phi_{c,p,t}$ on $\Phi^{channel}_{c,p,t}$. This residual term represents the part of the agnostic term that is not explained by the corresponding channel. That is:

$$\widetilde{\Phi^{channel}_{c,p,t}} = \Phi_{c,p,t} - \gamma \Phi^{channel}_{c,p,t}$$

where $\Phi^{channel}_{c,p,t}$ is a channel-specific density following Equation (5). γ represents the coefficient resulting from a linear regression of $\Phi_{c,p,t}$ on $\Phi^{channel}_{c,p,t}$.

Table 7 presents results using $\Phi^{HK}_{c,p,t}$ while Table 8 shows results using $\Phi^{EG}_{c,p,t}$. In the first column titled "None", both tables replicate the specification that is estimated in Table 3, except that the number of observations is limited to those for which the channel-defined densities are not missing. That is, in the

[14] The results from this exercise are presented in Appendix Section C. That is, we replicate Table 4 controlling for the agnostic density measures. Our overall findings are unchanged.

case where $\widetilde{\Phi_{c,p,t}} = \Phi_{c,p,t}$. Columns 2-6 in both tables use the residual term of running $\Phi_{c,p,t}$ on each $\Phi^{channel}_{c,p,t}$ (columns 2-6). Control variables are omitted to simplify clarity.

[Table 7 about here.]

[Table 8 about here.]

The point estimates reported in Tables 7 reveal that in the extensive margin (Panel A) customer linkages dominate over the agnostic relatedness measures; the coefficients of the agnostic density orthogonal to this channels is smallest. All other columns show results that are positive and stronger, implying that these "cleaned" agnostic measures continue to explain export emergence. Moreover, the point estimates shed light on the extent to which these channels can explain the agnostic relatedness measures. In particular, labor and supply linkages can explain about 16% of the agnostic measures (columns 4 and 5, respectively)[15], R&D and patents can explain about 23% (column 2 and 3) and customer linkages is the most explanatory of all measures, being able to explain about 29% (Column 6). When it comes to explaining future growth of already existing export products (Panel B), labor force and patent citation similarities best explain the agnostic measures. In particular, similarities based on labor force flows and patents and R&D can each explain about 40 percent.

Table 8 shows results using $\Phi^{EG}_{c,p,t}$. Two findings stand out. First, even though all residual densities prove statistically significant in the extensive margin, the proportion that the channels explain the EG agnostic relatedness measure is robust and similar to that of the HK agnostic relatedness measure: the R&D measures explain about 16% of the agnostic measures (Columns 2 and 3), patents and supply flows about 20% (Column 4 and 6) and customer linkages about 28% (Column 7). Second, at the intensive margin, labor flows

[15] To obtain the proportion explained by labor in the agnostic relatedness measure, we compute the percentage change in the point estimates for β_d in different specifications (e.g., $\frac{0.4028-0.3377}{0.4028}$)

explain the greatest proportion of the agnostic relatedness measure with now 70%, confirming the findings from using the agnostic HK measure.

Another way to see that customer linkages at the extensive margin and labor linkages at the intensive margin explain the greatest proportion of the agnostic relatedness measures is through figures 1 and 2. Both offer a visualization of the results in Tables 7 and 8, respectively. For example, in figure 1 in the extensive margin, the point estimates of the agnostic relatedness measured purged from customer linkages is smallest. This implies that it is the customer channel which correlates mostly with export emergence. Moreover, the comparison of both figures shows that the point estimates of the channels exhibit the same importance relative to the agnostic relatedness measures, both at the extensive and intensive margin. This confirms that the channels explain the same proportion of the agnostic measure, irrespective of its measurement with EG or HK densities.[16]

[Figure 1 about here.]

[Figure 2 about here.]

5 Concluding Remarks

New export products do not emerge randomly: they tend to be related to what countries already sell to the rest of the world. To understand the channels mediating this evolution from "old" to "new" export products, we use competing measures of relatedness across sectors based on technology, labor and input/output relationships. These measures allow us to explore the competing mechanisms that mediate both the emergence and growth of export products.

Our results suggest that technology linkages explain both the emergence of new export sectors and future growth of already existing sectors in very robust

[16]It is important to note that, statistically, there is often no difference between all estimators, each one using different measures of relatedness. Yet, we focus our interpretation on the point estimates.

ways. These measures can explain a big chunk of previously used agnostic relatedness measures in the dynamics of the export basket of countries. This is in fact consistent with the view that knowledge diffusion across industries plays an important role in explaining sectorial productivity shifts expressed in the form of changes in comparative advantage (Bahar et al., 2014).

Moreover, when looking at the emergence of new sectors, an additional and overlooked component plays a role that is as large as technology: the existence of competitive downstream sectors; suggesting the importance of an industry pulling through higher certainty of demand. Our global stylized fact generalizes the view by Javorcik (2004) that spillovers are more likely in backward linkages. This finding contrast with the view of theories of development that argue that countries should focus on downstream industries with more value added.

When it comes to driving export growth, we find that –on top of technology– labor linkages play an important role. The different channels driving the extensive and intensive margins coincide with the conclusion of Bresnahan (2001) who explore case studies to conclude that "*the economic factors that give rise to the start of a cluster can be very different from those that keep it going.*" We acknowledge that our findings are not necessarily causal, but reveal relationships between industries that are usually plagued by case-by case micro evidence. We are, at the best of our knowledge, the first paper trying to understand the channels behind the evolution of comparative advantage across related industries using a global sample. Identifying causality with exogenous sources of variation is left as an important challenge for future research.

References

Arbeláez, M. A., M. Meléndez, and N. León (2007). The emergence of new succesful export activities in Colombia.

Bahar, D., R. Hausmann, and C. A. Hidalgo (2014, jan). Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? *Journal of International Economics 92*(1), 111–123.

Bahar, D. and H. Rapoport (2017). Migration, knowledge diffusion and the comparative advantage of nations. *The Economic Journal (Forthcoming)*.

Balassa, B. (1965). Trade Liberalisation and Revealed Comparative Advantage. *The Manchester School 33*(2), 99–123.

Bernhofen, D. M. ., Z. El-Sahli, and R. Kneller (2016). The Impact of Technological Change on New Trade: Evidence from the Container Revolution.

Bresnahan, T. (2001, dec). 'Old Economy' Inputs for 'New Economy' Outcomes: Cluster Formation in the New Silicon Valleys. *Industrial and Corporate Change 10*(4), 835–860.

Cadot, O., C. Carrère, and V. Strauss-Kahn (2011, may). Export Diversification: What's behind the Hump? *Review of Economics and Statistics 93*(2), 590–605.

Carvalho, V. M., D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica 80*(5), 1977–2016.

Caselli, F., M. Koren, M. Lisicky, and S. Tenreyro (2015, aug). Diversification through Trade. Technical report, National Bureau of Economic Research, Cambridge, MA.

Costinot, A., D. Donaldson, and I. Komunjer (2012). What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas. *Review of Economic Studies 79*, 581–608.

Crespi, G., E. Fernández-Arias, and E. Stein (2014). *Rethinking Productive Development: Sound Policies and Institutions for Economic Transformation*. New York, NY: Palgrave Macmillan.

Cuñat, A. and M. J. Melitz (2012). Volatility, labor market flexibility, and the pattern of comparative advantage. *Journal of the European Economic Association 10*(2), 225–254.

Eaton, J. and S. Kortum (2002, sep). Technology, Geography, and Trade. *Econometrica 70*(5), 1741–1779.

Ellison, G. and E. L. Glaeser (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy 105*(5), 889.

Ellison, G., E. L. Glaeser, and W. R. Kerr (2010, jun). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review 100*(3), 1195–1213.

Feenstra, R. C., R. E. Lipsey, H. Deng, A. C. Ma, and H. Mo (2005). World Trade Flows: 1962-2000.

Feenstra, R. C. and J. Romalis (2014, may). International Prices and Endogenous Quality. *The Quarterly Journal of Economics 129*(2), 477–527.

Greenstone, M., R. Hornbeck, and E. Moretti (2010). Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy 118*(3), 536–598.

Griffith, R., R. Harrison, and J. Van Reenen (2006). How special is the special relationship? Using the impact of US R&D spillovers on UK firms as a test of technology sourcing. *The American Economic Review 96*(5), 1859–1875.

Grossman, G. M. and E. Helpman (1991, apr). Trade, knowledge spillovers, and growth. *European Economic Review 35*(2-3), 517–526.

Hall, B. H. and M. Trajtenberg (2006). *Uncovering general purpose technologies with patent data.*

Hallak, J. C. (2006, jan). Product quality and the direction of trade. *Journal of International Economics 68*(1), 238–265.

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

Hausmann, R., C. A. Hidalgo, S. Bustos, M. Coscia, A. Simoes, and M. A. Yildirim (2014). *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. Cambridge, MA: MIT Press.

Hausmann, R., J. Hwang, and D. Rodrik (2007). What you export matters. *Journal of economic growth 25*(1), 1–25.

Hausmann, R. and B. Klinger (2006). Structural Transformation and Patterns of Comparative Advantage in the Product Space. *{{}CID{}} Working Paper Series, Harvard University*.

Hausmann, R. and F. Neffke (2016). The workforce of pioneer plants.

Hausmann, R., F. Rodríguez, and R. Wagner (2006). Growth Collapses. *CID Working Paper Series* (RWP06-046).

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

Heckscher, E. and B. Ohlin (1991). *Heckscher-Ohlin Trade Theory*. Cambridge: MIT Press.

Hidalgo, C. A., B. Klinger, A. Barabási, and R. Hausmann (2007, jul). The product space conditions the development of nations. *Science 317*(5837), 482–7.

Hummels, D. and P. J. Klenow (2005, may). The Variety and Quality of a Nation's Exports. *American Economic Review 95*(3), 704–723.

Imbs, J. and R. Wacziarg (2003, feb). Stages of Diversification. *American Economic Review 93*(1), 63–86.

Javorcik, B. (2004). Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages. *The American Economic Review 94*(3), 605–627.

Kee, H. L. and H. Tang (2016, jun). Domestic Value Added in Exports: Theory and Firm Evidence from China. *American Economic Review 106*(6), 1402–1436.

Klinger, B. and R. Hausmann (2007, oct). The Structure of the Product Space and the Evolution of Comparative Advantage. *CID Working Paper Series* (128).

Koren, M. and S. Tenreyro (2007). Volatility and Development. *The Quarterly Journal of Economics 122*(1), 243–287.

Kremer, M. (1993). The O-ring theory of economic development. *The Quarterly Journal of Economics 108*(3), 551–575.

Krishna, P. and A. Levchenko (2009, may). Comparative Advantage, Complexity and Volatility. Technical report, National Bureau of Economic Research, Cambridge, MA.

Krugman, P. R. (1991). *Geography and trade.* Cambridge, MA: MIT Press.

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

Neffke, F., A. Otto, and A. Weyh (2016). Inter-industry Labor Flows.

Pietrobelli, C. and F. Saliola (2008, apr). Power relationships along the value chain: multinational firms, global buyers and performance of local suppliers. *Cambridge Journal of Economics 32*(6), 947–962.

Ricardo, D. (1821). *On the Principles of Political Economy and Taxation.* London: John Murray, Albemarle-Street.

Rodrik, D. (2012). Unconditional convergence in manufacturing. *The Quarterly Journal of Economics*, qjs047.

Rodrik, D. (2016, mar). Premature deindustrialization. *Journal of Economic Growth 21*(1), 1–33.

Romalis, J. (2004, feb). Factor Proportions and the Structure of Commodity Trade. *American Economic Review 94*(1), 67–97.

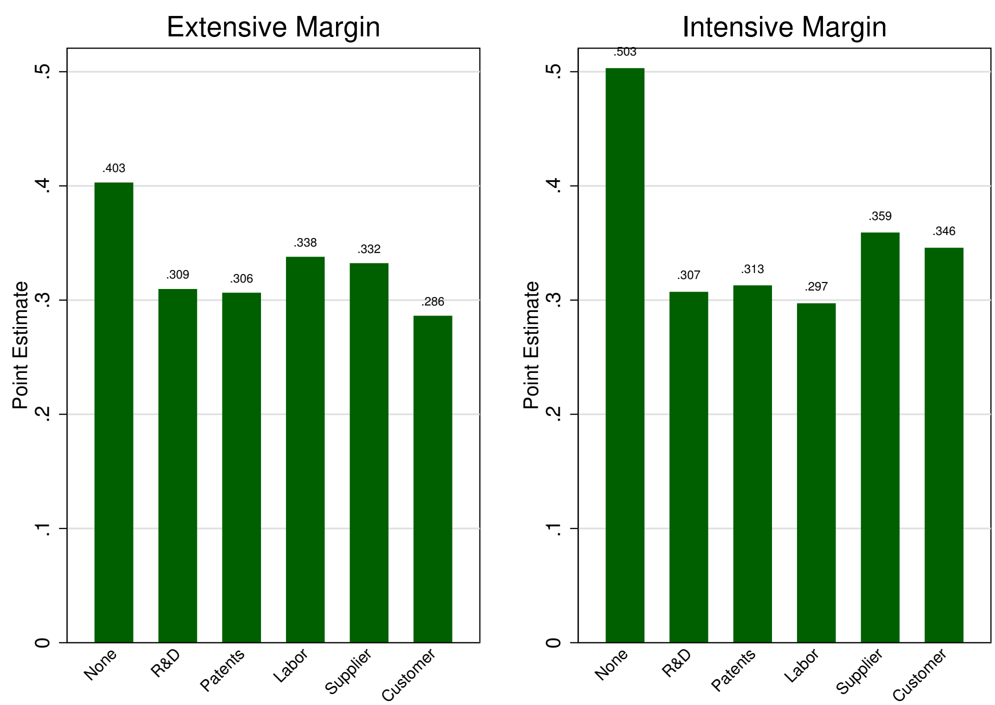
Sabel, C., E. Fernandez-Arias, R. Hausmann, A. Rodriguez-Clare, and E. H. Stein. (2011). *Self-Discovery as a Coordination Problem: Lessons from a Study of New Exports in Latin America.* Inter-American Development Bank.

Scherer, F. (1984). Using Linked Patent and R{&}D Data to Measure Interindustry Technology Flows. *R &D, Patents, and Productivity*.

Verhoogen, E. A. (2007). Trade, quality upgrading and wage inequality in the Mexican manufacturing sector. *The Quarterly Journal of Economics*.

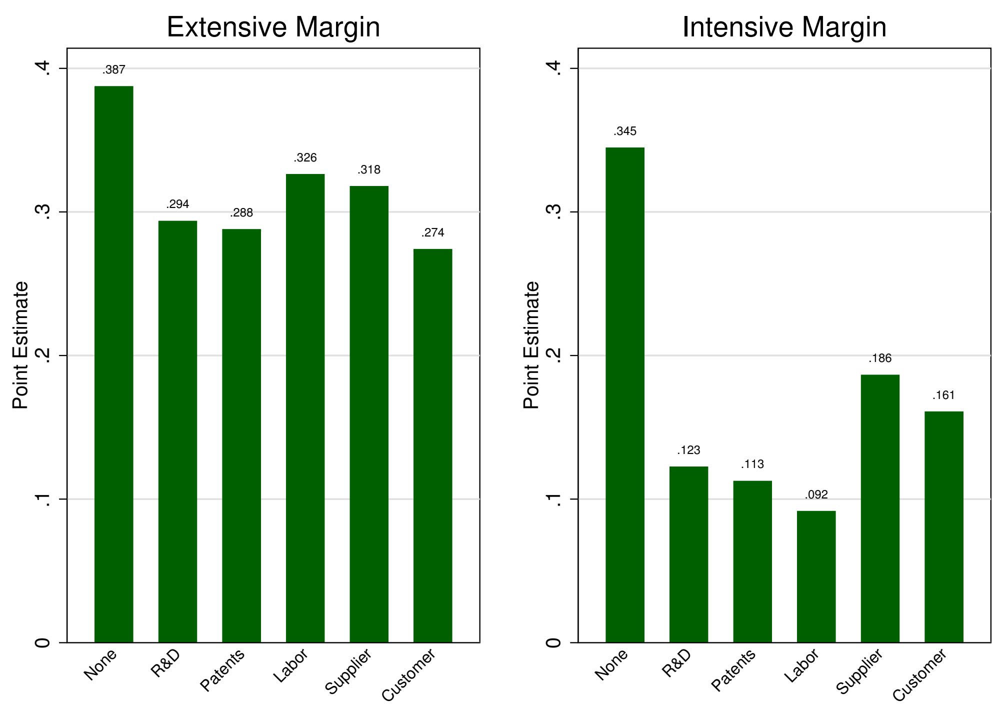
Wagner, R. and A. Zahler (2015). New exports from emerging markets: Do followers benefit from pioneers? *Journal of Development Economics 114*, 203–223.

Figure 1: Explaining the black box, HK densities



This figure compares the point estimate of the agnostic HK relatedness measure, called None, from specification 6 to those that are orthogonal to channel-defined density.

Figure 2: Explaining the black box, EG densities



This figure compares the point estimate of the agnostic EG relatedness measure, called None, from specification 6 to those that are orthogonal to channel-defined density.

Table 1: Correlations of Relatedness Measures

Variables	HK	EG	R&D	Patents	Consumer	Supplier	Labor
HK	1.000						
EG	0.498	1.000					
R&D	0.075	0.075	1.000				
Patents	0.158	0.137	0.341	1.000			
Consumer	0.102	0.101	0.431	0.289	1.000		
Supplier	0.120	0.113	0.315	0.364	0.457	1.000	
Labor	0.168	0.191	0.368	0.573	0.391	0.377	1.000

Table 2: Summary Statistics

Variable	**N**	**Mean**	**sd**	**Min**	**Max**
Panel A: Extensive Margin Sample ($RCA_{c,p,t} < 0.1$)					
New Product $RCA_{c,p,t+10} > 1$	93,260	0.019	0.14	0.0	1.0
$\Phi_{c,p,t}(HK)$	93,260	0.095	0.08	0.0	0.6
$\Phi_{c,p,t}(EG)$	93,260	0.102	0.08	0.0	0.6
$\Phi_{c,p,t}^{R\&D}$	79,769	0.097	0.08	0.0	0.5
$\Phi_{c,p,t}^{Patents}$	79,769	0.097	0.08	0.0	0.5
$\Phi_{c,p,t}^{Slinkages}$	79,769	0.096	0.08	0.0	0.6
$\Phi_{c,p,t}^{CLinkages}$	79,769	0.095	0.08	0.0	0.5
$\Phi_{c,p,t}^{Labor}$	79,769	0.094	0.08	0.0	0.5
Baseline RCA (log)	93,260	0.013	0.02	0.0	0.1
Panel B: Intensive Margin Sample ($Exports_{c,p,t} > 0$)					
CAGR 10 years	122,286	0.046	0.28	-0.9	2.7
$\Phi_{c,p,t}(HK)$	122,286	0.192	0.13	0.0	0.8
$\Phi_{c,p,t}(EG)$	122,286	0.186	0.12	0.0	0.8
$\Phi_{c,p,t}^{R\&D}$	108,643	0.181	0.11	0.0	0.5
$\Phi_{c,p,t}^{Patents}$	108,643	0.183	0.11	0.0	0.5
$\Phi_{c,p,t}^{Slinkages}$	108,643	0.182	0.11	0.0	0.6
$\Phi_{c,p,t}^{CLinkages}$	108,643	0.182	0.12	0.0	0.6
$\Phi_{c,p,t}^{Labor}$	108,643	0.182	0.12	0.0	0.6
Baseline Level of Exports (log)	122,286	14.394	3.40	7.6	25.4
Previous Period Growth Rate	122,286	1.106	2.60	-0.8	53.9
Previous Period Zero Exports	122,286	0.223	0.42	0.0	1.0

This table presents descriptive statistics for our key dependent variables: export emergence and export growth. It also includes statistics for agnostic and channel-specific density measures ($\Phi_{c,p,t}$) as well as control variables. The upper panel presents the sample used in the estimations of the extensive margin, where we limit the sample to those country-product observations that have RCA below 0.1 in the beginning of the 1990-2000 and 2000-2010 periods. The lower panel presents results used in the estimations of the intensive margin, where we limit our observations to those country-products with exports above zero at the beginning of the 1990-2000 and 2000-2010 periods.

Table 3: Emergence and growth of related industries

Panel A: Extensive Margin				
	HK		EG	
	(1)	(2)	(3)	(4)
Φ_{cpt}	0.3323	0.3377	0.1955	0.3332
	(0.045)***	(0.080)***	(0.033)***	(0.080)***
Baseline RCA (log)	0.2648	0.1099	0.2823	0.1123
	(0.033)***	(0.034)***	(0.034)***	(0.034)***
N	93260	73988	93260	73988
Adj. R2	0.04	0.01	0.04	0.01
$\alpha_{c,p}$	N	Y	N	Y

Panel B: Intensive Margin				
	HK		EG	
	(1)	(2)	(3)	(4)
Φ_{cpt}	0.6327	0.5040	0.3939	0.3389
	(0.079)***	(0.098)***	(0.056)***	(0.105)***
Baseline Level of Exports (log)	-0.0403	-0.1389	-0.0374	-0.1382
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Growth Rate	-0.0030	0.0061	-0.0039	0.0057
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Zero Exports	-0.0899	-0.0235	-0.0877	-0.0202
	(0.008)***	(0.010)**	(0.008)***	(0.010)**
N	122285	101698	122285	101698
Adj. R2	0.32	0.62	0.32	0.62
$\alpha_{c,p}$	N	Y	N	Y

This table estimates specification (4) with the agnostic density measures for 10 years. The upper panel estimates the specification for the extensive margin and the lower panel does so for the intensive margin. Columns 1 and 2 estimate results using HK proximity, and columns 3 and 4 use EG proximity as the main input for the right hand side variable of interest. All specifications include country-by-year and product-by-year fixed effects, with columns 2 and 4 controlling additionally for country-product fixed effects Standard errors are clustered at the country level and presented in parenthesis.

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 4: Emergence and growth of related industries, defined channels

Panel A: Extensive Margin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\Phi^{R\&D}_{c,p,t}$	0.0242					0.0091	
	(0.015)					(0.014)	
$\Phi^{Patents}_{c,p,t}$		0.0547					0.0590
		(0.019)***					(0.026)**
$\Phi^{Labor}_{c,p,t}$			0.0265			0.0077	-0.0158
			(0.013)*			(0.013)	(0.017)
$\Phi^{Slinkages}_{c,p,t}$				0.0085		-0.0080	-0.0152
				(0.009)		(0.009)	(0.009)
$\Phi^{CLinkages}_{c,p,t}$					0.0539	0.0499	0.0414
					(0.014)***	(0.014)***	(0.013)***
Baseline RCA (log)	0.1257	0.1215	0.1245	0.1266	0.1235	0.1231	0.1215
	(0.036)***	(0.036)***	(0.036)***	(0.037)***	(0.037)***	(0.036)***	(0.036)***
N	63232	63232	63232	63232	63232	63232	63232
Adj. R2	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$\alpha_{c,p}$	Y	Y	Y	Y	Y	Y	Y
$\Phi^{HK}_{c,p,t}$	.0287	.0287	.0287	.0287	.0287	.0287	.0287
$\Phi^{EG}_{c,p,t}$	.0277	.0277	.0277	.0277	.0277	.0277	.0277

Panel B: Intensive Margin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\Phi^{R\&D}_{c,p,t}$	0.1461					0.0807	
	(0.022)***					(0.025)***	
$\Phi^{Patents}_{c,p,t}$		0.2057					0.1479
		(0.036)***					(0.042)***
$\Phi^{Labor}_{c,p,t}$			0.1538			0.0956	0.0479
			(0.029)***			(0.031)***	(0.036)
$\Phi^{Slinkages}_{c,p,t}$				0.0773		-0.0098	-0.0193
				(0.023)***		(0.020)	(0.020)
$\Phi^{CLinkages}_{c,p,t}$					0.1520	0.0562	0.0499
					(0.033)***	(0.030)*	(0.029)*
Baseline Level of Exports (log)	-0.1387	-0.1390	-0.1389	-0.1386	-0.1387	-0.1390	-0.1390
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Growth Rate	0.0063	0.0064	0.0064	0.0062	0.0063	0.0065	0.0065
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Zero Exports	-0.0260	-0.0270	-0.0267	-0.0254	-0.0260	-0.0271	-0.0273
	(0.010)***	(0.010)***	(0.010)***	(0.010)**	(0.010)***	(0.010)***	(0.010)***
N	90798	90798	90798	90798	90798	90798	90798
Adj. R2	0.62	0.62	0.62	0.62	0.62	0.62	0.62
$\alpha_{c,p}$	Y	Y	Y	Y	Y	Y	Y
$\Phi^{HK}_{c,p,t}$	.0639	.0639	.0639	.0639	.0639	.0639	.0639
$\Phi^{EG}_{c,p,t}$	.0405	.0405	.0405	.0405	.0405	.0405	.0405

This table estimates specification (4) with the channel-specific density measures for 10 years. The upper panel estimates the specification for the extensive margin and the lower panel does so for the intensive margin. Columns 1 to 5 evaluate the impact of each channel-specific density measure on export emergence and export growth separately. Column 6 and 7 submit all measures jointly, distinguishing between R&D and patent channels because both capture technological linkages. All specifications include country-by-year, product-by-year and product-by-country fixed effects. All coefficients are standardized, including those for the agnostic densities in the bottom of each panel. Standard errors are clustered at the country level and presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 5: Correlations of Density Measures, controlling for fixed effects

Variables	Φ^{HK}	Φ^{EG}	$\Phi^{R\&D}$	$\Phi^{Patents}$	Φ^{Labor}	$\Phi^{Slinkages}$	$\Phi^{CLinkages}$
Φ^{HK}	1.000						
Φ^{EG}	0.634	1.000					
$\Phi^{R\&D}$	0.151	0.141	1.000				
$\Phi^{Patents}$	0.249	0.260	0.470	1.000			
Φ^{Labor}	0.285	0.302	0.474	0.754	1.000		
$\Phi^{Slinkages}$	0.204	0.225	0.401	0.556	0.523	1.000	
$\Phi^{CLinkages}$	0.200	0.221	0.411	0.545	0.547	0.330	1.000

Table 6: OECD vs. non-OECD countries, different channels

Sample	**N**	**HK**	**EG**	**R&D**	**Patents**	**Labor**	**Supplier**	**Customer**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Extensive Margin								
All Observations	63232	0.029***	0.028***	0.024	0.055***	0.026**	0.008	0.054***
Non OECD	56706	0.034***	0.029***	0.031*	0.066***	0.028*	0.005	0.060***
OECD	6526	0.011	0.024***	0.003	0.029*	0.031*	0.024**	0.034
Intensive Margin								
All Observations	30905	0.044***	0.036**	0.081**	0.180***	0.105**	0.061**	0.152***
Non OECD	25295	0.060**	0.053**	0.130***	0.244***	0.109*	0.036	0.187***
OECD	5610	0.001	0.004	0.030	0.106**	0.089*	0.099***	0.081

This table estimates specification (4) with the agnostic and channel-specific density measures for 10 years for different subset of the data. The upper panel estimates the specification for the extensive margin and the lower panel does so for the intensive margin. Columns 2 and 8 estimate the partial correlation of the agnostic and channel-specific density measures on export emergence and export growth separately. All specifications include country-by-year, product-by-year and product-by-country fixed effects. All coefficients are standardized. Standard errors are clustered at the country level and presented in parenthesis.

Table 7: Explaining the black box, HK densities

Panel A: Extensive Margin						
	(1)	(2)	(3)	(4)	(5)	(6)
	None	R&D	Patents	Labor	Supplier	Customer
$\widetilde{\Phi cpt}$	0.4028	0.3095	0.3063	0.3377	0.3321	0.2861
	(0.103)***	(0.092)***	(0.090)***	(0.096)***	(0.083)***	(0.096)***
N	63232	63232	63232	63232	63232	63232
Adj. R2	0.01	0.01	0.01	0.01	0.01	0.01
$\alpha_{c,p}$	Y	Y	Y	Y	Y	Y
Panel B: Intensive Margin						
	(1)	(2)	(3)	(4)	(5)	(6)
	None	R&D	Patents	Labor	Supplier	Customer
$\widetilde{\Phi cpt}$	0.5029	0.3070	0.3127	0.2970	0.3590	0.3456
	(0.100)***	(0.089)***	(0.093)***	(0.089)***	(0.088)***	(0.094)***
N	90798	90798	90798	90798	90798	90798
Adj. R2	0.62	0.62	0.62	0.62	0.62	0.62
$\alpha_{c,p}$	Y	Y	Y	Y	Y	Y

This table estimates specification (6) with the agnostic HK relatedness measure that is orthogonal to channel-defined density. The upper panel estimates the specification for the extensive margin and the lower panel does so for the intensive margin. Column 1 replicates the results with the original agnostic HK density, our reference case. Columns 2 to 6 evaluate the impact of each agnostic density that is cleaned from channel-defined densities. All specifications include country-by-year, product-by-year and country-by-product fixed effects. Control variables are not shown to simplify clarity. Standard errors are clustered at the country level and presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 8: Explaining the black box, EG densities

Panel A: Extensive Margin						
	(1)	(2)	(3)	(4)	(5)	(6)
	None	R&D	Patents	Labor	Supplier	Customer
$\widetilde{\Phi cpt}$	0.3875	0.2937	0.2879	0.3262	0.3179	0.2741
	(0.097)***	(0.087)***	(0.082)***	(0.092)***	(0.079)***	(0.088)***
N	63232	63232	63232	63232	63232	63232
Adj. R2	0.01	0.01	0.01	0.01	0.01	0.01
$\alpha_{c,p}$	Y	Y	Y	Y	Y	Y
Panel B: Intensive Margin						
	(1)	(2)	(3)	(4)	(5)	(6)
	None	R&D	Patents	Labor	Supplier	Customer
$\widetilde{\Phi cpt}$	0.3447	0.1225	0.1126	0.0916	0.1865	0.1609
	(0.108)***	(0.094)	(0.101)	(0.099)	(0.096)*	(0.101)
N	90798	90798	90798	90798	90798	90798
Adj. R2	0.62	0.62	0.62	0.62	0.62	0.62
$\alpha_{c,p}$	Y	Y	Y	Y	Y	Y

This table estimates specification (6) with the agnostic EG relatedness measure that is orthogonal to channel-defined density. The upper panel estimates the specification for the extensive margin and the lower panel does so for the intensive margin. Column 1 replicates the results with the original agnostic EG density, our reference case. Columns 2 to 6 evaluate the impact of each agnostic density that is cleaned from channel-defined densities. All specifications include country-by-year, product-by-year and country-by-product fixed effects. Control variables are not shown to simplify clarity. Standard errors are clustered at the country level and presented in parenthesis.

${}^{*}p < 0.10, {}^{**}p < 0.05, {}^{***}p < 0.01$

For Online Publication:
Appendix for *"The birth and growth of new export clusters"*

by Bahar, Stein, Wagner and Rosenow

A Exploring quality in surrounding products

Product quality may serve as an additional channel behind the emergence of a new export sector. Rich countries tend to export high quality goods (e.g. Hummels and Klenow, 2005; Hallak, 2006), and high quality is also instrumental for economic development and wage growth (Kremer, 1993; Grossman and Helpman, 1991; Verhoogen, 2007). Moreover, achieving higher quality in particular products can generate cross-industry productivity spillovers to related sectors. An anecdotical story on this is the case of the Colombian swimsuit industry. In the late 1980s the managers of *Leonisa* - a Colombian manufacturer of female underwear - visited a competitor in Spain. Among many observations, they discovered that unit prices of this competitor were about ten times larger than those of *Leonisa*. After that experience, Leonisa's management decided to invest in quality by upgrading its machinery, moving from $50,000 looms to $1 million looms capable of producing elastic lace as well as introducing changes in production techniques. Thus, by investing in both capital and organizational processes, Leonisa was able to produce higher quality underwear, making use of knitted fabrics and other top-of-the-line improvements. In a few years Leonisa went from exporting $2 products competing with Chinese producers to $30 products competing with Italian and French firms. Exports soared. Moreover, the accumulation of capabilities through quality upgrading allowed Leonisa to develop a new line of products: female swimsuits. At the same time, Leonisa was not the only beneficiary of its quality upgrading. Other Colombian firms, such as *Onda de Mar* and *Supertex,* joined the swimsuit industry (for a detailed account of the Leonisa case, see Arbeláez et al. (2007) as well as Sabel et al., 2011 and Crespi et al., 2014).

In this section we explore whether the data supports the link between quality upgrading and country's export diversification.

For this section our basic building block proves $Q_{c,p,t}$, a country and product-specific measure of export quality constructed by Feenstra and Romalis (2014). Varying over time, it captures the unit price of each product by each country when exported, adjusted by demand-side and supply-side considerations. In terms of demand side, the values for quality are calculated such that, conditional on prices, higher quality goods are exported more. In terms of supply side, the data is constructed following the "Washington apple" principle: high quality goods travel longer distances. We take the quality values by Feenstra and Romalis (2014) as given and normalize them to represent deviations from the average quality of that same product in the world in that particular year.[17]

Figure A1 shows the distribution of $Q_{c,p,t}$ for men shirts (SITC 8441) in the year 2000 plotted against GDP per capita for all countries in our sample.

[Figure A1 about here.]

Based on this, we use $Q_{c,p,t}$ as an additional weight in the $\Phi_{c,p,t}$ calculation for each combination of country c, product p and time t, such that:

$$\Phi^{Q}_{c,p,t} = \frac{\sum_{j \neq \mathrm{p}} \varphi_{p,j} \times R_{c,j,t} \times Q_{c,j,t}}{\sum_{j \neq p} \varphi_{p,j}}. \tag{7}$$

where the super index Q in $\Phi^{Q}_{c,p,t}$ means that it is a quality-based density, as opposed to the the raw density $\Phi_{c,p,t}$ in Eq (3). This measure reflects the average quality in the export basket of country c in time t weighted by relatedness to product p.

Notice that both measures $\Phi_{c,p,t}$ and $\Phi^{Q}_{c,p,t}$ are key in the empirical analysis. In order to establish the role of quality beyond the purely relatedness channel established in Table 3, we estimate run the following specification:

$$Y_{c,p,t \to T} = \beta_q \Phi^{Q}_{c,p,t} + \beta_d \Phi_{c,p,t} + Controls_{c,p,t} + \eta_{c,t} + \delta_{p,t} + \alpha_{c,p} + \varepsilon_{c,p,t} \tag{8}$$

[17]For example, $Q_{c,p,t} = 1.5$ means that country c exports product p in time t with quality that is 50% higher than average quality for that product in the World.

In order to compare the marginal return of the channels between each other and vis-a-vis the agnostic measures, we report standardized coefficients. The results are presented in Table A1.

[Table A1 about here.]

Two main findings stand out. First, without country-product fixed effects, quality-based density $\Phi^{Q}_{c,p,t}$ is associated with the the emergence and, in particular, the growth of related industries, controlling for the agnostic density measures. For example, an increase of one standard deviation in the quality-based density around a given product is associated with additional export growth of 8 to 10 percentage point, as seen in column 1 and 2 of panel B. Second, with additional country-product fixed effects, these results are no longer maintained; $\widetilde{\Phi^{Q}_{c,p,t}}$ can no longer explain the emergence or growth of new related industries. However, we argue that these additional fixed effects impose an exceedingly restrictive test on our identification strategy: we can only identify effects based on changes over time in the quality around a given product-country, not whether a product with greater quality around it tends to emerge compared to a product with less quality around it. Bearing in mind these caveats, we interpret the findings as suggestive evidence that quality upgrading in related sectors is associated with the emergence of a new product in the basket and growth of existing ones in particular.

Furthermore, to address concern that $\Phi_{c,p,t}$ and $\Phi^{Q}_{c,p,t}$ are multi collinear, we follow the same methodology described in Section 4.3 and, instead, estimate the following:

$$Y_{c,p,t\to T} = \beta_q \widetilde{\Phi^{Q}_{c,p,t}} + Controls_{c,p,t} + \eta_{c,t} + \delta_{p,t} + \alpha_{c,p} + \varepsilon_{c,p,t} \qquad (9)$$

where $\widetilde{\Phi^{Q}_{c,p,t}}$ is the part of $\Phi^{Q}_{c,p,t}$ that is orthogonal to $\Phi_{c,p,t}$.

The results are presented in Table A2. Control variables are not shown to simplify clarity.

[Table A2 about here.]

The findings from Table A2 confirm the patterns highlighted in the previous Table A1. Without country-product fixed effects, quality among related incumbent exports is associated with the emergence and, in particular, growth of a related product in the export basket. Yet again, more conservative country-product fixed effects render these results statistically insignificant.

B Using logarithmic growth as a measure for the intensive margin

Table A3 shows that the results for the intensive margin are robust to using a log-growth specification to construct $Y_{c,p,t \to T}$.

[Table A3 about here.]

C Estimating channels-specific measures controlling for agnostic ones

The following two tables replicate figure 4, controlling for the agnostic density measures by HK and EG, respectively. The results confirm the relative importance of customer linkages for the emergence of exports and patent linkages for export growth, respectively.

[Table A4 about here.]

[Table A5 about here.]

D Relatedness between products for agnostic and channel-specific measures

The following tables show the 15 most related products pairs for all 8 different relatedness measures used in this paper. The two agnostic relatedness measures are shown first.

[Table A6 about here.]

[Table A7 about here.]

[Table A8 about here.]

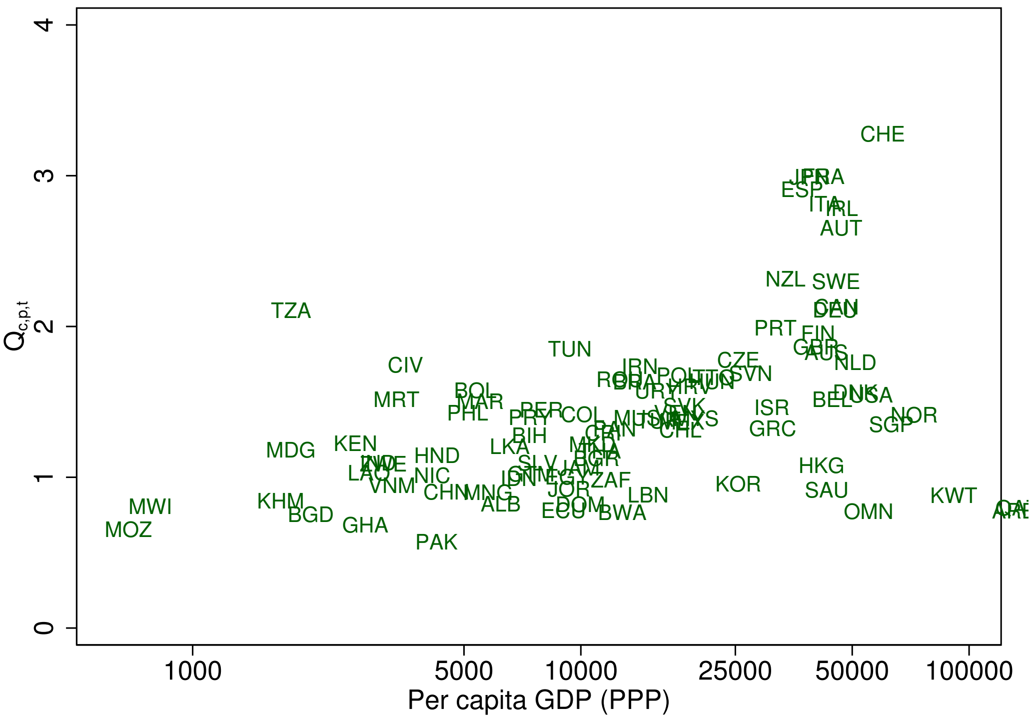
[Table A9 about here.]

[Table A10 about here.]

[Table A11 about here.]

[Table A12 about here.]

Figure A1: Quality for men's shirts (8441) in 2000



Product quality Q varies considerably across countries, as exemplified for men's shirts in 2000. The variation in product quality proves particularly pronounced as exporters' income per capita grows.

Table A1: Quality and Dynamic Comparative Advantage

Panel A: Extensive Margin				
	(1)	(2)	(3)	(4)
$\Phi_{c,p,t}(HK)$	0.0267		0.0316	
	(0.008)***		(0.014)**	
$\Phi_{c,p,t}(EG)$		0.0172		0.0379
		(0.007)**		(0.008)***
$\Phi^{Q}_{c,p,t}$	0.0087	0.0126	-0.0029	-0.0020
	(0.005)*	(0.005)**	(0.003)	(0.003)
Baseline RCA (log)	0.1878	0.1945	0.0969	0.0990
	(0.048)***	(0.049)***	(0.035)***	(0.036)***
N	101120	101120	81058	81058
Adj. R2	0.12	0.12	0.08	0.08
$\alpha_{c,p}$	N	N	Y	Y

Panel B: Intensive Margin				
	(1)	(2)	(3)	(4)
$\Phi_{c,p,t}(HK)$	0.0107		0.0707	
	(0.016)		(0.017)***	
$\Phi_{c,p,t}(EG)$		-0.0084		0.0340
		(0.008)		(0.013)**
$\Phi^{Q}_{c,p,t}$	0.0875	0.1017	-0.0092	0.0116
	(0.015)***	(0.011)***	(0.015)	(0.013)
Baseline Level of Exports (log)	-0.0413	-0.0411	-0.1389	-0.1383
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Growth Rate	-0.0031	-0.0032	0.0061	0.0058
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Zero Exports	-0.0882	-0.0870	-0.0237	-0.0204
	(0.008)***	(0.008)***	(0.010)**	(0.010)**
N	122285	122285	101698	101698
Adj. R2	0.33	0.33	0.62	0.62
$\alpha_{c,p}$	N	N	Y	Y

This table estimates specification (8) for 10 years, reporting standardized coefficients. The upper panel estimates the specification for the extensive margin and the lower panel does so for the intensive margin. All specifications include country-by-year and product-by-year fixed effects, with columns 3 and 4 controlling additionally for country-product fixed effects. Standard errors are clustered at the country level and presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A2: Quality and Dynamic Comparative Advantage, using $\widetilde{\Phi^{Q}_{c,p,t}}$

Panel A: Extensive Margin

	HK		EG	
	(1)	(2)	(3)	(4)
	est1	est2	est3	est4
Φ^{Q}_{cpt}	0.0860	0.0022	0.0280	0.0109
	(0.019)***	(0.020)	(0.015)*	(0.012)
Baseline RCA (log)	0.2808	0.1187	0.2897	0.1188
	(0.034)***	(0.035)***	(0.034)***	(0.035)***
N	93260	73988	93260	73988
Adj. R2	0.04	0.01	0.04	0.01
$\alpha_{c,p}$	N	Y	N	Y

Panel B: Intensive Margin

	HK		EG	
	(1)	(2)	(3)	(4)
	est1	est2	est3	est4
Φ^{Q}_{cpt}	0.6517	0.0268	0.3674	0.0573
	(0.072)***	(0.077)	(0.049)***	(0.057)
Baseline Level of Exports (log)	-0.0390	-0.1378	-0.0369	-0.1378
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Growth Rate	-0.0039	0.0055	-0.0042	0.0055
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Zero Exports	-0.0839	-0.0180	-0.0852	-0.0182
	(0.008)***	(0.010)*	(0.008)***	(0.010)*
N	122285	101698	122285	101698
Adj. R2	0.32	0.62	0.32	0.62
$\alpha_{c,p}$	N	Y	N	Y

This table estimates specification (9) for 10 years. The upper panel estimates the specification for the extensive margin and the lower panel does so for the intensive margin. Columns 1 and 2 use as independent variable quality-based density that is orthogonal to the agnostic HK relatedness measure. Conversely, Columns 3 and 4 use as independent variable quality-based density that is orthogonal to the agnostic EG relatedness measure. All specifications include country-by-year and product-by-year fixed effects, with columns 2 and 4 controlling additionally for country-product fixed effects. Control variables are not shown to simplify clarity. Standard errors are clustered at the country level and presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A3: Intensive Margin, Logarithmic Growth

Dependent Variable: Logarithmic Annual Export Value Growth Rate				
	HK		EG	
	(1)	(2)	(3)	(4)
	est1	est2	est3	est4
Φ_{cpt}	0.6327	0.5040	0.3939	0.3389
	(0.079)***	(0.098)***	(0.056)***	(0.105)***
Baseline Level of Exports (log)	-0.0403	-0.1389	-0.0374	-0.1382
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Growth Rate	-0.0030	0.0061	-0.0039	0.0057
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Zero Exports	-0.0899	-0.0235	-0.0877	-0.0202
	(0.008)***	(0.010)**	(0.008)***	(0.010)**
N	122285	101698	122285	101698
Adj. R2	0.32	0.62	0.32	0.62
$\alpha_{c,p}$	N	Y	N	Y

This table estimates specification (4) for the intensive margin with log-growth of exports as the dependent variable. Columns 1 and 2 estimate results using HK proximity, and columns 3 and 4 use EG proximity as the main input for the right hand side variable of interest. All specifications include country-by-year and product-by-year fixed effects, with columns 2 and 4 controlling additionally for country-product fixed effects Standard errors are clustered at the country level and presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A4: Emergence and growth of related industries, defined channels, agnostic HK controls

Panel A: Extensive Margin							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	est1	est2	est3	est4	est5	est6	est7
$\Phi_{c,p,t}^{Incoming\ R\&D}$	0.1985						-0.0009
	(0.167)						(0.165)
$\Phi_{c,p,t}^{Outgoing\ R\&D}$		0.3418					0.0573
		(0.195)*					(0.210)
$\Phi_{c,p,t}^{Patents}$			0.4339				0.5203
			(0.205)**				(0.328)
$\Phi_{c,p,t}^{Labor}$				0.1259			-0.3984
				(0.167)			(0.217)*
$\Phi_{c,p,t}^{Slinkages}$					0.0864		-0.0979
					(0.105)		(0.146)
$\Phi_{c,p,t}^{CLinkages}$						0.6089	0.5824
						(0.176)***	(0.194)***
$\Phi_{c,p,t}(HK)$	0.2716	0.2728	0.2524	0.2711	0.2767	0.2482	0.2637
	(0.138)*	(0.138)*	(0.141)*	(0.143)*	(0.139)**	(0.139)*	(0.143)*
Baseline RCA (log)	0.1054	0.1056	0.1031	0.1059	0.1059	0.1029	0.1018
	(0.037)***	(0.037)***	(0.037)***	(0.037)***	(0.037)***	(0.037)***	(0.037)***
N	69090	69090	69090	69090	69090	69090	69090
Adj. R2	0.07	0.07	0.07	0.07	0.07	0.07	0.07
$\alpha_{c,p}$	Y	Y	Y	Y	Y	Y	Y
Panel B: Intensive Margin							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	est1	est2	est3	est4	est5	est6	est7
$\Phi_{c,p,t}^{Incoming\ R\&D}$	1.0491						0.3333
	(0.199)***						(0.250)
$\Phi_{c,p,t}^{Outgoing\ R\&D}$		1.2344					0.4858
		(0.222)***					(0.254)*
$\Phi_{c,p,t}^{Patents}$			1.5373				1.1053
			(0.295)***				(0.391)***
$\Phi_{c,p,t}^{Labor}$				1.0634			0.1200
				(0.239)***			(0.320)
$\Phi_{c,p,t}^{Slinkages}$					0.4731		-0.3062
					(0.198)**		(0.180)*
$\Phi_{c,p,t}^{CLinkages}$						1.0665	0.2445
						(0.261)***	(0.254)
$\Phi_{c,p,t}(HK)$	0.4324	0.4356	0.3882	0.3892	0.4636	0.4328	0.3681
	(0.093)***	(0.095)***	(0.091)***	(0.091)***	(0.097)***	(0.093)***	(0.089)***
Baseline Level of Exports (log)	-0.1396	-0.1397	-0.1398	-0.1396	-0.1395	-0.1396	-0.1398
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Growth Rate	0.0067	0.0068	0.0068	0.0068	0.0067	0.0068	0.0069
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Zero Exports	-0.0301	-0.0305	-0.0308	-0.0304	-0.0300	-0.0303	-0.0309
	(0.010)***	(0.010)***	(0.010)***	(0.010)***	(0.010)***	(0.010)***	(0.010)***
N	90798	90798	90798	90798	90798	90798	90798
Adj. R2	0.62	0.62	0.62	0.62	0.62	0.62	0.62
$\alpha_{c,p}$	Y	Y	Y	Y	Y	Y	Y

This table estimates specification (4) with the channel-specific relatedness measures for 10 years, controling for the agnostic HK density measure. The upper panel estimates the specification for the extensive margin and the lower panel does so for the intensive margin. Columns 1 to 5 evaluate the impact of each channel-specific relatedness measure on export emergence and export growth separately. Columns 6 and 7 submit all measures jointly. All specifications include country-by-year, product-by-year and product-by-country fixed effects. Standard errors are clustered at the country level and presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A5: Emergence and growth of related industries, defined channels, agnostic EG controls

Panel A: Extensive Margin							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	est1	est2	est3	est4	est5	est6	est7
$\Phi^{Incoming\ R\&D}_{c,p,t}$	0.1566						-0.0283
	(0.167)						(0.166)
$\Phi^{Outgoing\ R\&D}_{c,p,t}$		0.3426					0.1170
		(0.201)*					(0.200)
$\Phi^{Patents}_{c,p,t}$			0.3959				0.5548
			(0.189)**				(0.332)*
$\Phi^{Labor}_{c,p,t}$				0.0656			-0.4724
				(0.153)			(0.218)**
$\Phi^{Slinkages}_{c,p,t}$					0.0596		-0.1038
					(0.106)		(0.148)
$\Phi^{CLinkages}_{c,p,t}$						0.5645	0.5467
						(0.169)***	(0.190)***
$\Phi_{c,p,t}(EG)$	0.3555	0.3570	0.3369	0.3588	0.3607	0.3284	0.3551
	(0.088)***	(0.090)***	(0.086)***	(0.089)***	(0.089)***	(0.088)***	(0.089)***
Baseline RCA (log)	0.1075	0.1074	0.1050	0.1082	0.1080	0.1048	0.1040
	(0.037)***	(0.037)***	(0.037)***	(0.038)***	(0.038)***	(0.037)***	(0.037)***
N	69090	69090	69090	69090	69090	69090	69090
Adj. R2	0.07	0.07	0.07	0.07	0.07	0.07	0.07
$\alpha_{c,p}$	Y	Y	Y	Y	Y	Y	Y

Panel B: Intensive Margin							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	est1	est2	est3	est4	est5	est6	est7
$\Phi^{Incoming\ R\&D}_{c,p,t}$	1.1764						0.3891
	(0.211)***						(0.250)
$\Phi^{Outgoing\ R\&D}_{c,p,t}$		1.3442					0.5074
		(0.218)***					(0.252)**
$\Phi^{Patents}_{c,p,t}$			1.6828				1.1083
			(0.302)***				(0.385)***
$\Phi^{Labor}_{c,p,t}$				1.2043			0.2115
				(0.244)***			(0.320)
$\Phi^{Slinkages}_{c,p,t}$					0.5508		-0.2874
					(0.199)***		(0.179)
$\Phi^{CLinkages}_{c,p,t}$						1.1676	0.2592
						(0.267)***	(0.254)
$\Phi_{c,p,t}(EG)$	0.2537	0.2626	0.1856	0.1838	0.2840	0.2428	0.1596
	(0.102)**	(0.104)**	(0.100)*	(0.102)*	(0.105)***	(0.102)**	(0.101)
Baseline Level of Exports (log)	-0.1390	-0.1391	-0.1392	-0.1391	-0.1389	-0.1390	-0.1393
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Growth Rate	0.0064	0.0065	0.0065	0.0065	0.0064	0.0064	0.0066
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Previous Period Zero Exports	-0.0272	-0.0277	-0.0280	-0.0277	-0.0271	-0.0274	-0.0283
	(0.010)***	(0.010)***	(0.010)***	(0.010)***	(0.010)***	(0.010)***	(0.010)***
N	90798	90798	90798	90798	90798	90798	90798
Adj. R2	0.62	0.62	0.62	0.62	0.62	0.62	0.62
$\alpha_{c,p}$	Y	Y	Y	Y	Y	Y	Y

This table estimates specification (4) with the channel-specific relatedness measures for 10 years, controling for the agnostic EG density measure. The upper panel estimates the specification for the extensive margin and the lower panel does so for the intensive margin. Columns 1 to 5 evaluate the impact of each channel-specific relatedness measure on export emergence and export growth separately. Columns 6 and 7 submit all measures jointly. All specifications include country-by-year, product-by-year and product-by-country fixed effects. Standard errors are clustered at the country level and presented in parenthesis.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table A6: Relatedness, agnostic HK

$SITC_i$	**Name**	$SITC_j$	**Name**
Ranking by Relatedness, Top 15			
8423	Men's trousers	8439	Other women outerwear
8459	Other knitted outerwear	8462	Knitted undergarments of cotton
8434	Skirts	8439	Other women outerwear
8433	Dresses	8452	Knitted women's suits & dresses
8433	Dresses	8435	Blouses
8439	Other women outerwear	8462	Knitted undergarments of cotton
8439	Other women outerwear	8459	Other knitted outerwear
7764	Electronic microcircuits	7768	Parts N.E.S. of electronic circuits
8439	Other women outerwear	8441	Men's undershirt
7361	Metal cutting machine-tools	7368	Dividing heads for machine-tools
8435	Blouses	8452	Knitted women's suits & dresses
1212	Wholly or partly stripped tobacco	1213	Tobacco refuse
8451	Knitted jerseys, pullovers & cardigans	8459	Other knitted outerwear
8423	Men's trousers	8459	Other knitted outerwear
2874	Lead ore	2875	Zinc

This table shows the 15 most related product pairs, based on HK's measure of agnostic relatedness, which is the minimum probability of co-exporting two given products.

Table A7: Relatedness, agnostic EG

$SITC_i$	**Name**	$SITC_j$	**Name**
Ranking by Relatedness, Top 15			
2655	Manila hemp	2659	Vegetable textile fibres N.E.S.
2613	Raw Silk	8994	Umbrellas & canes
4245	Castor oil	6545	Jute woven fabrics
2655	Manila hemp	4243	Coconut oil
8933	Plastic ornaments	8994	Umbrellas & canes
2613	Raw Silk	6597	Plaited products
6597	Plaited products	8994	Umbrellas & canes
2613	Raw Silk	8933	Plastic ornaments
2714	Crude natural potassium salts	2784	Asbestos
6597	Plaited products	8933	Plastic ornaments
2613	Raw Silk	8942	Toys
8942	Toys	8994	Umbrellas & canes
4245	Castor oil	6593	Kelem, schumacks & karamanie
6583	Travelling rugs & blankets	8994	Umbrellas & canes
2613	Raw Silk	8999	Manufactures N.E.S.

This table shows the 15 most related product pairs, based on EG's measure of agnostic relatedness, which is the co-location of export industries in the same country of origin.

Table A8: Relatedness, R&D Linkages

$SITC_i$	**Name**	$SITC_j$	**Name**
Ranking by Relatedness, Top 15			
3231	Solid fuels	3413	liquified hydrocarbons
3351	Petroleum jelly & mineral waxes	3413	liquified hydrocarbons
3231	Solid fuels	5111	Acyclic hydrocarbons
3351	Petroleum jelly & mineral waxes	5111	Acyclic hydrocarbons
3231	Solid fuels	3354	Petroleum bitumen N.E.S.
3351	Petroleum jelly & mineral waxes	3354	Petroleum bitumen N.E.S.
3354	Petroleum bitumen N.E.S.	3413	liquified hydrocarbons
3354	Petroleum bitumen N.E.S.	5111	Acyclic hydrocarbons
3413	liquified hydrocarbons	5111	Acyclic hydrocarbons
3231	Solid fuels	3345	Lubricating petroleum oils N.E.S.
3345	Lubricating petroleum oils N.E.S.	3413	liquified hydrocarbons
3345	Lubricating petroleum oils N.E.S.	5111	Acyclic hydrocarbons
3345	Lubricating petroleum oils N.E.S.	3354	Petroleum bitumen N.E.S.
7761	T.V. tubes & cathode rays	7768	Parts N.E.S. of electronic circuits
7762	Electronic valves & tubes	7763	Diodes & transistors

This table shows the 15 most related product pairs, based on the share of R&D expenditure in one industry used by another industry.

Table A9: Relatedness, Labor Linkages

$SITC_i$	**Name**	$SITC_j$	**Name**
Ranking by Relatedness, Top 15			
1221	Cigars	1222	Cigarretes
1221	Cigars	1223	Tobbacco, extract, essences & manufactures
1222	Cigarretes	1223	Tobbacco, extract, essences & manufactures
6413	Rolls/sheets of kraft paper	6417	Rolls/sheets of creped paper
6413	Rolls/sheets of kraft paper	6419	Converted paper N.E.S.
6417	Rolls/sheets of creped paper	6419	Converted paper N.E.S.
2518	Chemical wood pulp, sulphite	6415	Paper & paperboad in rolls or sheets
2519	Other cellulosic pulps	6412	Printing & writing paper in rolls or seets
2517	Chemical wood pulp, soda or sulphate	6412	Printing & writing paper in rolls or seets
6411	Newsprint	6415	Paper & paperboad in rolls or sheets
6412	Printing & writing paper in rolls or seets	6413	Rolls/sheets of kraft paper
6411	Newsprint	6422	Correspondence stationary
2512	Mechanical wood pulp	6411	Newsprint
2516	Chemical wood pulp, dissolving grades	6412	Printing & writing paper in rolls or seets
2516	Chemical wood pulp, dissolving grades	6415	Paper & paperboad in rolls or sheets

This table shows the 15 most related product pairs, based on labor flows industry pairs.

Table A10: Relatedness, Patent Linkages

$SITC_i$	**Name**	$SITC_j$	**Name**
Ranking by Relatedness, Top 15			
7931	Warships	7932	Ships & boats
5411	Provitamins & vitamins	5414	Vegetable alkaloids & derivatives
5411	Provitamins & vitamins	5416	Glycosides & vaccines
5414	Vegetable alkaloids & derivatives	5415	Bulk hormones
5415	Bulk hormones	5416	Glycosides & vaccines
5411	Provitamins & vitamins	5413	Antibiotics
5414	Vegetable alkaloids & derivatives	5416	Glycosides & vaccines
5413	Antibiotics	5414	Vegetable alkaloids & derivatives
5413	Antibiotics	5416	Glycosides & vaccines
5411	Provitamins & vitamins	5415	Bulk hormones
5413	Antibiotics	5415	Bulk hormones
5411	Provitamins & vitamins	5417	Medicaments
5416	Glycosides & vaccines	5417	Medicaments
5414	Vegetable alkaloids & derivatives	5417	Medicaments
5413	Antibiotics	5417	Medicaments

This table shows the 15 most related product pairs, based on patent citations between industry pairs.

Table A11: Relatedness, Customer Linkages

$SITC_i$	**Name**	$SITC_j$	**Name**
Ranking by Relatedness, Top 15			
112	Sheep & goat meat	116	Bovine & equine entrails
114	Poultry meat	121	Other animal entrails
118	Other animal meats	2117	Raw sheep skin with wool
112	Sheep & goat meat	115	Equine meat
112	Sheep & goat meat	113	Swine meat
115	Equine meat	2114	Raw goat skins
121	Other animal entrails	2911	Bones, horns, corals & ivory
115	Equine meat	118	Other animal meats
111	Bovine meat	115	Equine meat
113	Swine meat	2911	Bones, horns, corals & ivory
142	Sausages	2114	Raw goat skins
116	Bovine & equine entrails	2116	Raw sheep skin without wool
111	Bovine meat	2114	Raw goat skins
111	Bovine meat	113	Swine meat
111	Bovine meat	116	Bovine & equine entrails

This table shows the 15 most related product pairs, based on upstream linkages.

Table A12: Relatedness, Supplier Linkages

$SITC_i$	**Name**	$SITC_j$	**Name**
Ranking by Relatedness, Top 15			
7851	Motorcycles	7852	Bicycles
7831	Public transportation vehicles	7832	Tractors for semi-trailers
7810	Cars	7831	Public transportation vehicles
7810	Cars	7832	Tractors for semi-trailers
7911	Electric trains	7914	Not mechanically propelled railway for passengers
7914	Not mechanically propelled railway for passengers	7915	Not mechanically propelled railway for freight
7912	Rail tenders	7915	Not mechanically propelled railway for freight
7912	Rail tenders	7914	Not mechanically propelled railway for passengers
7913	Mechanically propelled railway	7914	Not mechanically propelled railway for passengers
7912	Rail tenders	7913	Mechanically propelled railway
7911	Electric trains	7915	Not mechanically propelled railway for freight
7913	Mechanically propelled railway	7915	Not mechanically propelled railway for freight
7911	Electric trains	7912	Rail tenders
7911	Electric trains	7913	Mechanically propelled railway
7931	Warships	7932	Ships & boats

This table shows the 15 most related product pairs, based on downstream linkages.