CESifo WORKING PAPERS

7436
2018

December 2018

Export Take-Offs and Acceleration: Unpacking Cross-Sector Linkages in the Evolution of Comparative Advantage

Dany Bahar, Samuel Rosenow, Ernesto Stein, Rodrigo Wagner

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7436*
Category 8: Trade Policy

Export Take-Offs and Acceleration: Unpacking Cross-Sector Linkages in the Evolution of Comparative Advantage

Abstract

The transition into non-traditional export activities attracts important policy and academic attention. Using international trade data, we explore how alternative linkages relate to the take-off and acceleration of export industries. Concretely, we run a horse-race among alternative Marshallian linkages across sectors: input-output relations, technology and labor. Technology has a predictive power depending on the specification used. We consistently find, however, that export take-offs are more likely to occur in sectors that are upstream to already competitive export industries. Our findings, which are mostly driven by developing economies, are consistent with Albert Hirschman's 60-year-old view that the forces behind upstream linkages fueled the growth of new competitive industries in the developing world.

JEL-Codes: O140, O330, F140.

Keywords: relatedness, comparative advantage, patents, labor, upstream, downstream.

Dany Bahar
Brookings Institution
Washington DC / USA
db21@post.harvard.edu

Samuel Rosenow
UNCTAD
Geneva / Switzerland
samuel.rosenow@googlemail.com

Ernesto Stein
IADB
Washington DC / USA
ernestos@iadb.org

*Rodrigo Wagner**
Universidad de Chile
Santiago / Chile
Rodrigo_Wagner@post.harvard.edu

*corresponding author

Latest Update: December 21, 2018
Authors acknowledge helpful comments from seminars at Harvard Business School, the Brookings Institution, the Inter-American Development Bank, George Washington University, LACEA Annual Meetings 2015, University of Chile, and Harvard's Center for International Development. All errors are our own. The views expressed by the authors in this paper do not necessarily reflect the opinions of their institutional affiliations.

1 Introduction

The transition into non-traditional export activities attracts important policy and academic attention.[1] By and large, the classical literature on international trade tends to focus on studying country and industry specific characteristics as the main determinants of comparative advantage, perceived to be independent of the existence of other industries, beyond the obvious general equilibrium effects.[2] Yet, at least since Hirschman (1958)[3], scholars have recognized that the evolution of competitive industries may evolve over time through *linkages* to other existing economic activities. Our paper adds to this literature by analyzing empirically how alternative types of linkages can explain dynamics in a country's export basket.

A large body of empirical literature shows that the emergence and growth of economic activities relates to the presence of incumbent industries in the same unit: a country (e.g., Hausmann and Klinger, 2006; Hausmann and Hi-

[1] In the academic literature, the diversification towards new export industries is commonly associated with economic development, macroeconomic stability as well as market access. 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). Further, Rodrik (2016) argues that reductions in diversification, like those induced by premature deindustrialization, can jeopardize the process of economic development, for example, because they prevent unconditional convergence in manufacturing (Rodrik, 2012). In this vein, Hartmann et al. (2017) show that export complexity can be associated with decreases in inequality. For important reviews of this policy debate see Hirschman (1968) and Rodrik (2008). For effects on macroeconomic stability, see for example, Krishna and Levchenko (2009); Koren and Tenreyro (2007); Caselli et al. (2018); Hausmann et al. (2006). For links between diversification and market access see Nicita and Rollo (2015); Hoekman and Nicita (2011); Fugazza and Nicita (2013).

[2] This 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). It typically takes the evolution of the comparative advantage of a single country-industry as independent of other industries within the same country.

[3] The "old" empirical literature trying to measure these Hirschmanian linkages, including a full special issue of the *Quarterly Journal of Economics in 1976*, focused more on the effects on growth across countries (Jones, 1976). This literature typically included few controls to account for alternative hypotheses, beyond linkages.

dalgo, 2011; Hidalgo et al., 2007; Boschma and Capone, 2015), a sub-national region (e.g., Ellison and Glaeser, 1997; Neffke et al., 2011; Delgado et al., 2016; Balland et al., 2018) or even a firm (e.g., Breschi et al., 2003; Fan and Lang, 2000). Papers also explore linkages based on traditional channels discussed by Marshall (1920): similarity in technologies (e.g., Scherer, 1984; Boschma and Capone, 2015; Breschi et al., 2003), similarity in workers and labor skills (e.g., Farjoun, 1994; Neffke and Henning, 2013; Neffke et al., 2017), or input-output relations (e.g., Fan and Lang, 2000).[4] While this literature explores various measures of relatedness across industries, its contributions tend to focus on studying one channel at the time. Recently, a large group of multi-disciplinary scholars working in this area highlighted the importance of unpacking the phenomenon of relatedness to understand the various channels behind it (Hidalgo et al., 2018). Our paper, precisely, fits into this call.

Our contribution focuses on understanding the relative importance of cross-industry linkages through various channels in explaining take-offs and accelerations of export industries at a global scale. In particular, we simultaneously study three channels in the same setting: input-output relations, pooling of workers and the sharing of technology and knowledge. Our study is conceptually similar to that of Ellison et al. (2010) who explore the role of different Marshallian linkages to study industry agglomeration in US regions, and that of Delgado et al. (2016) who do so with the purpose of defining regional clusters.[5] Yet, the focus of our paper is neither on production nor co-agglomeration of industries or firms within a country. Rather, our focus is on understanding how these linkages explain the evolution of comparative advantage, as measured by export dynamics across nations.[6]

[4]Recently, some of studies have looked at the role of workforce linkages on pioneer firms that enter new economic activities (e.g., Hausmann and Neffke, 2018; Jara-Figueroa et al., 2018).

[5]A related study is Farinha-Fernandes et al. (2018) who unpack relatedness between occupations to understand the evolution of the occupational structure of US urban areas.

[6]Our results are exclusively based on exports and not production data. Thus, when

When it comes to exports, there are several explanations for how the presence of related competitive industries can fuel take-offs or acceleration of export industries. First, 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. Second, the existence of competitive industries using a trained (and competitive) workforce similar to that required by a non-exported sector might play a role in explaining the take-off of the latter. Third, the existence of competitive industries producing goods that are intermediate inputs to sectors of a yet small or non-existent industry could help the latter to develop and become an exporter. Alternatively, the existence of a critical mass of firms in certain exporting industries could "pull" the development of a new upstream export sector.[7]

Our empirical analysis uses panel regressions of country-industry observations with multiple sets of interacted fixed effects. This allows us to control for particularities of a country in a sector, like natural advantages. It also controls for global trends in a sector as well as macroeconomic phenomena varying in each country every year, each industry every year, as well as country-industry time-invariant characteristics (such as fundamental comparative advantage determinants). Our measures of inter-industry linkages are taken from Ellison et al. (2010) and Greenstone et al. (2010), which are based on US data. Given the global character of our sample, which includes 114 countries, using US-specific data helps us to diminish concerns of biases due to endogeneity in our empirical approach.[8]

we refer to certain sectors as inexistent throughout the paper, unless otherwise noted, we refer to sectors that are not being exported competitively. That does not mean that such sector does not have domestic production. Conversely, when referring to sectors that do exist, they exhibit a RCA above 1 in the country's export basket, as described in section 2.2, and we refer to them as competitive sectors.

[7]Unlike specialized cross-industry demands, the effect of a sector-wide critical mass that shifts *all* other sectors (e.g. Krugman, 1991) is ultimately neutralized in our empirical approach, as we extract all country-year variation with our fixed effects.

[8]The identification assumption to interpret our findings as causal requires that the structure of the linkages in the US is related to the potential linkages in each country, but

Our findings highlight the significance of certain channels and downplay others. Our most salient result is that customer linkages support the take-off and acceleration of upstream export industries. In particular, a one standard deviation increase in the density of customer linkages increases by more than six times the probability of an export take-off in upstream sectors during the next decade. Under that same shock, the growth of existing exports expands 5 percentage points faster per year than the baseline. For example, a country is more likely to increase exports of fabric if it already exports garments. We also find that technology linkages matter: A one standard deviation increase in technology linkages across sectors make a related new export take-off almost ten times more likely , and is also associated with a subsequent additional annual export growth of 15 extra percentage points over the next decade.

However, for the average country in our sample, we find no robust evidence that labor linkages explain the evolution of comparative advantage. Similarly, in the same sample, the existence of supplier linkages do not tend to predict the emergence of new downstream export sectors.

It is worth noting that developing countries drive our findings on the relevance of customer linkages. This is consistent with Hirschman (1958), who suggested that backward linkages tend to be an important and natural path for new economic activities in developing countries. Hirschman's argument was that a customer industry helps build a larger market size for supplier firms while offering specific know-how on how to produce more competitively. Intuitively, the existing downstream firm has incentives to help in the development of local competitive procurement as a way to reduce its costs. All these elements might reduce risks on the side of the suppliers, incentivizing them to invest more in boosting competitiveness. A series of

orthogonal to the unobserved heterogeneity. This is somewhat similar to the approach taken by Ellison et al. (2010), who use the linkages based on UK data to instrument for US ones. While we acknowledge that our identification strategy is not perfect, we believe that our findings –even if taken as suggestive evidence– add to this long literature.

recent micro-level studies back up the evidence that firms "learn" from their customers. Javorcik (2004) finds evidence of productivity spillovers for firms that sell to new foreign plants located in Eastern Europe.[9] Consistently, Blalock and Veloso (2007) find that Indonesian firms competing with foreign suppliers of local producers were more likely to experience productivity gains. Also, Pietrobelli and Saliola (2008) find that selling to a multinational firm enhances supplier productivity. More recently, Kee and Tang (2016) show that China gained comparative advantage in those intermediate inputs used by pre-existing Chinese exporters. Our work complements these and other micro-level studies on backward linkages by exploring this mechanism in a global sample.[10]

We acknowledge that our findings are not necessarily causal, but they reveal systematic relationships between industries and across countries, in a research area that was mostly informed by case-by-case micro evidence. Our paper contributes to a growing literature that explores the relative importance of alternative channels behind the evolution of related economic activity, with our focus being on the dynamics of comparative advantage –as measured through export performance– in a global setting. Beyond the literature on cross-industry linkages summarized above, our study also contributes to a literature studying the broader role of the evolution o f industry competitiveness at different stages of economic growth and development (e.g., Imbs and Wacziarg, 2003; Cadot et al., 2011).

The remaining of the paper is structured as follows. Section 2 explains data sources and variable definitions, notably the various measures of density

[9] One possibility is that they mitigate the risks related to self-discovery costs. For evidence on self discovery costs in export, see Hausmann and Rodrik (2003). Javorcik (2004) explores backward linkages from FDI, not in exports. Other recent articles have focused on how FDI can help change comparative advantage (Amighini and Sanfilippo, 2014; Lectard and Rougier, 2018). Our work complements the previous analysis by showing the type of industry network that is most robustly associated to export take-offs.

[10] It is important to clarify that 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), but rather from structural changes in comparative advantage.

around each sector. Section 3 is the core of our paper: it reports results based on regression analysis on how the various linkages predict the take-off and growth of export industries. Section 4 offers a series of robustness tests. Finally, Section 5 concludes and discusses the implications of our results.

2 Data: New Exports and Relatedness

2.1 Data sources and basic procedures

Our empirical exercise requires combining international export data with measures of relatedness across industries. Thus, we need two main inputs: trade data and measures of relatedness across industries. Then we need a concordance to combine them.

Trade data

For bilateral exports we use UN COMTRADE 1984-2014 data with the adjustments indicated by Hausmann et al. (2014), with 786 sectors defined at 4-digit SITC revision 2.[11] Each 4 digit code represents a very specific export sector. Two examples of 4-digit SITC codes are "*Knitted/Crocheted Fabrics Elastic or Rubberized*" (SITC 6553), or "*Electrical Measuring, Checking, Analyzing Instruments*" (SITC 8748).

Following Hausmann et al. (2014), we exclude countries below 1 million citizens and total trade below US $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 of any sector for a particular year. This leaves us with 114 countries to construct the total value of exports per sector and country to the rest of the world for each year. The sample represents over 90% of world trade.

[11] We use product, good, sector and industry interchangeably throughout the paper. But given the level of granularity we use, they should be interpreted as sectors.

Marshallian relations among industries

The channel-based relations or linkages across industries come from Ellison et al. (2010), except for the labor flow measure taken from Greenstone et al. (2010). 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 prefer using these off-the shelf measures, already validated in the literature. An advantage of using these measures is that we follow the standard practice of using US data only to compute relatedness, rather than data from each individual country, to be used in a global sample (e.g., Romalis, 2004; Rajan and Zingales, 1998). This has a number of advantages. First, it bypasses the impossibility of using input-output relations from countries in which some of the industries have not yet emerged. Second, our measures are not affected by potential distortions in each economy. Third, and related to the previous point, using relatedness in the US yields should alleviate concerns of endogeneity in our empirical approach.

The measure of technological relatedness used by Ellison et al. (2010) comes from citation patterns in the NBER patent database (for 1975 to 1997). It captures the fraction of patents developed in industry i that cite patents developed in each industry j, taking the average of the bi-directional fractions between i and j.

The measures for supplier and customer linkages, also used by Ellison et al. (2010), came from the Bureau of Economic Analysis's Benchmark Table of Input-Output Accounts of 1987. That is, the fraction of each industry i's inputs purchased from industry j (a.k.a. supplier linkages); as well as the fraction of i's output that is sold to industry j (a.k.a. customer linkages). These relations were aggregated at the SIC3 digit level.

The measure of labor relatedness comes from Greenstone et al. 2010. It is based on the US Current Population Survey's, an outgoing rotation file published by the Bureau of Labor Statistics. It captures the fraction of workers separating from industry i that move to firms in industry j within

15 months.[12]

We normalize all these channel-specific measures to represent percentiles in the distribution of relatedness.[13] This normalization ensures that the channel-specific relatedness measures have the same range as the geographic-based (agnostic) relatedness measures, as explained below. For all measures higher relatedness means a closer relationship between sectors. See Online Appendix E for the top fifteen sector pairs in terms of relatedness for each one of these measures.

Concordance

The data on technological and input-output relatedness uses the SIC 3-digit industry classification, while the data on labor relatedness uses SIC 2-digit classification. In order to work at a comparable level of aggregation as our trade data, we match the cross-industries relatedness measures to SITC sectors following the approach described by Cuñat and Melitz (2012) in their footnote 24 and their subsequent documentation.[14]

[12]Note that the initial year of our regressions would be 1990. Thus the time period upon which these relatedness measures are constructed are appropriate as baseline for our exercise.

[13]For example, a relatedness of 0.3 implies that those two sectors are in the 30th percentile of relatedness.

[14]Cuñat and Melitz (2012) argue that "*[s]ince publicly available concordances from SITC rev.2 to US SIC do not indicate proportions on how individual SITC codes should be allocated to separate SIC codes, we construct our own concordance. We use export data for the United States, which is recorded at the Harmonized System (HS) level (roughly 15,000 product codes). For each HS code, both an SITC and an SIC code is listed. We aggregate up the value of US exports over all HS codes for the last ten available data years (1991–2000) across distinct SITC and SIC pairs. For each SITC code, we record the percentage of US exports across distinct SIC codes. We then concord exports for all countries from SITC to SIC codes using these percentage allocations. In most cases, this percentage is very high, so our use of US trade as a benchmark cannot induce any serious biases. For 50% of SITC codes, the percentage assigned to one SIC code is above 98%. For 75% of SITC codes, this percentage is above 76%.*" For the purposes of our work it is important to mention that, despite the differences in the aggregation levels of the export data (4-digit) and the relatedness data (3-digit), we maintain sufficient variation across industries when converting the latter to the 4-digit level. Using a coarser granularity of

Note that the Marshallian linkages are only available for manufacturing industries, and therefore, our analysis is limited to those (and not the whole set of SITC 4-digit sectors), as in Ellison et al. (2010). This makes us "lose" about 15 percent of the export sectors from the SITC exports data.

2.2 Variables construction

Dependent variables: take-off and acceleration of export sectors

We construct two dependent variables to explore the take-off and acceleration of export sectors. Generically labeled $Y_{c,p,t\to T}$, these quantify the *evolution* of exports and comparative advantage of a country in an industry between years t and T.

First, like many others (e.g., Hausmann and Klinger, 2006; Hidalgo et al., 2007; Boschma and Capone, 2015), we use the Revealed Comparative Advantage index (RCA) by Balassa (1965) as a main input to many of our measures, including the *take-off* of exports. RCA captures the share of a given industry in the country's total exports, divided by the share of the same sector in world's exports:

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

where $x_{c,p,t}$ is total export value of industry p from country c to the world in year t. [15]

We define the *take-off* of an export sector with a dummy variable that

trade data like 2-digit, on the other hand, would make us lose a lot of the variation. While in theory we could aggregate exports data to the 3-digit level, we would still suffer from an imperfect match. We therefore opted, consistent with other studies, to use 4-digit disaggregation levels in our study.

[15] 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.

takes the value one if the sector started with $RCA < 0.1$ in the initial year and ended with RCA >1 in the final year T. Formally,

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

While the choice of 0.1 is somewhat arbitrary, we choose it to include industries that are significantly small (in relative terms).[16] Thus, when studying take-offs we exclude from the regression all the sectors in which the country already had $RCA \geq 0.1$ in the initial year. We do so to avoid that export take-offs result from small improvements in competitiveness for a given country-sector throughout the decade, which could be explained by idiosyncratic reasons. Moreover, our definition of take-offs adds two extra conditions to address noise in the data. First, we ensure that the initial RCA is not transitorily low: we require $RCA \leq 0.1$ from $t-2$ to t. Second, we also define that the $RCA \geq 1$ condition is sustained for at least two years after the end of the decade (e.g., from T to $T+2$). These conditions aim to avoid confusions stemming from intermittent exporting (Bernini et al., 2016).

When studying accelerations of exports, we define our dependent variable $Y_{c,p,t\to T}$ as the compound annual growth rate (CAGR) in the export value, conditional on having positive exports initially $x_{c,p,t} > 0$.[17] That is:

[16] However, note that more than 60% of take-off events correspond to cases in which exports started the period with zero or almost zero (e.g. below USD $10,000) exports. While we cannot claim that this measures the emergence of a new sectors (given that it includes many sectors that, albeit small, already existed), we perform some robustness tests in Section 4 that use smaller initial thresholds and the main results remains robust.

[17] We note that the two measures are not completely mutually exclusive. There is a potential overlap in the support $0 < RCA_t < 0.1$. For our central point this is not a problem. While we want to make a distinction between take-off and growth of a new export industry, the reality is more continuous. In any case, most of the variation did not come from the overlapping part, otherwise we would have always the same results using both dependent variables, which is not the case. Overall, more than 60% of take-off events correspond to cases in which exports started the period with zero or almost zero (e.g. below USD $10,000) value.

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

A clarification of our two dependent variables may help interpret the results more easily. A take-off event implicitly relates to export diversification, because its occurrence potentially results in a larger share of a previously under represented export sector in the country's export basket.[18] When it comes to growth, since our estimations always use country-year, product-year and country-industry fixed effects, the results could also be understood as reflecting changes in competitiveness of that export sector.

To ensure that our results are not coming from overlapping observations, our analysis uses decade-long changes: from 1990 to 2000 and from 2000 to 2010. We prefer to take ten year periods to estimate the effect of medium to long term processes, as structural transformation is not an overnight process.

2.2.1 Independent variables

To construct variables that measure the extent to which an industry is exposed to other existing ones through cross-industry linkages, we construct "density" measures. We explain them step by step.

Co-location relatedness based on co-location of exports

On top of the Marshallian linkages explained above, we also construct a now standard cross-sector linkage based on co-location patterns. For this channel-agnostic measure we follow Hausmann and Klinger (2006), HK here onwards. It computes the probability that two sectors are co-exported competitively from the same country.[19] We follow HK, so the proximity variable $\varphi_{i,j}^{HK}$ uses

[18] We carefully use the word *potentially* because this will, of course, depend on the final distribution of all the shares of export sectors within the export basket of the country. If the rise of a sector is completely out-weighted by the decline in another, we would not be able to see diversification as measured by traditional concentration indices.

[19] The measure was also used in the seminal work of Hidalgo et al. (2007)

the minimum of the two conditional probabilities between pairs of industries.

$$\varphi_{i,j}^{HK} = \min\left\{ \Pr(RCA_i \geq 1 | RCA_j \geq 1) \quad ; \quad \Pr(RCA_j \geq 1 | RCA_i \geq 1) \right\} \tag{3}$$

Like Hidalgo et al. (2007), $\varphi_{i,j}^{HK}$ considers $RCA \geq 1$ as the threshold to define whether a sector is competitively exported in a particular country and year. The proximity variable $\varphi_{i,j}^{HK}$ is always distributed between 0 and 1. As an alternative index of co-location, we adapt the co-agglomeration index of Ellison and Glaeser (1997) –EG hence onwards– to export data. We label it $\varphi_{i,j}^{EG}$. The definition of the EG approach is in Online Appendix A.

Both measures of relatedness are geography-based. Yet, we refer to them as "agnostic" measures: while they respond to co-location patterns within the same geographic unit (a country, in our case), they provide no understanding on *how or why* these two industries tend to be co-located in the same country.

Correlation of various relatedness measures

Table 1 displays the correlation matrix between all the relatedness measures we have described. Note that these relatedness measures are defined for each industry pair and are symmetrical. Correlations between all the relatedness measures are positive and statistically different from zero. The correlation between the agnostic or co-location measures, φ^{HK} and φ^{EG}, is 0.49. The correlations of the agnostic relatedness measures with the channel-defined relatedness measures range between 0.10 to 0.19.

[Table 1 about here.]

Density around a sector

Next, we use these agnostic and Marshallian relatedness measures among industries to define densities, following previous work (e.g., Hausmann and Klinger, 2006; Hidalgo et al., 2007)). Density quantifies the extent to which a particular sector p is related to other already existing competitive sectors in a given country and year. Formally, density is defined as:

$$\Phi_{c,p,t} = \frac{\sum_{j \neq p} \varphi_{p,j} \times R_{c,j,t}}{\sum_{j \neq p} \varphi_{p,j}} \tag{4}$$

where $R_{c,j,t} = 1$ is a dummy variable indicating whether a neighboring industry is present (i.e. $RCA_{c,j,t} \geq 1$) , and 0 otherwise. The neighborhood proximities $\varphi_{p,j}$ correspond to the relatedness measures discussed before. $\Phi_{c,p,t}$ distributes between 0 and 1.

Depending on the proximity $\varphi_{p,j}$ used, the interpretation of density $\Phi_{c,p,t}$ would vary. Using any of the Marshallian proximity measures $\varphi^{m}_{p,j}$ would yield a different density $\Phi^{m}_{c,p,t}$. For example, using proximity $\varphi^{Labor}_{p,j}$ yields the *density* of Marshallian labor flows $\Phi^{Labor}_{c,p,t}$. A high value of $\Phi^{Labor}_{c,p,t}$ implies that country c in the initial year t was competitive in a large number of sectors which relate to sector p in terms of labor similarities. An analogous interpretation applies for all other density measures, including the agnostic ones.

2.3 Summary statistics of take-off and growth

Table 2 displays the summary statistics of our variables for the two ten-year changes. Panel A displays the take-off events sample (i.e., for all observations in a country, sector and year combination for which $RCA_{c,p,t} < 0.1$), while Panel B does so for the subgroup starts the period with positive exports ($x_{c,p,t} > 0$).

In our sample, a take-off event only happens in 0.6% of possible cases. That is, from all country-sector combinations that exhibit RCA below 0.1, on average, approximately one in 150 takes off within a decade. This means take-offs are unlikely events, on average. Existing export industries in a country grow at an average of 8 percent per year, in nominal US dollars. As a benchmark, US inflation in the 1990s and 2000s was 3 and 2.5 percent, respectively; suggesting that average real growth was around 5%.[20] Growth had massive heterogeneity, since the standard deviation was 22%; at least three times larger than average growth of exports.

Importantly, the relation between density and new exports is already apparent in the raw data, when we compare panels A and B in Table 2. Exported sectors, with an average density of about 0.19 (panel B), are twice as densely connected to other sectors than those of small and nonexistent exports in the sample for potential take-offs (average density is about 0.08 in panel A). This is true for all density measures we use, and even considering that the definition of density in equation (4) excludes the exporting of the own industry. This is a natural starting point to attempt to unpack the channels behind co-location. Note, as explained above, that the Marshallian linkages are available only for manufacturing industries.

The table also includes 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 four channel-specific ones. While the proximities of different channels were normalized, so their averages are uninteresting, the densities are also bound to be between zero and one, but they are not normalized. The density of a sector proxies for the existence of other sectors that share similar technologies, workers, customers or inputs. For example, values of $\Phi^{HK}_{c,p,t}$ closer to 1 indicate that a given sector 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 sector 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.

[Table 2 about here.]

[20]However, note that our regressions in the rest of the paper have country-year fixed effects (plus others), so the estimators can be interpreted in real, not nominal, terms.

3 Channels mediating take-off and growth of exports

This section, the core of our paper, presents findings on which Marshallian linkages are more strongly associated to changes in comparative advantage.[21]

3.1 Empirical model and main results

Our basic specification regresses a *change* in comparative advantage on different channel-based densities, namely:

$$Y_{c,p,t\to T} = \beta\Phi^{channel}_{c,p,t} + Controls_{p,c,t} + \eta_{c,t} + \delta_{p,t} + \alpha_{c,p} + \varepsilon_{c,p,t} \qquad (5)$$

The left hand side $Y_{c,p,t\to T}$ alternates between our two measures defined in Section 2.2: the binary take-off event and the continuous growth among existing exports. The right-hand-side densities $\Phi^{channel}_{c,p,t}$ measure the intensity with which industry p relates to the current export basket of the country. Thus, the main parameter of interest is the vector β, with one coefficient for each channel. The estimation method is a linear probability model for the binary variable and panel ordinary least squares (OLS) for growth of existing exports.

Importantly, we exploit the granularity of our data to control for all three feasible types of interacted fixed effects. This strongly reduces concerns about alternative explanations. First, $\eta_{c,t}$ represents country-by-decade effects capturing changes in income, institutions, exchange rates and population, among others. Second, $\delta_{p,t}$ represent industry-by-decade fixed effects, capturing variables like changes in global demand for industry p, common technological changes in an industry, among others. Importantly, we also include $\alpha_{c,p}$.

[21] To check the consistency with the existing literature, Section B in the Online Appendix confirms that the agnostic co-location densities predict changes in comparative advantage (Hausmann and Klinger, 2006; Hidalgo et al., 2007), even with our own strict set of fixed effects.

These fixed effects control for all possible country-industry interactions that might explain intrinsic comparative advantage driven by initial time invariant effects. For example, this controls for all the various alternative explanations taking the form of (industry intensity) $\times$ (country endowment), as in Romalis (2004) or Nunn (2007). It also captures most of the natural comparative advantage of a location in an industry.

Specification (5) includes additional controls that may vary by decade-long interval. These controls vary depending on whether we estimate the binary take-off or export growth. For the take-off of exports, we include the beginning of period RCA (e.g., $RCA_{c,p,t}$). For acceleration of exports, we control for the baseline level of exports. Both control for the convergence effect: growth is faster when starting from a lower base. For the binary take-off of exports, this relationship may be the other way around, since some exports, even if low, are predictive of more future exports as opposed to zero. Given the fat-tailed nature of these level-variables, we perform on them a log-like monotonic transformation: the inverse hyperbolic sine transformation.[22]

Furthermore, to estimate acceleration of exports, we also control for pre-existing "momentum": export growth (CAGR) in the previous period. We also include a dummy of zero exports at the beginning of the previous period to control for possible distortions when computing the momentum variable.[23]

[22] The inverse hyperbolic sine is defined at zero and behaves similarly to a log-transformation. The interpretation of regression estimators in the form of the inverse hyperbolic sine is similar to the interpretation of a log-transformed variable.

[23] We use the previous period 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 growth rates the CAGR of export value in the dependent variable will always be defined, given that we limit the sample only to sectors which are being exported at the beginning of the period (that is, $x_{c,p,t} > 0$). 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. We tried alternative specifications and the results remained robust.

Table 3 displays the estimates of Equation (5). Columns 1 to 4 consider the different channels $\Phi_{c,p,t}$ separately while Column 5 does it jointly. For comparison purposes, we report standardized coefficients by normalizing the regressors to have mean zero and unit standard deviation.

[Table 3 about here.]

Panel A shows that patent citations and customer linkages are relevant predictors of the take-off of an export industry, with a point estimate of roughly 0.05 (p-value < 0.01). The estimates and significance for these two channels remain mostly unchanged when all mechanisms are tested together. The interpretation of the coefficients in Column 5 is as follows. A small or inexistent export industry that has a one standard deviation higher density of customer linkages to the current basket is roughly 4 percentage points more likely to take off in the next decade. More precisely, the point estimate of 4.1 for customer linkages (Column 5), represents more than a sixfold increase vis-a-vis the unconditional probability of a take-off, which was 0.6% (see section 2.3). For patent citations the additional effect is ninefold increase of the unconditional take-off probability. To a much lesser extent, labor flows also predict take-offs. But the magnitude is weaker and it becomes insignificant when all channels are tested jointly.[24]

The results for export growth are in Panel B of Table 3. All channels are individually statistically significant in columns 1 to 4. Jointly, though, only patent linkages and customer linkages remain statistically significant at standard levels, as seen in Column 5. In particular, an increase of one standard deviation in a product's patent linkages based density is associated with 14.8 percentage points of additional export growth. This additional effect is almost twice the unconditional growth rate.

[24] An interesting benchmarking would be to compare these channel-based estimates with those of the agnostic co-location density, similar to the regressions in Hidalgo et al. (2007), using only the agnostic co-location HK. Such an estimation (see Online Appendix B) yields a coefficient $\Phi^{HK}_{c,p,t}$ of 0.028 . That means that the estimators of customer and of patents based densities shown in Table 3 are roughly twice as powerful as the agnostic density.

Discussion

When it comes to the take-off of new exports, our results emphasize the importance of technology in the process of cross-industry spillovers. In addition, we find that customer linkages also matter: exports are more likely to take off if that sector is an input to an already competitive sector in the country.

We emphasize this particular finding because it is consistent with a long-held view first put forward by Hirschman (1958). It is also consistent with more recent work. 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. 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 upstream supplier. Also, Amendolagine et al. (2019) argue that this effect might be stronger if the multinational is connected to global value chains. Together, these studies suggest that the existence of a competitive downstream industry might work both as a source of spillovers and as a mechanism to mitigate risks for entrepreneurs and investors to start off a new upstream sector – which then results in new exports. This is consistent with the idea that the emergence of a new sector is subject to a fixed cost associated with the uncertainty of markets (e.g. Wagner and Zahler, 2015).

When it comes to future growth of existing exports, customer linkages may play a role as well, but the effect is comparatively less strong. One more standard deviation in the value of customer-linkages density results in 5 percentage points faster export growth rate over the next decade. This is a 60% increase over the unconditional mean, but smaller than the effect of patent citations.

3.2 Effect in industrialized vs. developing countries

We now turn to explore the extent to which the economic channels exhibit differential impact across levels of development. Hirschman (1958) (Chapter 6) hypothesized that developing countries exhibit less connection among their activities and, therefore, the effects of linkages could be different in comparison to industrialized countries, at least for the take-off of a new competitive sector. Table 4 shows the results for sub-samples, split according to OECD membership. As expected, our findings indicate important heterogeneity in the effect, supporting Hirschman's hypothesis.

[Table 4 about here.]

The results highlight three findings. First, column 1 shows that none of the channels are statistically significant in the OECD subsample. This does not mean channels are unimportant. What it says is that, on average, we cannot tell apart which channel systematically predicts take-offs relatively more than the others. Note that part of this lack of significance may be due to the small sample of potential take-offs: for OECD economies, only 16% of the country-sector observations enter the take-off regression, because OECD economies already export more sectors in the first place with RCA above 0.1.[25]

Second, column 2 shows that for developing countries patents and customer linkages are important to explain export take-offs. This is consistent with our baseline results of Table 3. Importantly, here we confirm that customer linkages with existing activities tend to predict take-off, but not necessarily subsequent growth in developing economies, as shown in Column 4. This, again, coincides with Hirschman's view that a domestic buyer helps achieve a critical mass of demand for upstream sectors.

[25] In contrast to OECD economies, the sample of non-OECD countries splits roughly 50-50 between the potential take-off sample and the growth sample, because they have much fewer exports with RCA above 0.1.

Third, we also find some evidence for the divergent role of forward and backward linkages across development levels. One the one hand, supplier linkages do not contribute to explain new exports in developing countries. They even exhibit statistically significant negative coefficients in Columns 2 and 4. This, again, was argued by Hirschman: for developing countries it is likely that the final activity is performed in a richer country. For instance, cocoa powder exported from an African economy to Switzerland may lower the probability that exports of chocolate emerge in the same African economy in the subsequent decade. On the other hand, we find that supplier linkages do matter for export growth of existing sectors for developed countries, as seen in Column 3.

Further, our results confirm Hirschman's observation that for developing countries the effect of forward linkages is much less pronounced than that of backward linkages (Hirschman, 1958, pp. 116). The difference in coefficients between supplier and customer densities in Column 2 support this hypothesis (p-value <0.01). One way to think about these findings in the context of developing countries is the role that an existing competitive sector can have on its potential suppliers. It helps these emerging upstream sectors to get critical mass and reduce the costs of taking risks in a context of incomplete markets. In the presence of frictions in credit markets, for example, the mere existence of a local market for a sector 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. Beyond market incompleteness, we interpret having connections to local buyers as providing know-how about what and how to produce. Note that this is not just about having competitive customers *anywhere* -meaning, non locally- given that this effect is captured by our strict set of fixed effects, which would capture, for example, (static) transportation costs of inputs for a given country-industry pair.

Our findings for the acceleration of exports confirm the importance of patent linkages for developing countries (Column 4) and the lack thereof in developed economies (Column 3).[26] To interpret this finding, we should again avoid a narrow view of what the patent citation network means. At least since the seminal work by Jaffe et al. (1993), it is the conventional view that the spread of knowledge is highly localized. This reflects the short-ranged character of the mobility of inventors and knowledge-creators (Breschi and Lissoni, 2009). In that context, it makes sense that the technology based effect we document might be more predictive in developing countries, which have overall scarcity of this inventive expertise and/or poorer infrastructure to allow mobility. In OECD economies, the specific skills from other industries nearby may be less of a binding constraint, and therefore harder to produce a systematic prediction in the emergence and growth of export sectors. Overall, Columns 3 and 4 show that the factors that predict growth of exports in OECD economies are different from those in developing economies.

4 Robustness and additional tests

This section offers alternative tests to support our main results. In particular, we focus on the early stages of take-offs and on the possible biases arising from the estimation of linear probability models. We also present results aiming to to tell apart specific mechanisms from co-location based measures.

4.1 Addressing concerns about collinearity

We know that all densities Φ are based on the same matrix $R_{c,t}$, which contains the structure of existing exports with RCA above one. Although we multiply densities with different channel-based industrial proximities, there is still the potential concern of collinearity among Φ. This may complicate the interpretation of significance in the joint specification presented

[26] Similar sized standard errors suggest that the lack of significance is not due to the smaller number of observations in the sample across the different cuts.

in Column 5 of Table 3. Yet, even if the raw density measures correlate highly, the β parameters in Column 5 already correct for all the feasible fixed effect combinations. Thus, the empirical question becomes how much do the densities Φ correlate *after* controlling for fixed effects. Table 5 shows these correlations conditional on country-by-decade, product-by-decade and product-by-country fixed effects. The resulting correlations are positive but not strikingly large. This is consistent with the proximities in Table 1, which showed meaningful variation across channels. Additionally, we computed the Variance Inflation Factor (VIF), used to assess the availability of enough independent variation among correlated variables. The VIF value was 1.9; which is in the acceptable range. Therefore, multi-collinearity seems a less relevant concern for our empirical strategy. There is enough variation as to empirically tell apart the various channels.

[Table 5 about here.]

4.2 The birth of new competitive export sectors

Our baseline definition of binary export take-offs was based on country-sectors with a low initial RCA of 0.1. This condition, hence, includes country-sectors that were already existent. In this section we want to explore whether our results are robust to the birth or emergence of new exports. To do so, we limit our sample used to those country-sector pairs that not only have an initial RCA below 0.1 but also exhibit an initial export value below USD \$10,000.[27] Using that subset (which corresponds to about 60 percent of the original take offs sample), we explore the mechanisms that can explain export take-offs of these country-sectors in a decade time. While the argument posed by Hirschman (1958) was not only about the absolute birth of a new

[27]Given that there many sectors that are re-exported from third countries without having local production, and also considering that small amounts of trade have noise due to misclassification, we choose this threshold which is, de facto, equivalent to zero exports, as in Wagner and Zahler (2015).

industry –rather more about the take-off and the achieving of critical mass– it is still instructive to see how much of the effect in our baseline applies to completely new export sectors. The joint specification of Table 6 in Column 5 indicates that our main results remain robust to considering births of new exports. Both technology and customer linkages are significant, with similar magnitude to each other. The coefficients are between one half and two thirds of those reported in our baseline estimation in Table 3. An important difference appears when we look at the various channels individually in specifications 1 to 4: in that comparison only the existence of competitive buyers is significant. This and other robustness of customer linkages are the reason why we highlight this channel in our conclusions.

[Table 6 about here.]

4.3 Alternative estimation for the binary take-off and its evolution over time

Our baseline model for export take-offs of Table 3 follows recent contributions, such as that by Boschma and Capone, 2015, and sticks to the linear probability model for the binary outcome. There were many reasons to do so. First, non-linear models suffer from the so called incidental parameters problem with many fixed effects, as described by Greene (2004). In these models, the maximum likelihood estimator tends to be inconsistent when T, the length of the panel, is fixed, as in our case with two time intervals. Second, non-linear fixed effects models are not the most commensurate estimation technique, given that Angrist and Pischke (2009) argue that average effects from the linear probability resemble marginal effects of non-linear models. Third, and crucial for our case, non-linear fixed effects models impose a challenging computational complexity. Estimating three large groups of interacted fixed effects in our application proved untenable. Despite all these reasons, we ran additional specifications for robustness tests in Table

7.

For the nonlinear models of Columns 1 and 2 we use the complementary log-log specification. While also restricted to the unit interval, it is a more advisable option than logit or probit if the following conditions hold: (i) when the probability of take-off is small, as in our case; (ii) when the sample is restricted to only observations that are eligible to take-off (e.g., the sample excludes country-industry pairs that have already taken-off); and (iii) when the take-off event is associated to repeated exposure to a risk - as every year in a decade - but one observes data only at the end and changes are persistent. In that context, our decade-long propensity might not be symmetric as assumed in a logit or probit models. Then, the clog–log is preferred because it behaves like a hazard rate (Singer et al., 2003). For computational reasons we needed to limit the number of simultaneous fixed effects. Therefore, we estimate this clog-log model for the each of the two sub-periods separately. One for 1990 to 2000 and another for 2000 to 2010. The Table lists the initial year of each sample. The results of both time periods remark the robustness of having customer linkages with competitive sectors. These were positive and statistically significant. The patent linkages is positive and significant for the first decade, but not for the second.[28]

Alternatively, we also consider the trimmed approach of Horrace and Oaxaca (2006) for binary data. The approach suggests , after the first estimation of a linear probability model, dropping from the sample the observations for which the predicted value falls out of the unit interval, and use this subsample to re-estimate the linear probability model. Horrace and Oaxaca (2006) show that this approach may reduce the potential biases of the linear probability models. Column 3 reports the results of this approach. Despite having less observations, it is reassuring that the point estimate for customer linkages is very similar to our the baseline in Table 2, suggesting the potential

[28]Coefficients signs are comparable, but the magnitudes are not comparable to other models in this paper.

bias from linear probability models may not have been large. Importantly, customer linkages was the only variable that survived as significant in all this binary regression, reinforcing our emphasis in this channel.

[Table 7 about here.]

4.4 Unpacking the effect of co-location in its various channels

A central goal of our paper was to tell apart channels. Acknowledging that these channels are neither orthogonal to each other nor orthogonal to the agnostic co-location density, here we complement our findings of Section 3.1 by building additional support for the results.

There are at least two additional ways to explore whether our channel-based measures add meaningful information beyond co-location. First, one standard way is re-estimating the baseline Table 3, but this time adding the *co-location or agnostic* density $\Phi_{c,p,t}$ as additional control, to explore whether the channel-measures lose significance. We do this in the appendix and our baseline findings remain robust.[29]

Second, we unpack the effect of co-location and explore results when excluding a correlated channel, one at the time. To that end, we put forward a two-step approach. The first step is to get a simple linear regression of $\Phi^{agnostic}_{c,p,t}$ on a single $\Phi^{channel}_{c,p,t}$. Then we compute the residual, which we label $\widetilde{\Phi^{channel}_{c,p,t}}$. This means we are are extracting all the joint variance from the agnostic measure. The second step explores the strength of this residual $\widetilde{\Phi^{channel}_{c,p,t}}$ in explaining take-offs and growth of exports in a way similar to specification (5) but with a difference in the first term. Specifically, we estimate:

[29]The results from this exercise are presented in Online Appendix Section D.

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

[Table 8 about here.]

Table 8 shows the results when unpacking the HK agnostic co-location. The first column reports the results for the coefficient of the agnostic density without "extracting" the portion correlated to any other measure (hence, the title of the column is "-"). We report it to serve as benchmark.[30] As expected, this agnostic channel is a strong predictor of both take-off and growth of exports.

Column 2 reports the coefficient of the agnostic or co-location density "cleaned" from patent-based density. Columns 3 to 5 follow the trend, cleaning all other channels.

Overall, the cleaned coefficients in columns 2 to 5 are 20 to 40% smaller than the one in column 1 (with the "full" agnostic measure), depending on the channel and the outcome. For take-offs, the coefficient for the density orthogonal to customer linkages is the smallest (panel A, column 5), which confirms our findings so far. That is, the agnostic measure matters less once the role of customer linkages is stripped from it. When it comes to explaining future growth of already existing exports (Panel B), labor force and patent citation densities may matter more, as stripping these linkages from the agnostic measure result in lower coefficients. Table C1 in the appendix uses the alternative agnostic EG instead of the HK density, finding similar results.

Summing up, beyond other channels, the robustness tests in this section highlight the importance of customer linkages to ease the take-off of upstream exports. It is important to acknowledge - however - that the explanatory power of the agnostic co-location is strong and relevant, even if one

[30] It replicates the estimates of Table B1 in the Appendix, except that the number of observations is limited to those for which the channel-defined densities are not missing. That is, in the case where $\widetilde{\Phi_{c,p,t}} = \Phi_{c,p,t}$.

excludes specific correlated channels. This is a call for not over-interpreting one single channel as the only cause of the evolution of the export basket. In an unreported regression similar to Table 8, we controlled for all channels simultaneously and co-location was still significant and strong. Co-location in one benchmark country (the US) still contains additional information for new exports in the rest of the countries, although part is accounted by our measures of Marshallian channels.

5 Concluding Remarks

New exports do not emerge randomly. They tend to be related to pre-existing exports in the same country, as reviewed in Hidalgo et al. (2018). In this paper, we contribute to unpack this *relatedness,* shedding light on the channels mediating the take-off and growth of exports across industries and countries. To that end, we *simultaneously* explore the role of the alternative cross-sectoral linkages suggested by Marshall (1920): technology, labor and input-output relationships.

On average, we find that take-offs and growth of exports tend to increase if the country was already competitive in related sectors, the latter measured in terms of patent citations and customer linkages. Other channels tend to be less robust. Regarding patent citations, our results are consistent with the importance of technology adoption for export competitiveness. In that sense, patent linkages reflect the existence of knowledge and technology that, by being local, facilitates its diffusion across economic actors (e.g., Jaffe et al., 1993; Breschi and Lissoni, 2009; Boschma, 2005; Bahar et al., 2014). With respect to customer linkages, our findings imply that, for example, a country is more likely to become a competitive exporter of semi-conductors (an upstream product) if it already exports computer memory chips. Moreover, our global stylized fact generalizes micro-econometric evidence that showed spillovers are more likely in upstream linkages (e.g., Javorcik, 2004; Pietrobelli and Saliola, 2008; Kee and Tang, 2016; Amendolagine et al., 2019). Overall, we do not think that the role of upstream linkages reflects a pure flow of materials, but also a source of know-how, critical mass and lower risks (Hirschman, 1958).

Consistent with Albert Hirschman's view, the impact of linkages among sectors may greatly differ by the level of development. Our findings on customer linkages, for example, is driven by developing countries. Conversely, when looking at developed countries, our results suggest that supplier linkages explain export growth.

Finally, our findings provide some insights for the development policy debate. First, in many low-income countries people like to think about "adding value" through the "*vertical diversification into processing of primary commodities*" (see Cramer 1999; McMillan et al. 2003). That implies promoting diversification downstream from current competitive sectors. In contrast, our evidence for developing countries is that new export sectors tend to appear upstream, not downstream of current industries. Second, scholars have argued that distortions pile up in supply chains, providing a rationale to subsidize the most upstream sector, which otherwise would accumulate most of these distortions (Liu, 2018). However, our estimations suggest that even if a subsidy were to create that new upstream sector, this may not necessarily lead to the take-off of new downstream exports.

To be clear, our findings do not necessarily translate into a single normative advice. Under some circumstances, countries may decide to enrich the productive ecosystem against the revealed forces of comparative advantage within each market with particular characteristics. Clarifying those decisions is a matter of further discussion, as in Lin and Chang (2009). In any case, these interventions should respond to concrete and contextual market failures, instead of simply following the potential for agglomeration, as if clustering were a goal by itself (Rodriguez-Clare, 2007).

References

Amendolagine, V., A. F. Presbitero, R. Rabellotti, and M. Sanfilippo (2019). Local sourcing in developing countries: The role of foreign direct investments and global value chains. *World Development 113*, 73–88.

Amighini, A. and M. Sanfilippo (2014). Impact of south–south fdi and trade on the export upgrading of african economies. *World Development 64*, 1–17.

Angrist, J. and J.-S. Pischke (2009). *Mostly harmless econometrics: an empiricists guide*. Princeton University Press.

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.

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

Balland, P.-A., R. Boschma, J. Crespo, and D. L. Rigby (2018). Smart specialization policy in the european union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, 1–17.

Bernhofen, D. M. ., Z. El-Sahli, and R. Kneller (2016). The Impact of Technological Change on New Trade: Evidence from the Container Revolution. *Journal of International Economics 98*, 36–50.

Bernini, M., J. Du, and J. H. Love (2016). Explaining intermittent exporting: Exit and conditional re-entry in export markets. *Journal of International Business Studies 47*(9), 1058–1076.

Blalock, G. and F. M. Veloso (2007). Imports, productivity growth, and supply chain learning. *World Development 35*(7), 1134–1151.

Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional studies 39*(1), 61–74.

Boschma, R. and G. Capone (2015). Institutions and diversification: Related versus unrelated diversification in a varieties of capitalism framework. *Research Policy 44*(10), 1902–1914.

Breschi, S. and F. Lissoni (2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of economic geography 9*(4), 439–468.

Breschi, S., F. Lissoni, and F. Malerba (2003). Knowledge-relatedness in firm technological diversification. *Research policy 32*(1), 69–87.

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 (2018). Diversification through Trade. *Revise and Resubmit requested at Quarterly Journal of Economics*.

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.

Cramer, C. (1999). Can africa industrialize by processing primary commodities? the case of mozambican cashew nuts. *World Development 27*(7), 1247–1266.

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.

Delgado, M., M. E. Porter, and S. Stern (2016). Defining clusters of related industries. *Journal of Economic Geography*.

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.

Fan, J. P. and L. H. Lang (2000). The measurement of relatedness: An application to corporate diversification. *The Journal of Business 73*(4), 629–660.

Farinha-Fernandes, T., P.-A. Balland, A. Morrison, R. Boschma, et al. (2018). What drives the geography of jobs in the us? unpacking relatedness. *Papers in Evolutionary Economic Geography - Utrecht University*.

Farjoun, M. (1994). Beyond industry boundaries: Human expertise, diversification and resource-related industry groups. *Organization science 5*(2), 185–199.

Fugazza, M. and A. Nicita (2013). The direct and relative effects of preferential market access. *Journal of International Economics 89*(2), 357–368.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal 7*(1), 98–119.

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.

Hartmann, D., M. R. Guevara, C. Jara-Figueroa, M. Aristarán, and C. A. Hidalgo (2017). Linking economic complexity, institutions, and income inequality. *World Development 93*, 75–93.

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. *Harvard Kennedy School of Government, CID Working Paper, No. 128.*.

Hausmann, R. and F. M. Neffke (2018, nov). The workforce of pioneer plants. *Research Policy*.

Hausmann, R., F. Rodríguez, and R. Wagner (2006). *Growth Collapses.* Number RWP06-046. in Carmen Reinhart, Carlos Vegh and Andres Velasco (eds), Money, Crises and Transition, Cambridge, Mass.: MIT Press, pp.376-428.

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., P.-A. Balland, R. Boschma, M. Delgado, M. Feldman, K. Frenken, E. Glaeser, C. He, D. F. Kogler, A. Morrison, et al. (2018). The principle of relatedness. In *International Conference on Complex Systems*, pp. 451–457. Springer.

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.

Hirschman, A. O. (1958). *The strategy of economic development*. Yale University Press.

Hirschman, A. O. (1968). The political economy of import-substituting industrialization in latin america. *The Quarterly Journal of Economics 82*(1), 1–32.

Hoekman, B. and A. Nicita (2011). Trade policy, trade costs, and developing country trade. *World Development 39*(12), 2069–2079.

Horrace, W. C. and R. L. Oaxaca (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters 90*(3), 321–327.

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

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics 108*(3), 577–598.

Jara-Figueroa, C., B. Jun, E. L. Glaeser, and C. A. Hidalgo (2018). The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proceedings of the National Academy of Sciences 115*(50), 12646–12653.

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.

Jones, L. P. (1976). The measurement of hirschmanian linkages. *The Quarterly Journal of Economics 90*(2), 323–333.

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.

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

Krishna, P. and A. Levchenko (2009, may). Comparative Advantage, Complexity and Volatility. *Journal of Economic Behavior & Organization, 94, 314-329.*.

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

Lectard, P. and E. Rougier (2018). Can developing countries gain from defying comparative advantage? distance to comparative advantage, export diversification and sophistication, and the dynamics of specialization. *World Development 102*, 90–110.

Lin, J. and H.-J. Chang (2009). Should industrial policy in developing countries conform to comparative advantage or defy it? a debate between justin lin and ha-joon chang. *Development policy review 27*(5), 483–502.

Liu, E. (2018). Industrial Policies in Production Networks. *Princeton University - working paper*.

Marshall, A. (1920). *Principles of Economics*. Mc Millan.

McMillan, M., D. Rodrik, and K. H. Welch (2003). When economic reform goes wrong: cashews in mozambique. In S. Collins and D. Rodrik (Eds.), *Brookings Trade Forum*. Brookings Institution. Washington DC.

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

Neffke, F. and M. Henning (2013). Skill relatedness and firm diversification. *Strategic Management Journal 34*(3), 297–316.

Neffke, F., M. Henning, and R. Boschma (2011). How do regions diversify over time? industry relatedness and the development of new growth paths in regions. *Economic Geography 87*(3), 237–265.

Neffke, F., A. Otto, and A. Weyh (2017). Inter-industry Labor Flows. *Journal of Economic Behavior & Organization, 142, 275-292.*.

Nicita, A. and V. Rollo (2015). Market access conditions and sub-Saharan Africa' s exports diversification. *World Development 68*, 254–263.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics 122*(2), 569–600.

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.

Rajan, R. G. and L. Zingales (1998). Financial dependence and growth. *The American Economic Review*.

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

Rodriguez-Clare, A. (2007). Clusters and comparative advantage: Implications for industrial policy. *Journal of Development Economics 82*(1), 43–57.

Rodrik, D. (2008). Industrial policy for the xxi century. In D. Rodrik (Ed.), *One economics, many recipes*, Chapter 4. Princeton University Press.

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.

Scherer, F. (1984). Using linked patent and r&d data to measure interindustry technology flows. In *R&D, patents, and productivity*, pp. 417–464. University of Chicago Press.

Singer, J. D., J. B. Willett, J. B. Willett, et al. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford university press.

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

Table 1: Correlations of Relatedness Measures

Variables	HK	EG	Patents	Consumer	Supplier	Labor
HK	1.000					
EG	0.498	1.000				
Patents	0.158	0.137	1.000			
Consumer	0.102	0.101	0.289	1.000		
Supplier	0.120	0.113	0.364	0.457	1.000	
Labor	0.168	0.191	0.573	0.391	0.377	1.000

This table displays bivariate correlation coefficients for all proximity measures. It includes the two agnostic measures HK and EG, as well as the proximities of Marshallian channels.

Table 2: Summary Statistics

Variable	**N**	**Mean**	**sd**	**Min**	**Max**
Panel A - $RCA_{c,p,t} < 0.1$					
Take offs $RCA_{c,p,t+10} > 1$	63,232	0.006	0.08	0.0	1.0
$\Phi_{c,p,t}(HK)$	63,232	0.079	0.07	0.0	0.5
$\Phi_{c,p,t}(EG)$	63,232	0.087	0.07	0.0	0.5
$\Phi_{c,p,t}^{Patents}$	63,232	0.087	0.07	0.0	0.5
$\Phi_{c,p,t}^{SLinkages}$	63,232	0.087	0.07	0.0	0.6
$\Phi_{c,p,t}^{CLinkages}$	63,232	0.085	0.07	0.0	0.5
$\Phi_{c,p,t}^{Labor}$	63,232	0.085	0.07	0.0	0.5
Initial RCA	63,232	0.011	0.02	0.0	0.1
Panel B - $Exports_{c,p,t} > 0$					
10yrs Growth (CAGR)	90,798	0.080	0.22	-0.9	2.7
$\Phi_{c,p,t}(HK)$	90,798	0.201	0.13	0.0	0.8
$\Phi_{c,p,t}(EG)$	90,798	0.195	0.12	0.0	0.8
$\Phi_{c,p,t}^{Patents}$	90,798	0.194	0.11	0.0	0.5
$\Phi_{c,p,t}^{SLinkages}$	90,798	0.193	0.11	0.0	0.6
$\Phi_{c,p,t}^{CLinkages}$	90,798	0.193	0.12	0.0	0.6
$\Phi_{c,p,t}^{Labor}$	90,798	0.193	0.12	0.0	0.6
Initial Exports	90,798	14.956	3.30	7.6	25.4
Pre-period CAGR	90,798	0.791	2.46	-0.8	48.9
Pre-period zero exp	90,798	0.079	0.27	0.0	1.0

This table presents descriptive statistics for our key dependent variables: export take-offs and 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 export take offs, where we limit the sample to those country-industry 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 export growth rates, where we limit our observations to those country-industry pairs with exports above zero at the beginning of the 1990-2000 and 2000-2010 periods. Note that the sample used resembles the one used in the estimation of Specification (5) shown in Table 3.

Table 3: Take offs and growth of related industries

Panel A - Dependent Variable: Take off if initial $RCA_{c,p,t} < 0.1$					
	(1)	(2)	(3)	(4)	(5)
$\Phi^{Patents}_{c,p,t}$	0.0547				0.0590
	(0.019)***				(0.026)**
$\Phi^{Labor}_{c,p,t}$		0.0265			-0.0158
		(0.013)*			(0.017)
$\Phi^{Slinkages}_{c,p,t}$			0.0085		-0.0152
			(0.009)		(0.009)
$\Phi^{CLinkages}_{c,p,t}$				0.0539	0.0414
				(0.014)***	(0.013)***
Initial RCA	0.1215	0.1245	0.1266	0.1235	0.1215
	(0.036)***	(0.036)***	(0.037)***	(0.037)***	(0.036)***
N	63232	63232	63232	63232	63232
R2	0.53	0.53	0.53	0.53	0.53
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y	Y	Y	Y
Panel B - Dependent Variable: Growth (CAGR) if initial $Exports_{c,p,t} > 0$					
	(1)	(2)	(3)	(4)	(5)
$\Phi^{Patents}_{c,p,t}$	0.2057				0.1479
	(0.036)***				(0.042)***
$\Phi^{Labor}_{c,p,t}$		0.1538			0.0479
		(0.029)***			(0.036)
$\Phi^{Slinkages}_{c,p,t}$			0.0773		-0.0193
			(0.023)***		(0.020)
$\Phi^{CLinkages}_{c,p,t}$				0.1520	0.0499
				(0.033)***	(0.029)*
Initial Exports	-0.1390	-0.1389	-0.1386	-0.1387	-0.1390
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Pre-period CAGR	0.0064	0.0064	0.0062	0.0063	0.0065
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Pre-period zero exp	-0.0270	-0.0267	-0.0254	-0.0260	-0.0273
	(0.010)***	(0.010)***	(0.010)**	(0.010)***	(0.010)***
N	90798	90798	90798	90798	90798
R2	0.82	0.82	0.82	0.82	0.82
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y	Y	Y	Y

This table estimates specification (5) with the channel-specific density measures. The upper panel estimates the specification for export sectors take-offs and the lower panel does so for the export growth. Columns 1 to 4 evaluate the impact of each channel-specific density measure separately, while column 5 includes all channel-based measures jointly. All specifications include country-by-decade, product-by-decade and product-by-country fixed effects. All coefficients are standardized with zero mean and unit standard deviation. Standard errors are clustered at the country level and presented in parenthesis.

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

Table 4: Take-offs and growth of related industries, OECD vs. non-OECD

Dependent variables: Take-offs and growth of exports				
	Take Off		Growth	
	(1)	(2)	(3)	(4)
	OECD	nonOECD	OECD	nonOECD
$\Phi^{Patents}_{c,p,t}$	-0.0092	0.0804	0.0234	0.2354
	(0.031)	(0.032)**	(0.047)	(0.068)***
$\Phi^{Labor}_{c,p,t}$	0.0157	-0.0224	0.0683	0.0727
	(0.030)	(0.020)	(0.040)*	(0.053)
$\Phi^{SLinkages}_{c,p,t}$	0.0174	-0.0250	0.0439	-0.0632
	(0.016)	(0.010)**	(0.018)**	(0.036)*
$\Phi^{CLinkages}_{c,p,t}$	0.0224	0.0460	0.0389	0.0399
	(0.026)	(0.015)***	(0.027)	(0.052)
Initial RCA	0.1071	0.1314		
	(0.089)	(0.040)***		
Initial Exports			-0.1323	-0.1425
			(0.002)***	(0.002)***
Pre-period CAGR			0.0040	0.0083
			(0.001)**	(0.001)***
Pre-period zero exp			-0.0349	-0.0429
			(0.020)*	(0.012)***
N	6356	56706	37528	53268
R2	0.59	0.54	0.86	0.81
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y	Y	Y

This table estimates specification (5) using all channel-specific density measures, separating the sample between OECD and non-OECD economies. Columns 1 and 2 estimates the specification for export sectors take-offs while columns 3 and 4 do so for export growth. The specifications resemble column 5 of Table 3. All specifications include country-by-decade, product-by-decade and product-by-country fixed effects. All coefficients are standardized with zero mean and unit standard deviation. 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}	$\Phi^{Patents}$	Φ^{Labor}	$\Phi^{Slinkages}$	$\Phi^{CLinkages}$
Φ^{HK}	1.000					
Φ^{EG}	0.634	1.000				
$\Phi^{Patents}$	0.249	0.260	1.000			
Φ^{Labor}	0.285	0.302	0.754	1.000		
$\Phi^{SLinkages}$	0.204	0.225	0.556	0.523	1.000	
$\Phi^{CLinkages}$	0.200	0.221	0.545	0.547	0.330	1.000

This table presents correlations between density measures after conditioning on country-by-decade, product-by-decade and country-by-product fixed effects.

Table 6: Birth of new export sectors

Dependent Variable: Take offs if initial $RCA_{c,p,t} \leq 0.1$ **and** $exp_{c,p,t} \leq 10K$					
	(1)	(2)	(3)	(4)	(5)
$\Phi^{Patents}_{c,p,t}$	0.0196				0.0329
	(0.013)				(0.019)*
$\Phi^{Labor}_{c,p,t}$		0.0030			-0.0191
		(0.012)			(0.015)
$\Phi^{Slinkages}_{c,p,t}$			-0.0021		-0.0122
			(0.008)		(0.009)
$\Phi^{CLinkages}_{c,p,t}$				0.0268	0.0267
				(0.012)**	(0.013)**
Initial RCA	-0.0692	-0.0667	-0.0661	-0.0732	-0.0737
	(0.126)	(0.127)	(0.126)	(0.126)	(0.126)
N	32954	32954	32954	32954	32954
R2	0.54	0.54	0.54	0.54	0.54
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y	Y	Y	Y

This table estimates specification (5) with the channel-specific density measures. It estimates export sectors emergence, defined by country-sectors having RCA below 0.1 and export value below USD 10,000 at the beginning of the period, and achieving an RCA above 1 in a decade time. Columns 1 to 4 evaluate the impact of each channel-specific density measure separately, while Column 5 include all channel-based measures jointly. All specifications include country-by-decade, product-by-decade and product-by-country fixed effects. All coefficients are standardized with zero mean and unit standard deviation. Standard errors are clustered at the country level and presented in parenthesis.

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

Table 7: Non-linear estimation of take-off events

Dependent Variable: Take-off if initial $RCA_{c,p,t} < 0.1$			
	Non-linear		HO2006
	(1)	(2)	(3)
	1990	2000	All
$\Phi^{Patents}_{c,p,t}$	23.0454	15.8771	0.0543
	(13.730)*	(10.954)	(0.036)
$\Phi^{Labor}_{c,p,t}$	1.0126	5.8811	0.0141
	(9.453)	(8.880)	(0.024)
$\Phi^{Slinkages}_{c,p,t}$	-3.4025	-3.8452	-0.0158
	(6.374)	(4.813)	(0.017)
$\Phi^{CLinkages}_{c,p,t}$	16.8617	20.2757	0.0512
	(8.228)**	(8.731)**	(0.021)**
Initial RCA	12.2341	11.5037	0.1598
	(1.599)***	(1.995)***	(0.054)***
N	23416	15524	22194
Pseudo-R2	.05	.05	.
R2			0.56
η_c, δ_p	Y	Y	-
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	-	-	Y

This table re-estimates our main specification of Eq. (5) for the binary outcome of export take-off. But instead of using a standard linear probability model, it uses alternative approaches that deal with the binary nature of the dependent variable. Columns 1 and 2 display the coefficients of a complementary log-log model (cloglog) for deacades starting in 1990 and 2000, respectively, adding both country and sector fixed effects. Column 3 uses the trimmed methodology for linear probability models of Horrace and Oaxaca (2006), excluding observations for which the predicted values are outside of the unit interval. Column 3 includes country-by-decade, product-by-decade and product-by-country fixed effects. Coefficients in Column 3 are standardized to have mean zero and unit standard deviation. Standard errors are clustered at the country level and presented in parenthesis.

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

Table 8: Take offs and growth, unpacking agnostic density

Panel A - Dependent Variable: Take off if initial $RCA_{c,p,t} < 0.1$					
	(1)	(2)	(3)	(4)	(5)
	-	Patents	Labor	Supplier	Customer
$\widetilde{\Phi cpt}$	0.4028	0.3063	0.3377	0.3321	0.2861
	(0.103)***	(0.090)***	(0.096)***	(0.083)***	(0.096)***
N	63232	63232	63232	63232	63232
R2	0.53	0.53	0.53	0.53	0.53
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y	Y	Y	Y
Panel B - Dependent Variable: Growth (CAGR) if initial $Exports_{c,p,t} > 0$					
	(1)	(2)	(3)	(4)	(5)
	-	Patents	Labor	Supplier	Customer
$\widetilde{\Phi cpt}$	0.5029	0.3127	0.2970	0.3590	0.3456
	(0.100)***	(0.093)***	(0.089)***	(0.088)***	(0.094)***
N	90798	90798	90798	90798	90798
R2	0.82	0.82	0.82	0.82	0.82
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y	Y	Y	Y

This table estimates specification (6) with the agnostic (HK) relatedness measure orthogonal to every channel specific based density. The upper panel estimates the specification for take-off events and the lower panel does so for export growth. Column 1 is the benchmark: it reports results using the coefficient for the agnostic density HK, without extracting the effect of any channel. Columns 2 to 5 evaluate the impact of the agnostic density cleaned from each channel specific based densities. All specifications include country-by-year, product-by-year and country-by-product fixed effects. Control variables are not shown for expositional purposes. Standard errors are clustered at the country level and presented in parenthesis.

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

Online Appendix for

"Take-offs and acceleration of export industries: unpacking cross-sector linkages driving the evolution of comparative advantage"

by Bahar, Stein, Rosenow and Wagner

A Alternative agnostic relatedness

The Ellison and Glaeser (1997) EG index measures the intensity with which two given sectors are co-located in the same area, and in our case, co-exported by the same country. To compute the EG relatedness index between sectors i and j for a particular year, we use the formulation of the EG co-agglomeration index suggested by Ellison and Glaeser (1997):

$$\varphi_{i,j}^{EG} = \frac{\sum_{c=1}^{C}(s_{c,i} - x_c)(s_{c,j} - x_c)}{1 - \sum_{c=1}^{C} x_c^2} \tag{A1}$$

where $s_{c,i}$ and $s_{c,j}$ are, respectively, a country $c's$ share of sector i and j in world sector exports and x_c represents the share of country c's exports in global exports. The EG co-agglomeration index posits that two sectors are more related to each other the more similar their proportion in the export basket 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 sectors i and j is in the 90th percentile). The HK and EG indices are two different measures of the same underlying phenomenon. Namely, they reflect how much two sectors tend to be co-located. 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.[31]

B Take-off and growth of exports as explained by agnostic measures

Here we confirm in our sample that the emergence of new export sectors depends on incumbent exports as measured by co-location patterns. This is a known fact that has already been established by Hausmann and Klinger (2006); Hidalgo et al. (2007); Hausmann et al. (2014) at the country level. We just need to make sure it works in our sample as a sanity check. After this first step, section 3 of the paper unpacks this agnostic effect into the various channels, which are the center of our paper. Below is the detail of the exercise with the agnostic measures.

Here we follow a regression model that is equivalent from that in Section 3.1, but using the so called "agnostic" relatedness. Table B1 displays the results using also 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. Note that the coefficients of the regression tables in the main paper were standardized. Instead, the coefficients of the agnostic measures in Table B1 are not standardized. Therefore these are not directly comparable in their magnitude.

[Table B1 about here.]

The main finding from Table B1 is that the emergence of new sectors and the future growth of already existing sectors tend to be positively corre-

[31] See Tables E1 and E2 in Appendix E for the fifteen most related sector pairs based on the two relatedness measures HK and EG, respectively.

lated with the pre-existence of related exports ten years earlier, in the same country. These results are not new to the literature (e.g., Hausmann and Klinger, 2006; Hidalgo et al., 2007; Hausmann et al., 2014). In particular, the results in Column 1, which use HK proximities, imply that a sector is 3.4 percentage points, on average, more likely to take off if the baseline density is larger by one standard deviation.[32] This represents a fivefold increase in the unconditional probability of taking off (which was 0.6 percent). Using EG proximities (Column 2) the corresponding numbers represent an increase of, also, 3.3 percentage points. 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 also positively associates with faster export annual growth for both measures. Note that the results in both panels are robust to using the very conservative specification that includes country-year, product-year and country-industry fixed effects.

[32]The percentage point increase of 2.7 results from multiplying the densities' standard deviation of 0.08 with the estimated coefficient of 0.3323.

Table B1: Emergence and growth of related industries

Panel A - Dependent Variable: Take off if initial $RCA_{c,p,t} < 0.1$		
	(1)	(2)
	HK	EG
Φ_{cpt}	0.3377	0.3332
	(0.080)***	(0.080)***
Initial RCA	0.1099	0.1123
	(0.034)***	(0.034)***
N	73988	73988
R2	0.53	0.53
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y
Panel B - Dependent Variable: Growth (CAGR) if initial $Exports_{c,p,t} > 0$		
	(1)	(2)
	HK	EG
Φ_{cpt}	0.5040	0.3389
	(0.098)***	(0.105)***
Initial Exports	-0.1389	-0.1382
	(0.001)***	(0.001)***
Pre-period CAGR	0.0061	0.0057
	(0.001)***	(0.001)***
Pre-period zero exp	-0.0235	-0.0202
	(0.010)**	(0.010)**
N	101698	101698
R2	0.82	0.82
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y

This table estimates specification (5) using agnostic or co-location density measures. The upper panel estimates the specification for export sectors take-offs and the lower panel does so for the export growth. Columns 1 and 2 use the HK density measures, while Columns 3 and 4 use the EG ones. All specifications include country-by-decade and product-by-decade fixed effects. Columns 2 and 4 also include and product-by-country fixed effects. All coefficients are standardized with zero mean and unit standard deviation. Standard errors are clustered at the country level and presented in parenthesis.

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

C Alternative two-step unpacking of co-location

Table C1 repeats the exercise of Table 8, but now using the EG agnostic measure instead. Even though all residual densities prove statistically significant for export take-offs, the proportion that the channels explain the EG agnostic relatedness measure is robust and similar to that of the HK agnostic relatedness measure (see Section 4.4).

[Table C1 about here.]

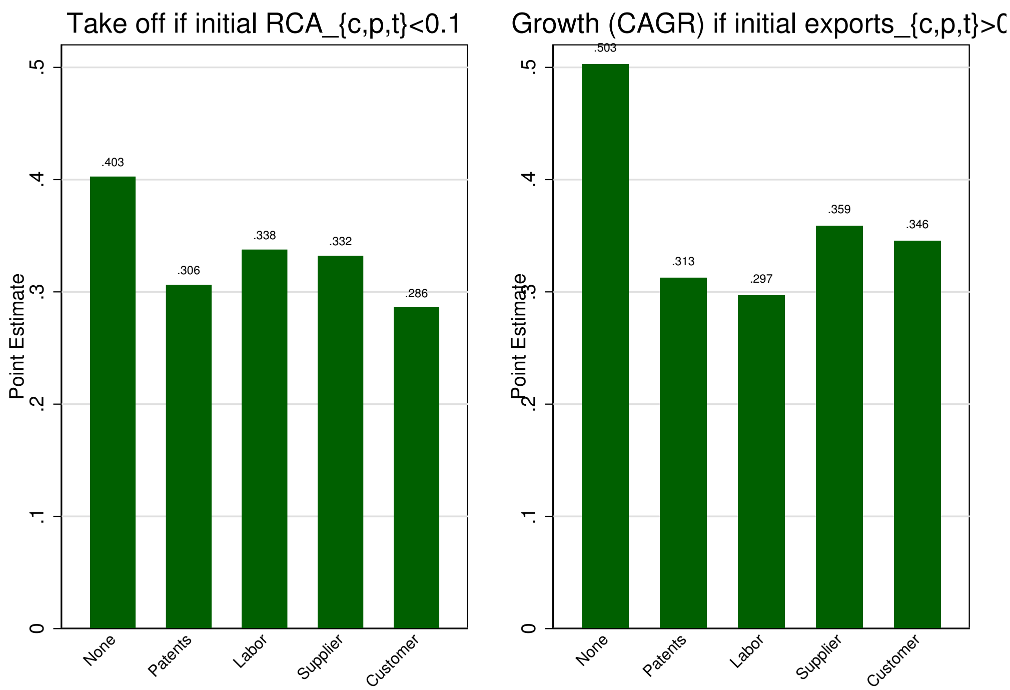
One way to see this results is through Figures C1 and C2. Both offer a visualization of the results in Tables 8 and C1, respectively. For example, in figure C1 the take-off panel, the point estimates of the agnostic relatedness measured purged from customer linkages is smallest. This implies that it is the customer channel which correlates mostly with export take off. 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 for export take off and growth.[33]

[Figure C1 about here.]

[Figure C2 about here.]

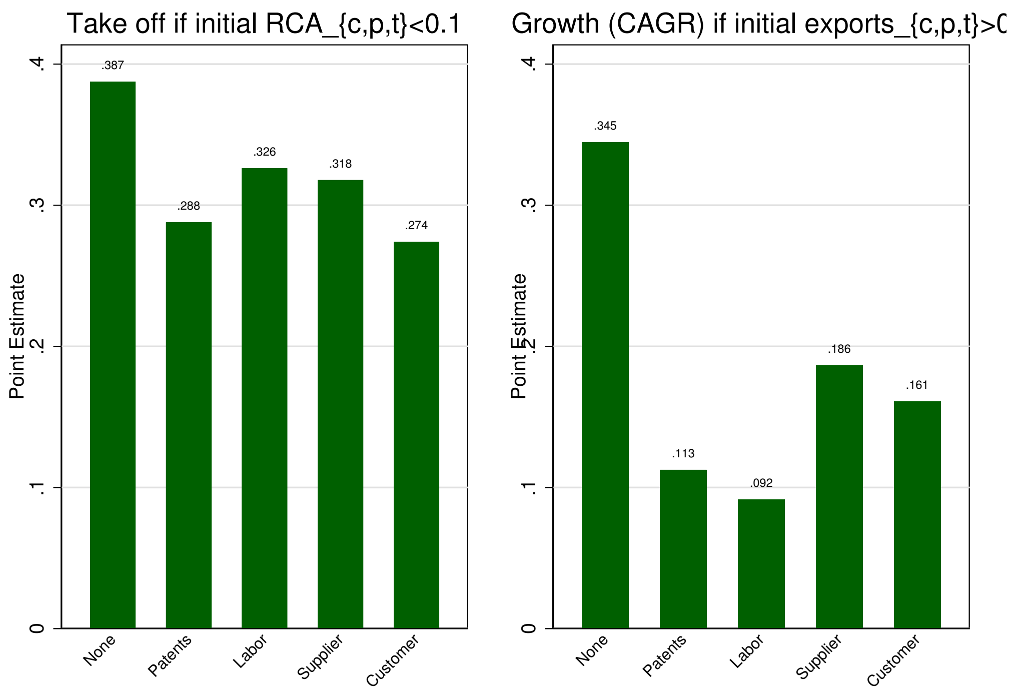
[33] It is important to note that, statistically, there is often no difference between all estimators, each one using different measures of relatedness. Yet, we focus our interpretation on the point estimates.

Figure C1: Take-offs and growth, unpacking agnostic (HK) density



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 C2: Take-offs and growth, unpacking agnostic (EG) density



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 C1: Take-offs and growth, unpacking agnostic (EG) density

Panel A - Dependent Variable: Take off if initial $RCA_{c,p,t} < 0.1$					
	(1)	(2)	(3)	(4)	(5)
	-	Patents	Labor	Supplier	Customer
$\widetilde{\Phi cpt}$	0.3875	0.2879	0.3262	0.3179	0.2741
	(0.097)***	(0.082)***	(0.092)***	(0.079)***	(0.088)***
N	63232	63232	63232	63232	63232
R2	0.53	0.53	0.53	0.53	0.53
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y	Y	Y	Y
Panel B - Dependent Variable: Growth (CAGR) if initial $Exports_{c,p,t} > 0$					
	(1)	(2)	(3)	(4)	(5)
	-	Patents	Labor	Supplier	Customer
$\widetilde{\Phi cpt}$	0.3447	0.1126	0.0916	0.1865	0.1609
	(0.108)***	(0.101)	(0.099)	(0.096)*	(0.101)
N	90798	90798	90798	90798	90798
R2	0.82	0.82	0.82	0.82	0.82
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y	Y	Y	Y

This table estimates specification (6) with the agnostic (EG) relatedness measure orthogonal to every channel specific based density. The upper panel estimates the specification for take-off events and the lower panel does so for export growth. Column 1 is the benchmark: it reports results using the coefficient for the agnostic density HK, without extracting the effect of any channel. Columns 2 to 5 evaluate the impact of the agnostic density cleaned from each channel specific based densities. All specifications include country-by-year, product-by-year and country-by-product fixed effects. Control variables are not shown for expositional purposes. Standard errors are clustered at the country level and presented in parenthesis.

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

D Estimating channels-specific measures controlling for agnostic ones

Tables D1 and D2 replicate Table 3, 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 D1 about here.]

[Table D2 about here.]

Table D1: Take-offs and growth incl. agnostic (HK) density control

Panel A - Dependent Variable: Take off if initial $RCA_{c,p,t} < 0.1$					
	(1)	(2)	(3)	(4)	(5)
$\Phi^{Patents}_{c,p,t}$	0.0406				0.0564
	(0.017)**				(0.026)**
$\Phi^{Labor}_{c,p,t}$		0.0119			-0.0252
		(0.013)			(0.018)
$\Phi^{SLinkages}_{c,p,t}$			-0.0001		-0.0175
			(0.008)		(0.009)*
$\Phi^{CLinkages}_{c,p,t}$				0.0430	0.0375
				(0.014)***	(0.013)***
$\Phi_{c,p,t}(HK)$	0.0256	0.0274	0.0287	0.0259	0.0262
	(0.007)***	(0.007)***	(0.007)***	(0.007)***	(0.007)***
Initial RCA	0.1146	0.1172	0.1182	0.1158	0.1148
	(0.036)***	(0.036)***	(0.036)***	(0.036)***	(0.036)***
N	63232	63232	63232	63232	63232
R2	0.53	0.53	0.53	0.53	0.53
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y	Y	Y	Y
Panel B - Dependent Variable: Growth (CAGR) if initial $Exports_{c,p,t} > 0$					
	(1)	(2)	(3)	(4)	(5)
$\Phi^{Patents}_{c,p,t}$	0.1744				0.1479
	(0.034)***				(0.042)***
$\Phi^{Labor}_{c,p,t}$		0.1236			0.0479
		(0.028)***			(0.036)
$\Phi^{SLinkages}_{c,p,t}$			0.0534		-0.0193
			(0.022)**		(0.020)
$\Phi^{CLinkages}_{c,p,t}$				0.1240	0.0499
				(0.030)***	(0.029)*
$\Phi_{c,p,t}(HK)$	0.0493	0.0495	0.0589	0.0550	
	(0.012)***	(0.012)***	(0.012)***	(0.012)***	
Initial Exports	-0.1398	-0.1396	-0.1395	-0.1396	-0.1390
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Pre-period CAGR	0.0068	0.0068	0.0067	0.0068	0.0065
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Pre-period zero exp	-0.0308	-0.0304	-0.0300	-0.0303	-0.0273
	(0.010)***	(0.010)***	(0.010)***	(0.010)***	(0.010)***
N	90798	90798	90798	90798	90798
R2	0.82	0.82	0.82	0.82	0.82
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y	Y	Y	Y

This table estimates specification (5) with the channel-specific density measures, controlling for the agnostic (HK) density. The upper panel estimates the specification for export sectors take-offs and the lower panel does so for the export growth. Columns 1 to 4 evaluate the impact of each channel-specific density measure separately, while Column 5 include all channel-based measures jointly. All specifications include country-by-decade, product-by-decade and product-by-country fixed effects. All coefficients are standardized with zero mean and unit standard deviation. Standard errors are clustered at the country level and presented in parenthesis.

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

Table D2: Take-offs and growth incl. agnostic (EG) density control

Panel A - Dependent Variable: Take off if initial $RCA_{c,p,t} < 0.1$					
	(1)	(2)	(3)	(4)	(5)
$\Phi^{Patents}_{c,p,t}$	0.0423				0.0598
	(0.017)**				(0.026)**
$\Phi^{Labor}_{c,p,t}$		0.0120			-0.0272
		(0.013)			(0.018)
$\Phi^{Slinkages}_{c,p,t}$			0.0006		-0.0171
			(0.008)		(0.009)*
$\Phi^{CLinkages}_{c,p,t}$				0.0431	0.0367
				(0.013)***	(0.013)***
$\Phi_{c,p,t}(EG)$	0.0246	0.0263	0.0276	0.0246	0.0254
	(0.006)***	(0.007)***	(0.007)***	(0.007)***	(0.007)***
Initial RCA	0.1168	0.1196	0.1207	0.1182	0.1170
	(0.036)***	(0.036)***	(0.036)***	(0.036)***	(0.036)***
N	63232	63232	63232	63232	63232
R2	0.53	0.53	0.53	0.53	0.53
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y	Y	Y	Y
Panel B - Dependent Variable: Growth (CAGR) if initial $Exports_{c,p,t} > 0$					
	(1)	(2)	(3)	(4)	(5)
$\Phi^{Patents}_{c,p,t}$	0.1909				0.1479
	(0.034)***				(0.042)***
$\Phi^{Labor}_{c,p,t}$		0.1399			0.0479
		(0.028)***			(0.036)
$\Phi^{Slinkages}_{c,p,t}$			0.0622		-0.0193
			(0.022)***		(0.020)
$\Phi^{CLinkages}_{c,p,t}$				0.1357	0.0499
				(0.031)***	(0.029)*
$\Phi_{c,p,t}(EG)$	0.0218	0.0216	0.0334	0.0286	
	(0.012)*	(0.012)*	(0.012)***	(0.012)**	
Initial Exports	-0.1392	-0.1391	-0.1389	-0.1390	-0.1390
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Pre-period CAGR	0.0065	0.0065	0.0064	0.0064	0.0065
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Pre-period zero exp	-0.0280	-0.0277	-0.0271	-0.0274	-0.0273
	(0.010)***	(0.010)***	(0.010)***	(0.010)***	(0.010)***
N	90798	90798	90798	90798	90798
R2	0.82	0.82	0.82	0.82	0.82
$\eta_{c,t}, \delta_{p,t}, \alpha_{c,p}$	Y	Y	Y	Y	Y

This table estimates specification (5) with the channel-specific density measures, controlling for the agnostic (EG) density. The upper panel estimates the specification for export sectors take-offs and the lower panel does so for the export growth. Columns 1 to 4 evaluate the impact of each channel-specific density measure separately, while Column 5 include all channel-based measures jointly. All specifications include country-by-decade, product-by-decade and product-by-country fixed effects. All coefficients are standardized with zero mean and unit standard deviation. Standard errors are clustered at the country level and presented in parenthesis.

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

E Relatedness between sectors for agnostic and channel-specific measures

Tables E1, E2, E3, E4, E5, and E6 show the 15 most related sector pairs for all different relatedness measures used in this paper. The two agnostic relatedness measures correspond to the first two tables.

[Table E1 about here.]

[Table E2 about here.]

[Table E3 about here.]

[Table E4 about here.]

[Table E5 about here.]

[Table E6 about here.]

Table E1: 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 sector pairs, based on HK's measure of agnostic relatedness, which is the minimum probability of co-exporting two given sectors.

Table E2: 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 sector pairs, based on EG's measure of agnostic relatedness, which is the co-location of export industries in the same country of origin.

Table E3: 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 sector pairs, based on labor flows industry pairs.

Table E4: 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 sector pairs, based on patent citations between industry pairs.

Table E5: 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 sector pairs, based on upstream linkages.

Table E6: 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 sector pairs, based on downstream linkages.