Public Disclosure Authorized

Public Disclosure Authorized

Policy Research Working Paper 10463

Export Diversification from an Activity Perspective

An Exploration Using Occupation Data

Hagen Kruse
Marcel Timmer
Gaaitzen de Vries
Xianjia Ye

WORLD BANK GROUP
Macroeconomics, Trade and Investment Global Practice
May 2023

Policy Research Working Paper 10463

Abstract

With international production fragmentation, countries specialize in activities along the production chain rather than particular products. This paper therefore analyzes export diversification taking an activity perspective. It measures export activities combining new data on the export income of workers in industries cross classified by occupational classes. Based on the panel data, the paper documents that countries initially specialize along the extensive margin (shifting activities across industries) but later on along the intensive margin (shifting activities across occupational classes). New activity specialization is found to be strongly related to the proximity of this activity to the initial export basket. Yet, countries that defy proximity appear to grow faster. The results show that an activity perspective delivers novel insights into trade development and structural change.

This paper is a product of the Macroeconomics, Trade and Investment Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at g.j.de.vries@rug.nl.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Produced by the Research Support Team

Export Diversification from an Activity Perspective: An Exploration Using Occupation Data

Hagen Kruse[a], Marcel Timmer[a], Gaaitzen de Vries[a], Xianjia Ye[a]

[a] Groningen Growth and Development Centre, Faculty of Economics and Business, University of Groningen, the Netherlands

Keywords: Export diversification; International trade; Structural change; Occupations;

JEL codes: F19, O14, O33

Acknowledgments

This paper is commissioned as part of the project 'Leveraging trade for more and better job opportunities in developing countries' by Maryla Maliszewska and Deborah Winkler, a product of the EFI-MTI-TIC-Trade and RI (ETIRI). The project is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at g.j.de.vries@rug.nl.

1. Introduction

Traditionally, trade performance and diversification are studied from a product perspective. This perspective lost focus due to large-scale offshoring trends with countries trading tasks rather than products, carrying out different activities in global value chains (Grossman and Rossi-Hansberg 2008; Baldwin and Robert-Nicoud 2014; Antrás and Chor 2022). The aim of this paper is to explore patterns of export specialization from an activity perspective. To aid intuition, table 1 provides an illustrative example. It shows the distribution of activities in exports of a particular product class, textiles, for four countries at different levels of development. Activities are identified in the data by cross classifying the occupational class of workers and their industry of work. It shows that activities of machine operators in the textile industry make up more than two-thirds of domestic value-added exports in Türkiye. Instead, in Italy much more value is added by managers, engineers and other professionals outside the textile industry. In Pakistan agricultural workers provide most value in textile exports as cotton is a major input in domestic textile production, whereas in Vietnam workers in wholesale trade services are major contributors. Activities in textiles exporting thus differ widely across countries, suggesting potential for new insights when using an activity perspective on trade.[1]

Exploiting a new panel data set on export activities covering countries at a wide range of income levels, we document four main empirical findings. First, production activities typically account for the majority of export value at low levels of GDP per capita. As countries grow richer, incomes from engineering, managerial, and services support activities grow and eventually account for the majority of export value. Second, poorer countries specialize along the extensive margin, shifting export activity across industries within occupational classes. Advanced countries mostly specialize along the intensive margin, shifting export activity between occupational classes within industries. Third, the probability that a country gains new comparative advantage in a particular activity is positively related to the proximity of this activity to the initial export basket. Proximity appears to be particularly predictive for new specializations along the extensive margin. Fourth, some countries specialize in new activities that are only weakly related to their initial basket. A higher

[1] The need and potential for an activity perspective has been stressed before, see for example Lederman and Maloney (2012). Wolff (2003) provides an early empirical example of the activity perspective on trade.

share of these path-defying specializations appears to be correlated with higher growth in GDP per capita.

Table 1: Top-3 activities in the exports of textiles, four selected countries, 2018

Pakistan		**Türkiye**	
1. Agricultural workers	45.0%	1. Machine operators in textile	67.8%
2. Machine operators in textile	20.3%	2. Managers in textile	10.0%
3. Sales workers in retail	8.0%	3. Agricultural workers	5.9%
All other activities	26.6%	All other activities	16.3%
Vietnam		**Italy**	
1. Machine operators in textile	56.4%	1. Machine operators in textile	29.9%
2. Sales workers in wholesale	7.2%	2. Other professionals in textile	7.2%
3. Other professionals in wholesale	2.9%	3. Managers in textile	6.7%
All other activities	33.5%	All other activities	56.2%

Notes: Entries show for each country the contribution of the top-3 activities in the exports of textiles in 2018 (in percentages of total export value). Activities are classified by industry of work and occupational class of workers. The last line reports the contribution of the other activities outside the top-3. Contributions to the textiles exports are based on the labour income of domestic workers involved in each activity. Own calculations, see section 2 for data sources.

Our main contribution is to the venerable tradition of studying patterns in structural change as incomes rise (Chenery et al., 1986; Syrquin, 1988; Herrendorf et al., 2014). Lack of export development is a traditional concern in this literature as it appears to be linked with slower structural change and productivity growth. Traditionally, countries are assumed to gradually develop manufacturing exports from "light" industries that are intensive in the use of unskilled labor, to "heavy" industries that are intensive in physical and human capital use (Syrquin, 1988; Hanson 2012). Hidalgo et al. (2007) and Hausmann et al. (2014) provided further detail, describing countries' development paths from exporting simple to increasingly complex goods. The usefulness of the product perspective depends crucially on the assumption that production technologies for a particular good are the same in all exporting countries. Yet, Hummels, Ishii and Yi (2001) documented vertical specialization as countries increasingly make use of imported inputs. With international fragmentation, countries specialize in different stages along the production chain such that they may carry out different activities in the same industry. Koopman, Wang and Wei (2012) showed for example that the domestic value-added content of Chinese exports of goods that are traditionally labelled as complex, such as electronic devices, was particularly low as it relied heavily on imported components. Johnson and Noguera (2012, 2017) quantified the scope of production sharing in the global economy, finding amongst others an increasing share of service activities in manufacturing exports. We continue this line of work breaking down the value-added content of exports and revisiting the relation between export development and income growth from an activity perspective.

To do so, we perform two empirical exercises. First, we redo the analysis of Hausmann and Klinger (2007) and Hidalgo et al. (2007) who identified development paths using the concept of "proximity". The degree of proximity between a pair of products is empirically inferred from co-occurrence of revealed comparative advantages (RCAs) for these products in a cross-country panel data set. Instead, we perform export analysis based on proximities of activity pairs rather than product pairs. The results show that countries first diversify predominantly through shifting export activities across industries and later in the development process through upgrading of activities within industries. The former diversification process is picked up in traditional industry analyses, see, e.g., Imbs and Wacziarg (2003). The latter process will not show up as export diversification when viewed from a product or industry-based perspective. Moreover, the activity perspective also provides a natural basis to probe further into the role of technology in export specialization as technological change is typically task specific rather than sector or product specific (Acemoglu and Autor, 2011). We find that proximity appears to be particularly predictive for new specializations in activities that are intensive in routine manual tasks, which differs across advanced and developing economies. For developing economies, a 1 standard deviation increase in average proximity relates to a 10.9 percentage point increase in the probability of specializing in a new routine manual activity. In advanced economies the effect is only 2.9 percentage points. In a next step, we also analyze trade specialization patterns of individual countries in the vein of Coniglio et al. (2021), adapting their product-based methodology to the activity perspective. We find that some countries acquire comparative advantages in activities that are not proximate, denoted as path-defying specialization in Coniglio et al. (2021). Parsimonious cross-country regressions suggest that path-defying activity specialization appears to be positively correlated with higher growth in GDP per capita. In contrast, Coniglio et al (2021) found that the effect turns negative for high-income countries when viewed from a product perspective.

With the new data set we also contribute to a recent wave of studies that use occupational statistics to describe structural change. For example, Newfarmer, Page and Tarp (2018) study various industries in agriculture and services that share important characteristics of manufacturing industries such as tradability, scalability and productivity growth, and name these "industries without smokestacks". Baccini et al. (2023) stress important variation in activities within the services sector from provision of informal personal services to much more productive formal business services and transport services that are used also for exporting. Duernecker and Herrendorf (2022) show how the share of service occupations within the good sectors increases in later stages of development. Using occupational statistics, Timmer et al. (2019) documented specialization of high-income countries in activities that are knowledge intensive. The *Jobs of the World Project* is another prominent example of the compilation of a large-scale data set to compare changes in occupational structures across countries and over time. Using this data, Bandiera et al. (2022) show how the nature of jobs vary across countries by stages of development, and within countries by household wealth and gender. In this study we focus specifically on the characteristics of jobs in exports. We show how adding an industry dimension enhances the ability of occupational statistics to capture different types of worker activities. We extend the database of Timmer et al. (2019) increasing the country coverage to 52 countries at all levels of development and adding new data on activities at a higher level of detail. More precisely, activities are identified in the data by cross-classifying 13 occupational classes and 35 industries of work, resulting in 455 distinct activities. This fine-grained dataset allows us to describe for the first time overall development patterns of export specialization from an activity perspective.[2]

As a final note we wish to emphasize that the nature of this study is explorative. Our findings are suggestive of important complementarities between various export activities, but we stay agnostic about their precise nature. Co-occurrence of specialization in particular activities might be driven by comparable developments in countries' endowments such as the buildup of general human capital or the business environment. But it also points to the possibility of spillovers and

[2] Relatedly, Diodato et al. (2022) enrich export product data with occupational statistics. They assume however that production technologies are the same across countries applying occupation structures found in U.S. production to all countries. We show however that production technologies are not constant around the world such that products will not map one to one into a set of activities.

complementarities between activities due to for example the need for specific job skills, shared infrastructure or need for specialized inputs and services (Hidalgo et al., 2007; Coniglio et al., 2021). In addition, it may also be related to less tangible spillovers in knowledge and soft technologies associated with the entrepreneurial discovery process (Hausmann and Rodrik, 2003). The remainder of this paper is organized as follows. Section 2 discusses data development. Section 3 explores basic patterns in activity exports as countries develop over time. Section 4 reports on the importance of proximity in determining the development of new specializations. Section 5 explores the degree of path dependence in specialization patterns of countries and relates it to economic growth. Section 6 concludes.

2. Data sources for exports of activities

The analysis of trade in terms of activities rather than in terms of products or industries poses major data challenges. In this section, we discuss how these challenges have been met through combining two main data sets: data on exports of value added and data on the activities of workers. To this end we introduce variable z_{jc}, the value added of activity type j in country c's exports. It is tracked by the labor income of the workers carrying out the activity j in the production chains of country c's exports. The first data set contains information on labor income of workers by occupational group and industry of work. The second data set contains information on the exports of value added by industry. These exports include value added in the exporting industry as well as other domestic industries that contribute more upstream in the production chain through delivery of inputs. The z_{jc}'s are derived for each country and year by joining the two data sets at the country-industry dimension. A crucial part in data construction is to make sure that occupations and industries are consistently defined and measured across countries and over time. Data is constructed for a set of 52 economies, ranging from low-income to high-income, see Appendix table C1 for a full list.

Data set on activities. Activities of workers are identified on the basis of their occupational group as well as the industry in which they work. In total 455 activities are distinguished based on 13 occupational groups in each of 35 industries. The data is taken from the *Occupations Database*

(OD), introduced in Reijnders and de Vries (2018), and extended with a set of lower income Asian economies by Bertulfo et al. (2022), see Appendix Table C2 for the sources. The database provides information on the number of workers by occupation-industry pair as well as their labor earnings. For each country national data has been harmonized using mappings from national occupation classifications to an international classification of thirteen occupational groups based on 2-digit codes in the ISCO88 (International Standard Classification of Occupations), see Appendix table C3. While some countries have adopted the ISCO in their administrative statistics, most follow their own classification and bridge tables needed to be developed to have consistent coding. Furthermore, national industry classifications were mapped to a common set of 35 industries covering the overall economy. These are mostly 2-digit industries in the ISIC (International Standard Industrial Classification) revision 3.1, see Appendix table C4.

Previous studies have compiled large-scale data sets to compare changes in occupational structures across countries and over time. The *Jobs of the World Project* is a prominent example (Bandiera et al., 2022). We elaborate by adding a cross-classification of occupations by industry. This is an important step to account for the variety in tasks required from workers of the same occupational group in different industries. Compare for example the so-called work activities performed by machine operators in the textile manufacturing industry and similar workers in the electronics manufacturing industry. Both occupations fall under ISCO 1-digit code 8. Data from O*NET indicates that "getting information" is a very important work activity in both industries, ranking first in electronics and second in textiles (out of forty-one work activities that are distinguished). But whereas "evaluating information to determine compliance with standards" and "making decisions and solving problems" are also very important activities in electronics (ranking third and second), they are much less important in textiles (11^{th} and 10^{th}). Conversely, "repairing and maintaining mechanical equipment" is barely important in electronics (34^{th}) but above average important in textiles (17^{th}).[3] More generally, one can statistically test for the significance of adding an industry dimension to the occupational class category. To this end, we make use of O*NET's importance indicators for the full set of industries, occupations and work activities.[4] More

[3] Based on work activity data from O*NET for job codes 51-2022 - Electrical and Electronic Equipment Assemblers and 51-6062 - Textile Cutting Machine Setters, Operators, and Tenders (assessed 22 July 2022).
[4] This is data for the U.S. only. Caunedo et al. (2021) document some cross-country differences

specifically, we test for the significance of adding the industry variable to the 1-digit occupation variable in a two-way multivariate analysis of variance (MANOVA) of averaged importance scores. The joint occupation-industry variable is highly significant indicating that it is informative to account for the industry of work beside a worker's occupational group in characterizing the work activities carried out.[5]

Data set on exports. In order to determine what activities in an economy are exported, information on the value added of a countries' exports is needed. The industry composition of value added in exports is determined following the method outlined in Hummels et al. (2001) and Koopman et al. (2012). This is based not only on the value added in the exporting industry, but also on value added further up the production chain in other domestic industries delivering intermediates. Computation requires the use of national input-output tables that contain information on inter-industry deliveries, as outlined in Appendix A1. National input-output tables are taken from the Multiregional Input–Output table (MRIO) Database of the Asian Development Bank available for the year 2000 and annually for the period from 2007 to 2018. The industry classifications used in these tables were harmonized and mapped to a common set of 35 industries akin to the set of industries defined in the activities data.[6] Value added consists of compensation for workers (labor income) and a gross operating surplus (capital income) that accrues as income to the owners of capital assets. Capital assets cannot be straightforwardly allocated to activities, in contrast to workers. For example, a computer can be utilized in many activities and there is no information on its particular use. Data used for the baseline regressions only includes labor incomes. In

---

in the task content of a particular occupation, suggesting that the task content of occupations is not necessarily constant across countries.

[5] O*Net provides a list of 41 work activities with importance indicator scores on a scale of 0-100 for each 6-digit SOC code. We weight up the indicators to our activities using data from the US BEA Industry and occupation matrix for 2020. We establish the composition of O*NET 6-digit SOC codes in each of our activities, mapping from 6-digit SOC to our 13 occupational groups and from BEA 2-digit SIC to our industry groups. Next, we test for significant differences in the averaged importance scores across different activity pairs, weighting with the number of workers in each detailed SOC occupation. For ease of computation the 41 indicators are reduced to 10 principal components first, capturing 83% of the variation. Wilks' test result for the joint occupation-industry variable is (F-value = 4.92 and p-statistic = 0.00), for the occupation dimension (F= 46.98, p=0.00) and for industry dimension (F=0.54 and p=1.00).

[6] ADB MRIOTs are not available for the years 2001-2006. The MRIO database vintage used in this study was accessed in October 2021.

robustness analysis capital income is also taken into account assuming its distribution over activities is equal to the labor income distribution.

More formally, data on exports is used for a particular country-year to derive a vector $\boldsymbol{l}$ of dimensions (i x 1) with typical element representing export income by workers in industry i as

(1) $$\boldsymbol{l} = \hat{\mathbf{v}}\,(\mathbf{I} - \mathbf{A}^{\boldsymbol{D}})^{-1}\mathbf{e} \quad ,$$

with $\mathbf{e}$ a vector of exports (i x 1), $\mathbf{A}^{\boldsymbol{D}}$ the domestic input coefficients matrix (i x i) and $\mathbf{I}$ an identity matrix (i x i) such that $(\mathbf{I} - \mathbf{A}^{\boldsymbol{D}})^{-1}$ is the Leontief inverse matrix, and $\hat{\mathbf{v}}$ the diagonal matrix of labor income to gross output ratios (i x i). The values of export activities are derived by multiplying export income by industry with the occupation income distribution for each industry from the activity data set. Appendix A1 provides a more detailed technical exposition. Data for all variables vary across countries as well as over time allowing for meaningful cross-country and temporal comparisons.

3. Export activities along the development path

How do export baskets evolve in terms of activities as countries grow richer? Figure 1 shows the development of export activity over GDP per capita. It is estimated using a non-parametric LOWESS smoother on data for 59 countries over a time span of 20 years.[7] For parsimony, panel A aggregates the 455 activities into 5 broad groups: engineering, managerial, production, support, and other activities across all industries. Panel A suggests that at lower levels of economic development, production activities account for a major part of export income. The production share steadily declines when countries grow richer: from more than 50 percent of total exports at GDP per capita levels below 5,000 US$ to about 30 percent at levels above 40,000 US$. Engineering, managerial, and support activities account for the majority of labor income from exporting at higher levels of income. Panel B further details production activities in four broad industry groups. It shows that income from production activities within agriculture, mining, and "light" manufacturing industries (including food and textiles industries) shifts towards production

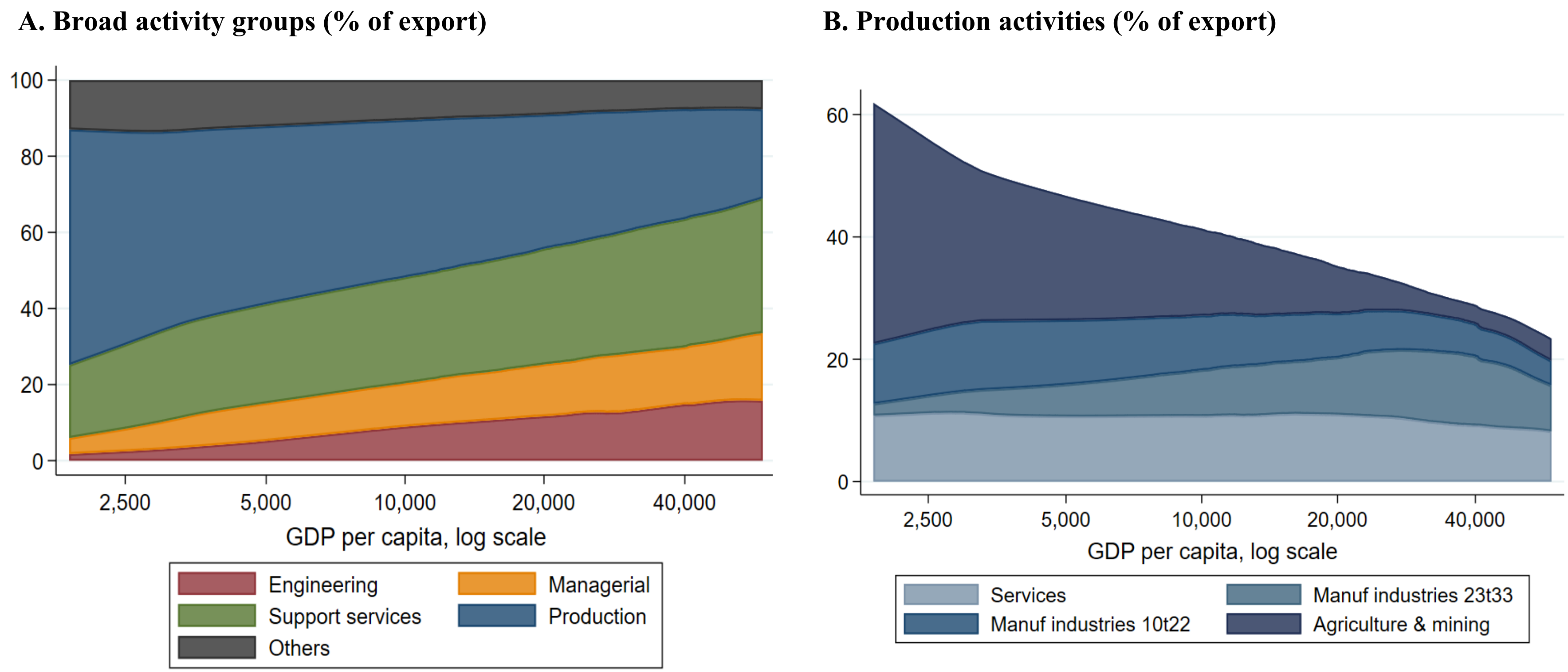
[7] For the purpose of Figure 1, data for eight countries has been added to the dataset, in particular data for some low-income countries, see Appendix G. However, the data for these eight countries have a lower level of detail and hence are not used in the rest of the paper.

activities in "heavy" manufacturing industries (including electrical and transport equipment industries). Remarkable is the increasing relative importance of production activities in services. These activities include, for example, work by drivers in the transport sector or cargo handlers in wholesaling, contributing indirectly to exports (Baccini et al, 2023). Production activities in services industries make up around one-tenth of exports across all income levels, while production activities in other industries rapidly decline during development.

The average cross-country development pattern is broadly reflected in individual country experiences. Figure 2 shows the changing activity content of exports in two rapidly growing countries, China and Vietnam, over the period from 2000 to 2018. Export of production activities decreased rapidly in China after 2007, whereas exports of engineering and support occupations increased. Export of production activities remained relatively high in Vietnam throughout this period. McCaig and Pavcnik (2018) stress the importance of reallocation of activity from informal firms toward formal firms during the export boom following the 2001 United States-Vietnam Bilateral Trade Agreement. Most noticeable is the shift towards production in heavy manufacturing, almost doubling from 7% in 2007 to 13% of export income in 2018.[8] In section 5, we will more formally investigate the extent to which countries follow a common development path in export specialization.

[8] See Pahl et al. (2022) and Winkler et al. (2023) for a complementary analysis of the job content of exports for a large set of countries.

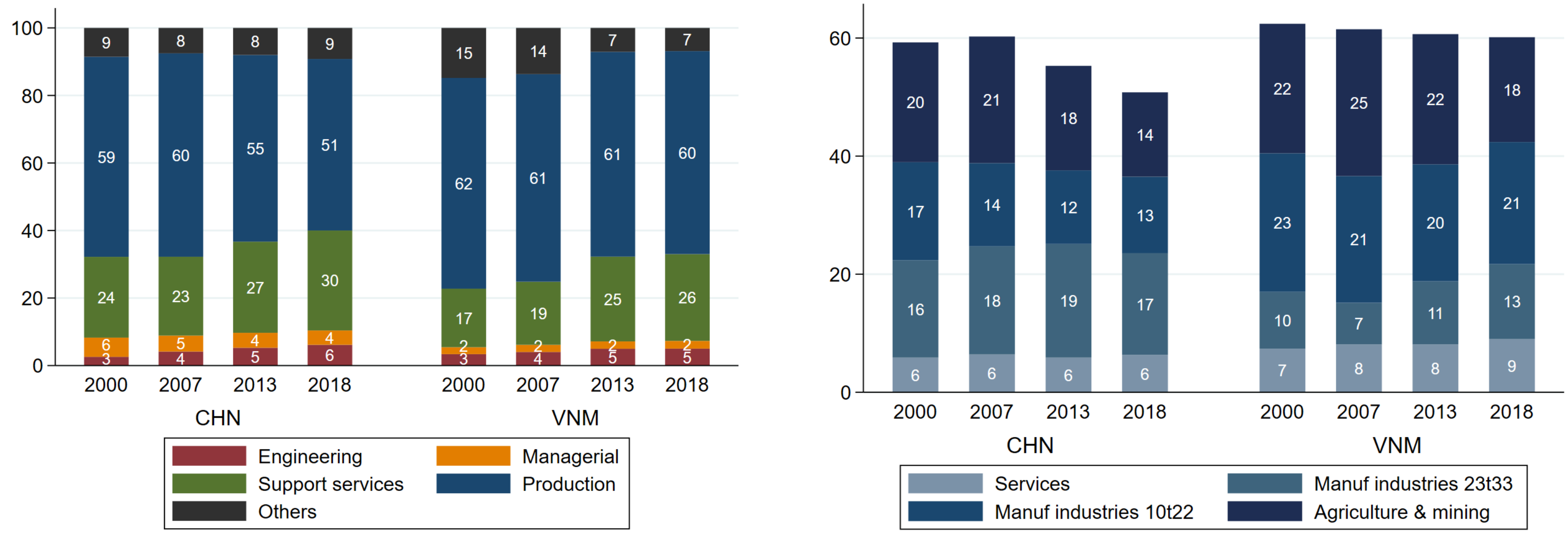
Figure 1: Export of activities over levels of economic development



Notes: Based on percentage shares of activities in overall export income for 59 countries and 20 years. Shares are plotted against GDP per capita (in 2017 US$, log scale) using a non-parametric LOWESS smoother with bandwidth 0.5. Broad groups are aggregated up from detailed activities and summed over all industries in panel A. Support services include: other professionals, clerical support workers, and sales workers; Production includes: craft workers and machine operators, agricultural workers, and drivers; Others include: legislators, health professionals, teachers, personal support workers; and other workers, see Appendix table C3. Further breakdown of production activities in panel B by industry in which production activity takes place: agriculture and mining refer to ISIC rev. 4 codes A and B, manufacturing industries to code C and services industries to codes D to U.
Source: Own calculations, see section 2, Appendix A and G. GDP per capita from Penn World Tables (Feenstra et al., 2015).

Figure 2: Export of activities, China and Vietnam

A. Broad activity groups (% of export) **B. Production activities (% of export)**



Notes: Percentage shares of activities in export incomes for China (CHN) and Vietnam (VNM). For further notes, see Figure 1 *Source*: Own calculations.

Changes in the activity basket of exports may be driven by changes along the extensive margin, that is shift of activity across industries within occupational classes, and changes along the intensive margin, that is shift of activity across occupational classes within industries. We quantify the relative importance of these shifts conducting a standard shift-share decomposition. The overall change in S_o, the share of occupation o in overall export income of a country in a particular period can be expressed as

(2) $$\Delta S_o = \sum_i \overline{S_i}\, \Delta S_{io} + \sum_i \overline{S_{io}}\, \Delta S_i,$$

where S_{io} denotes the share of workers with occupation o in industry i in overall export income, and S_i denotes the share of industry i in overall export income in the country. Δ denotes the change during the period and a bar over a variable indicates the period average of this variable. The first term on the right-hand side represents the change in the overall occupation share that is attributable to changes in the intensive (occupation) margin, while the second term reflects changes in the extensive (industry) margin.

Table 2 shows the decomposition for each of five broad occupational groups between 2000 and 2018. The set of economies is split into a group of advanced and a group of developing economies according to their income level in 2000 (listed in Appendix Table C1). Results are reported as simple averages across all economies in a group. The decline in export of production activities in developing economies is mostly accounted for by shifts along the extensive margin as workers move towards less production-intensive industries. As countries grow richer, shifts along the intensive margin start to dominate, in line with the production outsourcing hypothesis, also known as "servicification" of goods industries highlighted in Duernecker and Herrendorf (2022). Export shares of engineering activities continue to grow at all levels of development, mainly accounted for by shifts along the extensive margin. In contrast, the increase in export share of managerial activities is exclusively through the increasing shares within industries, both in developing and advanced economies. The relative importance of shifts along the intensive and extensive margins is further analyzed in the next section.

Table 2: Decomposition of changes in activity export shares, 2000-2018.

	Advanced economies			Developing economies		
Occupational group	Within industry	Between industry	Total	Within industry	Between industry	Total
Managerial	104.4	-4.4	100	138.2	-38.2	100
Support services	84.0	16.0	100	47.3	52.7	100
Production	57.0	43.0	100	22.0	78.0	100
Other occupations	41.8	58.2	100	-10.3	110.3	100
Engineering	32.6	67.4	100	39.4	60.6	100

Notes: The change in the share of a broad occupation group in export incomes is decomposed into between-industry and within-industry effects according to equation (2) for five broad aggregations of occupational groupings. Results are standardized by the total change within each broad aggregation and ordered by the within-industry component for advanced economies. Simple country averages are given for advanced and developing economies (listed in Appendix Table C1).

4. Development patterns of activity specializations

We make use of the full detail in our data set to probe particular patterns in the development of new export activities. To this end we make use of the concept of 'proximity' introduced by Hidalgo et al. (2007) and adapt this from a product to an activity perspective.

4.1 Measuring activity specialization

A standard way to describe a country's development in exporting is through revealed comparative advantage (RCA) indices introduced by Balassa (1965). The standard RCA index is calculated on the basis of export values of products as in Hidalgo et al. (2007). Instead, we define an activity specialization (AS) index based on activity incomes in exports. Denote z_{cj} the income of activity type j in country c's exports in a particular year. The AS index for activity j in country c is

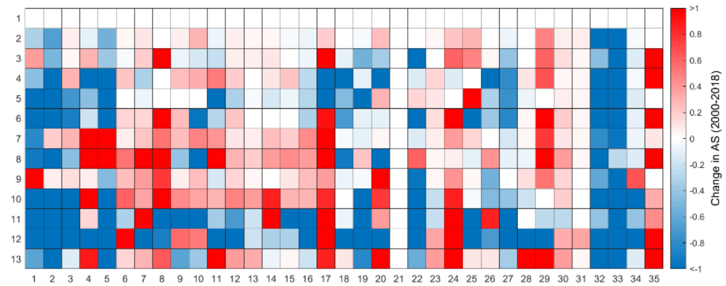
(3) $$AS_{cj} = \frac{\left(z_{cj}/\sum_j z_{cj}\right)}{\sum_c z_{cj}/\sum_c \sum_j z_{cj}} .$$

The numerator measures the share of activity j in overall activity income in country c's exports. The denominator calculates the same share across all countries. If the index is above one for a particular activity j, country c is said to be specialized in exporting that activity.

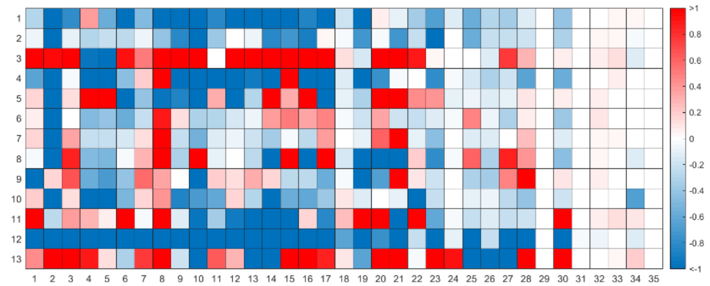
The AS indices can be used to track activity specialization in a country over time. Figure 3 provides heat maps for the change in the AS indices during the period 2000-2018 for four countries at different levels of development. Changes in the AS indices for Cambodia are relatively minor compared to the changes in the other countries. The AS for various occupations in hotels and restaurants (industry number 22) and in inland transport (23) industries went up. But in many manufacturing industries, AS indices barely changed for any occupational class. In contrast, AS indices in Vietnam changed frequently with major declines in mining (2), food (3) and leather manufacturing (5) industries, while in many other manufacturing industries and in utilities (17) and water transport (24) AS indices increased. In particular, AS for machine operators (occupation class 10) increased, but also for clerical support workers (7) and personal service workers (8). In Mexico AS developments diverged across occupations: AS for various occupations went up, for example in food manufacturing (3), petroleum refining (8) and transport equipment (15), but at the same time there were strong declines in AS for legislators (1) and managers (2) in the same industries. In China, a clear shift in specialization away from activities in goods producing industries (1-16) towards activities in services production (17-35) is visible. The strong increase in the AS of sales workers (9) is remarkable, also in most goods producing industries. Overall, it is noteworthy that in all countries AS indices within the same industry do not necessarily move in the same direction for all occupations. Both increases and declines in AS for particular classes are frequently found within the same industry illustrating that shifts along the extensive and intensive margins both play a role in determining export activity specializations.

Figure 3: Heat maps of changes in activity specialization indices

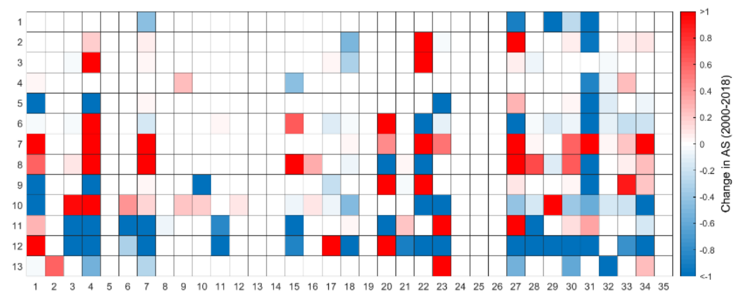
A. Vietnam



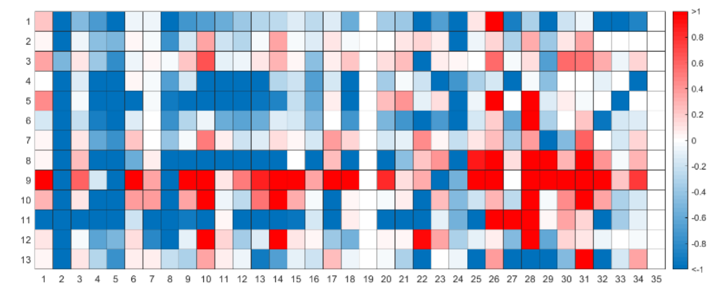
B. Mexico



C. Cambodia



D. China



Notes: Heat maps depicting the change in the activity specialization (AS) index for an activity over the period 2000-2018 for a particular country. Activities are cross-classified by 13 occupational classes (vertical axis) and 35 industries (horizontal axis; industries 3 to 16 are in the manufacturing sector). Red cells indicate a positive change in the AS index, while blue cells indicate a decline. The AS index calculated according to equation (3).

4.2 Regression setup

Hidalgo et al. (2007) found that the development of a particular product specialization depended strongly on its proximity to the initial product basket of a country. We adapt the concept of 'proximity' from a product to an activity perspective to investigate the importance of proximity for activity specializations. More specifically, we define proximity between two activities *x* and *y*, denoted as φ_{xy}, as the empirical probability that a country is specialized in activity x, conditional on being specialized in activity y:

$$(4) \qquad \varphi_{xy} = min\{\, P\big(AS_x \; > \; 1 \big|\, AS_y \; > \; 1\big), P\big(AS_y \; > \; 1 \big|\, AS_x \; > \; 1\big) \,\},$$

where probabilities are derived from the cross-country data as

$$(5) \qquad P\big(AS_y \; > \; 1 \big|\, AS_x \; > \; 1\big) = \frac{(\text{Number of Countries with } AS_x \; > 1 \text{ and } AS_y \; > 1)}{(\text{Number of Countries with } AS_x \; > 1)} \; .$$

A high φ_{xy} means that activity y is frequently found to be a specialization of countries that also have a specialization in activity x and vice versa.[9] Additionally, we define $\omega_{c,j,t}$ the average proximity of a particular activity *j* to the activities in which country *c* currently is specialized at time *t*, as

$$(6) \qquad \omega_{c,j,t} = \frac{\sum_r \big\{\mathbf{1} |\; AS_{c,r,t} > 1 \,\big\} \varphi_{j,r,t}}{\sum_r \varphi_{j,r,t}}$$

with $\big\{\mathbf{1} |\; AS_{c,r,t} \; > \; 1\,\big\}$ an indicator function being 1 when the condition that a particular activity *r* in which country *c* currently is specialized is met and 0 otherwise. A higher value of ω for say activity *j1* than for activity *j2* indicates that the country is currently specialized in activities that are on average more proximate to activity *j1* than *j2*. Subsequently, we will formally test whether it is also more likely for *j1* to develop into a new specialization.

The baseline linear probability regression is given by:

$$(7) \qquad x_{c,j,t+T} = \; \beta_0 + \beta_1 \omega_{c,j,t} + \beta_2 AS_{c,j,t} + \boldsymbol{\beta_z Z} + \varepsilon_{c,j,t},$$

[9] Using conditional probabilities as in (4) isolates the degree of activity proximity from their overall prevalence in trade. The proximity measure is symmetric which is a key requirement for using the measure in subsequent analysis. Instead of taking a minimum in (4) an average could be used.

where $x_{c,j,t+T}$ is binary being 1 when $AS_{c,j,t+1} > 1$ and 0 otherwise, ω the average proximity of j (see equation 6), t and $t+T$ indicate begin and end year of the period and **Z** is a vector of control variables (discussed below).[10] The regression only considers activities where the country initially does not have a comparative advantage in (i.e. $AS_{c,j,t} < 1$). The β_2 coefficient is expected to be positive as a higher initial AS level for an activity is likely to increase the possibility that it develops into a new specialization, that is, that it passes the threshold of AS=1. The main coefficient of interest is β_1. If it is positive, it indicates that a new activity which is more proximate to the current specialization basket of a country has a higher likelihood of developing into a new specialization.

4.3 Empirical results

The results of the baseline regression are reported in Table 3, with initial specialization and average proximity lagged with five years.[11] Country-year and industry-year dummies were added to control for any time-varying country or industry characteristics. In the baseline regression, a positive and significant coefficient on average proximity is found (column 1). A one standard deviation increase in average proximity of a new activity increases the probability of specializing in this activity by 6.1 percentage points in the full sample of economies. This positive and significant relation is found for both advanced and developing economies, with a stronger relation for the latter group (3.3 and 6.3 percentage points respectively, see columns 2 and 3).

[10] See Hausmann and Klinger (2007) for a comparable regression setup.

[11] Note that the period is 7 years for the initial period 2000-2007 and 5 years for the periods after using overlapping periods 2007-2012, 2008-2013, …, 2013-2018. This is because ADB MRIOTs are not available for the years 2001-2006. Robustness of the results to alternative lags and non-overlapping periods is considered below.

Table 3: Probability regressions for new activity specializations, baseline.

	Dependent variable: whether country c has a comparative advantage in activity j at time t+5 ($x_{c,j,t+5}$= 1)		
	(1)	(2)	(3)
$AS_{c,j,t}$	0.075***	0.190***	0.055**
	(0.025)	(0.006)	(0.022)
$\omega_{c,j,t}$	0.061***	0.033***	0.063***
	(0.006)	(0.003)	(0.005)
Country-Year FE	X	X	X
Industry-Year FE	X	X	X
Sample	Total	Advanced	Developing
N	137,723	56,919	80,804
Adjusted R^2	0.069	0.087	0.065

Notes: Results from linear probability regression using equation (4). The independent variables are activity specialization $AS_{c,j,t}$ and average proximity ω, defined in equations (3) and (6). ω is normalized by subtracting the mean and dividing by the standard deviation to ease interpretation. Robust standard errors (clustered by country) are reported in parentheses. * p<0.1, ** p<0.05 and *** p<0.01. Own calculations based on annual data for 2000-2018, see section 3.

Extensive and intensive margins We next investigate the predictive power of proximity for specializations along the extensive and intensive margins. Columns (1-3) in Table 4 report on activity specializations along the intensive margin by adding country-industry fixed effects. Columns (4-6) report on specializations along the extensive margin, adding a country-occupation dummy instead. The role of average proximity is highly significant for both types of specializations. For developing countries proximity appears slightly more relevant for specialization along the extensive margin than the intensive margin, with one standard deviation impact size of 5.2 and 5.0 percentage points respectively. At later stages of development, proximity for specialization along the extensive margin becomes much less important, with impact of only 1.4 percentage points for the extensive margin, and 3.3 percentage points for the intensive margin for the group of advanced economies.

Table 4: Probability regressions for new activity specializations at the intensive and extensive margins

	Dependent variable: whether country c has a comparative advantage in activity j at time t+5 ($x_{c,j,t+5}$= 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
$AS_{c,j,t}$	0.068***	0.171***	0.049***	0.083***	0.255***	0.058**
	(0.023)	(0.006)	(0.019)	(0.029)	(0.006)	(0.023)
$\omega_{c,j,t}$	0.050***	0.033***	0.050***	0.045***	0.014***	0.052***
	(0.005)	(0.003)	(0.004)	(0.005)	(0.003)	(0.004)
Year FE	X	X	X	X	X	X
Country-Industry FE	X	X	X			
Country-Occupation FE				X	X	X
Sample	Total	Advanced	Developing	Total	Advanced	Developing
N	137,723	56,919	80,804	137,723	56,919	84,804
Adjusted R^2	0.145	0.150	0.148	0.124	0.165	0.105

Notes: see Table 3.

Routine-intense activities Advances in information technology have changed the way in which certain tasks are performed. In particular, Autor et al. (2003) argue that computers and robots tend to displace labor in the performance of routine and non-cognitive tasks. This is nowadays typically referred to as routine-biased technological change. Lewandowski et al. (2020) and Caunedo et al. (2022) find that all countries experienced a shift away from routine to non-routine jobs, but its pace was slower in developing countries than in developed countries. Reijnders and de Vries (2018) show that relocation of routine occupations from advanced countries accounted for a major part of this difference, moderating the effect of the technology bias for developing countries. At the same time, innovation generates new tasks and activities and allows for the development of new specializations (Acemoglu and Restreppo 2019).

To investigate possible differences in the development of activity specializations, we constructed for each of our 455 activities a measure of routine task intensity. We closely follow the approach by Acemoglu and Autor (2011) and make a distinction between two types of routine tasks: manual and cognitive tasks based on O*NET measures.[12] Routine manual tasks are more prevalent in

[12] In this approach, O*NET work activities and work context importance scales are combined and standardized. However, whereas Acemoglu and Autor (2011) construct task measures by broad occupation classes, we match to detailed occupation-industry data from the U.S. Bureau of Labor Statistics and then collapse using labor supply weights to create task measures by activity. See Appendix F for details.

production and operative occupations, and routine cognitive tasks are most intensively used in clerical and sales occupations. We split the activity data accordingly into two samples and redo the baseline regressions for each sample, see Table 5. For activities with a high routine manual intensity, we indeed observe a major difference across advanced and developing economies. New specializations in routine manual activities strongly relate to proximity of the activities to current specialization. For developing economies, a 1 standard deviation increase in average proximity relates to a 10.9 percentage point increase in the probability of specializing in a new routine manual activity. In advanced economies the effect is only 2.9 percentage points. The impact of proximity on specialization in routine cognitive activities however is comparable to the impact for all activities found in the baseline (columns 4-6).

The qualitative results presented in Tables 3, 4 and 5 appear robust in a battery of additional tests, as shown in Appendix D. Taken together, the regression results show a strong commonality in the way activity export baskets develop over time, especially in earlier stages of development. Proximity to the initial product basket plays a strong role in the development of new product specializations. Low-income countries initially specialize across the extensive margin, exporting new activities carried out by workers of the same occupational grouping but in different industries. Proximity to the initial specialization basket is particularly important for the development of new routine manual activities in this phase. As income levels rise, specialization along the intensive margins becomes more prevalent, and workers shift towards different activities in the same industries. Proximity to the initial basket becomes less important for new specializations in this later phase of development.

Table 5: Probability regressions for new specializations in routine task intense activities

	Dependent variable: whether country c has a comparative advantage in activity j at time t+5 ($x_{c,j,t+5}$= 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
$AS_{c,j,t}$	0.071**	0.236***	0.051**	0.248***	0.218***	0.267***
	(0.030)	(0.009)	(0.024)	(0.009)	(0.012)	(0.014)
$\omega_{c,j,t}$	0.089***	0.029***	0.109***	0.053***	0.055***	0.055***
	(0.008)	(0.005)	(0.007)	(0.004)	(0.005)	(0.005)
Routine manual	0.001	0.009***	-0.001			
	(0.002)	(0.003)	(0.003)			
Routine cognitive				-0.016***	-0.030***	-0.008**
				(0.003)	(0.006)	(0.004)
Country-Year FE	X	X	X	X	X	X
Industry-Year FE	X	X	X	X	X	X
Sample	Total	Advanced	Developing	Total	Advanced	Developing
N	56,697	23,982	32,715	40,227	15,901	24,326
Adjusted R^2	0.074	0.098	0.078	0.112	0.093	0.132

Notes: see Table 3. Results in columns 1-3 only include observations where the routine manual task intensity of the activity is above the mean, and results in columns 4-6 where the routine cognitive task intensity is above the mean. Accordingly, the regressions control for manual (columns 1-3) and cognitive (columns 4-6) routine task intensity of the activity. Routine task intensity constructed following Acemoglu and Autor (2011), see Appendix F.

5. Path-dependence in activity specialization and growth

The cross-country average patterns found in the previous section are used to benchmark trade specialization patterns of individual countries in the vein of Coniglio et al. (2021), adapting their product-based methodology to the activity perspective. Coniglio et al. (2021) found that economic growth is weaker in countries with a higher degree of path dependence in their export specializations. We revisit this relationship using our activity export data. Following Coniglio et al. (2021) a country is denoted as "path-defying" when its new specializations are weakly related to its initial basket. For a particular country, the cumulative distribution function (CDF) of proximity for new activity specializations between time t and time $t + T$ is compared with a hypothetical random CDF of proximity for all possible new specializations, that is, activities in which the country did not have a comparative advantage yet.[13] Subsequently, stochastic

[13] Specialization is also shaped by the exit of activities in exports, i.e., de-specialization. This paper focuses on the entry of activities in which countries gain specialization.

dominance of the former over the latter distribution is tested to determine whether a country is path-defying.

To derive the distributions, we define the activity content of the initial export basket of country *c* at time *t*, $\boldsymbol{I_{ct}}$, as the set of activities in which the country has a comparative advantage, more formally:

$$(8) \qquad \boldsymbol{I_{ct}} \equiv \{j = 1, \dots, M \mid AS_{c,j,t} > 1\},$$

with *M* the total number of all activities.[14] The so-called option set $\boldsymbol{O}$ is the complement of set $\boldsymbol{I}$. It includes all activities in which a country does not have a specialization yet, but might develop one[15]:

$$(9) \qquad \boldsymbol{O_{ct}} \equiv \{j = 1, \dots, M \mid AS_{c,j,t} \leq 1\}.$$

Next, define new entries $\mathbf{N_{ct}}$ as activities in which a country actually acquires a new comparative advantage during [t, t + T] :

$$(10) \qquad \boldsymbol{N_{ct}} \equiv \{j = 1, \dots, M \mid AS_{c,j,t+T} > 1 \text{ and } AS_{c,j,t} \leq 1\},$$

such that $\mathbf{N_{ct}}$ is a subset of $\mathbf{O_{ct}}$.

For each entry ($n \in \boldsymbol{N_{ct}}$), we define the set $\boldsymbol{D}$ containing the proximity of *n* with all activities belonging to the initial export basket as $\boldsymbol{D_{nct}} = \{\varphi_{n,1,t}; \varphi_{n,2,t}; \dots; \varphi_{n,|I_{ct}|,t}\}$, with φ as defined in equation (4) and $|\boldsymbol{I_{ct}}|$ the number of elements in set $\boldsymbol{I_{ct}}$. The highest proximity of the entry *n* with any activity in the initial export basket is $d_{n,c,t} \equiv max\{\boldsymbol{D_{n,c,t}}\}$. A key variable of interest is the share of so-called path-defying entries. We classify for each country-year the entry as path-defying if the proximity to the initial export basket is lower than the mean proximity of the activities in the option set, $\mu_{ct} \equiv \frac{1}{|\boldsymbol{O_{ct}}|}\sum_{r \in \boldsymbol{O_{c,t}}} d_{r,c,t}$. The share of path defying entries in the set of entries (PDShare) during [t, t+T] is thus given by:

---

[14] Bold upper-case letters are used for sets.

[15] Note that the option set differs across countries as it is the complement of the set of activities already in a country's specialization basket.

(11) $$PDShare = \frac{\sum_{j \in N_{ct}} \{\mathbf{1} \mid \{AS_{j,c,t+T}>1 \text{ and } AS_{j,c,t} \leq 1\} \text{ and } \{d_{j,c,t} < \mu_{c,t}\}\}}{|\boldsymbol{N_{ct}}|}.$$

In addition, we can define a time-country specific empirical CDF of proximity for entries as:

(12) $$F_{ct}^{N}(d) \equiv \frac{1}{|\boldsymbol{N_{ct}}|} \sum_{n \in \boldsymbol{N_{ct}}} \mathbf{1}\{d_{n,c,t} \leq d\}$$

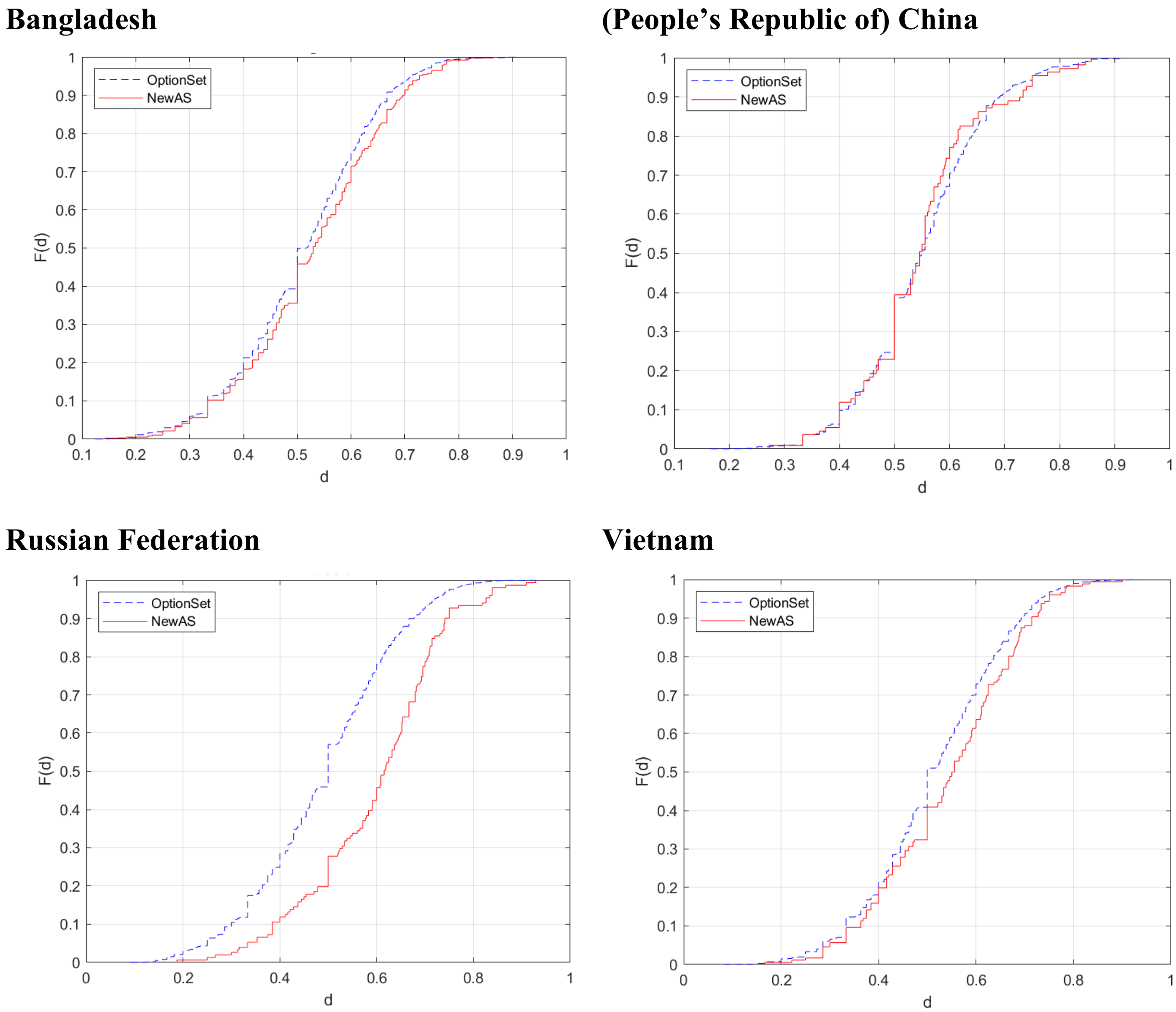
The hypothetical distribution of proximity is based on all activities that belong to the country's option set ($\boldsymbol{O_{ct}}$) rather than only the actual entries:

(13) $$F_{ct}^{O}(d) \equiv \frac{1}{|\boldsymbol{O_{ct}}|} \sum_{r \in \boldsymbol{O_{ct}}} \mathbf{1}\{d_{r,c,t} \leq d\}.$$

When the CDF of proximity of the actual entries, $F^N(d)$, is equal or larger than the hypothetical CDF, $F^O(d)$, for all $d \in [0,1]$ a country is denoted as path defying. The null hypothesis of path-defiance is tested in a two-sample one-sided Kolmogorov–Smirnov tests for first order stochastic dominance of the distribution of (12) over the distribution of (13). The intuition of this test is based on the insight that at any point in time a country has a large number of activities in its option set for which it has not (yet) developed a comparative advantage. These potential new entries differ in proximity to the initial specialization basket. Path-defiance is rejected when new specializations are significantly more concentrated at higher levels of proximity.

Figure 4 illustrates that the actual and hypothetical CDFs are very close in the case of China. The null hypothesis cannot be rejected (at the 1 percent significance level), indicating that the Chinese specialization pattern in the period 2000 to 2018 was path-defying. In contrast, the test strongly rejects the null for the Russian Federation, indicating that its export development path is heavily influenced by the "pull" of its initial specialization pattern. During this period the share of path-defying entries in new specializations was 52.3% in China and only 19.9% in Russia. The hypothesis of path defiance is also (marginally) rejected for Bangladesh as well as for Vietnam, although their share of path-defying entries is much larger than for Russia (45.8 and 40.9 percent). Appendix Table E provides path-defying entry shares and test results for all countries in the dataset.

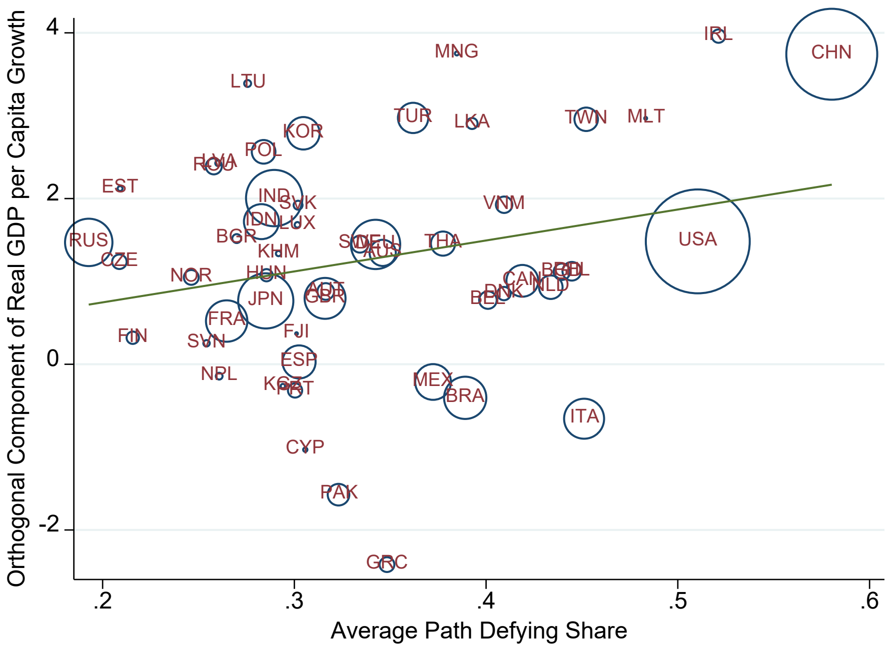
Figure 4: Path defiance: comparing proximity of new entries and option set to initial basket



Notes: Cumulative distribution functions of proximity of new entries to initial export baskets: actual entries **N** versus option set **O**, see equations (12) and (13).

Is there a relation between the pace at which countries introduce new activity specializations and their overall economic growth performance? Figure 5 plots GDP per capita growth against the share of path-defying entries for the period 2000-2018 (controlling for initial GDP per capita levels). The regression line is suggestive of a positive relationship but variance around it is high. Following Coniglio et al. (2021), a parsimonious growth model is subsequently estimated including various standard growth covariates. We like to emphasize that the results are only indicative and cannot be considered as causal evidence for a growth relationship. The estimation mainly serves to investigate to what extent results from an activity perspective will differ from results based on a product perspective.

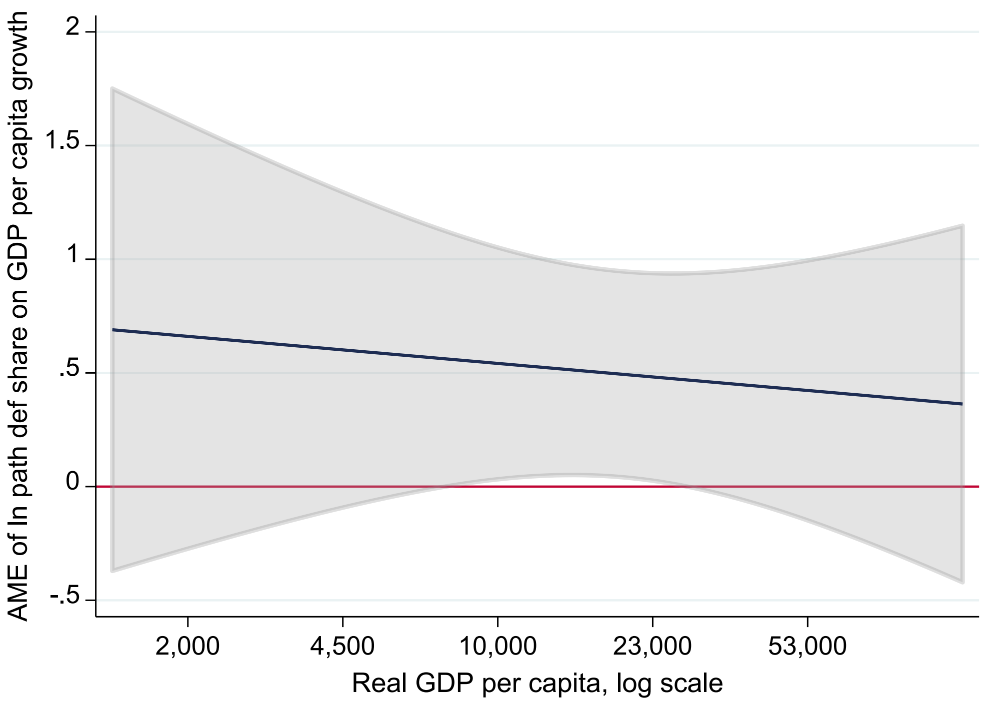
Figure 5: Economic growth and degree of path-defying specialization



Notes: Figure plots the orthogonal component of average real GDP per capita growth rate against the average share of path defying entries (as defined in equation 11) across the eight overlapping periods for each country. Regression control for initial (log) GDP per capita. Slope (standard error) of the linear fit is 3.73 (2.18). Size of circles represent country size measured as average real GDP in 2017 US$.
Sources: Authors' calculations based on Penn World Table version 10.0, ADB MRIOTs and Occupations Database.

Additional data sources and regression results are given in Appendix B. The dependent variable is the average growth rate of GDP per capita in the period [t, t + 5]. The key variable of interest is the share of path-defying entries which appears to be strongly significant and positively related to growth. Interestingly, the positive impact remains even after controlling for factor endowments in terms of human capital levels and financial development. This suggests that the capabilities needed to undertake new path-defying activities go beyond the mere expansion of educational attainment levels and general development of financial markets in a country. It is also found that the impact of path defiance on growth is moderated by the initial level of GDP per capita: it appears to be stronger for poorer countries than for richer countries. The average marginal effect (AME) is positive at all levels of development albeit only significantly different from zero (at 95% confidence) for GDP per capita levels between about 9,000 and 30,000 US$ (Figure 6). This qualifies the finding of Coniglio et al (2021, Figure 5) who also find that the marginal effect is positive for the poorest countries and declining with higher incomes, yet in contrast establish that the effect is small and turns negative for middle-income countries. Confidence intervals are too large to draw firm conclusions, but the different findings do illustrate the potential for analyses based on the activity perspective to generate new insights beyond those based on the product perspective.

Figure 6: Average marginal effects of path-defying entries on economic growth



Notes: Figure shows average marginal effects (AME) of path-defiance on GDP per capita growth at various levels of economic development. Effects calculated based on regression estimates reported in column 2 of Appendix Table B1. Point estimates (in blue line) and 95 confidence intervals (in grey) are shown over levels of (log) GDP per capita.

6. Concluding remarks

This study documented for the first-time patterns in export specialization from an activity perspective and showed how it can generate additional insights beyond those based on the product perspective. First, export incomes from production activities decline and engineering, managerial, and services support activities grow as countries develop. Second, countries initially specialize along the extensive margin (shifting activities across industries) but later along the intensive margin (shifting activities across occupational classes). Third, new activity specialization is strongly related to the proximity of this activity to the initial export basket, in particular for specializations along the extensive margin and in routine intensive occupations. Fourth, countries that defy proximity and diversify quicker appear to grow faster in GDP per capita. This is correlation however and no claim for causation is made.

We see at least two promising avenues for further research. One avenue is in the modeling of structural change and the role of international trade. The canonical macro-structural change framework focuses on the sectoral composition of the economy in terms of employment and value-added. Trade can shape the sectoral composition in various ways (Alessandria, Yi and Johnson 2021). Lower trade barriers facilitate specialization, for example through shifting comparative advantage and promoting economies of scale. Sectoral specialization will consequently affect the sectoral composition of the economy. And given a set of trade barriers, policy changes or technology shocks to the economy may also affect specialization patterns and consequently sector composition. Trade barriers are typically related to products whereas technological change such as automation affects particular activities rather than products or sectors. Modeling the composition of the economy in terms of activities in addition to sectors appears therefore to be a promising way forward as for example in Bárány and Siegel (2018) and Duernecker and Herrendorf (2022).

A second avenue for research is in further developing an integrated product and activity data set. Our data tracks exports at a rather aggregate industry level compared to the detailed product level export data that is available. Suppose that production technologies for a particular detailed product are the same around the world. In that case, cross-country differences in the occupational composition of the exporting industry (as documented in this study) are due to a different set of products being exported. In that case, activity specialization across the extensive margin can also be studied from a product perspective as in Diodato et al. (2022). If production technologies are not constant around the world, however, detailed products will no longer map one to one into a set of activities. Caunedo et al. (2021) document some cross-country differences in the task content of a particular occupation, suggesting that a description of production technologies in terms of occupational structures is only a first step towards better understanding of activity specializations.

These avenues are worthwhile to pursue also for the purpose of policy. Our findings re-open the debate on the appropriate target of development and trade policies. The product space paradigm by Hidalgo et al. (2007) has been used to guide countries in the design of industrial policies. According to this paradigm, policy makers should follow a gradualist approach and focus on introducing new products that are close to their current product mix to avoid failure, as introduction of radically new products purportedly requires capabilities that are scarce and difficult to create (Coniglio et al., 2021). Our results highlight that products that might seem far apart (close) in product space might actually be rather close (far apart) in terms of activities. As such, an activity-based analysis provides a clearer link with the export capabilities that are required. This study and the associated database hopefully provide a fruitful steppingstone for deriving well-founded policy implications for trade development and structural change.

References

Acemoglu, D., Autor, D.H., 2011. Skills, tasks and technologies: Implications for employment and earnings. In: Ashenfelter, O., Card, D.E. (Eds.), Handbook of Labor Economics, vol. 4. Elsevier, Amsterdam, pp. 1043–1171.

Acemoglu, D., Restrepo, P., 2019. Automation and new tasks: How technology displaces and reinstates labor. J. Econ. Perspect. 33 (2), 3-30.

Alessandria, G.A., R. C. Johnson, and K.-M. Yi, 2021. Perspectives on Trade and Structural Transformation. NBER Working Paper 28720

Antrás, P. and Chor, D., 2022. Global value chains. In G. Gopinath, E. Helpman and K. Rogoff (eds.) Handbook of International Economics: International Trade, Vol. 5, Amsterdam: Elsevier.

Autor, D., Murnane, R.J., Levy, F., 2003. The skill content of recent technological change: An empirical exploration. Q. J. Econ. 118 (4), 1279–1334.

Baccini, L, Matteo Fiorini, Bernard Hoekman, Marco Sanfilippo, 2023. Services, Jobs, and Economic Development in Africa, The World Bank Research Observer, Volume 38, Issue 1, Pages 147–178, https://doi.org/10.1093/wbro/lkac006.

Baldwin, R., Robert-Nicoud, F., 2014. Trade in goods and trade in tasks: An integrating framework. J. Int. Econ. 92, 51–62.

Balassa, B., 1965. Trade liberalisation and 'revealed' comparative advantage. Manch. Sch. 33, 99–123.

Bandiera, O., Elsayed, A., Heil, A., Smurra, A., 2022. Economic Development and the Organisation Of Labour: Evidence from the Jobs of the World Project, Journal of the European Economic Association, 20 (60), 2226 – 2270.

Bárány, Z., Siegel, C. 2018. Job polarization and structural change. Am. Econ. J.: Macroecon. 10 (1), 57 – 89.

Bertulfo, D., E. Gentile, and G. J. de Vries, 2022. "The employment effects of technology, trade, and consumption in global value chains: Evidence for Developing Asia." Asian Development Review, vol. 39(2), pp 1-44. https://doi.org/10.1142/S011611052250010X.

Caunedo, J., Keller, E., Shin, Y., 2021. Technology and the task content of jobs across the development spectrum. NBER Working Paper No. 28681.

Chenery, H., Robinson, S. and Syrquin, M., 1986. Industrialization and Growth A Comparative Study. Oxford University Press, Oxford.

Coniglio, N., Vurchio, D., Cantore, N., Clara, M., 2021. On the evolution of comparative advantage: path dependent versus path defying changes. J. Int. Econ. 133, 103522.

Darvas, Z. (2012). Real effective exchange rates for 178 countries: a new database. Working Paper. Department of Mathematical Economics and Economic Analysis, Corvinus University of Budapest, Budapest.

Diodato, D. R. Hausmann, U. Schetter, 2022, A Simple Theory of Economic Development at the Extensive Industry Margin. CID Faculty Working Paper No. 416.

Duernecker, G., Herrendorf, B., 2022. Structural transformation of occupation employment. Economica 356 (89), 789-814.

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer, 2015. The Next Generation of the Penn World Table, Am. Econ. Rev., 105(10), 3150-3182.

Grossman, G.M., Rossi-Hansberg, E., 2008. Trading tasks: A simple theory of offshoring. Am. Econ. Rev. 98 (5), 1978–1997.

Hanson, G. H., 2012. The rise of middle kingdoms: Emerging economies in global trade. Journal of economic perspectives, 26(2), 41-64.

Harrison, A., Rodríguez-Clare, A., 2010. Trade, foreign investment, and industrial policy for developing countries. In: Rodrik, D., Rosenzweig, M. (Eds.), Handbook of Development Economics, vol. 5. Elsevier, Amsterdam, pp. 4039-4214.

Hausmann, R., Klinger, B., 2007. The structure of the product space and the evolution of comparative advantage. CID Working Paper No. 146.

Hausmann, R., Rodrik D., 2003. Economic development as self-discovery. J. Dev. Econ. 72, 603–633.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., & Simoes, A., 2014. The atlas of economic complexity: Mapping paths to prosperity. Mit Press.

Herrendorf, B., Rogerson, R., & Valentinyi, A., 2014. Growth and structural transformation. Handbook of economic growth, 2, 855-941.

Hidalgo, C., Klinger, B., Barabási, A., Hausmann, R., 2007. The product space conditions the development of nations. Science 317, 482 487.

Hummels, D., Ishii, J., Yi, K.-M., 2001. The nature and growth of vertical specialization in world trade. J. Int. Econ. 54, 75–96.

Imbs, J., and R. Wacziarg, 2003. Stages of Diversification. Am. Econ. Rev., 93 (1), 63-86.

Johnson, R. C., & Noguera, G., 2012. Accounting for intermediates: Production sharing and trade in value added. Jour. of Int. Econ., 86(2), 224-236.

Johnson, R.C., Noguera, G., 2017. A portrait of trade in value added over four decades. Rev. Econ. Stat. 99, 896–911.

Kali, R., Reyes, J., McGee, J., & Shirrell, S., 2013. Growth networks. Journal of Development Economics, 101, 216-227.

Kaufmann, D., Kraay, A., & Mastruzzi, M., 2011. The worldwide governance indicators: Methodology and analytical issues1. Hague journal on the rule of law, 3(2), 220-246.

Koopman, R., Wang, Z., Wei, S.J. , 2012. Estimating domestic content in exports when processing trade is pervasive. J. Dev. Econ. 99 (1), 178-189.

Koopman, R., Wang, Z., & Wei, S. J., 2014. Tracing value-added and double counting in gross exports. Am. Econ. rev., 104(2), 459-494.

Kruse, H., Mensah, E., Sen, K., & de Vries, G., 2022. A manufacturing (re) naissance? Industrialization in the developing world. IMF Economic Review, 1-35.

Lashitew, A.A., Ross, M.L., Werker, E., 2021. What drives successful economic diversification in resource-rich countries? World Bank Res. Obs. 36 (2), 164-196.

Lederman, D., & Maloney, W., 2012. Does what you export matter?: In search of empirical guidance for industrial policies. World Bank Publications.

Lewandowski, P., Park, A., Hardy, W., Du, Y., Wu, S., 2022. Technology, skills, and globalization: Explaining international differences in routine and nonroutine work using survey data. World Bank Econ. Rev. 36 (3), 687-708.

McCaig, B., Pavcnik, N., 2018. Export markets and labor allocation in a low-income country. Am. Econ. Rev. 108 (7), 1899-1941.

Mensah, E.B., de Vries, G. J., 2023. How important are exports for job growth in sub-Saharan Africa? A demand side analysis based on new time series Supply and Use tables. Mimeo.

Newfarmer, R.S., Page, J.M., Tarp, F. (eds.), 2018. Industries without smokestacks: Industrialization in Africa reconsidered. Oxford University Press, Oxford, pp. 170-190.

Pahl, S., Timmer M.P., Gouma, R., Woltjer, P.J., 2022. Jobs and productivity growth in global value chains: New evidence for twenty-five low- and middle-income countries. World Bank Econ. Rev. 36 (3), 670-686

Reijnders, L.S.M., de Vries, G.J., 2018. Technology, offshoring and the rise of non-routine jobs. J. Dev. Econ. 135, 412-432.

Rodrik, D., 2018. New technologies, global value chains, and developing economies. NBER Working Paper No. 25164.

Syrquin, M., 1988. Patterns of Structural Change. In: Chenery, H. and Srinivasan, T.N., Eds., Handbook of Development Economics, Elsevier, New York, 203-273.

Timmer, M.P., Miroudot, S., de Vries, G.J., 2019. Functional specialisation in trade. J. Econ. Geogr. 19 (1), 1-30.

UNDP, 2017. National logistics strategy. UNDP Ethiopia, Addis Ababa.

Winkler, Deborah, Luis A. Lunar, Hagen Kruse, and Maryla Maliszewska. 2023. "Linking trade to jobs and activities: new stylized facts for low- and middle-income countries", mimeo World Bank.

Wolff, E., 2003. Skills and changing comparative advantage. Rev. Econ. Stat. 85: 77–93.

World Bank, 2020. World Development Report 2020: Trading for development in the age of global value chains. World Bank, Washington, DC.

Appendix A. Measuring the activity content of exports

Let **e** be a vector of exports (of dimension i x 1) with i the number of industries in a particular country.[16] Let $\mathbf{A}^{\boldsymbol{D}}$ be the i x i domestic coefficient matrix with typical element $a_{i_1 i_2}$ indicating output from domestic industry i_1 used in production of one unit output of industry i_2 (all in nominal terms). We can then derive a vector **y** (i x 1) which represents the total gross output needed in each industry to produce exports as:

(A1) $$\mathbf{y} = (\mathbf{I} - \mathbf{A}^{\boldsymbol{D}})^{-1}\mathbf{e},$$

where **I** is a i x i identity matrix with ones on the diagonal and zeros elsewhere. $(\mathbf{I} - \mathbf{A}^{\boldsymbol{D}})^{-1}$ is the well-known Leontief inverse matrix which ensures that all output related to exports is taken into account, not only by the industry that is exporting, but also output of other domestic sectors that contribute through the delivery of intermediate inputs. We define a vector $\boldsymbol{l}$ of dimensions (i x 1) with typical element representing export income by workers in industry i. It can be derived by pre-multiplying **y** as given in equation (A1):

(A2) $$\boldsymbol{l} = \hat{\mathbf{v}}\mathbf{y} = \hat{\mathbf{v}}(\mathbf{I} - \mathbf{A}^{\boldsymbol{D}})^{-1}\mathbf{e}\ ,$$

where $\hat{\mathbf{v}}$ is the matrix (i x i) with diagonal element v_{ii} representing the labor income to gross output ratios for industry i and zeroes elsewhere. [17]

Next, let **B** be a matrix of dimension o x i with typical element b_{oi} denotes the labor income of workers having occupation o in industry i, expressed as a share of labor income in industry i. We derive matrix **Z** of dimensions (o x i) with a typical element z_{oi} representing exported labor income by workers having occupation o in industry i as

(A3) $$\mathbf{Z} = \mathbf{B} \odot [\mathbf{1} \otimes \boldsymbol{l}^{\boldsymbol{T}}],$$

with $\odot$ element-wise multiplication (Hadamard Product), $\mathbf{1}$ is the column vector whose elements are all equal to one, $\otimes$ is the Kronecker product and T indicating vector transposition. The elements of **Z** are used to calculate the activity specialization indices defined in the main text.

[16] In the setup of models of value-added trade, each product is associated with a sector, so i is also the number of products.

[17] Where the "^" indicates it is a matrix with the values from the vector on the main diagonal.

Note that the calculations are country- and year-specific. Put otherwise, $\mathbf{B}, \mathbf{v}, \boldsymbol{A^D}$ and $\mathbf{e}$ in (A2) and (A3) vary across countries as well as over time.

Appendix B. Cross country growth regressions for activity specialization

To test the relationship between the degree of path defiance and economic growth, we follow Coniglio et al. (2021) and regress GDP per capita growth on the share of path-defying entries plus a variety of economic growth controls. Appendix Table B1 reports on cross-country growth regressions for the sample of 52 economies. The dependent variable is the average growth rate of GDP per capita in the period [t, t + 5]. The share of path defying entries in the set of entries (PDShare) during [t, t+T] is given in equation (11) in the main text. We follow Coniglio et al. and use ln(PD_share*100) as the main explanatory variable. Results in column (1) suggest that a higher path-defiance in a country's specialization pattern is associated with higher GDP per capita growth. Results in column (2) are based on a regression which includes an interaction term with the level of GDP per capita. The average marginal effect of path-defiance is comparable to the marginal effect in the regression without interaction. It suggests that countries with a higher degree of path-defying specialization have faster growth in GDP per capita, albeit the relationship is not significant at conventional significance levels for any level. Various standard growth covariates such as the level of trade openness and human capital are included. Also, we control for the total number of new activities specializations, to account for changes in the denominator for the degree of path dependence. See Appendix Table B2 for the data sources of the growth covariates. All covariates appear to be significant, except for human capital and political stability (column 2). Interestingly, the positive impact of path-defiance remains even after controlling for factor endowments in terms of human capital levels and financial development. The relationship appears to be moderated by the initial level of GDP per capita. Average marginal effects (AME) of path defiance based on the results in column 2 are given in Figure 6.

For robustness, two alternatives for the measurement of the path-defying entries are considered. Results in columns 3 and 4 are based on a less stringent definition of path defying, using the mean proximity value plus one standard deviation as an alternative threshold. Alternatively, values of the percentile rank of the entry, denoted as $\mathrm{p}_{\mathrm{n,c,t}} \equiv \left(\mathrm{F}^{\mathrm{O}}_{\mathrm{c,t}}\right)^{-1}\left(\mathrm{d}_{\mathrm{n,c,t}}\right)$ are used as explanatory variable (using the average of all new entries in each period). This third measure informs in which percentile of the hypothetical distribution of proximity (derived from the option set) each entry falls. Low percentiles are associated with entries that are poorly related to the country's initial export basket and are therefore path-defying. Results using this alternative measure are given in columns 5 and 6. The average marginal effect at the mean is higher for both alternative measures than in the base line.

Appendix Table B1: Cross-country growth regressions with path-defying specialization

	Dep variable: average GDP per capita growth rate [t, t+5]					
	(1)	(2)	(3)	(4)	(5)	(6)
Definition of path defying / *Independent variables*	*Option 1*	*Option 1*	*Option 2*	*Option 2*	*Option 3*	*Option 3*
ln path def share [t, t+5]	0.596 (0.228)***	1.212 (1.804)	1.999 (0.519)***	5.671 (5.696)	1.554 (0.453)***	1.814 (4.211)
ln path def share [t, t+5] x ln GDP per capita [t]		-0.073 (0.181)		-0.400 (0.563)		-0.053 (0.426)
ln GDP per capita [t]		-1.053 (0.662)		0.508 (2.382)		-1.044 (1.511)
ln total number of new entries	-0.872 (0.199)***	-0.789 (0.206)***	-0.809 (0.191)***	-0.750 (0.194)***	-0.870 (0.193)***	-0.800 (0.198)***
ln Population [t]	0.420 (0.077)***	0.321 (0.077)***	0.400 (0.077)***	0.308 (0.078)***	0.400 (0.077)***	0.308 (0.078)***
Resource Rents / GDP [t]	0.109 (0.018)***	0.097 (0.016)***	0.118 (0.018)***	0.107 (0.017)***	0.115 (0.018)***	0.104 (0.017)***
Human Capital [t]	-0.557 (0.198)**	0.098 (0.278)	-0.512 (0.197)**	0.110 (0.283)	-0.491 (0.198)**	0.135 (0.279)
Financial Development [t]	-8.287 (0.734)***	-5.565 (0.972)***	-7.933 (0.709)***	-5.439 (0.934)***	-8.050 (0.722)***	-5.473 (0.958)***
Political Stability [t]	0.050 (0.199)	0.080 (0.191)	0.078 (0.193)	0.069 (0.191)	-0.039 (0.199)	-0.014 (0.195)
Rule of Law [t]	0.933 (0.245)***	1.048 (0.231)***	0.872 (0.234)***	0.990 (0.222)***	0.861 (0.238)***	0.976 (0.225)***
ln Trade Openness [t]	0.362 (0.210)*	0.531 (0.203)***	0.425 (0.210)**	0.583 (0.201)***	0.417 (0.210)*	0.582 (0.201)***
AME of PD share		*0.494* *(0.230)***		*1.710* *(0.553)****		*1.288* *(0.472)****
Period FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	405	405	408	408	408	408
Economies	51	51	51	51	51	51
Adj-R^2	0.528	0.554	0.540	0.563	0.537	0.560

Notes: Regressions on GDP per capita growth rate across the eight overlapping periods using period fixed effects. Path-defying entries in columns (1) and (2) are defined as entries with a lower proximity to a country's initial export basket than the average proximity in its option set (option 1). Path defying entries in columns (3) and (4) have proximity lower than the average proximity value plus one standard deviation (option 2) and in columns (5) and (6) the average percentile of the random hypothetical distribution of proximity (derived from the option set) is used (option 3). AME of PD share is the average marginal effect of PD share on GDP per capita growth. Chinese Taipei is not included in the analysis due to missing observations for several control variables. Robust standard errors are reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Appendix table B2: Additional variables description and sources for cross-country growth regressions

Short name	Full description	Source
GDP per capita	Rgdpna/pop; levels in logs; average growth rate in percentages	PWT 10 (Feenstra et al. 2015)
Human capital	Index measure [1, 4]; SD = 0.56; as 5-year change variable [-0.02, 0.15] & SD = 0.02	PWT 10 (Feenstra et al. 2015)
Financial development	Index measure [0, 1] (no observations for Chinese Taipei); SD = 0.25; as 5-year change variable [-0.04, 0.04] & SD = 0.01	IMF
Terms of trade	Net barter terms of trade index; in regressions only used as 'forward-looking' five year average annual growth rate; [-0.1, 0.1] & SD = 0.20	World Bank World Development Indicators
Political stability	WGI 'Political stability and absence of violence/terrorism' indicator [-3, 2]; SD = 0.92; as 5-year change variable [-0.25, 0.25] & SD = 0.06	World Bank World Development Indicators; Kaufmann, Kraay, and Mastruzzi (2010)
Rule of law	WGI 'Rule of law' indicator [-1.4, 2.1]; SD = 0.97; as 5-year change variable [-0.13, 0.12] & SD = 0.03	World Bank World Development Indicators; Kaufmann, Kraay, and Mastruzzi (2010)
Economic complexity	Index measure of a country's export sophistication [-1.5, 3]; SD = 0.85; as 5-year change variable [-0.14, 0.10] & SD = 0.03	Harvard Growth Lab; Hausmann et al. (2014)
Resource rents / GDP	Total natural resources rents (of GDP) including oil, gas, coal, mineral, and forest [0, 42.3] (no observations for Chinese Taipei) & SD = 6.38	World Bank World Development Indicators
REER	Real effective exchange rate, "broad annual index" considering 172 trading partners; in regressions only used as 'forward-looking' five year average annual growth rate; [0.06, 0.09] & SD = 0.02	Bruegel; Darvas (2012, data update 2021)
Trade openness	(Exports + Imports) / GDP, all in current PPPs [0.04, 4.56]; SD = 0.69; as 5-year change variable [-0.17, 0.25] & SD = 0.04.	PWT 10 (Feenstra et al. 2015)

Appendix C. Detailed sources and classifications

Appendix table C1: Economies included in the analysis, split by advanced and developing

	Advanced		Advanced		Developing		Developing		Developing
1	Australia	12	Japan	1	Bangladesh	12	Indonesia	23	Romania
2	Austria	13	Luxembourg	2	Brazil	13	Kyrgyzstan	24	Russian Federation
3	Belgium	14	Netherlands	3	Bulgaria	14	Latvia	25	Slovak Republic
4	Canada	15	Norway	4	Cambodia	15	Lithuania	26	Slovenia
5	Denmark	16	Portugal	5	China	16	Malta	27	Sri Lanka
6	Finland	17	Korea, Rep.	6	Cyprus	17	Mexico	28	Thailand
7	France	18	Spain	7	Czechia	18	Mongolia	29	Türkiye
8	Germany	19	Sweden	8	Estonia	19	Nepal	30	Vietnam
9	Greece	20	Chinese Taipei	9	Fiji	20	Pakistan		
10	Ireland	21	United Kingdom	10	Hungary	21	Philippines		
11	Italy	22	United States	11	India	22	Poland		

Appendix Table C2: Sources Occupations Database

Country	Source(s)	Years
Australia	Labor Force: Employed Persons Quarterly Large Source Dataset	2000-2011
Bangladesh	Labor Force Survey (LFS)	2006, 2010, 2013, 2016
Brazil	National Household Sample Survey (PNAD)	2001-2015
Cambodia	Cambodia Socio-Economic Survey (CSES)	2003/2004, 2007-2017
Canada	Canadian Labor Force Survey	2000-2014
China	Population census, IZA wage indicator survey [a]	2000, 2010
EU members and Norway[b]	Labour force survey [c], Structure of earnings survey [d]	2000-2013
Fiji	Employment and Unemployment Survey (EUS)	2004, 2005, 2010, 2011, 2015, 2016
India	National Sample Survey – Employment Unemployment Survey (NSS-EUS)	1999/2000, 2004/2005, 2011/2012
Indonesia	National Labor Force Survey (SAKERNAS)	2000, 2003, 2005, 2008, 2010-2017
Japan	Population census, wage structure surveys	1995, 2000, 2005, 2010, 2015
Kyrgyz Republic	Kyrgyzstan Integrated Household Survey (KIHS)	2012-2018
Mexico	Population census	2000, 2010
	Encuesta Nacional de Ocupacion y Empleo	2010-2018
Mongolia	Labor Force Survey (LFS)	2002, 2003, 2006-2018
Nepal	Nepal Labor Force Survey (NLFS)	1999, 2008, 2017/2018
Pakistan	Labor Force Survey (LFS)	2001/02, 2003/04, 2005/06, 2006/07, 2008/09, 2009/10, 2010/11, 2012/13, 2013/14, 2014/15, 2017/18
Philippines	Labor Force Survey (LFS), quarterly releases	2001-2008, 2010-2015, 2017
Russian Federation	Labor force survey	2000, 2008
Sri Lanka	Labor Force Survey (LFS)	2002-2007, 2009-2017
Korea, Rep.	Korea Labor and Income Panel Study (KLIPS)	1999-2017
Chinese Taipei	Manpower survey	2000-2018
Thailand	Labor Force Survey (LFS)	2000, 2005, 2010-2018
Türkiye	Labor force survey	2000-2018
United States	Population census [f]	2000
	American community surveys	2000-2017
Vietnam	Labor Force Survey (LFS)	2007, 2009, 2010, 2012-2014, 2016

[a] The IZA wage indicator survey with data for Chinese workers is available for 2010. [b] The 27 countries member of the EU on January 2007 and Norway. [c] Poland from 2004 onwards. [d] Structure of Earnings Surveys for 2002, 2006, 2010. [e] We drop Indonesian data for 2000-2002 because of anomalies in the data. [f] Data from the 2000 US population census refer to 1999.

Appendix table C3: Classification of occupations

Occupational grouping	Description	ISCO 88 codes
1	Legislators	[11]
2	Managers	[12–13]
3	Engineering professionals	[21, 31]
4	Health professionals	[22, 32]
5	Teaching professionals	[23, 33]
6	Other professionals	[24, 34]
7	Clerical support workers	[41–42]
8	Personal service workers	[51, 910, 912–916]
9	Sales workers	[52, 911]
10	Craft workers and machine operators	[71–74, 81–82, 93]
11	Agricultural workers	[60-61, 92]
12	Other, including armed forces	[01-03, 90,99]
13	Drivers	[83]

Notes: based on Reijnders and de Vries (2018). In section 3, five broad aggregations of occupational groupings are considered: managerial (occupational grouping 2); engineering (3); production (10, 11, and 13); support services (6, 7, and 9); other occupations (1, 4, 5, 8, and 12).

Appendix table C4: Classification of industries

#	ISIC rev 3 code	Industry
1	AtB	Agriculture, hunting, forestry, and fishing
2	C	Mining and quarrying
3	15t16	Food, beverages, and tobacco
4	17t18	Textiles and textile products
5	19	Leather, leather products, and footwear
6	20	Wood and products of wood and cork
7	21t22	Pulp, paper, paper products, printing, and publishing
8	23	Coke, refined petroleum, and nuclear fuel
9	24	Chemicals and chemical products
10	25	Rubber and plastics
11	26	Other non-metallic minerals
12	27t28	Basic metals and fabricated metal
13	29	Machinery, nec
14	30t33	Electrical and optical equipment
15	34t35	Transport equipment
16	36t37	Manufacturing, nec; recycling
17	E	Electricity, gas, and water supply
18	F	Construction
19	50	Sale, maintenance, and repair of motor vehicles and motorcycles; retail sale of fuel
20	51	Wholesale trade and commission trade, except of motor vehicles and motorcycles
21	52	Retail trade, except of motor vehicles and motorcycles; repair of household goods
22	H	Hotels and restaurants
23	60	Inland transport
24	61	Water transport
25	62	Air transport
26	63	Other supporting and auxiliary transport activities; activities of travel agencies
27	64	Post and telecommunications
28	J	Financial intermediation
29	70	Real estate activities
30	71t74	Renting of M&Eq and other business activities
31	L	Public administration and defence; compulsory social security
32	M	Education
33	N	Health and social work
34	O	Other community, social, and personal services
35	P	Private households with employed persons

Notes: 35 industries in ADB MRIOTs.

Appendix D. Robustness analysis

This appendix examines the robustness of key results on the importance of proximity for the development of new activity specializations. We re-run the regressions presented in section 4.3 in Tables 3-5 and consider: *i*) regressions based on the subset of activities with a lower threshold for the initial specialization index (i.e. observations where the initial $AS_{c,j,t}<0.9$ or $AS_{c,j,t}<0.8$ instead of $AS_{c,j,t}<1$ used in the baseline); *ii)* using the network proximity measure proposed by Kali et al. (2013) instead of average proximity $\omega_{c,j,t}$; *iii)* a measure of average proximity where new activities that relate to current specialization with a higher share in a country's export basket receive a greater weight[18]; *iv)* inclusion of capital income, which is the remainder when labor income is subtracted from industry gross value added, proportionally allocated to the labor incomes of activities in the industry. Briefly, the additional analysis suggests that the key findings remain significant, while the size of the effect varies depending on the approach considered.[19]

Appendix Table D1 presents robustness analysis for the relevance of proximity for future specialization. Panel A considers observations with a threshold for initial specialization below 0.9. Panel B is similar, but sets the threshold at 0.8. By successively lowering the threshold, only new specializations that experienced a stronger increase in specialization are considered, which is less likely to occur. Hence, the coefficient for $AS_{c,j,t}$ is expected to be lower, which is observed in panels A and B. The coefficient remains positive and significant. Panels C and D consider alternative measures for proximity. New activity specialization's proximity to initial specialization remains positive and significant. Panel C suggests this pattern is stronger for developing countries if the network proximity measure by Kali et al. (2013) is used. In contrast, if capital is proportionally distributed (Panel E), the pattern is stronger for advanced economies.

---

[18] Average proximity defined in (5) is modified, such that $\widetilde{\omega}_{c,r,t} = \frac{\sum_j \{\mathbf{1} | AS_{c,j,t} > 1\} \theta_{c,j,t} \varphi_{j,r,t}}{\sum_j \theta_{c,j,t} \varphi_{j,r,t}}$, where $\theta_{c,j,t} = z_{c,j,t} / \sum_j z_{c,j,t}$, i.e. the share of activity *j* in domestic value added exports of country *c* in year *t*. The intuition for this modified measure is that specialization in a new activity receives larger weight if the new activity has higher proximity to activities that are economically important.

[19] We also examined robustness of the results to changes in the data sample. We dropped observations for 2000 and only considered the period from 2007 onwards, and we considered three non-overlapping periods (2000-2007, 2008-2012, and 2013-2018). These changes to the data do not qualitatively alter the key findings.

Appendix Table D2 considers robustness of the results regarding specialization in activities at the intensive and extensive margins. The finding for advanced economies whereby proximity is more relevant for the introduction of new specializations within the same industry is confirmed in the alternative regressions. For developing countries, the baseline finding that proximity matters more for specialization within the same occupation (across industries), is confirmed in all robustness regressions. For example, the impact size is 3.4 against 3.1 percentage points across occupations if a weighted average proximity measure is considered (see panel D), which compares to the baseline impact size of 5.2 against 5.0 percentage points.

Finally, robustness of the results for routine manual tasks is explored in Appendix Table D3. The baseline finding is confirmed in the alternative regressions, namely that especially in developing countries new specializations in activities with high routine manual intensity strongly relate to proximity of the activity to current specialization. The difference in average proximity ($\omega_{c,j,t}$) between advanced and developing countries tends to be less pronounced in the robustness regressions compared to the baseline results. In the baseline, the effect is about four times larger for developing countries (cf. Table 5). In the robustness, it is around two to four times larger for developing countries depending on the approach considered.

Appendix Table D1: Probability regressions, robustness of baseline results

	Dependent variable: whether country c has a comparative advantage in activity j at time $t+5$ ($x_{c,j,t+5}$= 1)		
	(1)	(2)	(3)
A. $AS_{c,j,t}$ <0.9			
$AS_{c,j,t}$	0.051***	0.124***	0.040**
	(0.018)	(0.006)	(0.016)
$\omega_{c,j,t}$	0.052***	0.036***	0.053***
	(0.004)	(0.003)	(0.004)
N	132,867	54,365	78,502
Adjusted R^2	0.058	0.067	0.058
B. $AS_{c,j,t}$ <0.8			
$AS_{c,j,t}$	0.031***	0.070***	0.027**
	(0.012)	(0.007)	(0.011)
$\omega_{c,j,t}$	0.045***	0.037***	0.045***
	(0.003)	(0.003)	(0.003)
N	127,480	51,704	75,776
Adjusted R^2	0.052	0.057	0.054
C. Network proximity (Kali et al. 2013)			
$AS_{c,j,t}$	0.081***	0.222***	0.058**
	(0.028)	(0.006)	(0.023)
$Network\ proximity_{c,j,t}$	0.026***	0.000	0.031***
	(0.004)	(0.002)	(0.003)
N	137,723	56,919	80,804
Adjusted R^2	0.063	0.084	0.062
D. Weighted proximity			
$AS_{c,j,t}$	0.089***	0.212***	0.064***
	(0.028)	(0.006)	(0.024)
$\widetilde{\omega}_{c,j,t}$	0.037***	0.020***	0.042***
	(0.003)	(0.002)	(0.003)
N	137,723	56,919	80,804
Adjusted R^2	0.063	0.086	0.060
E. Proportional distribution capital income			
$AS_{c,j,t}$	0.227***	0.196***	0.248***
	(0.004)	(0.006)	(0.006)
$\omega_{c,j,t}$	0.024***	0.033***	0.018***
	(0.002)	(0.003)	(0.002)
N	138,000	57,139	80,861
Adjusted R^2	0.095	0.094	0.098
Country-Year FE	X	X	X
Industry-Year FE	X	X	X
Sample	Total	Advanced	Developing

Notes: Results from linear probability regression using equation (4). The independent variables are activity specialization $AS_{c,j,t}$ and average proximity ω, defined in equations (3) and (6). ω is normalized (by subtracting the mean and dividing by the standard deviation) to ease interpretation. Robust standard errors (clustered by country) are reported in parentheses. * p<0.1, ** p<0.05 and *** p<0.01. Own calculations based on annual data for 2000-2018, see section 4.3.

Appendix Table D2: Specialization in activities at the intensive and extensive margins, robustness

	Dependent variable: whether country c has a comparative advantage in activity j at time $t+5$ ($x_{c,j,t+5}$= 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
A. $AS_{c,j,t}$ <0.9						
$AS_{c,j,t}$	0.045***	0.108***	0.036**	0.059***	0.199***	0.044***
	(0.016)	(0.007)	(0.014)	(0.022)	(0.007)	(0.018)
$\omega_{c,j,t}$	0.043***	0.040***	0.040***	0.039***	0.016***	0.044***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
N	132,867	51,636	81,231	132,867	51,636	81,231
Adjusted R^2	0.134	0.135	0.137	0.117	0.148	0.103
B. $AS_{c,j,t}$ <0.8						
$AS_{c,j,t}$	0.028**	0.057***	0.024**	0.041**	0.153***	0.031**
	(0.010)	(0.007)	(0.010)	(0.016)	(0.007)	(0.013)
$\omega_{c,j,t}$	0.038***	0.041***	0.033***	0.033***	0.015***	0.038***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
N	127,480	49,049	78,431	127,480	49,049	78,431
Adjusted R^2	0.130	0.128	0.134	0.115	0.140	0.103
C. Network proximity (Kali et al. 2013)						
$AS_{c,j,t}$	0.071***	0.195***	0.050**	0.082***	0.261***	0.058***
	(0.024)	(0.006)	(0.020)	(0.029)	(0.007)	(0.023)
$Network\ proximity_{c,j,t}$	0.029***	0.010***	0.030***	0.029***	0.007***	0.033***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
N	137,723	54,114	83,609	137,723	54.114	83,609
Adjusted R^2	0.143	0.150	0.146	0.124	0.166	0.107
D. Weighted proximity						
$AS_{c,j,t}$	0.082***	0.201***	0.058***	0.088***	0.263***	0.063***
	(0.026)	(0.006)	(0.022)	(0.031)	(0.006)	(0.025)
$\widetilde{\omega}_{c,j,t}$	0.026***	0.014***	0.029***	0.027***	0.010***	0.034***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
N	137,723	54,114	83,609	137,723	54,114	83,609
Adjusted R^2	0.140	0.150	0.142	0.122	0.166	0.105
E. Proportional distribution capital income						
$AS_{c,j,t}$	0.203***	0.174***	0.223***	0.273***	0.265***	0.278***
	(0.004)	(0.007)	(0.006)	(0.004)	(0.007)	(0.006)
$\omega_{c,j,t}$	0.021***	0.039***	0.011***	0.013***	0.012***	0.016***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
N	138,000	54,275	83,725	138,000	54,275	83,725
Adjusted R^2	0.168	0.164	0.172	0.167	0.172	0.145
Year FE	X	X	X	X	X	X
Country-Industry FE	X	X	X			
Country-Occupation FE				X	X	X
Sample	Total	Advanced	Developing	Total	Advanced	Developing

Notes: Results from linear probability regression using equation (4). The independent variables are activity specialization $AS_{c,j,t}$ and average proximity ω, defined in equations (3) and (6). ω is normalized (by subtracting the mean and dividing by the standard deviation) to ease interpretation. Robust standard errors (clustered by country) are reported in parentheses. * p<0.1, ** p<0.05 and *** p<0.01. Own calculations based on annual data for 2000-2018, see section 4.3.

Appendix Table D3: Activities split by manual routine task intensity, robustness

	Dependent variable: whether country c has a comparative advantage in activity j at time $t+5$ ($x_{c,j,t+5}$= 1)		
	(1)	(2)	(3)
A. $AS_{c,j,t}$ <0.9			
$AS_{c,j,t}$	0.049**	0.161***	0.038**
	(0.021)	(0.009)	(0.018)
$\omega_{c,j,t}$	0.077***	0.030***	0.097***
	(0.005)	(0.004)	(0.005)
Routine manual	0.000	0.007**	-0.002
	(0.002)	(0.003)	(0.003)
N	54,115	22,641	31,474
Adjusted R^2	0.063	0.072	0.069
B. $AS_{c,j,t}$ <0.8			
$AS_{c,j,t}$	0.030**	0.103***	0.025**
	(0.014)	(0.010)	(0.012)
$\omega_{c,j,t}$	0.069***	0.031***	0.087***
	(0.004)	(0.004)	(0.004)
Routine manual	0.003	0.010***	-0.001
	(0.002)	(0.003)	(0.002)
N	51,166	21,191	29,975
Adjusted R^2	0.057	0.060	0.065
C. Network proximity (Kali et al. 2013)			
$AS_{c,j,t}$	0.076**	0.257***	0.054**
	(0.032)	(0.009)	(0.025)
$Network\ proximity_{c,j,t}$	0.036***	-0.003	0.046***
	(0.005)	(0.003)	(0.004)
Routine manual	0.003	0.010***	0.004
	(0.002)	(0.003)	(0.003)
N	56,697	23,982	32,715
Adjusted R^2	0.066	0.097	0.067
D. Weighted proximity			
$AS_{c,j,t}$	0.081**	0.243***	0.059**
	(0.033)	(0.009)	(0.027)
$\widetilde{\omega}_{c,j,t}$	0.047***	0.024***	0.052***
	(0.004)	(0.003)	(0.004)
Routine manual	-0.004	0.007**	-0.006*
	(0.003)	(0.003)	(0.003)
N	56,697	23,982	32,715
Adjusted R^2	0.066	0.098	0.062
E. Proportional distribution capital income			
$AS_{c,j,t}$	0.263***	0.240***	0.281***
	(0.006)	(0.009)	(0.008)
$\omega_{c,j,t}$	0.041***	0.025***	0.056***
	(0.003)	(0.004)	(0.004)
Routine manual	0.006***	0.008***	0.004
	(0.002)	(0.003)	(0.003)
N	57,102	24,296	32,806
Adjusted R^2	0.112	0.105	0.123
Country-Year FE	X	X	X
Industry-Year FE	X	X	X
Sample	Total	Advanced	Developing

Notes: Results from linear probability regression using equation (4). The independent variables are activity specialization $AS_{c,j,t}$ and average proximity ω, defined in equations (3) and (6). ω is normalized (by subtracting the mean and dividing by the standard deviation) to ease interpretation. Routine task intensity constructed following Acemoglu and Autor (2011), see text. Robust standard errors (clustered by country) are reported in parentheses. * $p<0.1$, ** $p<0.05$ and *** $p<0.01$. Own calculations based on annual data for 2000-2018, see section 4.3.

Appendix table E. Test of path defiance in activity export specialization

Economy	Number of new activity specializations	Number of path-defying new activity specializations	Share of path-defying new activity specializations	P-value Kolmogorov-Smirnov test
China	109	57	52.3	0.685
Chinese Taipei	247	128	51.8	0.278
Ireland	205	104	50.7	0.785
United States	210	104	49.5	0.687
Denmark	342	158	46.2	0.120
Bangladesh	371	170	45.8	0.051
Malta	299	137	45.8	0.423
Netherlands	390	173	44.4	0.010
Mongolia	298	126	42.3	0.000
Italy	339	143	42.2	0.061
Philippines	244	102	41.8	0.000
Sri Lanka	223	93	41.7	0.001
Canada	210	86	41.0	0.005
Vietnam	176	72	40.9	0.014
Thailand	260	104	40.0	0.000
Bulgaria	299	119	39.8	0.000
Türkiye	289	113	39.1	0.000
Brazil	219	83	37.9	0.000
Greece	283	106	37.5	0.000
Belgium	360	134	37.2	0.035
Pakistan	164	58	35.4	0.000
Australia	210	74	35.2	0.000
Mexico	222	78	35.1	0.000
Portugal	357	125	35.0	0.000
Sweden	294	102	34.7	0.000
United Kingdom	293	98	33.4	0.000
Spain	316	104	32.9	0.000
Germany	205	67	32.7	0.000
Cyprus	179	58	32.4	0.000
Luxembourg	198	64	32.3	0.000
Fiji	250	80	32.0	0.000
Austria	375	117	31.2	0.000
India	169	52	30.8	0.000
Romania	254	78	30.7	0.000
Slovak Republic	290	89	30.7	0.000
Poland	278	84	30.2	0.000
Korea, Rep.	192	58	30.2	0.000

Indonesia	233	70	30.0	0.000
Kyrgyzstan	212	62	29.2	0.000
Latvia	384	108	28.1	0.000
Slovenia	350	98	28.0	0.000
Japan	77	21	27.3	0.001
Lithuania	308	83	26.9	0.000
Cambodia	117	31	26.5	0.000
Norway	219	58	26.5	0.000
Hungary	260	67	25.8	0.000
Estonia	275	67	24.4	0.000
Nepal	165	40	24.2	0.000
France	323	75	23.2	0.000
Finland	225	46	20.4	0.000
Russian Federation	151	30	19.9	0.000
Czechia	224	44	19.6	0.000

Notes: First column shows the total number of new specializations across the eight overlapping periods. Second column shows the total number of path defying new specialization across the periods, where we use the baseline definition for path-defying as those that have a proximity to the initial export basket which is lower than the mean proximity of the activities in the option set. Third column is second column divided by first. Fourth column shows the p-value from a two-sample one-sided Kolmogorov-Smirnov test whether the cumulative distribution of the option set proximities stochastically dominates that of the actual data. Sorting by share of path-defying new activity specializations from high to low.

Appendix F. Measuring the routine task intensity of activities

The O*NET task measures used in this paper are composite measures of O*NET Work Activities and Work Context Importance scales. We closely follow the approach by Acemoglu and Autor (2011) to construct task measures by occupation-industry pair.

Task measure	O*NET Work Activities or Work Context Importance scales
Routine cognitive	4.C.3.b.7 Importance of repeating the same tasks 4.C.3.b.4 Importance of being exact or accurate 4.C.3.b.8 Structured v. Unstructured work (reverse)
Routine manual	4.C.3.d.3 Pace determined by speed of equipment 4.A.3.a.3 Controlling machines and processes 4.C.2.d.1.i Spend time making repetitive motions

The O*NET scales use the O*NET-SOC occupational classification scheme, which is directly matched to the industry by occupation data from the BLS.[20] Each scale is standardized to have mean zero and standard deviation one, using labor supply weights for the 728 occupations from the BLS. The composite task measures listed in the Table above are equal to the summation of their respective constituent scales, then also standardized to mean zero and standard deviation one using labor supply weights for the 728 occupations from the BLS. In the final step, the task measures are merged with the industry by occupation data from the BLS and then collapsed (using the labor supply weights) to the 35 industries and 12 occupational groupings used for the empirical analysis.[21]

[20] O*NET scales are obtained from www.onetonline.org, accessed November 2022. Occupation by industry data is from Employment Projections program, U.S. Bureau of Labor Statistics (https://www.bls.gov/emp/tables/industry-occupation-matrix-industry.htm).

[21] We obtain 12 occupational groupings instead of 13 in the baseline, because one occupational grouping, #12 "Other" is not observed in the BLS data. Hence, the regression results in Table 3 exclude occupational grouping #12.

Appendix G. Additional data sources for seven countries

Figure 1 makes use of a larger country sample, which besides the 52 economies in the main analysis, also includes Chile, Colombia, Costa Rica, Croatia, Ethiopia, Ghana, and Switzerland. This better captures trends for the least developed countries, such as Ethiopia, but data quality is lower. We briefly describe the sources and the approach. Detailed information, including crosswalks are provided in the replication package. Note the replication package includes estimates of both jobs and income by activity. We alter **b** in equation (A3) to jobs instead of labor income to estimate the job content of exports. Jobs by activity from trade is analyzed in detail by Winkler et al. (2023).

Chile, Costa Rica, Colombia, Croatia, and Switzerland. We use *i*) the OECD Inter-Country Input-Output Database, November 2022 release[22]; *ii*) the labor compensation shares and employment from the OECD's Trade in Employment database, November 2021 release. Occupation data for Chile is from CASEN, for the years 2000, 2003, 2006, 2009, 2011, 2013, 2015, and 2017. Occupation data for Colombia is from Gran Encuesta Integrada de Hogares (GEIH) for the years 2009-2018. The occupation data is accessed using datalibweb at the World Bank and we used the harmonization code provided by through the repository for the Global Labor Database. Occupation data for Costa Rica is from the 2000 and 2011 population census, extrapolated to 2018 using trends from ENAHO for 2011-2018. Occupation data for Croatia and Switzerland are from the EU Labour Force Surveys for the period from 2000-2018 for Switzerland and 2002-2018 for Croatia. Estimates of the job and income content of exports by activity are obtained for 45 sectors of the total economy, subsequently aggregated to 20 sectors.

Ethiopia and Ghana. We make use of time series input-output tables introduced by Mensah and de Vries (2023). Labor shares for Ghana are based on estimates in the IFPRI's Social Accounting Matrices, years 2005, 2013, and 2015. For Ethiopia we use IFPRI's Social Accounting Matrices for the years 2005, 2011, and 2018. Employment is from the GGDC/UNU-WIDER Economic Transformation Database, release February 2021, combined with 2-digit manufacturing employment data developed by Kruse et al. (2022). Occupation data for Ethiopia is from labor survey for the years 2005, 2013, and 2021. Occupation employment and relative median wage data for Ghana is from the Ghana Living Standard Survey (GLSS), years 1998, 2006, 2013, and 2017.

[22] Accessed and downloaded from oe.cd/tiva in November 2022.

Estimates of the job and income content of exports by activity are for 20 sectors of the total economy.