Maastricht University

The Innovation and Structural Transformation Database

Citation for published version (APA):

Foster-McGregor, N., Nomaler, Ö., & Verspagen, B. (2024). *The Innovation and Structural Transformation Database: A Guide*. UNU-MERIT. UNU-MERIT Working Papers No. 027 https://unu-merit.nl/publications/wppdf/2024/wp2024-027.pdf

Document status and date:
Published: 25/10/2024

Document Version:
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-SA

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 10 Nov. 2026

Working Paper Series

#2024-027

The Innovation and Structural Transformation Database: A Guide

Neil Foster-McGregor, Önder Nomaler & Bart Verspagen

Published 22 October 2024

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: **info@merit.unu.edu** | website: **http://www.merit.unu.edu**

Boschstraat 24, 6211 AX Maastricht, The Netherlands
Tel: (31) (43) 388 44 00

UNU-MERIT Working Papers
ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology
UNU-MERIT

UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at UNU-MERIT to stimulate discussion on the issues raised.

The Innovation and Structural Transformation Database: A Guide

October 2024

Neil Foster-McGregor, Önder Nomaler & Bart Verspagen

Abstract

This document summarizes the content of the Innovation and Structural Transformation database provides a collection of indicators for the global economy, emphasizing the dynamic nature of development and economic growth. The database is publicly available from the web (see link in the main text). This document summarizes the content of the database, further providing information on the data sources and the construction of the variables included in the database.

Keywords: Structural change, innovation, economic development, economic growth.

JEL Codes: O14, O30, O40, O50, O53

The Innovation and Structural Transformation Database and this guide are provided under a Creative Commons BY 4.0 license. This license allows users to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.

1. Introduction

The Innovation and Structural Transformation database provides a collection of indicators for the global economy, emphasizing the dynamic nature of development and economic growth. This document summarizes the content of the database, further providing information on the data sources and the construction of the variables included in the database. The database is publicly available from the web at

https://dataverse.nl/dataverse/innovation_and_structural_transformation_database/

Any future updates of the database will be available at the same link. The database and this document are provided under a CC BY 4.0 license. This means that the database can be used freely, as long as appropriate credits are given.[1] This credit can be given by citing the forthcoming book that illustrates the use of the database:

> Felipe, J., Foster-McGregor, N., Nomaler, Ö and B. Verspagen, 2025, *Innovation and Structural Transformation in Asia*, Abingdon (UK): Taylor & Francis.

The aim of the database is to provide a resource for researchers, policymakers and others that provides information on a set of variables and indicators that are fundamental to understand long-run growth and development. These indicators revolve around the topic of structural transformation, development, and international competitiveness. Underlying the database is the premise that development is a process of structural transformation, i.e., that increasing the standard of living in an economy involves changing what the economy produces and consumes and involves changes in the way actors in the economy interact. Traditionally, this process of structural transformation was considered to involve a movement of production (and consumption) across broad sectors (e.g., from agriculture to industry to services). With the availability of new data, however, a more nuanced perspective has emerged. One dimension of this is the observation that much structural transformation is occurring within sectors, through for example the changing structure of occupations and functions within sectors, as well as changes in the composition of production and activities within sectors.

Related to these changes have been developments of the global structure of production, with structural transformation depending to a great extent on the way the economy interacts with foreign markets in today's global economy. The position of an economy in the international structure of production depends on the competitiveness of firms and other actors in the economy. In this context, competitiveness, or comparative advantage, determines where economies, or rather

[1] For details of this license see https://creativecommons.org/licenses/by/4.0/.

the firms located in economies, can contribute value added in global production networks, which in turn will help determine overall levels of well-being in the economy.

The competitiveness, or comparative advantage, of an economy is further a function of the capabilities that exist in the economy. Development can further be seen as a process of developing competitiveness in a broader range of products (i.e., diversification), including a move into more complex products within and across sectors (i.e., upgrading). An important dimension of an economy's capabilities relates to its performance in innovation. In this context, there has been a huge amount of interest in recent years in the development of digital technologies and how innovation in these technologies is changing the nature of production and work.

Based upon these arguments, the database is formed around four main pillars. The first pillar covers data on structural change and productivity, presenting indicators that document how fast, and into which direction the structure of the economy changes and the resulting impact of these changes on productivity growth. The second pillar focuses on competitiveness and comparative advantage, which is approached through the notion of product complexity, and which in turn can be seen as a measure of development potential. This pillar focuses on developing a detailed picture of the so-called product space and how the economies in the database occupy this space. The approach involves a large data set of exported products and shows how economies are specialized in these different products. The third pillar of the database is concerned with innovation, with the database making use of patent data to identify the extent of innovation as well as the direction of that innovation, further providing an overall picture of patented inventions and how they relate to the economy, as well as an overview of trends in technologies related to the so-called fourth industrial revolution. The final pillar considers the global economy and the role of individual economies in the global structure of production through an examination of so-called Global Value Chains (GVCs), which highlight that production is fragmented geographically and firms may contribute (small) parts of total value added to a variety of GVCs. The data for this final pillar is heavily based on global input-output tables. The remainder of this document describes in turn the indicators that fall under each of the four pillars, along with the data sources used to construct them.

2. Pillar 1: Structural Change

2.1. Basic structural change indicators

The basic indicators for structural change consider the sectoral composition of the economy, as well as the structure of international trade (exports and imports). For the sectoral structure of the economy, the database considers the structure of value added, household consumption demand and total demand, all in constant 2010 US$. These data are drawn from the ADB input output tables. The following variables are defined:

$$\sigma_{tij}^{VA} = \frac{VA_{tij}}{\sum_k VA_{tik}}$$

$$\sigma_{tij}^{HC} = \frac{HC_{tij}}{\sum_k HC_{tik}}$$

$$\sigma_{tij}^{TD} = \frac{TD_{tij}}{\sum_k TD_{tik}}$$

where the indices i, j, and t denote economy, sector and time period, respectively, VA is value added produced in the sector/economy, HC is final consumption by households in economy i of products produced by sector j (either imported or produced domestically), and *TD* is the sum of household consumption, government consumption and gross fixed capital formation by residents of economy i aimed at products produced by sector *j* (again domestically or imported).

These shares are used to produce an indicator of structural change, which is termed the NAV indicator (Norm of Absolute Value changes). This is defined as follows:

$$NAV_i^K = \frac{\sum_j \left|\sigma_{Tij}^K - \sigma_{0ij}^K\right|}{2T}$$

where $K \in \{VA, HC, TD\}$ is an index that refers to the particular indicator of structure being considered and *T* is the number of years considered. The greater the degree of structural change between period 0 and T, the higher the value of the NAV indicator.

Figure 1 shows the sectoral shares of total demand (TD) and the associated NAV indicator for Nepal.[2] Note that the database does not have data for the years 2001-2006, so the period between 2000 and 2007 is interpolated. We see that the recent period 2015-19 is one of relatively rapid structural change in total demand in Nepal.

[2] This figure, as those below, is copied from the graphical interface on the website. In this interface, users can select specific (combinations of) indicators, economies and sectors to be displayed. These figures are often linked to each other, e.g., clicking/selecting an observation in one figure provides more detailed information for this observation in a related figure. All figures can be saved and used in proprietary documents.

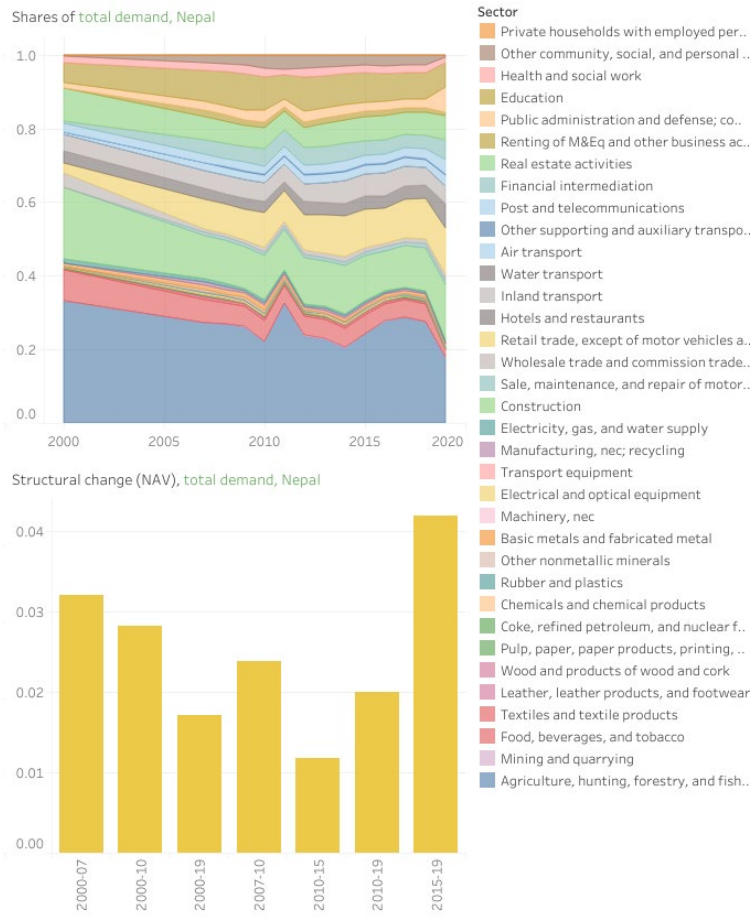

Figure 1. Structural change in total demand – the case of Nepal

The sectors that are available in this pillar of the database are identical to the sectors used in the GVC pillar, with a selection of these sectors also used in other sections of the database. These sectors are documented in Table 1.

Table 1. Sectors used in basic structural change and GVC sections

Sec num	ISIC Rev. 4	Description
1	A	Agriculture, hunting, forestry, and fishing
2	B	Mining and quarrying
3	C10-C12	Food, beverages, and tobacco
4	C13, C14	Textiles and textile products
5	C15	Leather, leather products, and footwear
6	C16	Wood and products of wood and cork
7	C17, C18, J58	Pulp, paper, paper products, printing, and publishing
8	C19	Coke, refined petroleum, and nuclear fuel
9	C20, C21	Chemicals and chemical products
10	C22	Rubber and plastics
11	C23	Other nonmetallic minerals
12	C24, C25	Basic metals and fabricated metal
13	C28, C33	Machinery, nec
14	C26, C27	Electrical and optical equipment
15	C29, C30	Transport equipment
16	C31, C32	Manufacturing, nec; recycling
17	D, E36	Electricity, gas, and water supply
18	F	Construction
19	G45	Sale, maintenance, and repair of motor vehicles and motorcycles; retail sale of fuel
20	G46	Wholesale trade and commission trade, except of motor vehicles and motorcycles
21	G47	Retail trade, except of motor vehicles and motorcycles; repair of household goods
22	I	Hotels and restaurants
23	H49	Inland transport
24	H50	Water transport
25	H51	Air transport
26	H52, N79	Other supporting and auxiliary transport activities; activities of travel agencies
27	H53, J61	Post and telecommunications
28	K	Financial intermediation
29	L	Real estate activities
30	J58-J60, J62-J63, M, N77-N78, N80-N82	Renting of M&Eq and other business activities
31	O	Public administration and defense; compulsory social security
32	P	Education
33	Q	Health and social work
34	E37-E39, R, S	Other community, social, and personal services
35	T	Private households with employed persons

Turning to indicators of the structure of international trade (both exports and imports), the database includes two groups of indicators. The first is related to global value chains (GVCs), while the other refers to innovation, and in particular the group of technologies that is referred to as 4th industrial revolution (4IR) technologies. The indicators on GVCs distinguish between different types of products, namely consumption goods, intermediate goods and capital goods. According to this distinction, consumption goods and capital goods are considered as the outputs of a value

chain, while intermediate goods are the inputs. Thus, the relative specialization or relative use of these types of goods provides a rough indication for how an economy is integrated in GVCs and can be considered as an alternative to the input-output based measures of GVCs that are used elsewhere (including the later pillar on GVCs).

Detailed product-level data on imports and exports from the UN Comtrade database are first used to calculate the shares of the different categories of products:

$$\sigma_{tih}^{Mgvc} = \frac{Mgvc_{tih}}{\sum_k Mgvc_{tik}}$$

$$\sigma_{tih}^{Xgvc} = \frac{Xgvc_{tih}}{\sum_k Xgvc_{tik}}$$

where $Mgvc$ and $Xgvc$ are the value of imports and exports in a particular GVC category, respectively, h is an index representing the *GVC* category, and i and t denote economy and the time period, respectively. The *GVC* categories can be defined in different levels, with three sets of categories used in the database. These are:

$$gvc_1 = \{consumption, intermediate, capital\}$$

$$gvc_2 = \left\{ \begin{array}{c} consumption, \\ generic\ intermediate, specific\ intermediate, mixed\ intermediate, \\ non-classified\ intermediate, \\ generic\ capital, specific\ capital, mixed\ capital \end{array} \right\}$$

$$gvc_3 = \left\{ \begin{array}{c} primary\ consumption, \quad processed\ consumption, \\ primary\ intermediate, processed\ intermediate, \\ non-classified\ intermediate, \\ capital \end{array} \right\}$$

These definitions of the *GVC* categories are based on the BEC classification and are explained in more detail in the Annex. The data are presented in the form of shares in the database and are also used to calculate indicators of relative specialization. For exports, this takes the form of the well-known Revealed Comparative Advantage (RCA) indicator:

$$RCA_{tih}^{gvc} = \frac{\sigma_{tih}^{Xgvc}}{\sigma_{tWh}^{Xgvc}}$$

where the index *W* in the denominator on the righthand side refers to the "world", i.e., total exports over all economies in the database. A value of $RCA_{tih}^{gvc} > 1$ indicates that an economy has a comparative advantage in the particular *GVC* category h.

For imports, a similar indicator is calculated, but is referred to as the Comparative Import Intensity (CII) in the database:

$$CII_{tih}^{gvc} = \frac{\sigma_{tih}^{Mgvc}}{\sigma_{tWh}^{Mgvc}}$$

Where again a value of $CII_{tih}^{gvc} > 1$ indicates that an economy imports products in the *GVC* category h relatively intensively.

Figure 2 shows the shares and *RCA* time series for exports, split by the gvc_1 set, for Argentina and Pakistan. We see that the share of capital goods is small in both economies, while consumer goods is the largest category, and increasing over time in Pakistan, and intermediate goods is the largest category in Argentina. In the case of Pakistan, an *RCA* larger than one is found for consumer goods only, while in Argentina an RCA greater than one is observed for intermediate goods, and in some years for consumer goods.

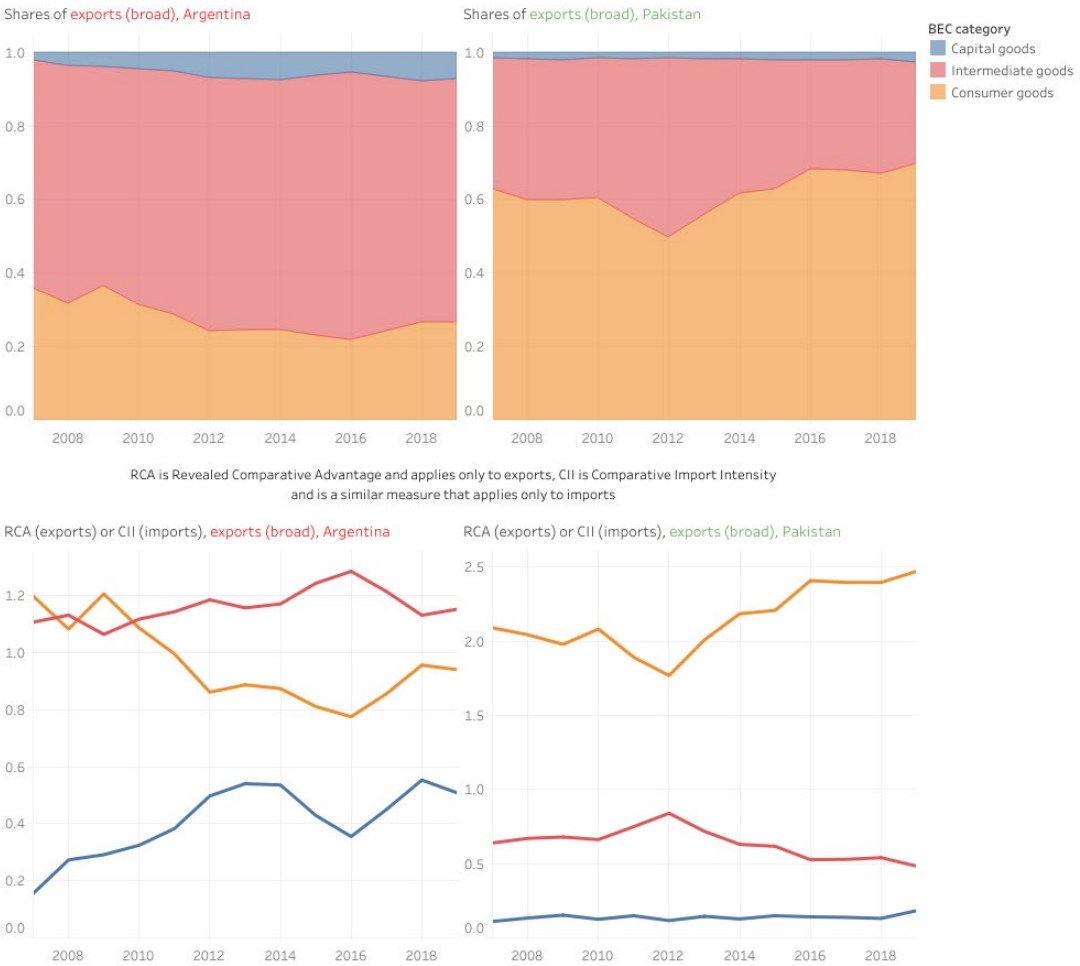

Figure 2. The composition of exports by GVC category for Argentina and Pakistan

Table 2. 4IR fields in international trade

4IR field	Product codes (Harmonized System)
CADCAM[3]	845811; 845891; 845921; 845931; 845951; 845961; 846011; 846021; 846031; 846221; 846231; 846241
Robots[4]	847950; 847989
Automated welding	851521; 851531
3D printing	847780; 847710; 847720; 847730; 847740; 847751; 847759; 847790
Regulating instruments	903210; 903220; 903281; 903289; 903290
ICT	844351; 847050; 847110; 847130; 847141; 847149; 847150; 847160; 847170; 847180; 847190; 847220; 847290; 847330; 847350; 851721; 851722; 900911; 900912; 851711; 851719; 851730; 851750; 851780; 851790; 852510; 852520; 852790; 853110; 851810; 851821; 851822; 851829; 851830; 851840; 851850; 851890; 851910; 851921; 851929; 851931; 851939; 851940; 851992; 851993; 851999; 852010; 852020; 852032; 852033; 852039; 852090; 852110; 852190; 852210; 852290; 852530; 852540; 852712; 852713; 852719; 852721; 852729; 852731; 852732; 852739; 852812; 852813; 852821; 852822; 852830; 950410; 852330; 852460; 853400; 854011; 854012; 854020; 854040; 854050; 854060; 854071; 854072; 854079; 854081; 854089; 854091; 854099; 854110; 854121; 854129; 854130; 854140; 854150; 854160; 854190; 854212; 854213; 854214; 854219; 854230; 854240; 854290; 854890; 852390; 852410; 852491; 852499; 852910; 852990; 854381; 901320

For trade in products related to the 4th industrial revolution, a classification of exported products that is based on Foster-McGregor et al., (2019) and Acemoglu and Restrepo (2018) is developed and used. The approach results in a definition of six subfields of 4IR products that can be found in the data on international trade. Table 2 provides an overview of the 4IR fields and the Harmonized System (HS) product codes that are assigned to them. Given the imperfect overlap between these technologies and the HS codes, it is inevitable that we will also be capturing earlier vintages of technology (e.g., third industrial revolution technologies) in these classifications. Despite this, the data provide an insight into the use of advanced technologies in these domains and into the means of identifying countries with the capabilities to use these technologies (and therefore potentially benefit from them).

[3] Termed numerically controlled machines in Acemoglu and Restrepo (2018).
[4] In addition to the HS code on industrial robots, this category further includes the category "dedicated machinery" from Acemoglu and Restrepo (2018), which includes both industrial robots and dedicated automated machinery.

For the 4IR trade data, the same indicators are defined as for the GVC trade data, namely the share of exports and imports in these products and the indicators of RCA and CCI in these products:

$$\sigma_{tih}^{M4ir} = \frac{M4ir_{tih}}{\sum_k M4ir_{tik} + Mnon4ir}$$

$$\sigma_{tih}^{X4ir} = \frac{X4ir_{tih}}{\sum_k X4ir_{tik} + Xnon4ir}$$

$$RCA_{tih}^{4ir} = \frac{\sigma_{tih}^{X4ir}}{\sigma_{tWh}^{X4ir}}$$

$$CII_{tih}^{4ir} = \frac{\sigma_{tih}^{M4ir}}{\sigma_{tWh}^{M4ir}}$$

where $M4ir$ and $X4ir$ are the value of imports and exports, respectively, in the designated 4*ir* category, $Mnon4ir$ and $Xnon4ir$ are the values of imports and exports of non-4IR products respectively, and the subscript h denotes one of the 4IR fields:

$$4ir = \begin{Bmatrix} CADCAM, Robots, Automated\ welding, \\ 3D\ printing, Regulating\ instruments, ICT \end{Bmatrix}$$

Note that the definition of the shares of 4IR categories in total trade is slightly different from the definition for GVC categories. The reason is that the total of the GVC categories is equal to total trade (there is no trade that is not included in any GVC category), while 4IR trade is only a small part of total trade. Therefore, for the 4IR trade data, we also define within-4IR versions of these indicators, as follows:

$$\tilde{\sigma}_{tih}^{M4ir} = \frac{M4ir_{tih}}{\sum_k M4ir_{tik}}$$

$$\tilde{\sigma}_{tih}^{X4ir} = \frac{X4ir_{tih}}{\sum_k X4ir_{tik}}$$

$$\widetilde{RCA}_{tih}^{4ir} = \frac{\tilde{\sigma}_{tih}^{X4ir}}{\tilde{\sigma}_{tWh}^{X4ir}}$$

$$\widetilde{CII}_{tih}^{4ir} = \frac{\tilde{\sigma}_{tih}^{M4ir}}{\tilde{\sigma}_{tWh}^{M4ir}}$$

where the tilde (~) denotes within-4IR versions of the variables.

Figure 3 shows the 4IR indicators in the case of exports for Thailand and Japan. Thailand has a much higher share of exports in 4IR fields than Japan, but in both economies, ICT is by far the largest field within 4IR technologies. When considering all 4IR fields combined the RCA is above one in both economies for the entire period. In the case of Japan, an RCA above one is found in all 4IR fields except ICT, while in the case of Thailand an RCA above one is found for ICT only

across the entire period, with CADCAM and regulating instruments developing an RCA above one in the latter years of the sample. Moving to the within-4IR RCAs, greater differences between the two economies are found. The one 4IR category in which both economies perform similarly is CADCAM, which is high and above one for both economies. For the remaining 4IR categories, Thailand shows no RCA above one, while Japan has RCA above one in all 4IR countries.

Figure 3. The composition of exports by 4IR category for Japan and Thailand

Table 3 lists the indicators that are found in the Basic structural change section of the database.

Table 3. Indicators in the Basic structural change section

Sectoral shares of value added
Sectoral shares of household consumption
Sectoral shares of total demand
Amount of structural change in value added (NAV)
Amount of structural change in household consumption (NAV)
Amount of structural change in total demand (NAV)
Share of exports by GVC categories, various aggregation levels
Share of imports by GVC categories, various aggregation levels
RCA of exports by GVC categories, various aggregation levels
CII of imports by GVC categories, various aggregation levels
Share of exports by 4IR categories, various aggregation levels
Share of imports by 4IR categories, various aggregation levels
RCA of exports by 4IR categories, various aggregation levels
CII of imports by 4IR categories, various aggregation levels
Share of exports by 4IR categories, various aggregation levels, within 4IR
Share of imports by 4IR categories, various aggregation levels, within 4IR
RCA of exports by 4IR categories, various aggregation levels, within 4IR
CII of imports by 4IR categories, various aggregation levels, within 4IR

2.2. Structural Change and labor productivity growth

Moving beyond the basic indicators in Pillar 1, the second set of indicators focus on productivity growth, and more specifically labor productivity growth, which is defined as value added in constant prices divided by employment. For the analysis of structural change, this indicator needs to be further broken down by economic sector. There are two main bottlenecks in the construction of a data set for labor productivity for a broad set of economies: the availability of comparable employment data for a large group of economies and the availability of value-added and employment data at a detailed sectoral breakdown.

The construction of the database requires certain compromises in dealing with these bottlenecks. For the sectoral breakdown, this involved settling for a dataset with just seven sectors.[5] This is a relatively broad and aggregated level when compared to data that are available for many (advanced) economies. However, many economies – particularly lesser developed economies – typically do not have these detailed data. To allow for the inclusion of a broad set of countries and to maintain comparability, the dataset thus relies upon the seven-sector breakdown.

[5] This is somewhat smaller than comparable databases, such as the Groningen 10-sector database (https://www.rug.nl/ggdc/structuralchange/previous-sector-database/10-sector-2014?lang=en).

In the case of employment data, it is very difficult to obtain actual data for many (most) of the economies in the database. To overcome this constraint, therefore, estimated data from the International Labor Organization was used. This data is presented in their online ILOSTAT database[6], specifically the section entitled ILOEST (ILO modelled estimates and projections). The data used are from the table called *Employment by sex and economic activity*, which presents employment in 1,000 persons at the ISIC Rev4 sectoral breakdown.

Value added at the sectoral level is drawn from a database on national accounts held at the UN-DESA Statistical Division.[7] Data on value added by economic activity provided by this database is used, with the data measured in constant 2015 prices and US dollars. The database uses market exchange rates, i.e., it makes no attempt to take into account real exchange rates that "correct" for price differentials between economies. Although using real exchange rates (or purchasing power parity exchange rates, PPP) is customary for comparing living standards based on aggregate GDP, correcting for price differentials at the sectoral level is impossible due to a lack of data. Value added data are expressed in billions of US$.

The data on value added are collated according to the ISIC Rev. 3 classification, whereas the employment data are reported according to the ISIC Rev. 4 classification. At the aggregation level used in the construction of the database (i.e., the seven sectors), these sectors are comparable across the two revisions however. Table 4 documents the seven sectors.

Table 4. Sectoral breakdown in the structural change and productivity growth section of the database

Sector number	Description	ISIC codes	
		Rev. 4	Rev. 3
1	Agriculture; forestry and fishing	A	A, B
2	Mining, Utilities	B, D, E	C, E
3	Manufacturing	C	D
4	Construction	F	F
5	Trade; Hotels and restaurants	G, I	G, H
6	Transport, Storage, Communication	H, J	I
7	All other activities	K – U	J – P

These data are used to construct several indicators. The first is the NAV indicator that was already defined above for the case of value-added. In this case it is calculated using employment shares and defined as:

[6] https://ilostat.ilo.org/data/
[7] https://unstats.un.org/unsd/snaama/Downloads

$$NAV_i = \frac{\sum_j |S_{Tij} - S_{0ij}|}{2T}$$

where i refers to economies and S_{tij} is the share of sector j in total employment in period t in economy i.

Labor productivity is the main focus of the other indicators in this second set of indicators. Labor productivity is defined at the sectoral as well as the aggregate level and is calculated as value added (GDP) divided by employment. We provide the trend for both overall labor productivity, as well as the two components, value added (GDP) and employment, with the trend defined as a series of index numbers with base year 2015 (2015 = 100). This index is constructed simply as 100 times the value in a particular year divided by the 2015 value.

The indicators for the impact of structural change on labor productivity growth are based on a decomposition of labor productivity growth. This can be represented as follows (we drop the economy subscript *i*):

$$\frac{\hat{P}}{T} = \frac{P_T - P_0}{TP_0} = \frac{\sum_j S_{Tj}P_{Tj} - \sum_j S_{0j}P_{0j}}{T\sum_j S_{0j}P_{0j}} =$$

$$\underbrace{\frac{\sum_j P_{0j}(S_{Tj} - S_{0j})}{TP_0}}_{\text{SSRE}} + \underbrace{\frac{\sum_j (P_{Tj} - P_{0j})(S_{Tj} - S_{0j})}{TP_0}}_{\text{DSRE}} + \underbrace{\frac{\sum_j S_{0j}(P_{Tj} - P_{0j})}{TP_0}}_{\text{WS}}$$

where P_T and P_0 are aggregate (labor) productivity in year T and 0, respectively, and P_{tj} is labor productivity in period t and sector j. The three terms into which total productivity growth (between years T and 0) is decomposed are on the righthand side of this expression, and are labelled SSRE for Static Structural Reallocation Effect, DSRE for Dynamic Structural Reallocation Effect and WS for Within Sector. SSRE measures the effect on productivity growth of reallocating resources (i.e., labor) to sectors with higher productivity levels at the start of the period (year 0), DSRE measures the effect of reallocating resources to sectors with a faster growth rate of productivity between years T and 0, and WS measures the contribution of productivity growth that occurs within sectors.

Figure 4 shows a set of diagrams that illustrate the structural change indicators for Indonesia. The top part of the diagram shows the value of the *NAV* indicator for various periods, which are displayed on the horizontal axis. We see that structural change in Indonesia was strongest during the 1991-1995 period, and weakest during the 2001-05 period. Note that not all periods on the horizontal axis are equal in length, but the *NAV* indicator is scaled to yearly changes, i.e., it is comparable between periods of unequal length.

The bottom part of the figure shows the productivity decompositions, again for different periods on the horizontal axis (where again the effects are scaled for differences in period length). Here the black line indicates total labor productivity growth over the different time periods. This growth rate is equal to the sum of the stacked bars, which represent the three effects introduced above. The figure reveals that labor productivity growth in Indonesia was negative during the period 1996-2000, which encompasses the "Asian crisis". The WS effect dominates in most periods, while the DSRE is always negative. The SSRE peaks in the period just before the Asian crisis (1991-95).

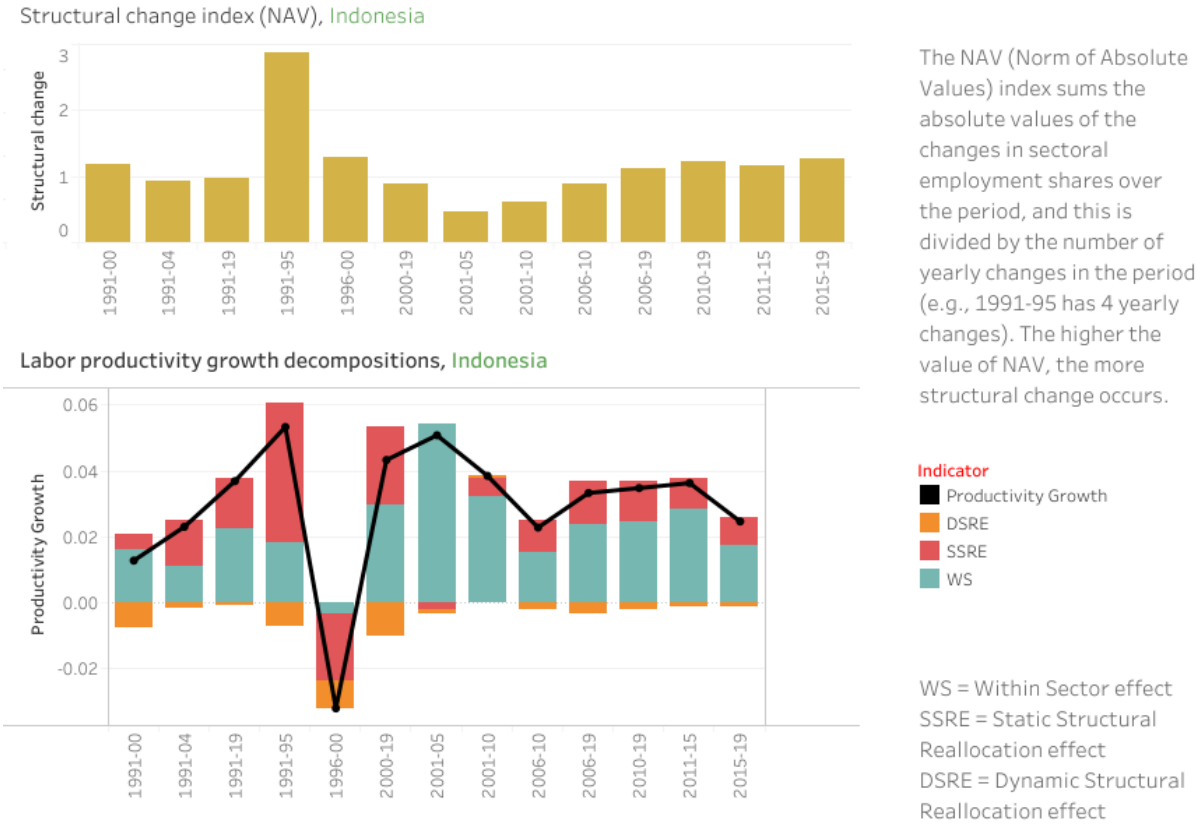

Figure 4. Structural change and labor productivity growth in Indonesia

Figure 5 shows a further feature of the website, with the figure reporting a breakdown of the three effects in the decomposition (WS, SSRE and DSRE) into their components at the sectoral level. This uses the following definitions (we omit the time scaling variable T):

$$change\ in\ labor\ productivity_j = \frac{P_{Tj} - P_{0j}}{P_0}$$

$$initial\ employment\ share_j = S_{0j}$$

$$change\ in\ employment\ share_j = S_{Tj} - S_{0j}$$

$$initial\ relative\ productivity_j = \frac{P_{0j}}{P_0}$$

It is easily seen that these four components can be used to construct the three effects in the decomposition:

- SSRE is the sum across sectors of $initial\ relative\ productivity_j \times change\ of\ employment\ share_j$
- DSRE is the sum across sectors of $change\ of\ productivity_j \times change\ of\ employment\ share_j$
- WS effect is the sum across sectors of $initial\ employment\ share_j \times change\ of\ productivity_j$

Using the example of Indonesia, Figure 5 plots for each of the three decomposition terms the two constituent parts of the decomposition term for each sector against each other. Effects that are positive are plotted in blue and negative contributions are plotted in red, while the size of the bubbles indicates the magnitude of the absolute value of the effect. In the case of Indonesia, and focusing on the period 1991-2000, it can be seen that the Mining & utilities sector makes a strong contribution to the WS effect, which arises largely due to rapid productivity change, despite the relatively small employment share. Similar conclusions can be drawn for the other two effects, DSRE and SSRE.

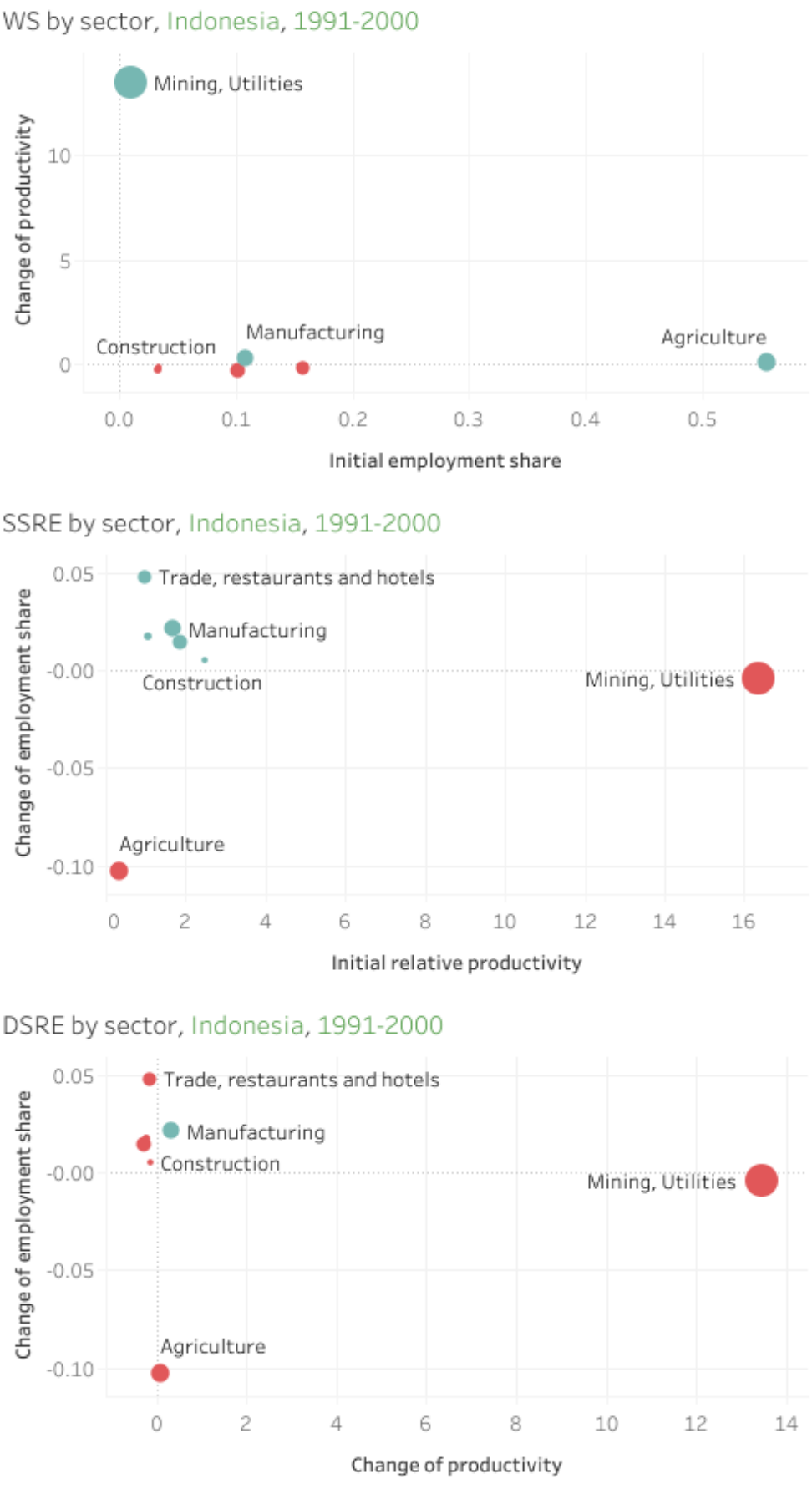

Figure 5. Sector labor productivity decompositions view in the database

The database further reports relative productivity levels for pairs of economies. An example of this is provided in Figure 6 for Japan and India. These relative productivity indicators per sector are calculated as follows:

$$RelP_{kljt} = \frac{P_{kjt}XR_k}{P_{ljt}XR_l}$$

where P is labor productivity as before (measured in 2015 US$ per worker), XR is a real exchange rate factor, and the subscripts k, l, and j denote two countries and a sector respectively (j can also denote the total economy). The real exchange rate factor *XR* is the ratio between the World Bank's PPP exchange rate for international $ and the market exchange rate for US$, both in 2015 (the base year for the price index used in the productivity calculations). Note that *XR* converts the productivity data into units that reflect domestic purchasing power. In the graph we see that relative productivity in Japan compared to India is gradually falling over the period 1991-2019.

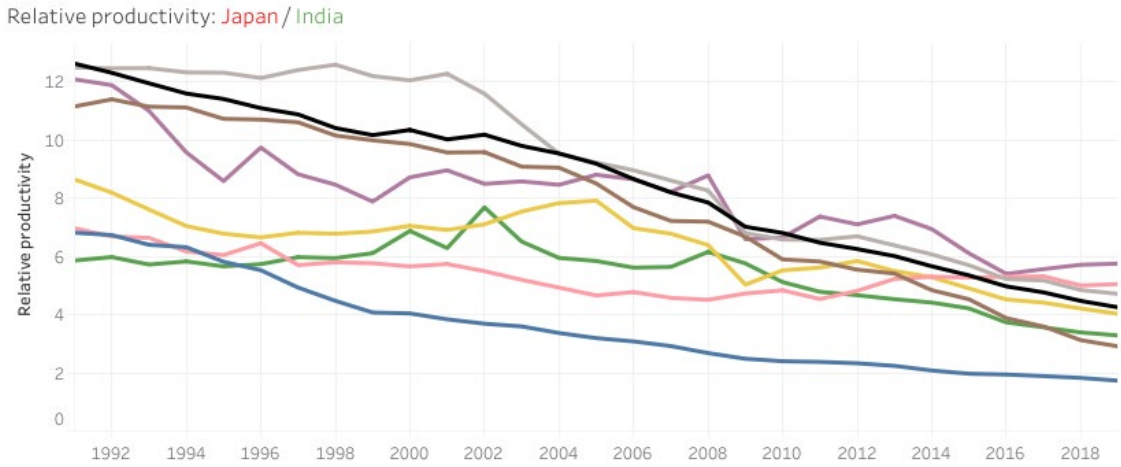

Figure 6. relative labor productivity view in the database

Table 5 documents the variables that are available in the structural change part of the database.

Table 5. Indicators in structural change and productivity section

Index (2015=100) of employment, by sector and for total economy
Index (2015=100) of value added (GDP), by sector and for total economy
Index (2015=100) of labor productivity, by sector and for total economy
Sectoral shares of employment
Sectoral shares of value added (GDP)
Aggregate labor productivity growth for various (sub)periods
Within Sector (WS) effect in aggregate labor productivity growth for various (sub)periods
Static Structural Reallocation Effect (SSRE) in aggregate labor productivity growth for various (sub)periods
Dynamic Structural Reallocation Effect (DSRE) in aggregate labor productivity growth for various (sub)periods
$\boldsymbol{change\ of\ productivity_j}$ contribution to the decomposition effect in the sector
$\boldsymbol{initial\ employment\ share_j}$ contribution to the decomposition effect in the sector
$\boldsymbol{change\ of\ employment\ share_j}$ contribution to the decomposition effect in the sector
$\boldsymbol{initial\ relative\ productivity_j}$ contribution to the decomposition effect in the sector

3. Pillar 2: Product Complexity

The pillar on product complexity builds on the idea that the performance of economies in international trade is determined by productive capabilities, and that detailed data on exports can therefore provide an indication of the capabilities that are present in the productive structure of an economy. The starting point for the construction of most of the indicators in this pillar is information on whether two (exported) products are jointly produced with comparative advantage by a single economy, information that is taken as an indication that the production of these two products share certain capabilities.

The indicators in this group are based on detailed product level export data (with information on more than 5,000 products). Using this data an established method from the existing literature, known as the ECI (Economic Complexity Index) method, is adopted to calculate an indicator of product quality from the product-level export data (Hidalgo and Hausmann, 2009; Hidalgo, 2021). This indicator of product quality is referred to as *product complexity* in the database. Conceptually, product complexity is an indicator of the production capabilities that are needed to sell the product with comparative advantage in international markets. Thus, each of the 5,000+ products in the database has its own value of product complexity, with the complexity of the products that an economy exports with comparative advantage further being an indicator of the capabilities of firms located in a particular economy. The formal definition of the product complexity indicator is provided in the annex.

3.1. Basic product complexity indicators and upgrading

In this part of the database, a number of basic indicators of product complexity are presented, along with information on the way in which complexity is distributed over economies and how this distribution affects the upgrading potential of economies. The basic idea behind the upgrading potential is that it is generally advantageous to move into products with higher complexity because these markets are generally more exclusive and require higher capabilities, and hence offer a higher potential for creating value added.

Product complexity indicators are based on the notion of comparative advantage, which is calculated as follows

$$c_{ij} = 1 \; if \; \frac{x_{ij} \Big/ \sum_q x_{qj}}{\sum_p x_{ip} \Big/ \sum_q \sum_p x_{qp}} > 1 \; and \; 0 \; otherwise$$

where c_{ij} is the (binary) comparative advantage of economy i in product j, and x_{ij} is the value of exports of economy i in product j. In words, the indicator of comparative advantage is one (i.e., the economy has a comparative advantage in product j) if the economy's share in total exports of the product is larger than its share in exports of all products together, and zero otherwise.[8] The values of comparative advantage c_{ij} can be arranged in a matrix, *R*, with each row representing a product, and each column an economy. The sum over rows in a column is the number of products in which the economy has comparative advantage, and we refer to this as the diversification level of the economy:

$$diversification_i = \frac{1}{m} \sum_q c_{iq}$$

where m is the number of products in the database, such that the diversification index represents the share of all products in which economy i has comparative advantage. Similarly, the ubiquity of a product is defined as:

$$ubiquity_j = \sum_p c_{pj}$$

Which is simply the number of economies that export a particular good, j, with comparative advantage. Using the diversification and ubiquity indicators, we follow Hidalgo and Hausmann (2009) in defining a measure called standardness:

[8] An equivalent definition of comparative advantage would be that the export share of product j in country i is greater than the export share of product j across all countries (i.e., the world).

$$standardness_i = \frac{\sum_q c_{iq} ubiquity_q}{m \times diversification_i} \Big/ Max_q(ubiquity_q)$$

Note that the expression in the numerator is the average ubiquity of the products in which the economy has comparative advantage. The denominator is the maximum value of ubiquity over all products, which we use (contrary to the definition in Hidalgo and Hausmann, 2009) to scale the standardness measure (which is thus bound by $\langle 0..1]$).

The database also contains sectoral versions of standardness and diversification. For this, we assign each of the products in the export database to an economic sector.[9] This implies that this part of the database is limited to sectors that produce tradeable goods, which rules out, for example, many of the services sectors.[10] The sectors for which we present indicators are listed in Table 6 below. The indicators are calculated by summing over the products that belong to a sector:

$$diversification_{ik} = \frac{1}{m_k} \sum_{q \in k} c_{iq}$$

and

$$standardness_{ik} = \frac{\sum_{q \in k} c_{iq} ubiquity_q}{m_k \times diversification_{ik}} \Big/ Max_{q \in k}(ubiquity_q)$$

where k indicates a sector, and m_k is the number of products in the sector.

The calculation of product complexity follows the ECI method of Hidalgo (2021). For this, we first create a row-normalized and a column-normalized version of the comparative advantage matrix *R*. The row-normalized version is created by dividing each element of the matrix by diversity of the column, and the column-normalized version is created by dividing each element by ubiquity of the row. We then create a new matrix by pre-multiplying the column-normalized version of the matrix by the transpose of the row-normalized version. Product complexity is equal to the eigenvector that belongs to the second-dominant eigenvalue of this matrix.[11]

Based upon the constructed indicator of product complexity, economy fitness levels are calculated as the average value of product complexity of the products that the economy has comparative advantage in:

[9] This assignment is done using concordances between ISIC (for the input-output sectors) and the Harmonised System (HS) for the export database, or, as introduced below, concordances between HS and BEC. These concordances are available on the World Bank's WITS server https://wits.worldbank.org/.

[10] The data used to construct the various indicators in this pillar is the UN Comtrade database, which reports information on goods trade only. Data on services trade does exist – and is also available from the UN – but consistent data across economies is only available at a much higher level of aggregation.

[11] The eigenvectors have ambiguous sign, and we choose the sign such that products with low ubiquity tend to have high complexity.

$$fitness_i = \frac{\sum_q c_{iq} C_q}{m \times diversification_i}$$

The sectoral version of fitness is

$$fitness_{ik} = \frac{\sum_{q \in k} c_{iq} C_q}{m_k \times diversification_{ik}}$$

Figure 7 provides an illustration of the indicators described above, with data reported on the Basic metals and fabricated metal sector in 2019. Economies' fitness in the sector is on the horizontal axis, while diversification is reported on the vertical axis. The size of the bubbles in this figure is proportional to the level of standardness. We observe a generally positive association between the two indicators, but this is not a linear relation. At first, diversification increases along with fitness, but at a certain threshold value, fitness does not increase any further. We also see that economies with high diversification and high fitness tend to be the ones with low standardness, while economies with low fitness and low diversification also have high standardness.

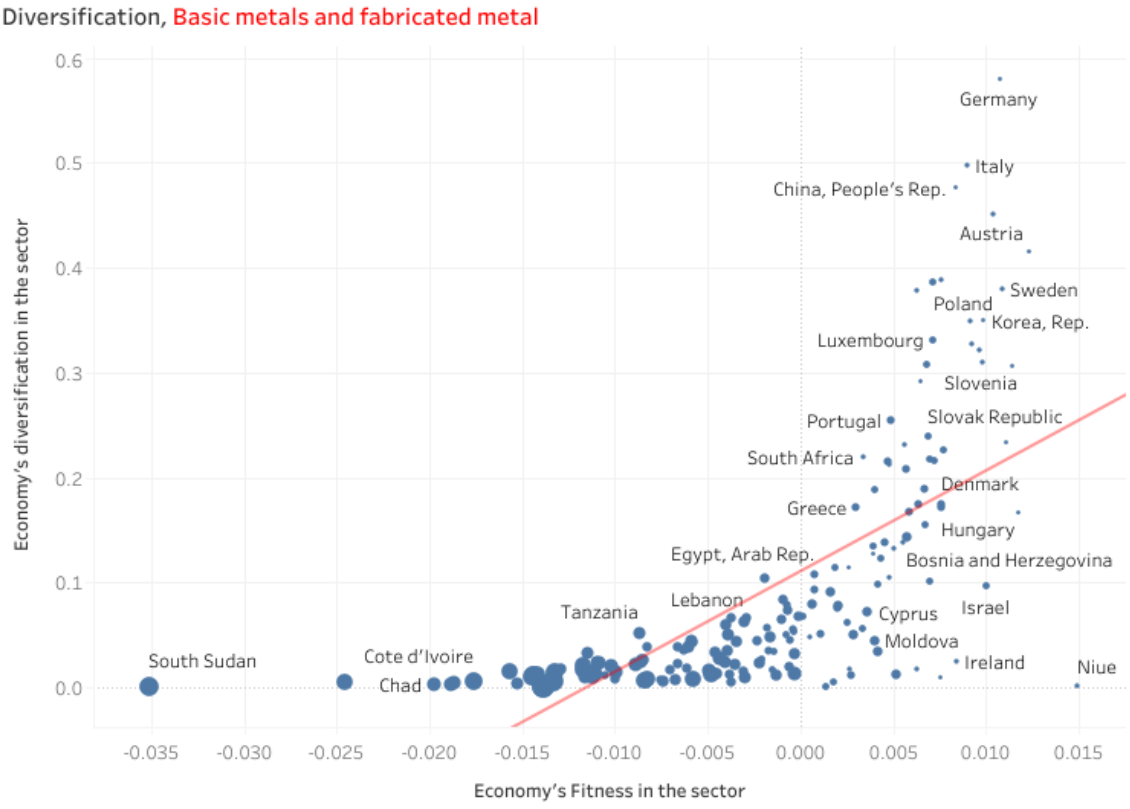

Figure 7. Diversification and fitness in basic metals and fabricated metal, 2019

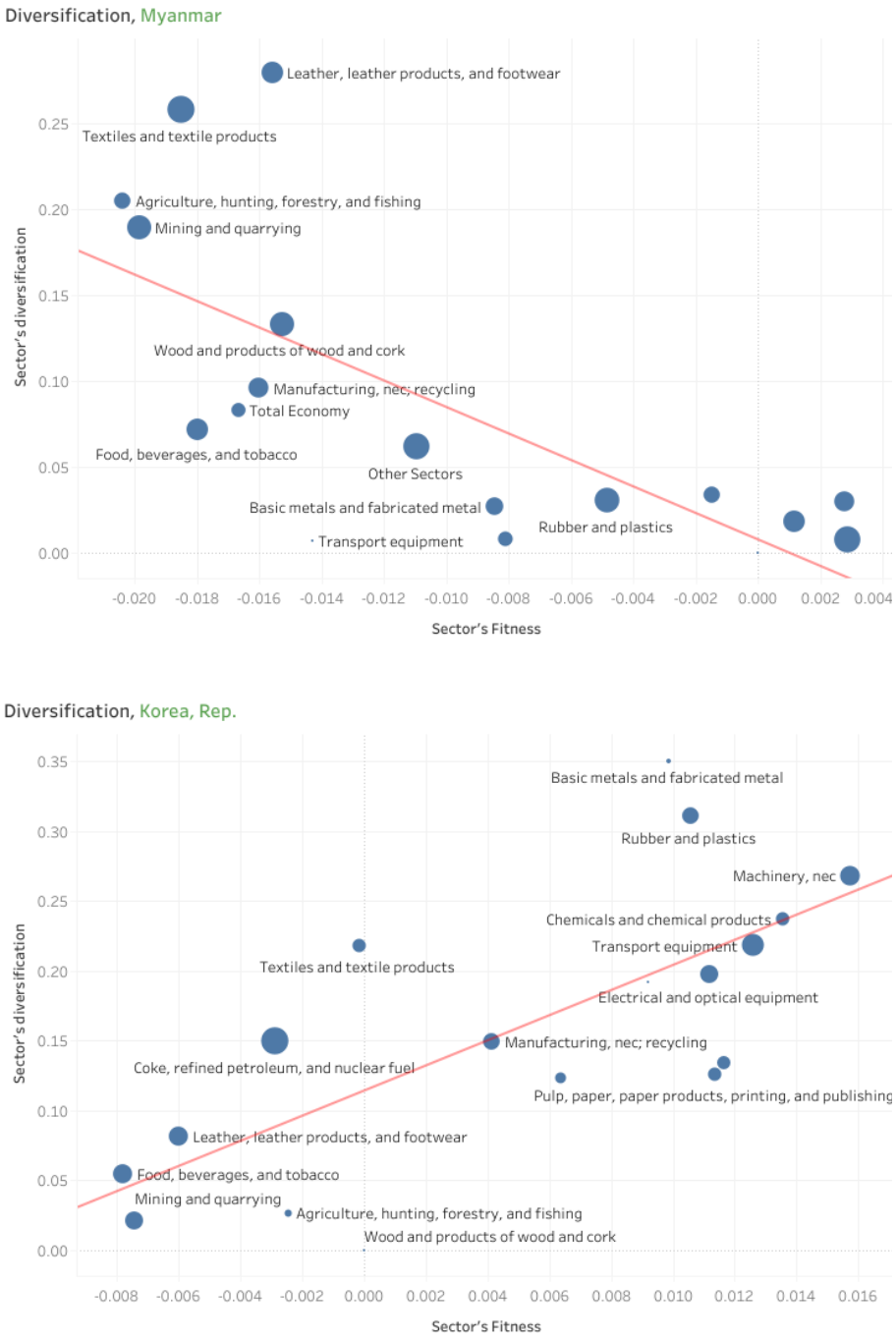

Figure 8. Sectoral diversification and fitness for Myanmar and the Republic of Korea, 2019

These indicators can also be graphed for sectors within countries, as in Figure 8. This figure reports for the Republic of Korea and Myanmar sectoral observations for fitness and diversification, with the size of the bubble again proportional to standardness. The figure reveals that in Myanmar, the relationship between the two indicators is negative, while in the Republic of Korea it is positive. This is a generally observed difference between developed economies and

less developed economies. Developed economies achieve relatively high fitness values in sectors where they have strong diversification, whereas less developed economies tend to have high fitness in sectors where they have low diversification.

A further dimension of the indicators in this pillar involves identifying products and sectors that economies may diversify into. This information is presented through the development of upgrading diagrams, which make use of one more concept from the literature, namely the concept of related variety. This means that the opportunities for developing new comparative advantages (i.e., exporting products with comparative advantage in which the economy did not have a previous comparative advantage) depends on the existing structure of comparative advantages of the economy. The reason for this is that products may share capabilities that are required for them to be exported with comparative advantage. Thus, if an economy has comparative advantage (is specialized) in a product that share capabilities with the product of interest, then the economy will likely have a higher probability of developing comparative advantage in the product of interest.

The approach involves the use of a probability measure for related variety that is explained in detail in Nomaler and Verspagen (2022), and which is further summarized in the Annex. This is first calculated at the product level, i.e., for every product in the database, we calculate the probability that an economy is specialized in this product. This probability is expressed as an offset from an autonomous probability that is constant between economies for a given product, hence this probability can be negative (indicating a low likelihood that the economy will be or become specialized in this product) or positive (a high likelihood). We call this the upgrade probability that the economy has in the specific product.

These measures are visualized in a so-called upgrade diagram, which appears in the database in both a short run and a long run version. The upgrade diagram displays sectors, more specifically a subset of economic sectors according to the ISIC rev. 4 classification (these are also the sectors for which *fitness*, *standardness* and *diversification* are available), combined with two broad BEC categories: intermediate or final goods. BEC is a classification of exports into different categories according to the stage of the global value chain that a particular product belongs to. The database aggregates BEC to (only) make a distinction between final products (used in consumption or investment) and intermediate products. Thus, each of the 5,000+ products of the export database are allocated to a sector and a BEC class (final/intermediate).

On the horizontal axis of the upgrade diagram, a measure for the economy's potential upgrade gain in the sector is reported. This is defined as the average complexity of products belonging to the sector in which the economy does not have a comparative advantage, minus the average complexity of products of the sector in which the economy has comparative advantage. The potential upgrade gain thus measures the extra complexity that the economy will achieve by gaining an "average" comparative advantage in the sector.

The potential upgrade gain indicator is also available in the general part of the database, where it can be plotted along with economy fitness (in a sector or in the total economy). This is displayed in Figure 9, which shows the observed relationship between economy fitness and potential upgrade gains in the Basic metals and fabricated metal sector in 2019. We see a tight and clearly negative relationship: economies with low (high) fitness have a high (low) upgrade potential.

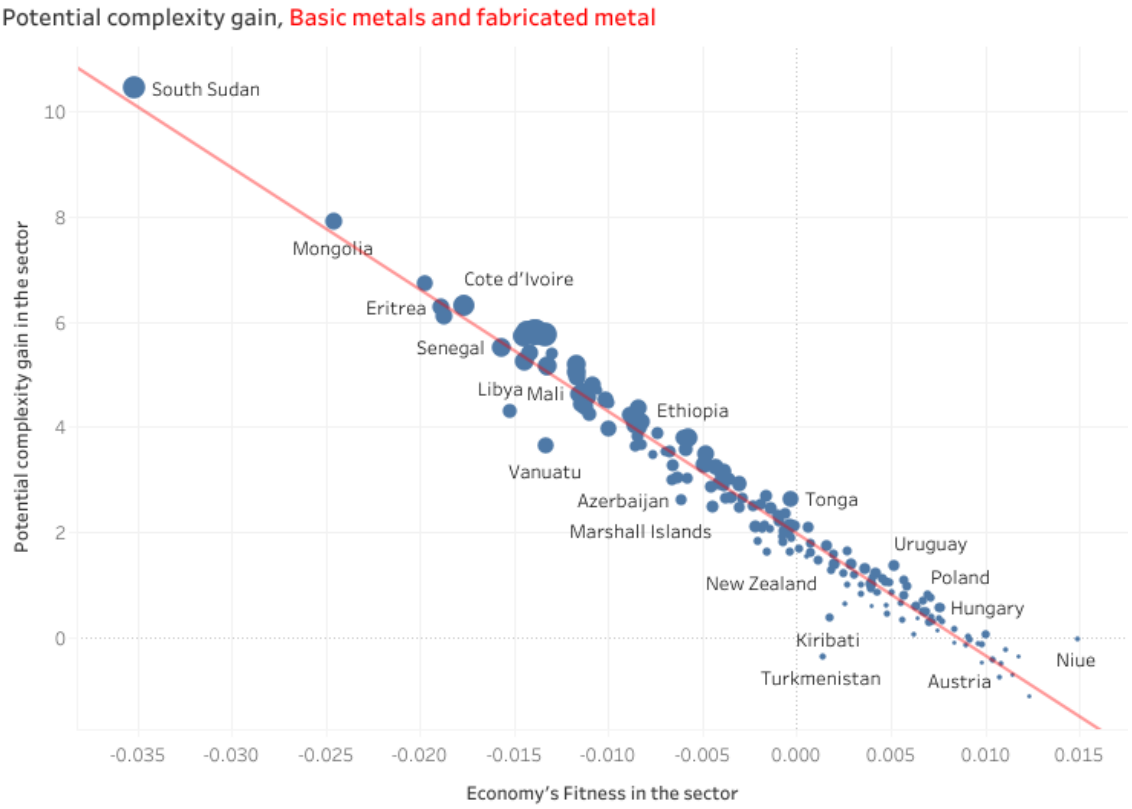

Figure 9. Potential complexity gains in Basic metals and fabricated metal, 2019

The vertical axis of the upgrade diagrams measures the economy's upgrade probability in the sector. This is the average upgrade probability of the products in the sector in which the economy does not yet have a comparative advantage. Finally, the size of the bubble in the upgrade diagram reflects the fraction of products in the sector in which the economy does not yet have a comparative advantage (the larger the bubble, the larger this fraction is).

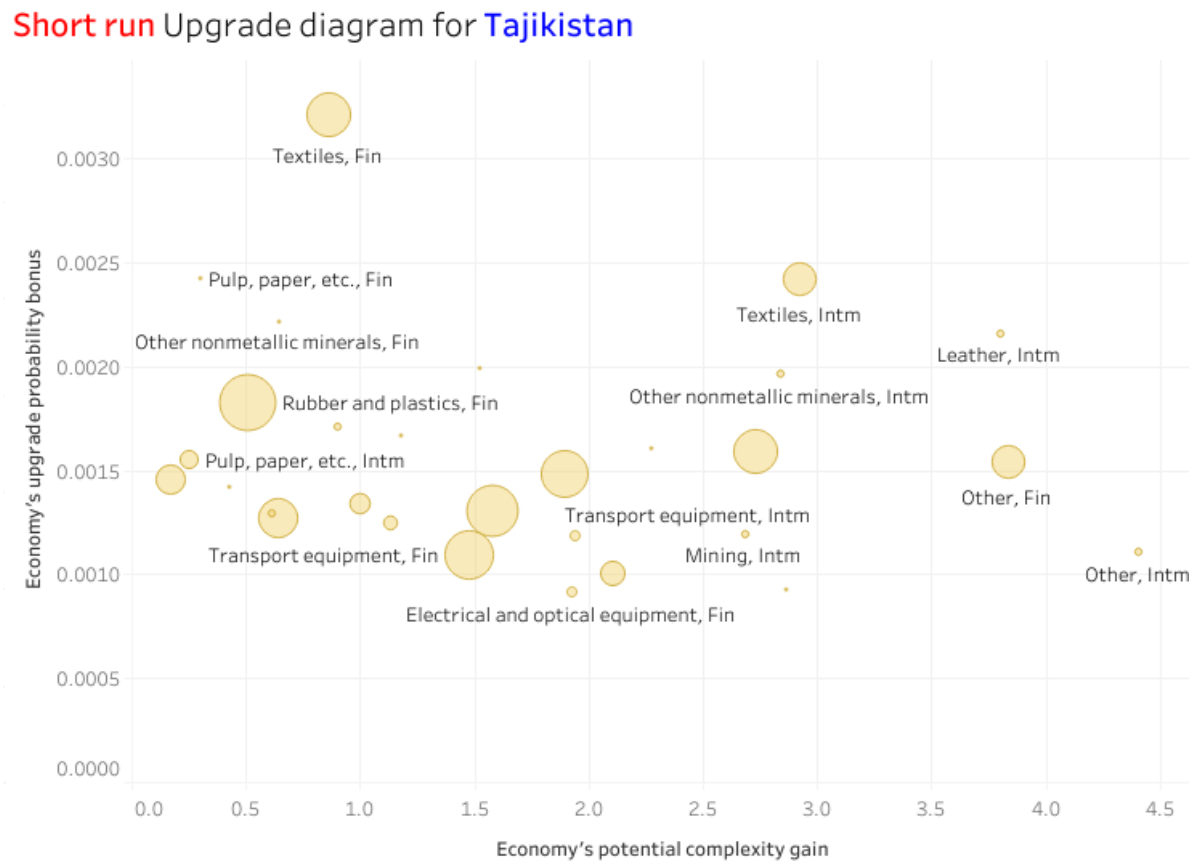

Figure 10. Upgrade diagram for Tajikistan

Figure 10 shows an example upgrade diagram for Tajikistan. In the sector labels, "Intm" and "Fin" refer to the BEC categories intermediate and final goods, respectively. Thus, "Textiles, intm" refers to intermediate goods produced by the textiles sector (ISIC C13+C14) for example. A combination of high upgrade probability and high potential complexity gain would be ideal, but as can be seen in Figure 10, these combinations (found in the upper-right corner of the diagram) do not arise easily. Most economies do not have observations in this part of the diagram, but instead show a declining envelop of the diagram that stretches from the upper left corner to the bottom right corner. For Tajikistan, this envelop can be seen to stretch from "Textiles, Fin" to "Leather, Intm". Observations on or near to this envelop show the basic tradeoff between likelihood and complexity gain of the upgrade. Observations below the envelop have either relatively low gain, or low probability, or both.

The upgrade diagrams in the database are available for three separate periods: 2011, 2016 and 2019. For each of those years, we use the sum of exports in the year itself and the previous year (i.e., 2010+2011, 2015+2016 and 2018+2019) to calculate comparative advantages and product

complexity.[12] The set of exported products differs slightly between those three periods, since the version of the Harmonised System (HS) that is native to the different years (HS2007, HS1012 and HS2017, respectively) is used in the construction of the indicators.

The short run upgrade diagram is based on the empirical observations (that is, on comparative advantages that the economy *has* in the specific time-period, which in turn lead to upgrade probabilities, and to product complexity) in the data for the selected period. Conversely, the long-run upgrade diagram assumes that the economy has already gained additional comparative advantages in the 20% of products with the highest upgrade probability. Assuming that these upgrades have actually happened, but no upgrading took place in other economies, the upgrade probabilities and potential complexity gains are re-calculated for the economy under consideration, which then comprise the long-run upgrading opportunities.

Table 6. Sectors used in the basic complexity and upgrading section

Sec #	ISIC Rev 4 code	Sector
1	A	Agriculture and fishing
2	B	Mining
3	C10-C12	Food, beverages, and tobacco
4	C13, C14	Textiles
5	C15	Leather
6	C16	Wood and products
7	C17, C18, J58	Pulp, paper, etc.
8	C19	Refined petroleum etc.
9	C20, C21	Chemicals
10	C22	Rubber and plastics
11	C23	Other nonmetallic minerals
12	C24, C25	Basic and fabricated metals
13	C28, C33	Machinery, nec
14	C26, C27	Electrical and optical equipment
15	C29, C30	Transport equipment
16	C31, C32	Other manufacturing
17	All other tradeable goods	Other

Note: in the upgrade diagrams, all sectors are available with two subclasses: "Intm" for intermediate products, and "Fin" for final products.

Table 7. Indicators available in the basic complexity and upgrading section

Diversification, total economy and sector
Standardness, total economy and sector
Fitness, total economy and sector
Potential complexity gain, total economy and sector (yearly)
Upgrade probability bonus, short run
Potential complexity gain, short run, as in upgrade diagrams
Fraction of products in sector without comparative advantage, short run

[12] In the calculation of product complexity used in the upgrade diagram, we use non-binary RCA, which is scaled between 0 and 1 with 0.5 as the "neutral" value. This makes product complexity less volatile over time.

Upgrade probability bonus, long run
Potential complexity gain, long run, as in upgrade diagrams
Fraction of products in sector without comparative advantage, long run

3.1. Product complexity in the context of global value chains

As explained in Section 5 below, an input-output accounting framework is used to construct indicators for global value chains. In order to relate product complexity to global value chains, the assignment of products to sectors as described in Table 1 is used. Based upon this assignment the following indicators for average product complexity of a sector j in the input-output database can be defined (note that j denotes a combination of sector and economy, e.g., the Japanese textiles industry):

$$C_j^F = \sum\nolimits_{k \in F} \frac{x_{kj}}{x_j^F} c_k$$

$$C_j^U = \sum\nolimits_{k \in U} \frac{x_{kj}}{x_j^U} c_k$$

where C_j^F is the average product complexity of final goods exported by sector j, C_j^U is the average product complexity of intermediate goods exported by sector j, F denotes the set of all final goods (as classified in BEC) produced in sector j, U denotes the set of all intermediate goods (as classified in BEC) produced in sector j, x_{kj} is the export value of product k produced by sector j, x_j^F is the value of all final goods exported by sector j, x_j^U is the value of all intermediate goods exported by sector j, and c_k is product complexity of product k. In other words, the complexity of either final or intermediate goods exported by a country-sector is a weighted average of the complexity of the products exported by that country-sector, with the weights being the share of a product's exports in the final or intermediate exports of that country-sector.

Based on the latter of the above definitions, a further GVC-related indicator for product complexity is constructed, which is referred to as the input-complexity of GVCs:

$$W_j = \sum\nolimits_i \frac{v_{ij}}{f_j} C_i^F$$

where W_j is the input-complexity of the value chain of sector j, v_{ij} is the value-added contribution of sector i to the total value of the value chain of sector j, and f_j is the total value of all (final goods) deliveries of the value chain. Note that the vector v_{ij} for sector j will be defined in Section 5 below. While W_j denotes total input complexity of the value chain, it can also be split into a domestic and foreign part:

$$W_j^{dom} = \sum_{i \in dom} \frac{v_{ij}}{\sum_{k \in dom} v_{kj}} C_i^F$$

$$W_j^{for} = \sum_{i \in for} \frac{v_{ij}}{\sum_{k \in for} v_{kj}} C_i^F$$

Where the superscripts dom and for denotes domestic and foreign, or the set of domestic or foreign input sectors to the chain, respectively. In addition to input complexity of the value chain, output complexity of the value chain can simply be defined as C_j^F.

Figure 11 documents input complexity versus output complexity in the Food and beverages sector, for 2019. We see a clear positive relationship: economies with higher input complexity also tend to have higher output complexity. However, there are some economies that are clearly above (or below) the line that summarizes this relationship (i.e., the line of best fit). Those economies that are significantly above (below) the line can be considered to add comparatively much (little) complexity to the output of the value chain.

Table 8 provides the list of indicators that are available in the section of the database on product complexity in global value chains. The list of sectors that is included in the calculations, and for which data on input- and output complexity are available is documented in Table 6 above.[13]

[13] Though the sector "all other sectors" is excluded from the analysis here.

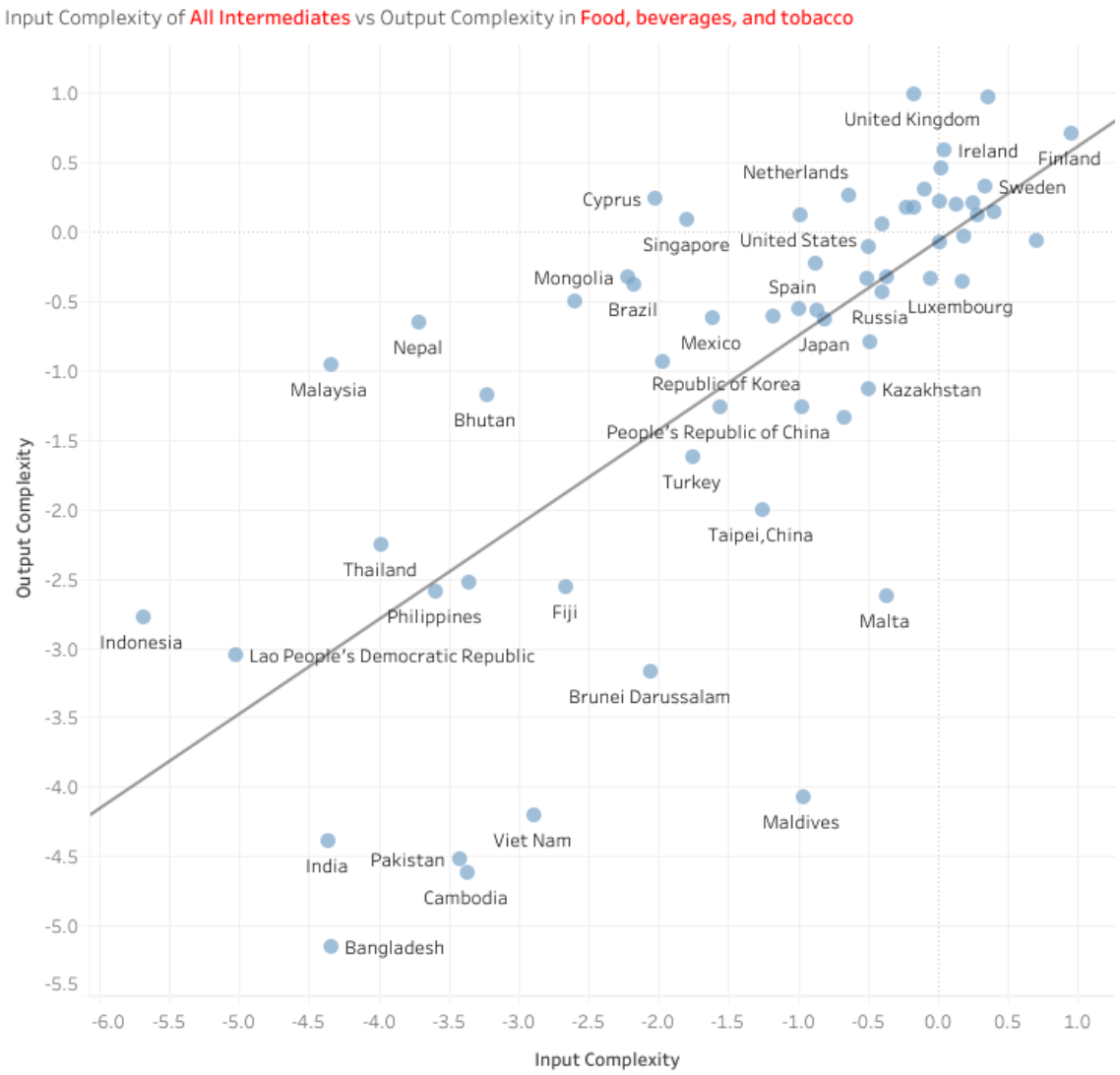

Figure 11. Input and output complexity of intermediates in the Food, beverages and tobacco sector in 2019

Table 8. Indicators in the product complexity in global value chains section
Output complexity of the chain (final goods)
Complexity of intermediate goods of the sector (output)
Complexity of all goods (final and intermediate) of the sector (output)
Input complexity all domestic and foreign value (total) into the chain
Input complexity of domestic value into the chain
Input complexity of foreign value into the chain

4. Pillar 3: Innovation

Innovation is a multi-faceted process, in which many different kinds of inputs and outputs can be found. Technology is an important input into the innovation process, but innovation can also be organizational, or be related to marketing and design, among other things. In our database, we

consider only indicators related to technological innovation, and within this restriction we cover only patented inventions. There are many well-known drawbacks of patents as indicators of (technological) innovation, such as the fact that not all patented inventions lead to actual innovations that appear in the market. Also, the propensity to patents inventions differs greatly by industry. The value of patents also differs widely, with a majority of patents being of very little value, and a few patents accounting for the large majority of total value of patents. In spite of all these drawbacks, patent indicators are developed and used in the database since they are available for a large number of economies and can be used to construct indicators that have a high degree of comparability between economies.

The patent indicators used in the database are constructed using the PATSTAT database. This is a database with information on individual patent applications from a variety of patent offices around the world. PATSTAT is provided by the European Patent Office (EPO), and contains, among other things, information on so-called patent families. A patent family is defined as a set of related patent documents (applications and grants) that originate from different patent offices around the world but cover a single invention. Each of the different documents, or family members, provide protection in a specific geographical jurisdiction, which is often an economy, but can also be multiple economies. The patent family is the basic unit that is used to construct the patent indicators in the database, particularly the use of the so-called DOCDB definition of families.

The basic approach for constructing the indicators involves counting patent families, where only families that have members in at least two patent jurisdictions (areas where protection is sought) or at a regional patent office such as the EPO are considered. By eliminating patent families with just one jurisdiction, the intention is to include only patents above a certain threshold of originality and expected economic value. When constructing the indicators, all members of patent families – whether they be patent applications or grants – are included.

PATSTAT also provides an indication of the most likely industry of origin of the patent application. This is based on a concordance between the patent class (a classification of patents into technological domains) and the NACE industry classification. This information is used to link some the patent indicators to the input-output tables that are used in the global value chain section of the database. Location of patent families (i.e., the origin of the invention covered by the patent family) is determined in a fractional way on the basis of the location of inventors.

4.1. Patent indicators for 4th industrial revolution technologies

The term 4th industrial revolution has been used to describe a set of emerging technologies in which digital connectivity, automation and artificial intelligence play a central role. These technologies are applied to a range of technological domains and productive sector (in other words, they are

pervasive). It is generally believed that these technologies have great economic potential but will also transform the workplace and society more generally.

Using the PATSTAT database 4th industrial revolution (4IR) patents are identified, with the database providing counts of the patent families that are considered to be 4IR, further distinguishing between a number of technological fields in which these patent families are classified. In order to identify 4IR patents, a methodology proposed by the European Patent office (EPO, 2020) is applied. The approach adopted does not follow the exact EPO methodology, but this approach is approximated through the PATSTAT database query that was designed to extract information on 4IR patents.[14]

Because there are relatively few 4IR patent families defined in this way (in total, there are about 200,000 4IR patent families), the database uses 10-year cumulative numbers. Thus, the data documented in the database for year T always include the numbers of patents families from year $T-9$ to T.

Figure 12 shows 4IR patenting trends for Hong Kong, China. In the upper panel we see a steady rise of the share of 4IR patenting in total global patenting, against which we can compare the share of 4IR patenting in Hong Kong, China. We see that this share is higher than the global share (for the entire period) and that it rises more steeply after 2010, suggesting that Hong Kong, China is a relatively intensive producer of 4IR technologies. The lower panel of the figure reports information on the performance of Hong Kong, China in total global patenting, with the figure showing that the share of Hong Kong, China in total global patenting is increasing over time, albeit at a slower pace after 2010.

[14] The exact EPO query could not be replicated because the Query language used by EPO is not generally available outside patent offices, and because PATSTAT does not contain the full text of the patents. The query used for the database is fully documented in Menéndez et al. (2023).

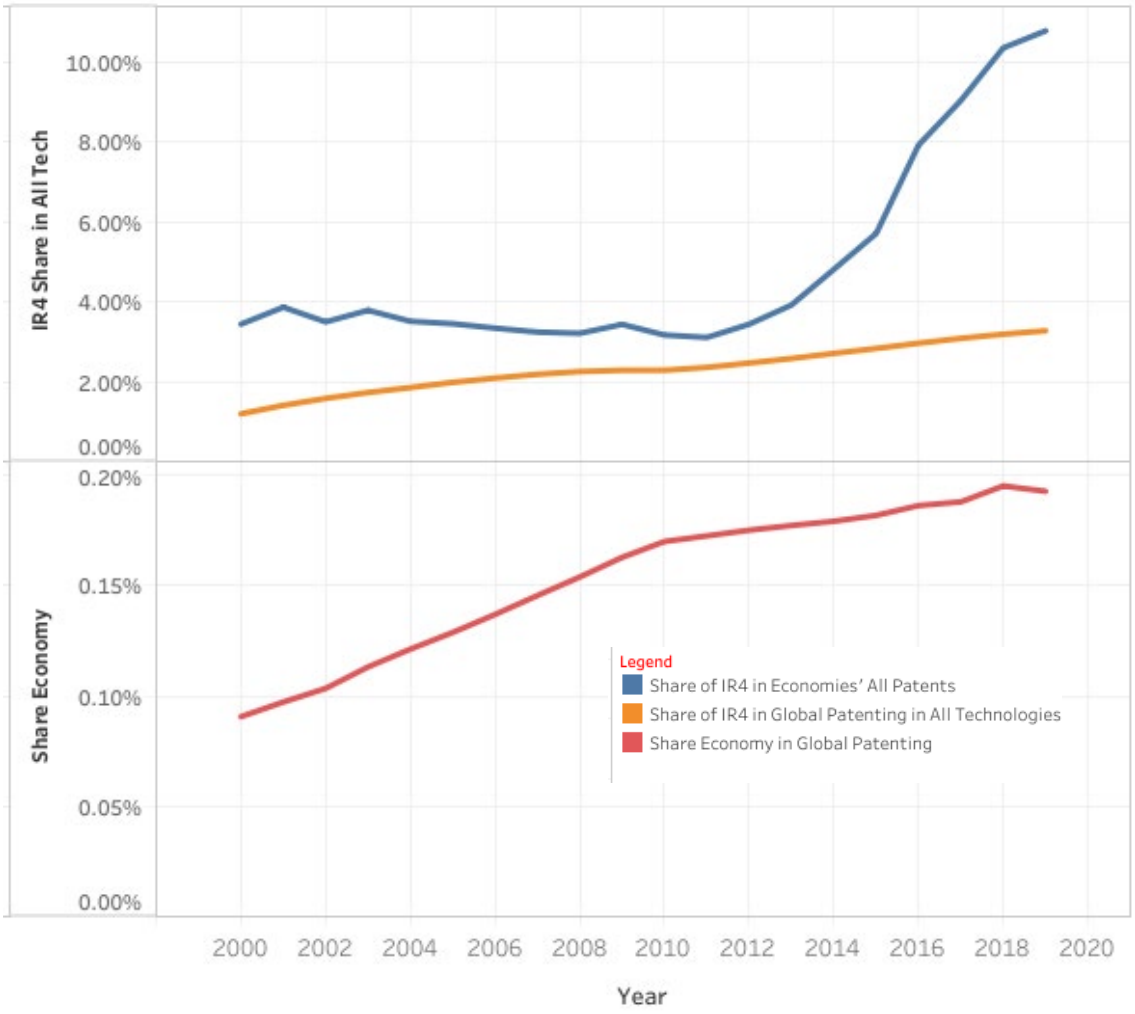

Figure 12. Total- and 4IR-patenting view in the database, Hong Kong, China

Figure 13 shows the total number of patents in 4IR in Hong Kong, China, split by subfield. Here the steep rise after 2010 is also visible. We also see that by the end of the period, connectivity, consumer goods, data security, (smart) home, IT hardware, services and software are the largest 4IR subfields in Hong Kong, China patenting. The 4IR subfields that are available in the database are documented in Table 9, while Table 10 lists the indicators in this part of the database.

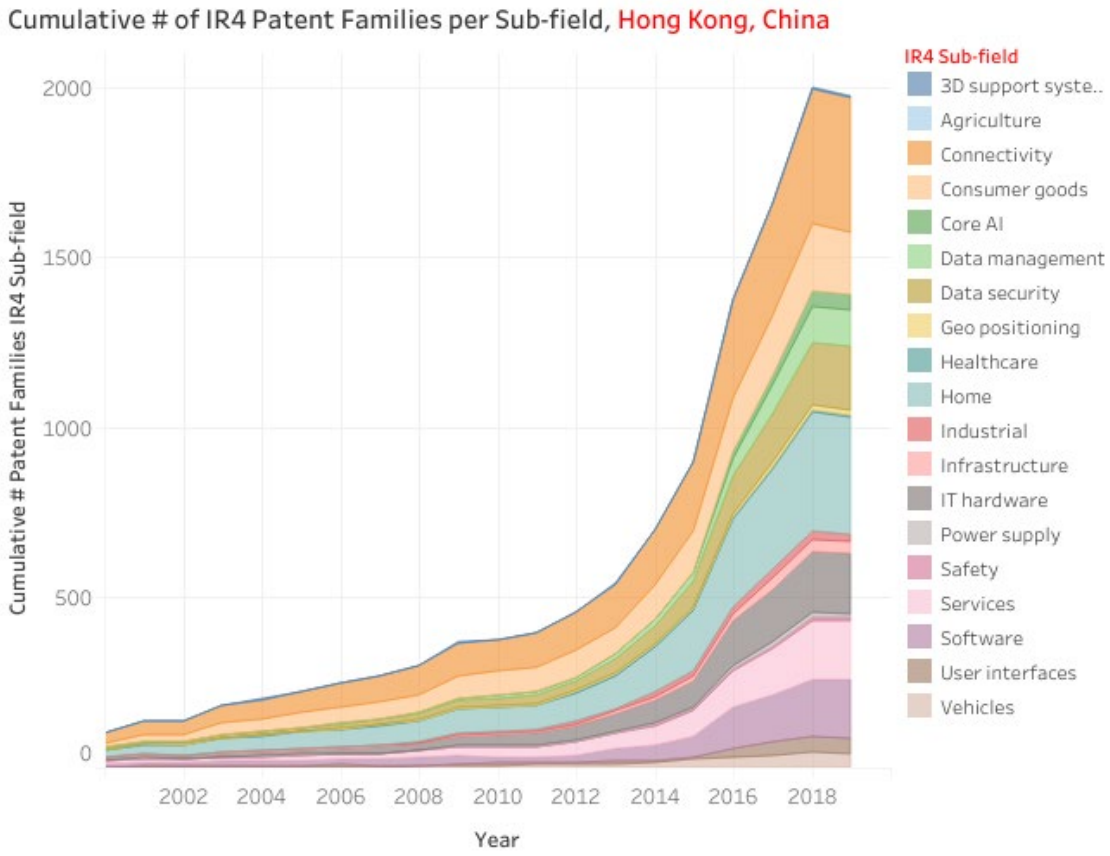

Figure 13. 4IR patenting by subfields in Hong Kong, China

Table 9. Subfields of 4IR technologies

Subfield #	Description
1	3D support systems
2	Agriculture
3	Connectivity
4	Consumer goods
5	Core artificial intelligence
6	Data management
7	Data security
8	Geo positioning
9	Healthcare
10	Home
11	Industrial
12	Infrastructure
13	IT hardware
14	Power supply
15	Safety
16	Services
17	Software
18	User interfaces
19	Vehicles
20	Total

Table 10. Variables in the 4IR patent indicators section

Patent families in all 4IR fields
Patent families by 4IR subfields
Patent families in all technology fields
Share of 4IR patent families in all patent families, per economy and globally
Share of economy in all patent families

4.2. Patent indicators in the context of global value chains

In order to relate patent counts to global value chains, a similar approach to that used to relate product complexity to global value chains (section 3.2 of this document) is adopted. Initially, an indicator called the *patent content of value added* is define, with this indicators being defined at the level of sectors in economies:

$$Q_j = \frac{Pat_j}{VA_j}$$

where Q_j is the patent content of value added in sector j (remember that this is a sector in a particular economy, e.g., the Japanese basic metals sector, or the German chemicals sector), VA_j is value added in the sector, and Pat_j is the number of patent families assigned to the sector, again

cumulated over 10 years. Two versions for the patent content indicator are defined: patents in all technology fields and patents in 4IR technology fields only, as defined in the previous section.

The indicators Q_j can thus be considered analogous to C_j^F in Section 3.1. An analogue of the W_j indicator can further be defined as:

$$O_j = \sum_i \frac{v_{ij}}{f_j} Q_i$$

Where O_j is the patent content of the value chain of sector *j* (where again this indicator is constructed for either all technology fields or 4IR fields only) As in section 3.1, this indicator can further be separated into two parts:

$$O_j^{dom} = \sum_i \frac{v_{ij}}{\sum_{k \in dom} v_{kj}} Q_i$$

$$O_j^{for} = \sum_i \frac{v_{ij}}{\sum_{k \in for} v_{kj}} Q_i$$

Where the superscripts dom and for again refer to domestic and foreign, respectively, or the set of domestic or foreign input sectors to the chain. In other words, O_j^{dom} captures the patent content of the value chain from domestic sources, while O_j^{for} captures the patent content from foreign sources.

The indicators O_j , O_j^{dom} and O_j^{for} focus on value chains, i.e., they take a production point of view. We can also use the patent content of value added to consider final demand (i.e., consumption and investment). In this way, we can construct the indicator

$$J_j = \sum_i \frac{w_{ij}}{\psi_j} Q_i$$

where w_{ij} is the value-added contribution of sector i to the final demand of sector j products in the economy to which j belongs and ψ_j is total final demand of sector *j* products in the economy to which j belongs. An example of the J_j indicator is that it could reflect the average patent content of value added delivered to final demand (i.e., consumption and investment) of transport equipment in Viet Nam (in this case, j is the Vietnamese transport equipment sector).

Figure 14 plots for all economies in the database the patent intensity of own value added against patent intensity of the value chain ("all value added used") for 2015, for the total economy. We see a strong correlation between these indicators: economies that have patent-intensive production also tend to use more patent-intensive inputs in their value chains. The Asian economies Tapei,China, Republic of Korea and Japan top the list on both indicators.

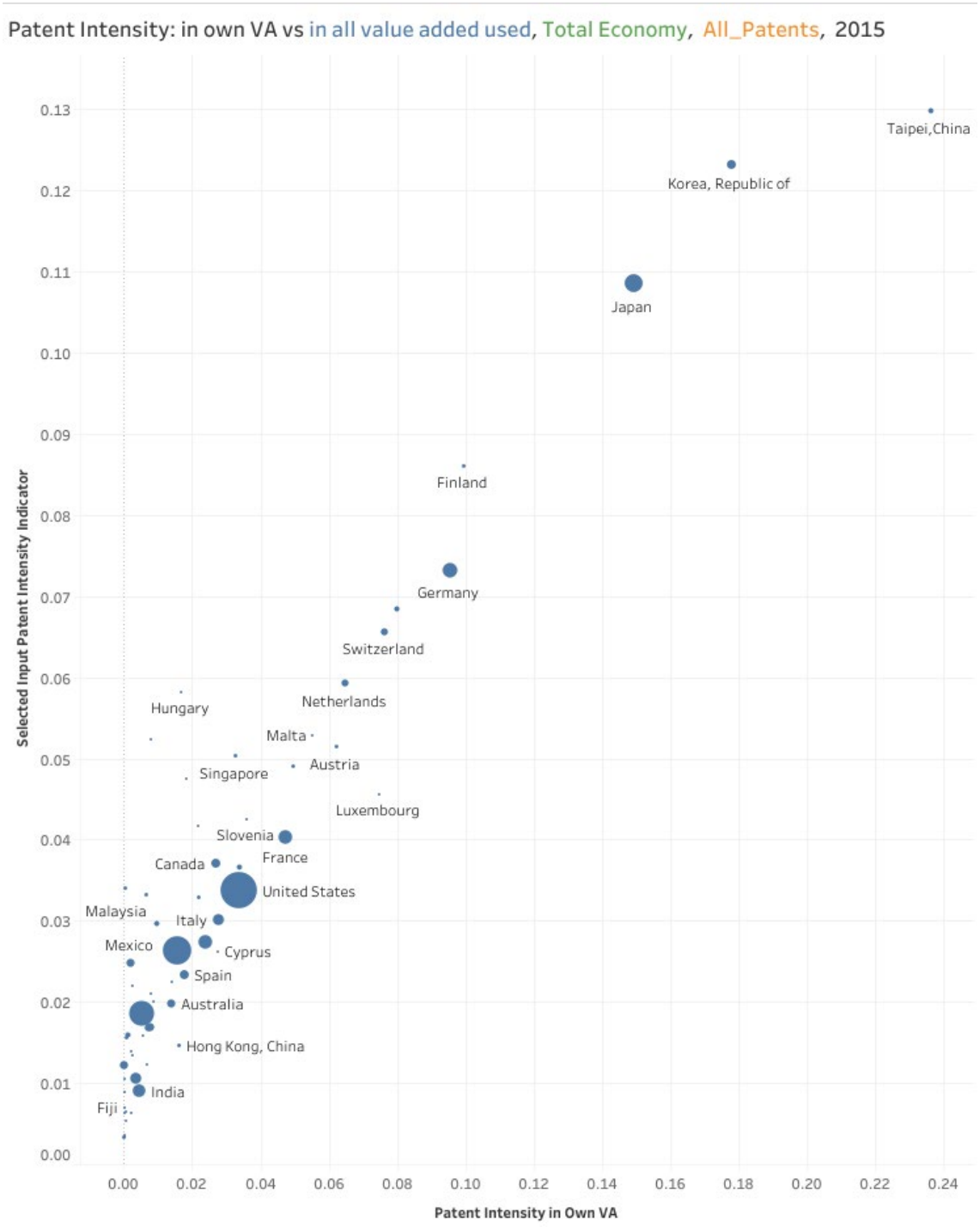

Figure 14. Patent intensity of GVCs in 2015

Table 11 lists the indicators in this part of the database.

Table 11. Indicators available in innovation and global value chains section

Patent content of value added
Patent content of the value chain
Domestic patent content of the value chain
Foreign patent content of the value chain
Patent content of final demand
Patent content of value added, only 4IR patents
Patent content of the value chain, only 4IR patents
Domestic patent content of the value chain, only 4IR patents
Foreign patent content of the value chain, only 4IR patents
Patent content of final demand, only 4IR patents

5. Pillar 4: Indicators for Global Value Chain participation

Indicators in this pillar are heavily based on input-output tables, which requires a framework that allows for the definition and conceptualization of GVCs using such data. This involves making a distinction between so-called intermediate goods and final goods. Final goods are goods that are either used for consumption or investment (e.g., a car), while intermediate goods are goods that are used in the construction of final goods, i.e., steel that is used to build a car.

The distinction between final and intermediate also holds for services. For example, a haircut is a service delivered to final demand, while a courier service delivering business documents is an intermediate delivery. Some goods or services can either be delivered as final or as intermediary (i.e., electricity, insurance), and they may appear in two places in the input-output tables that we use.

The distinction between intermediate and final is used to conceptualize the value chain. Value chains are initially defined by sectors or industries, e.g., the steel industry or the food industry. These sectors are further defined by the economy in which they are located, i.e., we have the German car industry as well as the Italian car industry, and the Japanese food sector as well as the Brazilian food sector. The value chains in each of those economy-sector combinations are defined as the entire (so-called vertically integrated) chain that is needed to produce the final output of the sector. In the example of a car sector, it includes the cars themselves as final deliveries, and all the intermediate deliveries (goods and services) needed to produce these cars.

In this way, trade, and hence the global dimension of the value chains appears in two forms. First, there are exports (and the associated imports) of the final good itself. This is fully in line with the traditional idea of trade in the economic literature, e.g., exchanging cloth for wine. Second, intermediate goods are traded, and hence there will be an international component to the value chain in terms of the inputs that are used in the production of final goods.

One primary aim of our indicators is to measure what global value chains imply for domestic value added (GDP). This can be seen from two angles. One is from the point of view of the value chain of a sector in an economy. In this case, we can ask how much value, or employment, the sector itself contributes to the chain, and how much is contributed by producers in other sectors, either domestically or abroad. This is what is called the backward perspective below. The other perspective, the forward one, asks how much value added or employment a sector contributes to other value chains, either domestically or abroad. By nature of this distinction, a forward contribution to other value chains is always in the form of intermediate deliveries, as is the backward contribution of sectors other than that for which we define the value chain.

Using input-output accounting, we can formulate several representations of global value chains. The most basic one is as follows:

$$V = \Omega(I - A)^{-1}\Phi$$

where V is the GVC matrix, Ω is a diagonal matrix with value added coefficients (value added divided by gross output) on the main diagonal, and zeros otherwise, I is the identity matrix, A is a matrix of input coefficients (intermediate deliveries from the row sector to the column sector, divided by gross output of the column sector), and Φ is a diagonal matrix with deliveries to final demand on the main diagonal and zeros otherwise. All these matrices have dimension $mn \times mn$, where m is the number of sectors and n is the number of economies. The rows and columns of these matrices refer to a specific combination of a sector and an economy, e.g., the agriculture sector in Japan or the mining sector in Australia.

In the matrix V, each row sums to the total value added (contribution to GDP) of the sector, and each column sums to the final demand deliveries of the sector. If we consider a single column in the matrix V, i.e., a vector v_{ij} with a fixed value for j, then the cells of this vector represent value added contributions of the sector i to the value chain of sector j. These contributions are related to deliveries of intermediate goods by the sectors i to the value chain of the sector j. For example, German transport equipment value chain (represented as sector j), cells related to Australian iron ore (where the index i represents the Australian mining industry), to Indian steel (where i represents the Indian basic metals industry), to American insurance (where i represents the US services industry), and so on could be identified. Many of the entries in the vector v_{ij} will involve foreign contributions to the sector j, which is why we speak of a *global* value chain, but some elements will also be domestic, including the sector j itself as well as other domestic sectors. The columns of the matrix V are used also in other parts of the database, specifically in the sections on product complexity and global value chains (section 3.2) and patenting indicators in global value chains (section 4.2).

5.1. GVC positioning indicators

The indicators for GVC positioning are also based on the matrix V. These indicators are either classified as backward or forward. Backward refers to the column-view of the matrix, as explained above, i.e., this looks at a given value chain and asks which other sectors (domestically and abroad) contribute value to it. The forward perspective, on the other hand, looks at rows of the matrix V, and asks how a given sector contributes to a range of value chains, both domestically and abroad.

The basic backward indicator splits the total value (final demand) of the chain into three parts:

$$\beta_j^o = \frac{v_{jj}}{\sum_i v_{ij}}$$

$$\beta_j^{dom} = \frac{\sum_{i \in dom, i \neq j} v_{ij}}{\sum_i v_{ij}}$$

$$\beta_j^{for} = \frac{\sum_{i \in for} v_{ij}}{\sum_i v_{ij}}$$

where β_j^o is the share of own-sector value in the chain, β_j^{dom} is the share of other domestic sectors in the chain's total value, β_j^{for} is the share of foreign sectors in the chain's total value, and as before dom and for denote the sets of domestic and foreign sectors, respectively.

Similarly, for the forward perspective, we define the following three indicators:

$$\varphi_j^o = \frac{v_{jj}}{\sum_i v_{ji}}$$

$$\varphi_j^{dom} = \frac{\sum_{i \in dom, i \neq j} v_{ji}}{\sum_i v_{ji}}$$

$$\varphi_j^{for} = \frac{\sum_{i \in for} v_{ji}}{\sum_i v_{ji}}$$

where φ_j^o is value added contributed to a sector's own value chain as a share of total value added of the sector, φ_j^{dom} is value added contributed to other domestic value chains as a share of total value added of the sector, and φ_j^{for} is value added contributed to foreign value chains as a share of total value added of the sector.

These indicators are used to define two further indicators of domestic integration:

$$\eta_j^{bw} = \frac{\beta_j^{dom}}{\beta_j^{dom} + \beta_j^o}$$

$$\eta_j^{fw} = \frac{\varphi_j^{dom}}{\varphi_j^{dom} + \varphi_j^{o}}$$

A high value for η_j^{bw} indicates that the sector has strong linkages to other domestic sectors, rather than to itself. Similarly, η_j^{fw} will be high if the sector has strong forward linkages to other domestic sectors.

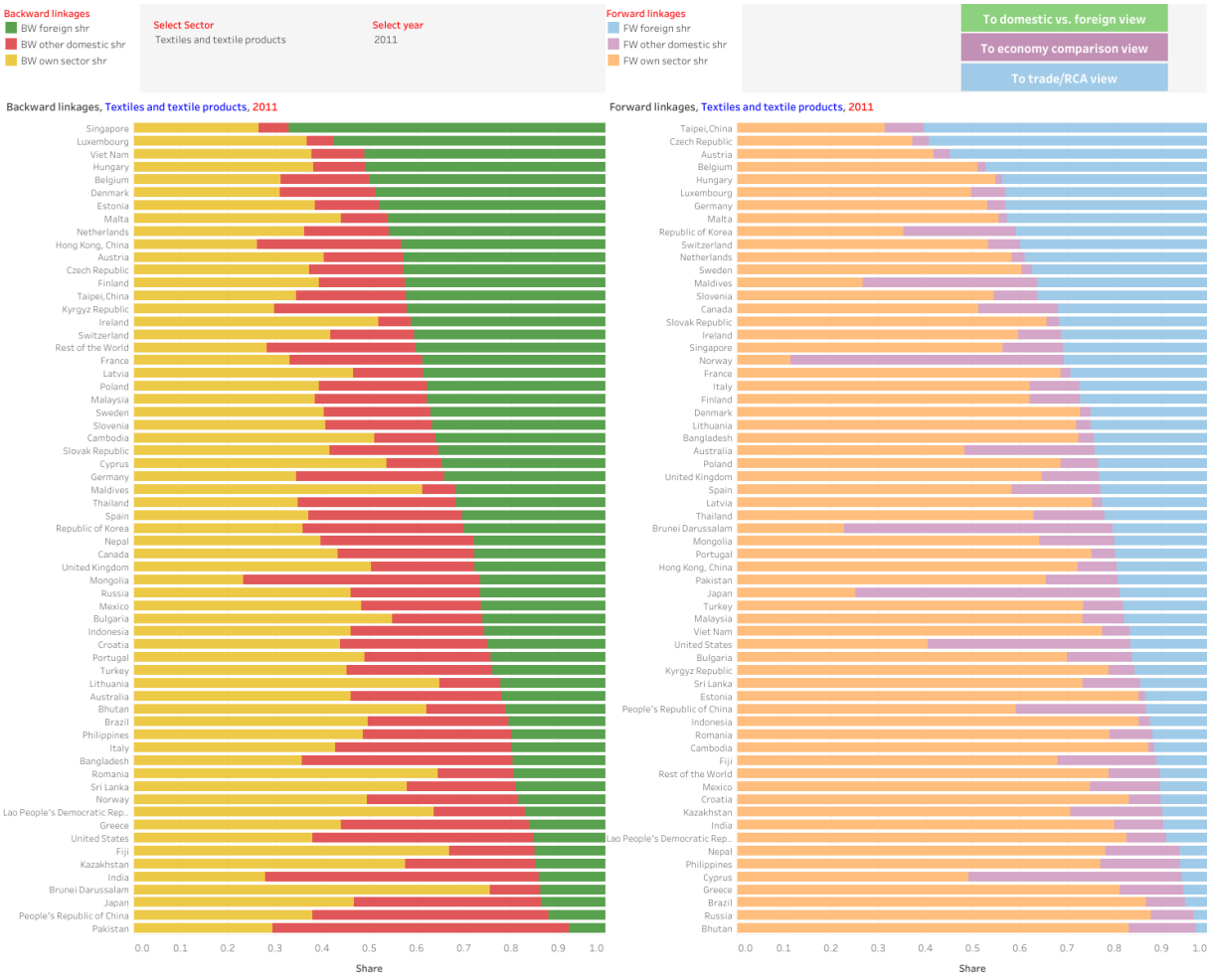

Figure 15. Backward and forward linkages in the textiles sector in 2019

Figure 15 shows information on backward and forward linkages for the Textiles and textile products sector for all economies in the dataset in 2019. The PR China and Bangladesh are the two economies that have the lowest shares of foreign value added in the final demand served by the chain (backward linkage), while Belgium has the highest share. Other domestic sectors have an important backward role in many economies, including PR China and Bangladesh, but also, for example, in India and Thailand. The own sector has a large backward role in Ireland, Brunei Darussalam, and Fiji.

In terms of forward linkages, Bhutan stands out as the economy with the smallest share of value added supplied to foreign chains, while the Czech Republic has the highest share. Most economies have a fairly small share of value added supplied to other domestic sectors, with the own value chain usually being the largest share of value added supplied.

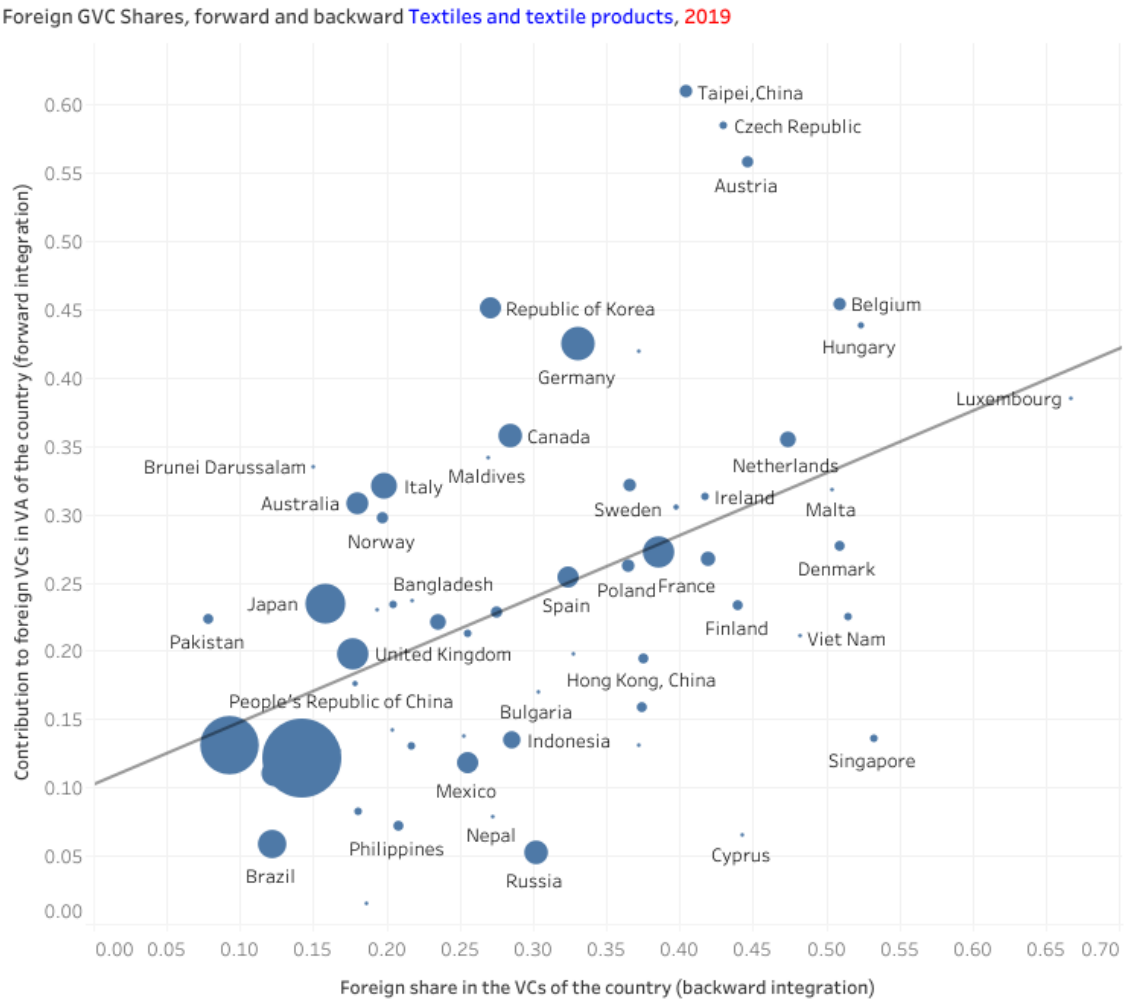

Figure 16. Foreign forward and backward integration in the textiles sector in 2019

Figure 16 shows, again for the textiles and textile products sector in 2019, foreign backward integration against foreign forward integration. The two indicators are positively correlated. The size of the dots indicates the size of the value chain in terms of total value added. Economies which have high foreign shares across both dimensions appear to the somewhat smaller open economies like Taipei,China, Austria and the Czech Republic, while larger economies like the PR China and the United States do not have large foreign shares in either dimension. To some extent, this reflects the more general observation that larger economies tend to be more diversified and have lower trade shares when compared with smaller economies.

In addition to this, the GVC integration pillar reports a number of specialization indicators based on data on international trade (imports and exports). These indicators can be split into a set of indicators for gross trade (exports or imports measured in terms of gross output), and a set for value added trade. The measures for gross trade include total gross trade and trade in intermediates. Specifically, we the following gross trade flows indicators are defined:

Gross intermediate exports of economy i to global sector $j = \sum_{p \in i} \sum_{q \notin i, q \in j} u_{pq}$

Gross intermediate exports of economy-sector $ij = \sum_{p \in i, p \in j} \sum_{q \notin i} u_{pq}$

Gross final product exports of economy-sector $ij = \sum_{p \in i, p \in j} \sum_{q \notin i} f_{pq}$

Gross total exports of economy-sector $ij = \sum_{p \in i, p \in j} \sum_{q \notin i} u_{pq} + \sum_{p \in i, p \in j} \sum_{q \notin i} f_{pq}$

Gross intermediate imports of economy i from global sector $j = \sum_{p \in i} \sum_{q \notin i, q \in j} u_{qp}$

Gross intermediate imports of economy-sector $ij = \sum_{p \in i, p \in j} \sum_{q \notin i} u_{qp}$

Gross final product imports of economy i from global sector $j = \sum_{p \notin i, p \in j} \sum_{q \in i} f_{qp}$

Gross total imports of economy i from global sector $j \ = \sum_{p \in i} \sum_{q \notin i, q \in j} u_{qp} + \sum_{p \notin i, p \in j} \sum_{q \in i} f_{qp}$

where u_{kl} is the row-k and column-l element of the matrix of intermediate deliveries of the input-output table (matrix U); f_{gh} is the row-g and column-h element of the final demand matrix F, of which the rows represent country-sectors and the columns represent countries; the subscript i denotes a country, j denotes a sector; $p \in i$ means that the row or column p belongs to country i, $p \in j$ means that the row or column p belongs to sector j; and by "economy-sector" we refer to a specific sector in a specific economy (e.g., the basic metals industry in Japan).

For value added trade, we use the matrix V, and the indicators are defined as follows:

Value added exports of economy i to global sector $j = \sum_{p \in i} \sum_{q \notin i, q \in j} v_{pq}$

Value added imports of economy i from global sector $j = \sum_{p \in i} \sum_{q \notin i, q \in j} v_{qp}$

Value added exports of economy-sector $ij = \sum_{p \in i, p \in j} \sum_{q \notin i} v_{pq}$

Value added imports of economy-sector $ij = \sum_{p \in i, p \in j} \sum_{q \notin i} v_{qp}$

This gives twelve trade measures, of which six are import measures and six are export measures, though the import and export indicators are not symmetrical. This latter observation is in the nature

of input-output accounting. The input output table (which does not contain a so-called capital flow matrix), specifies only the imports of intermediate products of the economy-sector, but not the investment goods that it imports. Final product imports, including investment, are specified in the final demand part of the table (which in addition to investment contains household consumption and government purchases). Similarly, one can account for the final product exports of economy-sectors to global economies, but not the final product exports of an economy to global sectors, since destination sectors of investment goods are unknown.

The database further reports the Balassa index for each of the 12 metrics defined above. This yields an RCA measure for exports, and a CII measure for imports (see section 2.1 for the definitions of RCA and CII). Figure 17 shows the RCA in the chemicals sector for value added exports on the vertical axis and for total gross exports on the horizontal axis. These two indicators are obviously correlated, but the correlation is far from perfect (R = 0.71).

Table 12 reports the list of indicators in this part of the database.

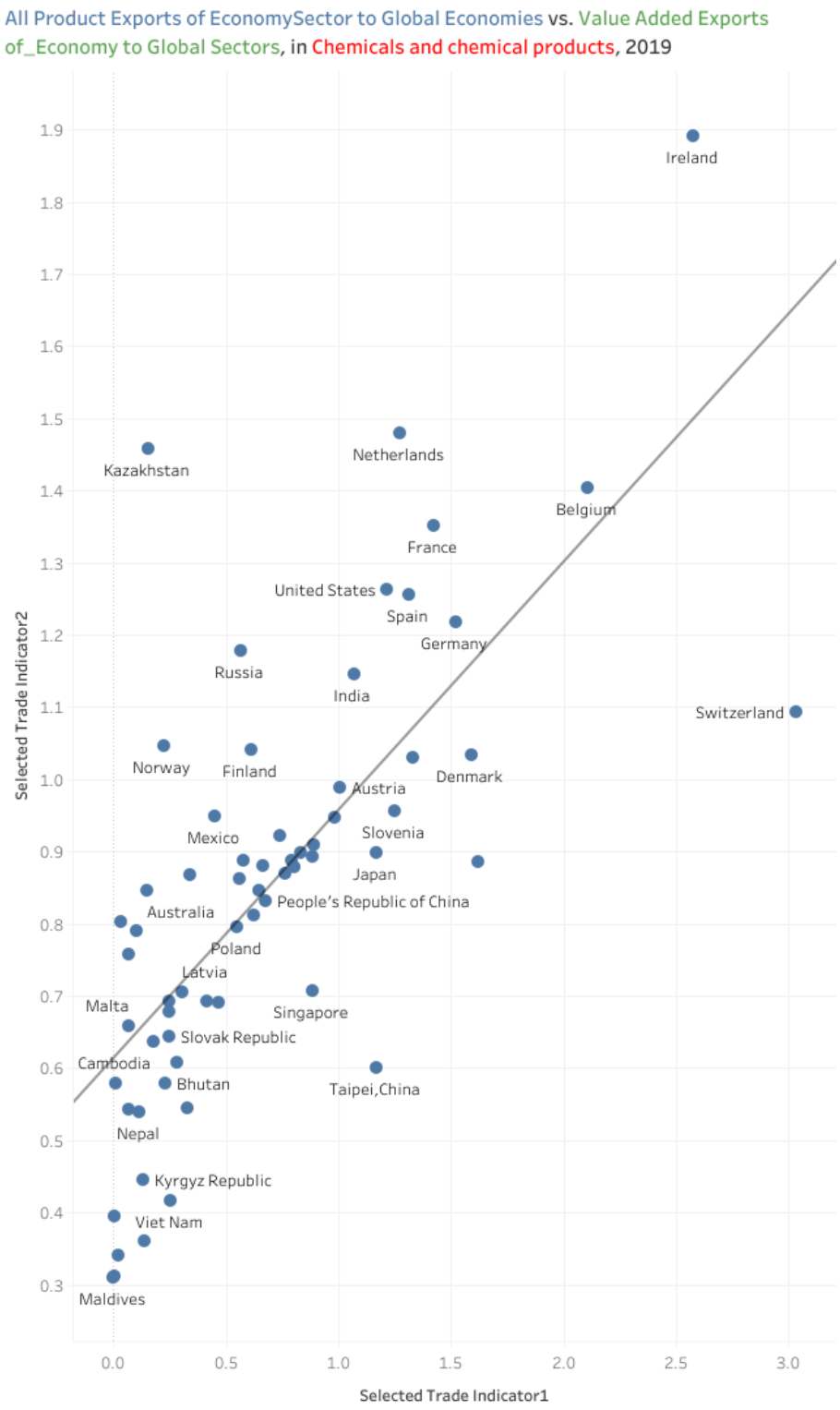

Figure 17. RCA in gross and value-added exports in Chemicals in 2019

Table 12. Indicators used in GVC positioning indicators section

Backward: share of own value added in the value chain
Backward: share of domestic other sectors in value added of the chain
Backward: share of foreign sectors in value added of the chain
Forward: share of own value chain in value added of the sector
Forward: share of value delivered to other domestic value chains as share of value added of the sector
Forward: share of value delivered to foreign value chains as share of value added of the sector
Domestic backward integration
Domestic forward integration
Value added exports of the economy to a specific global sector, RCA
Value added imports of the economy from a specific global sector, CII
Value added exports, RCA (of the exporting sector)
Value added imports, CII (of the imported sector)
Gross intermediate exports of economy to a specific global sector, RCA
Gross intermediate imports of economy from a specific global sector, CII
Gross intermediate exports of economy-sector, RCA
Gross intermediate imports from economy-sector, CII
Gross final demand exports of economy-sector, RCA
Gross final demand imports from economy-sector, CII
Gross total exports of economy-sector, RCA
Gross total imports of economy-sector, CII

5.2. Regional and global value chains

This section of the database aims to provide an impression of how "global" global value chains are. It uses the notion of geographical distance to operationalize this. This starts from a distance matrix that was obtained from CEPII[15], and which contains the geographical distance in km between the capitals of the economies in the input-output database. This distance is taken as representative of the general distance between economy-pairs. With this distance matrix and the matrix V that was introduced above the following two basic measures are constructed:

$$R_j^{bw} = \frac{\sum_{i \in for} d_{ij} v_{ij}}{\sum_{i \in for} v_{ij}} \Big/ \sum_{i \in for} d_{ij}$$

$$R_j^{fw} = \frac{\sum_{i \in for} d_{ij} v_{ji}}{\sum_{i \in for} v_{ji}} \Big/ \sum_{i \in for} d_{ij}$$

where R_j^{bw} is the backward GVC Geo Radius of the sector j, R_j^{fw} is the forward GVC Geo Radius of the sector j, and $d_{ij} = d_{ji}$ is the distance between the economy to which sector i belongs and the economy to which sector j belongs. The numerator in both expressions is the weighted average of the distance to GVC trading partners for sector j, using either backward or forward foreign linkages as weights (domestic linkages are disregarded in the construction of these indicators). The denominator in both expressions is the unweighted distance to trading partners. Hence the

[15] http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=6

indicator is a relative distance, where the GVC-weighted distance is expressed relative to the unweighted distance. A value smaller (larger) than 1 indicates that GVC trading partners tend to be relatively close-by (far away). As it turns out the values that we find in the database are almost exclusively less than one, which is consistent with the general finding in so-called gravity trade models that trade is more intensive between economies that are close to each other.

Figure 18. Backward and forward geo radius in Rubber and plastics in 2019

Figure 18 shows a plot of the backward geo radius against the forward geo radius for the Rubber and plastics sector in 2019. A positive correlation between the two dimensions exists. Generally,

the geo radiuses are observed to be below unity, which indicates that trade (within global value chains) occurs over relatively short distances. There are exceptions to this general pattern, however, with the forward radius of Indonesia and India being above one, for example.

Figure 19 shows the time series for the forward and backward geo radius of Indonesia in the Rubber and plastics sector. Here it can be observed that both of these indicators, but especially the forward radius, are rising over time, with a jump around 2014-15. Since 2017 the forward radius increases to a value above one.

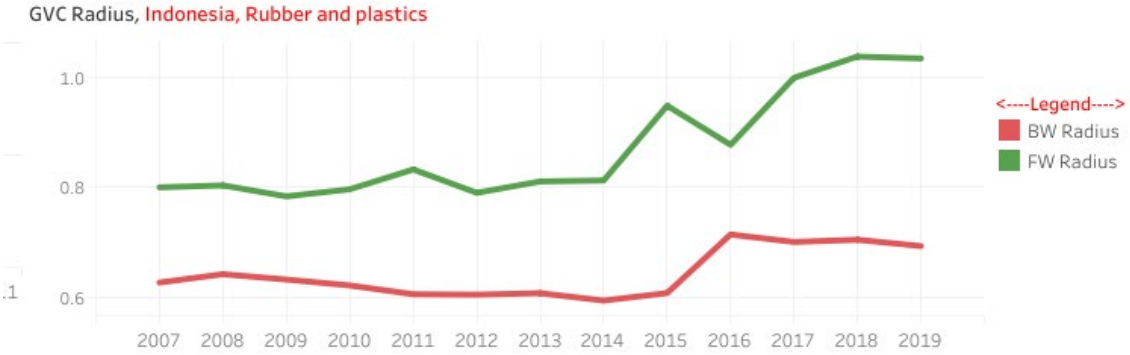

Figure 19. Sectoral backward and forward geo radius in Rubber and plastics in Indonesia, 2007-2019

The list of indicators provided in this section are listed in Table 13, while the sectors for which these indicators are available are the same as in previous sections (Table 1).

Table 13. Indicators in GVC positioning section

Backward Geo Radius of GVC integration
Forward Geo Radius of GVC integration

References

Acemoglu, D. and P. Restrepo, 2018, Demographics and automation, NBER Working Paper no. 24421, National Bureau of Economic Research.

EPO (European Patent Office), 2020, Patents and the fourth industrial revolution. The global technology trends enabling the data-driven economy, Munich.

Foster-McGregor, N., Nomaler, Ö. and B. Verspagen, 2019, Measuring the creation and adoption of new technologies using trade and patent data, UNIDO Working Paper

Hidalgo, C.A., 2021, Economic complexity theory and applications. Nat Rev Phys 3, 92–113. https://doi.org/10.1038/s42254-020-00275-1

Hidalgo, C., and R. Hausmann, 2009, The Building Blocks of Economic Complexity, Proceedings of the National Academy of Sciences 106(26):10570–75.

Menéndez De Medina , Mercedes, Önder Nomaler & Bart Verspagen, 2023, Identification of Fourth Industrial Revolution technologies using PATSTAT data, UNU-MERIT Working Paper 2023-023. https://unu-merit.nl/publications/wppdf/2023/wp2023-023.pdf

Nomaler, Ö and B. Verspagen, 2022, Some new views on product space and related diversification, UNU-MERIT Working Paper no. 2022/11, UNU-MERIT.

Annex

A1. The use of the BEC classification

The approach to construct the data on global value chains in the structural change pillar (section 2 of the main text) is based upon the latest version of the Broad Economic Categories (BEC) classification, BEC Rev. 5.[16] This classification allows for a split of HS products into three broad end use categories: intermediate consumptions (intermediates); gross fixed capital formation (capital); and final consumption (consumer goods). A correspondence between the 2012 revision of the HS classification and BEC Rev. 5 has been provided by the United Nations Statistics Division.[17] In addition to these three dimensions, the BEC Rev. 5 classification has a number of mixed categories (e.g., intermediate/capital, consumer/capital, etc.). These mixed categories are allocated to one of the three aggregates, with intermediate/capital and intermediate/consumption being allocated to intermediates, consumer/capital and consumer/intermediates to consumer goods, and capital/consumer and capital/intermediates to capital goods.

In addition to the split between the three end use categories, the BEC classification further splits intermediate and consumer goods into primary and processed goods. Primary goods are those with characteristics of primary sectors of the economy, such as farming, forestry, fishing, and extractive industries, as well as goods that undertake only minor modifications and whose value is thus still largely from primary sectors (e.g., ginned cotton). Processed goods, conversely, are those that involve extensive processing. This distinction between primary and processed has been used to help identify global value chain trade, with producers of primary goods considered higher upstream (i.e., strong forward linkages), while producers of processed goods tend to be further downstream (i.e., strong backward linkages).

A novelty of BEC Rev. 5 over the earlier BEC Rev. 4 relates to the conclusion that the definition of intermediates – including the processed and primary split – is too broad to be useful in the context of global value chains. Processed intermediate goods, for example, have been found to comprise many generic products with published reference prices or commonly sold at auction as well as intermediates with a more specific use in a particular industry. BEC Rev. 5, therefore, introduces an additional split between generic and specific intermediate goods (as well as capital goods) to better identify trade within global value chains.

[16] https://unstats.un.org/unsd/trade/classifications/SeriesM_53_Rev.5_17-01722-E-Classification-by-Broad-Economic-Categories_PRINT.pdf

[17] http://unstats.un.org/unsd/statcom/47th-session/documents/BG-2016-11-Manual-of-the-Fifth-Revision-of-the-BEC-E.pdf

A2. Product complexity and upgrading probability

The probability measure for upgrading is based on the idea that a country's current specialization structure (partly) determines which products are likely targets for developing new comparative advantages. The calculations start by defining X as the familiar binary matrix of revealed comparative advantage (RCA), with elements:

$x_{ij} = 1$ if $\frac{E_{ij}/E_j}{E_i/E} \geq 1$ and $x_{ij} = 0$ otherwise,

where E_{ij} denotes the value of exports of product i by country j and the absence of a subscript indicates a summation over the relevant dimension. The matrix X has dimensions $m \times n$, where m is the number of products and n is the number of countries. Typically, $m >> n$. We assume that each country exports at least one product, and each product is exported by at least one country. A given X matrix contains a total of $r = \sum_{\forall i \in \{1,2,..m\}} \sum_{\forall j \in \{1,2,..n\}} x_{ij}$ revealed comparative advantages.

The so-called density metric for related variety (Klinger and Hausman, 2006) draws on the notion of conditional probability, computed on the basis of observed co-occurrences. The conditional probability captures the idea that having a comparative advantage in one product provides information about the likelihood that a country has a comparative advantage in another product. If the conditional probability is high (low), the two products are likely to require similar capabilities to export them with comparative advantage.

Formally, let k_{pq} denote the number of countries which have comparative advantage both in product q and in product p, and s_p denote the number of countries which have comparative advantage in product p (what is usually called the 'ubiquity' of a product). Then $c_{qp} = k_{qp}/s_p$ denotes the probability that a country has comparative advantage in product q, conditional on the country having comparative advantage in product p. In matrix notation, these conditional probabilities are given by:

$$C = S^{-1}XX^T,$$

where the superscript T indicates a transposed matrix and S is the matrix with the corresponding row-sums of X on the main diagonal and zeros elsewhere. Note that the diagonal of S thus contains the ubiquity of respective products, i.e.,

$$s_{ij} = \begin{cases} \sum_{\forall k \in \{1,2,..n\}} x_{ik} & if\ i = j \\ 0 & otherwise \end{cases}$$

The database involves the use of a new measure that is also based on the idea of conditional probability, but which differs from the density measure. For this new measure, a matrix with a complementary conditional probability measure is used, which is defined as:

$$B = U^{-1}ZX^T,$$

where $Z = O - X$ and U is a matrix with the row-sums of Z on the main diagonal. Thus, Z is a matrix of "anti-RCA", in which the elements are $z_{ij} = 1$ if $x_{ij} = 0$ and $z_{ij} = 0$ if $x_{ij} = 1$. The diagonal elements of U are then defined as $u_{ii} = n - s_{ii}$, which is the number of countries that have no RCA in product j. The elements b_{pq} of matrix B reflect the probability that a country has RCA in product q conditional on not having RCA in product p. We interpret these conditional probabilities as follows: if b_{pq} is low (high), products p and q are likely to share a high (low) degree of capabilities needed to export them with comparative advantage. Thus, the conditional probabilities b_{pq} are complementary to c_{pq}.

The proposed measures are then defined, in matrix notation as:

$$E = \frac{C^T X + B^T Z}{m},$$

which is an $m \times n$ matrix. Note that for a country j that has actual comparative advantage in f_j products, each entry in the j^{th} column of CX is the sum of f_j alternative conditional probabilities, while each entry in the j^{th} column of BZ is the sum of $m - f_j$ other conditional probabilities. Thus, through the division by m, each entry in (the j^{th} column of) E is the sum of exactly m alternative conditional probability estimations, which means that the elements of E are average conditional probabilities.

Matrix E can be seen as a probabilistic estimation of X. It can be rearranged to obtain separate measures of related and unrelated variety. For this, we proceed as follows:

$$E = \frac{C^T X + B^T Z}{m} = \frac{C^T X - B^T (O - X)}{m} = \frac{B^T O + K^T X}{m},$$

where $K \equiv C - B$ is a matrix of *marginal* conditional probabilities. Written this way, the matrix E has two constituent (additive) elements. The first of these, $E_1 \equiv B^T O/m$ is an $m \times n$ matrix in which all rows are equal to each other, i.e., where there is no country-variation. In each of the country-columns of this matrix, the i^{th} element indicates an *autonomous* probability of being specialized in product i, which we define as the probability for a hypothetical country without any prior comparative advantages (a country that does not trade and begins exporting just product i).

We characterize this component of matrix E as the part that corresponds to unrelated diversity. The component $E_2 \equiv K^T X/m$ (also $m \times n$) results from the particular specialization profile of the country in terms of both product relatedness (reflected in K) and the matrix X. E_2 is the *induced* part of E that corresponds to related diversity. In the database, we use E_2 as the measure for upgrading probabilities.

In order to calculate product complexity, we also use the matrix X that reports comparative advantages. This follows the method of ECI (Hidalgo, 2021). For this, we need the "raw" diversification measure that has already been defined in the main text, and which can be re-defined here as the diagonal matrix D with elements:

$$d_i = \sum_p X_{ip}$$

on the main diagonal. We then calculate a new matrix as follows:

$$(U^{-1}X)^T X D^{-1}$$

Product complexity is calculated as the eigenvector that corresponds to the second largest eigenvalue of this matrix.

The UNU-MERIT WORKING Paper Series

2024-01 *The green transformation as a new direction for techno-economic development* by Rasmus Lema and Carlota Perez
2024-02 *Evidence on an endogenous growth model with public R&D* by Thomas H.W. Ziesemer
2024-03 *Higher Educational Systems and E-resilience* by Mindel van de Laar, Julieta Marotta and Lisa de Graaf
2024-04 *Estimating the wage premia of refugee immigrants: Lessons from Sweden* by Christopher F. Baum, Hans Lööf, Andreas Stephan and Klaus F. Zimmermann
2024-05 *Green window of opportunity through global value chains of critical minerals: An empirical test for refining copper and lithium industries* By Jorge Valverde-Carbonell and Alejandro Micco A.
2024-06 *Migration and Consumption* by Roberta Misuraca and Klaus F. Zimmermann
2024-07 *The power of narratives: How framing refugee migration impacts attitudes and political support towards refugees* by Tobias Hillenbrand, Bruno Martorano, Laura Metzger and Melissa Siegel
2024-08 *Strategies of search and patenting under different IPR regimes* by Robin Cowan, Nicolas Jonard and Ruth Samson
2024-09 *From drains to bridges: The role of internationally mobile PhD students in linking non-mobile with foreign scientists* by Rodrigo Ito, Diego Chavarro, Tommaso Ciarli, Robin Cowan and Fabiana Visentin
2024-10 *Expanding horizons: The impact of post-doctoral mobility grants on research diversification* by Yağmur Yıldız, Diego Chavarro, Fabiana Visentin and Tommaso Ciarli
2024-11 *Scorching heat and shrinking horizons: The impact of rising temperatures on marriages and migration in rural India* by Manisha Mukherjee, Bruno Martorano and Melissa Siegel
2024-12 *Labour market integration for refugees: A social justice framework* by Tamara Kool and Zina Nimeh
2024-13 *The evolution of development with trade in global value chains* by Önder Nomaler and Bart Verspagen
2024-14 *The impacts of the Microfinance Multiplied approach on seasonal food insecurity: Evidence from a high-frequency panel survey in Uganda* by Ricardo Morel, Franziska Gassmann, Bruno Martorano, Nyasha Tirivayi and John Kamau
2024-15 *Building CoPS Capability for Catching Up During Transitions: The Case of ACWA Power in Saudi Arabia* by Kaori Mita and Michiko Iizuka
2024-16 *The KSTE+I approach and the AI technologies: evidence from the European regions* by Francesco D'Allesandro, Enrico Santarelli and Marco Vivarelli
2024-17 *Theory and measurement in SFC models: the role of the financial sector* by Huub Meijers and Joan Muysken
2024-18 *Is artificial intelligence generating a new paradigm? Evidence from the emerging phase* by Giacomo Damioli, Vincent Van Roy, Daniel Vertesy and Marco Vivarelli
2024-19 *Industrializing Africa - a science, technology, and innovation policy perspective* by Bertha Vallejo
2024-20 *Innovation in the wind energy sector* by Rasmus Lema, Davide Bonaglia and Ulrich Elmer Hansen

2024-21 *The Investment Landscape in the Lithium Triangle: Implications for the EU's Critical Raw Materials Act Agenda* by Beatriz Calzada Olvera and Melissa Vergara-Fernández

2024-22 *3G internet diffusion and secondary education attainment: Evidence of opportunity cost in Viet Nam* by Trang Thi Pham and Bernardo Caldarola

2024-23 Influence of the Sectoral Composition of Production on Occupational and Sectoral Wage Structures: Evidence from Multiple Economies by Nelson Marconi, Danilo Spinola, Tiago Porto and Eliane Araújo

2024-24 *Transition from Information and Communications Technology to Educational Technologies and Game-Based Learning in the Republic of Moldova: The Case of The Mobility Era Game* by Irina Burlacu de Paoli

2024-25 *When one teaches, two learn: The bidirectional learning process supervisor-PhD student* by Pierre Boutros, Michele Pezzoni, Sotaro Shibayama and Fabiana Visentin

2024-26 *Cloud technologies, firm growth and industry concentration: Evidence from France* by Bernardo Caldarola and Luca Fontanelli

2024-27 *The Innovation and Structural Transformation Database: A Guide* by Neil Foster-McGregor, Önder Nomaler & Bart Verspagen