The evolution of development with trade in global value chains

Önder Nomaler and Bart Verspagen

Published 5 June 2024
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 evolution of development with trade in global value chains

Önder Nomaler* & Bart Verspagen**

4 June 2024

Abstract

We propose to use canonical correspondence analysis (CCA) as a way to summarize the main trends in the dynamics of trade, global value chains and development over the period 1995 – 2018. CCA is a descriptive method that extends the algorithm (non-canonical correspondence analysis) that is widely used for calculating the economic complexity index. Both techniques (CCA and economic complexity) are aimed at reducing the dimensionality of large cross-country datasets on international trade. CCA has the advantage that the correlation between the derived indicator(s) to a set of underlying economic variables (in our case at the country level) is included in the derivation of the summary indicators. This facilitates the use of >1 dimensions to summarize the trade dataset. We illustrate this by relating the summary trade indicators (CCA dimensions) to a set of variables about integration of countries in global value chains, as well as a number of general indicators about development. The results indicate a trade-off between general GVC integration and a specialization in supplying intermediates to the global economy. We construct dynamic trajectories that show how individual countries or groups of products (such as high-, medium- and low-tech) navigate this trade-off over time.

Keywords: economic complexity index; global value chains; trade specialization; development

JEL codes: O11; F14; F63

We thank Maria Savona and other participants at the “Catching up and global value chains in times of transformation” Final CatChain Symposium, 19-20 February 2024 in Maastricht, and at the GGDC conference in Groningen, 28-29 May 2024, for comments on preliminary versions. The views expressed and any remaining error remain solely our responsibility.

* UNU-MERIT, Maastricht, the Netherlands, nomaler@merit.unu.edu

** Maastricht University, School of Business and Economics, the Netherlands, b.verspagen@maastrichtuniversity.nl
1. Introduction

The world economy in the 21st century is characterized by a number of “big trends”. We see a tendency towards (and desire for more) greening of the economy, we see the further rise of global value chains, development of (some) developing nations, the introduction of radical new technologies, as well as other tendencies that have long-run structural effects. While economists and other scholars address all these tendencies from different perspectives, there is also relatively little attention to incorporating a number of these tendencies together into one line of analysis.

This paper wants to contribute towards such broad analysis of the structural changes occurring in the global economy. We take a descriptive empirical perspective, leaving theoretical interpretation for later work. Our aim is to provide a quantitative overview of a number of tendencies in the global economy by summarizing a large amount of data. The prime data source that we use is a detailed database on exports, which makes international trade our primary focus. Our analysis is aimed at linking these trade data to global developments that can be measured at the country level, such as growth (changes in GDP per capita), greenhouse gas (GHG) emissions, and income inequality. A special role is reserved for indicators (at the country level) about participation in global value chains (GVC), i.e., the slicing-up of production chains over different locations on the globe, thereby offering countries the opportunity to specialize in small parts of value production.

The emphasis on trade data and their relation to country level development in a large sample of countries is something that we share with the literature on economic complexity (e.g., Hidalgo, 2021). We adopt a method that is somewhat similar to the methods used in this economic complexity field, but we provide two broad extensions to those methods. First, we will provide a multi-dimensional indicator that summarizes global developments, instead of the one-dimensional economic complexity indicator. Second, we use a statistical method that, although closely related to the economic complexity method, provides a much more direct linkage between the underlying trade data and the country-level variables that we include.

The method that we will implement is known as canonical correspondence analysis (CCA). It will be presented and explained in the next section, where we also explain how it relates and differs to the economic complexity method, which is (non-canonical) correspondence analysis (CA). Both CCA and CA are well known methods in quantitative ecology, where they are used for “ordination”, a type of exploratory data analysis used to reduce high-dimensional data into a few latent dimensions.

The main part of the paper explores how the CCA ordination method can help to form an informative quantitative overview of broad global development in the 21st century economy. For this, we select a number of indicators that we introduce in Section 3 of the paper. Section 4 then presents the results of the CCA in terms of so-called biplots, which are graphical representations of the descriptive summary that CCA provides for the main correlations in the data. In these biplots, we graph countries and traded products in terms of their mutual relationship, and their relationship to the country variables, including GVC participation that we use.

While we think that the paper has interesting insights to offer about the role of GVCs in development, and the role of specific products and countries in this, we also feel that the CCA method has potential to be applied to a wider set of issues in economics. We therefore also want to present the paper as an introduction to CCA for economists and complexity scholars, hoping for further applications in the future.
2. Canonical Correspondence Analysis

To derive our stylized facts, we use a method that is relatively new to the economics discipline, but which is commonly used in the quantitative ecology literature: canonical correspondence analysis (CCA). This is a method for exploratory data analysis that summarizes high-dimensional data in a low-dimensional latent space. For example, the ecologist may have data on the frequency of species of birds in different sites, as well as data about the physical characteristics of the sites (e.g., environment variables such as average temperature, extent of tree coverage). While there may be many species, many sites and many variables that distinguish them, CCA will provide a low-dimensional space in which both the environment variables and species can be mapped, thus providing a compact and generalizable interpretation of the relationships between the species, the sites and the environment variables. This "ordination space", as it is called, describes relations between sites and environment variables, between species and environment variables, and between sites and environment variables.

In our application of the CCA method to economic development in GVCs, traded products (such as rice, microprocessors or steel bars) take the role of species, countries take the place of sites, and a number of aggregate (country-level) variables such as GDP per capita, inequality, GHG emissions and various GVC indicators derived from MRIOs are used as the environment variables. Thus, we are seeking an impressionistic representation of the relationships between countries, some of their macro-development characteristics, and their detailed trade specialization patterns.

The concept of the product complexity index (PCI) and the economic complexity index (ECI) as proposed by Hidalgo (2021) is a related use of correspondence analysis (Mealy et al, 2019; van Dam et al., 2021). Here there are no environment variables included in the analysis, and the CA is not of the canonical type. PCI and ECI are the first-dimension product- and country-scores of correspondence analysis on our dataset of trade specialization by products (a product-by-country matrix of revealed comparative advantages). Our analysis extends this "complexity framework" by including the macro-development country variables, and by considering >1 dimensions that result from the CCA.

Our implementation of CCA starts with export value data in US$ for \( n \) products exported by \( m \) countries. The revealed comparative advantage (RCA) of country \( p \) in product \( q \) is calculated as:

\[
R_{qp} = \frac{s_{qp}}{s_{p}},
\]

where \( s_{qp} \) is the share of country \( p \) in total export value (over all countries) of product \( q \), and \( s_{p} \) is the share of country \( p \) in the sum of total export value (over all products and countries). Based on these RCA values, we construct a product-by-country matrix \( X \) of dimensions \( n \times m \), in which element \( x_{qp} \) is equal to 1 if \( R_{qp} > 1 \) and 0 otherwise. Thus, \( X \) is a binary RCA matrix that represents the information on trade performance of the countries.

The product space idea as proposed by Hidalgo et al. (2007) portrays the products that make up the rows of matrix \( X \) as a high-dimensional space that countries can occupy by having (attaining) comparative advantages (\( x_{qp} = 1 \)) in each of the products. Their application of correspondence analysis (Hidalgo and Hausmann, 2009; Mealy et al, 2019; van Dam et al., 2021) reduces the dimensionality of the product space to just one (complexity), and the Economic Complexity Index (ECI) of a country measures the position of the country in product space. The original idea in Hidalgo and Hausmann (2009) was to derive complexity and the ECI by an iterative procedure called the ‘method of reflections’. It starts from the \( X \) matrix and, in the first step, calculates both row- and column-sums. The sum of a column is the number of products for which the country has a comparative advantage, and this is called ‘diversity’ of the country. The sum of the row is the number of countries that have comparative advantage in the product, and this is called ‘ubiquity’ of the product.
Each iteration of the method of reflections produces a country indicator and a product indicator. The initial values are diversity and ubiquity, respectively. Then in each next iteration, the country indicator is updated to the average of the product indicator of the previous iteration, over all products for which the country has comparative advantage. Similarly, the product indicator is updated to the average of the country indicator of the previous iteration, over all countries that have comparative advantage in the product. Both indicators are standardized and the procedure starts again.

Hidalgo and Hausmann (2009) observe that at the initial step of the reflection algorithm, countries with high diversity tend to have comparative advantage in products with low ubiquity (and vice versa). They take this as evidence for the hypothesis that products with high ubiquity require a lower level of productive capabilities than products with low ubiquity. Each step of the method of reflections is supposed to capture more information about this relation between productive capabilities and the indicators.

Later, e.g., Hidalgo (2021), the method of reflections was formulated as an eigenvalue problem, where product complexity and the ECI can be calculated as eigenvectors. This is equivalent to correspondence analysis (Mealy et al., 2019; van Dam et al., 2021). The procedure starts with two normalized versions of $\mathbf{X}$: a ubiquity-normalized version $\mathbf{X}^u$ where each element of $\mathbf{X}$ is divided by ubiquity of the row, and a diversity-normalized version $\mathbf{X}^d$ where each element is divided by diversity of the column. Then an $n \times n$ matrix is created as $\mathbf{C}^p = \mathbf{X}^d \mathbf{X}^u$, where the prime denotes a transposition, and the $p$ superscript indicates that we have a product-by-product matrix. This matrix carries information about the (pair-wise) similarity between all products in terms of the communality of the countries that are specialized in them. This proximity metric can be seen as the opposite counterpart of the dissimilarity metric known as $\chi^2$ distance, in the sense that the contribution of a country (where both products are located commonly) to the metric is penalized (divisively) by the diversity of the country.

The first eigenvalue of matrix $\mathbf{C}^p$ is 1, and the corresponding eigenvector is constant. The second eigenvector is used as a measure for product complexity. The third and higher-order eigenvectors are ignored, although they contain information (Legendre and Legendre, 1998). The sum of the eigenvalues, excluding the trivial first one, is equal to the trace of matrix $\mathbf{C}^p$, which represents the total variance of the trade data. Thus, considering a larger number of eigenvectors will capture a larger share of the variance of the trade data. Also, Euclidean distances in the space that covers all non-trivial eigenvectors correspond exactly to $\chi^2$ distances between products in the full data, and spaces that use less eigenvectors provide an approximation to these distances (the more eigenvectors are used, the better the approximation).

The Economic Complexity Index (ECI) of a country can either be calculated as the second eigenvector of the alternative country-by-country matrix $\mathbf{C}^c = \mathbf{X}^d \mathbf{X}^u$, or as the average of the complexity of the products in which the country has a comparative advantage. The two methods are equivalent up to a multiplicative factor of the ECI, and assuming $n > m$, the $m$ eigenvalues of $\mathbf{C}^p$ are also identical to the leading $m$ the eigenvalues of $\mathbf{C}^c$. Although the literature uses a single ECI indicator, we can also consider the higher-order eigenvectors of $\mathbf{C}^p$ as higher-order ECIs.

ECI and similar derived indicators (e.g., the Economic Complexity Outlook Indicator, Hausmann et al., 2014) is often used as an independent variable in regression models aimed at explaining

---

1 Just as the method of reflections is identical to the iterative ‘reciprocal averaging’ algorithm used by ecologists in the early days of correspondence analysis.

2 Note that $\mathbf{C}^c$ carries information about the similarity between countries in terms of their specialization patterns. The contribution of a co-occurring product is (divisively) penalized by its ubiquity, thus this proximity metric can, again, be seen as the counterpart of a $\chi^2$ distance.
economic growth. Such a procedure essentially consists of what Ter Braak (1986) calls “interpretation of the ordination axes with the help of external knowledge and data” (the ordination axes are the reduced dimensions, i.e., product complexity and ECI in our application). Canonical correspondence analysis as proposed by Ter Braak directly includes the “external knowledge and data” in the dimension reduction procedure: “the ordination axes are chosen in the light of known environmental variables by imposing the extra restriction that the axes be linear combinations of environmental variables” (Ter Braak, 1986, p. 1167).

In the ecological tradition of Ter Braak, canonical correspondence analysis is typically used to interpret the relation between physical attributes of a site, and the relative occurrence frequency of a number of species that are under investigation. The non-canonical form of correspondence analysis uses the co-occurrence of species at individual sites to quantify similarities between species. The canonical form adds to this picture a relation between the species and characteristics of the site, such as average temperature or humidity. Thus, the environmental site variables can be seen as an exogenous determinant of the relative species frequencies, and canonical correspondence analysis summarizes this relationship into a low dimensional space into which both the species and the sites can be visualized.

Canonical correspondence analysis will construct a set of country-level indicators which in the Hidalgo and Hausmann tradition are similar to ECI, and correspond to what in the ecological tradition are site scores. It will also construct a set of product-level indicators, which are similar to the ecological species scores, and to Hidalgo and Hausmann’s complexity. While in the ecological tradition the environmental variables can be seen as explanatory variables, our counterpart of these variables, which will include input-output based variables on GVC participation, bears no such interpretation. All that canonical correlation analysis will identify in our case, are associations between different types of outcome variables: comparative advantage in products, GVC indicators, and development-related variables such as GDP per capita, greenhouse gas emissions, and inequality.

As prescribed by Ter Braak (1986), the country level CCA indicators will be linear combinations of the country variables, plus a constant. The crucial idea of canonical correspondence analysis is that we wish to maximize the correlation between these linear combinations and a country indicator where each country’s observation is equal to the average of the product indicator over all products for which the country has comparative advantage.

The country variables are standardized in advance such that their weighted mean is 0 and their weighted standard deviation is 1, where the weight for each country is proportional to its diversity (i.e., the number of products where it has RCA>1). The weights should add up to 1. These standardized country variables are stored in a matrix $Y$ with dimensions $m \times (z + 1)$ where $z$ is the number of country variables (in our case, $z = 6$). One column of $Y$ that is to be used as a constant in a regression later is populated with 1s, hence the number of columns of $Y$ is $z + 1$.

CCA can be implemented as an iterative procedure that is similar to the method of reflections, but includes the economic variables (Ter Braak, 1986, p. 1169). The procedure starts by initializing scores on the country indicator by arbitrary but distinct values. Step 1 calculates product scores as the average of the country scores over all countries that have comparative advantage in the product. In step 2, new country scores are calculated as the average of product scores of the products for which the country has comparative advantage. A multivariate linear regression is performed in step 3 where the country scores are the dependent variable and the country variables (plus a constant) are the explanatory variables. This is a weighted regression where for each country, the weight is proportional to its diversity while all weights add up to 1. In step 4, the predicted values of this regression are adopted as the new country scores. Finally step 5
standardizes the country scores such that their weighted average is 0 and their weighted variance is 1. After this step, we return to step 1, unless changes in country scores were smaller than a threshold, in which case the procedure stops. The regression in step (3) and the use of the regression results in step 4 are what distinguishes this algorithm from the method of reflections.3

As with the method of reflections, and as suggested by Ter Braak (1986), instead of this iterative procedure, we can obtain the ordination by solving an eigenvalue problem. The crucial part of the canonical element of CCA is the weighted regression in step (3) of the above description of Ter Braak’s algorithm. To represent this in the eigenvalue analysis that we set up, we need an \( m \times m \) diagonal weight matrix \( W \) with diversity of the country divided by the sum of diversity of all countries on the main diagonal, and zeros otherwise. And we need a matrix \( T = [Y'WY]^{-1}Y'W \) which enables the inclusion of the weighted regressions in the eigen-analysis.

Whereas the iterative procedure extracts one canonical dimension at a time, the eigenvalue version of the problems extracts all canonical dimensions at the same time. Thus, the regression of step (3) uses a set of country scores (coming from step 2) represented in a matrix \( V \) (instead of a vector of country scores in each instance of the iterative procedure). With our definition of matrix \( T, B = TV \) will provide the regression coefficients, where a \( B \) is a matrix of dimensions \((z + 1) \times z\). The \( z \) columns of this matrix specify separate sets of regression coefficient, one for each canonical dimension that we extract. In other words, each column of \( B \) or \( V \) corresponds to one instance of the iterative procedure. The matrix \( E = YB = YTV \) will provide the weighted predicted country scores of step (4).

The core of canonical correspondence analysis is formed by eigen-analysis of the matrix \( \Phi = YTC \). Note that \( C \) is the same matrix as used in the complexity calculations. The premultiplication of \( C \) with \( YT \) introduces the weighted regressions on the environment variables (in our case, country variables), which is exactly what turns ordinary CA into the so-called ‘canonical’ form. The first (trivial) eigenvalue of \( \Phi = YTC \) is 1, as it is in the method of reflections (i.e., non-canonical CA), and the corresponding eigenvector is ignored. We extract the next \( z \) eigenvectors to form a matrix \( E \) in which each column contains one eigenvector. Thus matrix \( E \) has dimensions \( m \times z \), i.e., with as many columns as there are non-trivial positive eigenvalues. Each column forms a canonical axis, or canonical dimension, for the countries. As was the case with the correspondence analysis version of the method of reflections, complexity is a multi-dimensional indicator, but the number of dimensions (i.e., \( z \)) is much smaller than the number of countries \( n \) and/or the number of products \( m \). Each column of \( E \) corresponds to the vector that is computed at the 4th step of the iterative algorithm by Ter Braak (1986) as summarized above. In other words, the matrix \( E \) contains the ‘predicted’ country scores.

Note that the scale (norm) of the eigenvectors is arbitrary in principle, and different algorithms may deliver eigenvectors with different scales (many algorithms will settle for eigenvectors with norm equal to 1). Similar to step (5) in the above representation of Ter Braak’s (1986) algorithm, we use country diversity weights to standardize each column of the eigenvector matrix \( E \) such that its weighted mean is 0 and weighted variance is 1. This standardized eigenvector matrix is denoted as \( E_{Std} \), and it contains the ‘predicted’ country scores of step (5) of the above described iterative procedure. We then follow the iterative procedure, moving to step (1) to calculate product scores as RCA-weighted averages of predicted country scores: \( U = XuE_{Std} \) where \( U \) (with dimensions \( n \times z \)) contains the product scores. Each column of \( U \) represents product scores on one canonical axis.

---

3 The iterative procedure as described here calculates one canonical axis at a time. To calculate multiple canonical axes, we need to repeat the procedure. For the second and higher-order canonical axis, an additional step needs to be inserted after step (5), aimed at making the fitted country scores orthogonal to the previous axes.
The procedure finalizes with the computation of the country scores (in line with the 2nd step of Ter Braak's algorithm) as \( \mathbf{V} = \mathbf{X}^d \mathbf{U} = \mathbf{X}^d \mathbf{X}^e \mathbf{E}_{\text{std}} = \mathbf{C}^c \mathbf{E}_{\text{std}} \), which highlights the relation (and difference) between 'predicted country scores' \( \mathbf{E}_{\text{std}} \) and 'country scores' \( \mathbf{V} \).

To see the equivalence of the iterative algorithm and the eigenvalue problem, we can follow the equations so far and imagine that step (2) of the \( p \)-th iteration of the iterative algorithm led to \( \mathbf{V}(p) = \mathbf{X}^d \mathbf{U}(p) = \mathbf{X}^d \mathbf{X}^e \mathbf{E}_{\text{std}}(p - 1) = \mathbf{C}^c \mathbf{E}_{\text{std}}(p - 1) \). Note that we use \( (p) \) to denote the state of a vector variable at the \( p \)-th iteration. Following through, step (4) will yield \( \mathbf{E}(p) = (\mathbf{YTC}^c)\mathbf{E}_{\text{std}}(p - 1) \) at the \( p \)-th iteration. Also considering step (5) (i.e., the standardization of \( \mathbf{E}(p) \) into \( \mathbf{E}_{\text{std}}(p) \), which at the point of convergence, reduces to a mere matter of rescaling per column), the convergence criterion of the algorithm requires that \( \mathbf{E}_{\text{std}}(p) = \mathbf{E}_{\text{std}}(p - 1) \). In matrix algebra, this condition can conveniently be written as \( \mathbf{E}(p) = \mathbf{E}_{\text{std}}(p - 1) \mathbf{\lambda} \), where \( \mathbf{\lambda} \) is a diagonal matrix. This immediately leads to \( (\mathbf{YTC}^c)\mathbf{E}_{\text{std}}(p - 1) = \mathbf{E}_{\text{std}}(p - 1) \mathbf{\lambda} \) which is obviously an eigenvalue problem where \( \mathbf{\lambda} \) is a diagonalized matrix of the eigenvalues of the matrix \( \mathbf{\Phi} = \mathbf{YTC}^c \).

The quantitative ecology literature typically aims at depicting site (country) scores and species (product) scores together in 2-dimensional scatter plots (i.e., using two of the canonical axes) along with indications (in terms of rays/arrows that emanate from the origin) of how these site/species ordinations relate to the 'environmental' (in our case, 'country') variables. These scatterplots are referred to as 'biplots'. This paper will also present its results in terms of such biplots.

Under the topic of 'scaling', the quantitative ecology literature suggests alternative ways to depict countries and products (sites species) in the biplot. An important question is whether one plots country scores as the RCA-weighted average of product scores, or vice versa. The technical term here is the concept of a 'barycenter' which specifies whether countries will show up in the biplot as the inner space of the products or vice versa. Among the alternatives suggested (see Legendre and Legendre, 1998, for a thorough discussion), the particular scaling scheme we adopt in the making of our biplots uses \( \mathbf{U} \) as the product score matrix, along with the country scores matrix rescaled as \( \tilde{\mathbf{V}} = \mathbf{V} / \mathbf{\lambda}^{-1} \), where \( \mathbf{\lambda} \) is a diagonalized matrix of the leading \( z \) non-trivial eigenvalues of the matrix \( \mathbf{\Phi} = \mathbf{YTC}^c \). Observe that all non-trivial eigenvalues are less than 1, thus division by an eigenvalue scales the country scores up. This results in countries occupying the outer-space in each of our biplots, while products forming the barycenter (i.e., the inner-space).

Canonical correspondence analysis maintains the interpretation of the eigenvalues representing a share of the variance of the trade data. Each \( i \)-th eigenvector (where \( i \in [1, z] \)) captures a share of variance of the trade data given by the metric \( \text{Inert}_i \), (so-called 'inertia'), which is the \( i \)-th non-trivial eigenvalue as divided by \( \text{tr}(\mathbf{C}^p) - 1 \). In the interpretation of the results (biplots), the relative values of \( \text{Inert}_i \) can be related to the product and country scores on dimension \( i \). As we present the eigenvalues in decreasing order of their corresponding eigenvalue, higher-order canonical dimensions will capture ever smaller shares of the total variance of the trade data. Note also that all \( z \) canonical axes together represent a total share of this variance that is smaller than 1, i.e., there remains a part of the variance that is not captured by the canonical analysis. Finally, we observe that the first \( p \) canonical dimensions usually capture a smaller share of the variance in the trade data than the first \( p \) dimensions of non-canonical CA, because the correlations to the country variables in canonical analysis present a restriction on the product and country scores.

\[ ^4 \text{Note that an alternative scaling scheme proposed in the quantitative ecology literature places the countries at the barycenter by deflating the country scores (i.e., the respective columns of } \mathbf{V} \text{) by multiplication by the square root of the associated eigenvalue, while inflating the product scores (i.e., the respective columns of } \mathbf{U} \text{) by division by the square root of the associated eigenvalue.} \]
3. Data and Indicators

We will use data for the years 1995, 2000, 2005, 2010, 2015 and 2018. Our input-output data are not available beyond 2018, which is why we break with the 5-year periodization after 2015. For each of the years, we use 6-digit Harmonized System (HS) data from COMTRADE for the value of exports. We use the HS version that is native for the year, and this implies that the list and the number of product codes differs per year. We use export data for about 150 countries that include all major economies of the world, and calculate (binary) RCA on the basis of this complete dataset. Then we limit the set of countries to 64 on the basis of the availability of the input-output data, but use RCA as calculated for the larger set of countries.

We use six country-level variables for the canonical part of the CA. The first of these is the log of GDP per capita (in constant 2017 international PPP$), which we draw from the World Development Indicators database of the World Bank. The second is greenhouse gas emissions in tons of CO$_2$ equivalents per capita, which we take from the EDGAR database at EUROSTAT. The third indicator is the Gini coefficient for income distribution. We use the Gini coefficient for disposable household income from the SWIID database (v9.6) for this (Solt, 2020).

The remaining three country Global Value Chain (GVC) indicators are derived from the OECD ICIO database (2023 version). We use the multi-regional table from this database to calculate our own GVC indicators. The definition of these indicators rests on two matrices see, e.g., Los et al. 2015, Foster, 2019):

$$S^d = \hat{A}L^{-1}\hat{F}^d$$
$$S^f = \hat{A}L^{-1}\hat{F}^f$$

where $\hat{A}$ is a diagonal matrix with value added coefficients (value added/gross output) on the main diagonal and zeros elsewhere, $L^{-1}$ is the common inverse Leontief matrix calculated from the multi-regional table, $\hat{F}^d$ is a diagonal matrix with final demand deliveries to the domestic economy on the main diagonal and zeros elsewhere, and $\hat{F}^f$ is a diagonal matrix with final demand deliveries to the foreign economy on the main diagonal and zeros elsewhere. The sum of matrices $S^d$ and $S^f$ represents total value produced and used in the global economy:

$$S = S^d + S^f = \hat{A}L^{-1}\hat{F}$$

where $\hat{F} = \hat{F}^d + \hat{F}^f$ is the diagonal matrix of total final demand. The rows of matrix $S$ will sum to value added (contribution to GDP) of the country-sector, while the columns of $S$ will sum to (total) final demand served by the country-sector.

Then we define a new matrix $\hat{S}^d$ that is equal to matrix $S^d$ with all elements where the row- and column-country is the same, set to zero. We set these block-diagonal elements to zero because they represent value that is produced and used in the same country, i.e., it is value that never crosses a border and hence we consider it to be outside global value chains.

In order to explain the construction of our first two GVC indicators, we use a simplified 2 country and 2 sector example version of these matrices, displayed below:

---

5 Detailed product codes for each year are available on request. There are two exceptions to the 6-digit aggregation level: we use the 4-digit code 2710 instead of the collection of 6-digit codes within it, and we use the 5-digit code 81125 instead of the 6-digit codes within it. The reason for these exceptions is that the underlying 6-digit codes seem incomplete.

6 For 1995, we have only 62 countries, because Belgium and Luxemburg are combined in the HS trade data, and we drop this combined country from the analysis.
<table>
<thead>
<tr>
<th></th>
<th>Matrix $\tilde{S}^d$</th>
<th></th>
<th>Matrix $S^f$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cAs1 cAs2 cBs1 cBs2</td>
<td>cAs1 cAs2 cBs1 cBs2</td>
<td></td>
</tr>
<tr>
<td>cAs1</td>
<td>0</td>
<td>Part $\alpha_{AB}$</td>
<td>Part $\beta_{AB}$</td>
</tr>
<tr>
<td>cAs2</td>
<td></td>
<td></td>
<td>Part $\gamma_{AB}$</td>
</tr>
<tr>
<td>cBs1</td>
<td>Part $\alpha_{BA}$</td>
<td>0</td>
<td>Part $\gamma_{BA}$</td>
</tr>
<tr>
<td>cBs2</td>
<td></td>
<td></td>
<td>Part $\beta_{BA}$</td>
</tr>
</tbody>
</table>

In this diagram, the labels cXsY denote country X-sector Y, where X denotes countries A and B, and Y denotes sectors 1 and 2. Each of the colored parts denotes a 2-by-2 sub-matrix. Note that we have already set the parts denoted by “0” to zero, and the other parts will generally contain semi-positive numbers. Parts $\alpha_{AB}$ and $\alpha_{BA}$ contain intermediate inputs. Depending on whether one uses a “forward” or a “backward” perspective, these are either intermediate inputs supplied to other countries (forward perspective), or intermediate inputs used for domestic deliveries of final demand (backward perspective). Parts $\gamma_{AB}$ and $\gamma_{BA}$ denote inputs supplied to foreign countries (the forward perspective) or inputs sourced from foreign countries used for exported final demand (the backward perspective). In the parts $\beta_{AB}$ and $\beta_{BA}$, the forward and backward perspectives coincide. Here we find the country’s “own” parts of exported final demand, which includes both actual final demand deliveries and domestic intermediate deliveries. Once again, it is important to observe that all non-zero parts in the diagram correspond to value added that crosses at least one border as embedded in some final or intermediate product.

The parts $\alpha$, $\beta$ and $\gamma$ are used to construct two indicators of revealed comparative advantage of the possible “functions” in GVCs, where the functions are either delivery or use of inputs (represented by the $\alpha$ and $\gamma$ parts), or “own value” (corresponding to the $\beta$ part). Our GVC-RCA indicators measure to what extent countries “specialize” (i.e., have an RCA) in these roles. By definition, a country that specializes in providing own value, cannot specialize in using or producing (depending on whether we use a forward or backward perspective) intermediate value, i.e., a comparative advantage in intermediates corresponds to a comparative “non-advantage” in own value delivery, and vice versa. Hence, we only need one of the two RCAs. We will use the indicator for intermediate inputs, (i.e., the $\alpha$ and $\gamma$ parts). We will define both a backward (use) and a forward (delivery) version of each RCA.

First, we look at a backward indicator. Final demand delivered by country A in GVCs consists of parts $\alpha_{BA}$ and $\gamma_{BA}$, which are foreign inputs that country A used, and part $\beta_{AB}$ which is domestically ("own") produced value delivered by country A to GVCs. For country B, these will be parts $\alpha_{AB}, \gamma_{AB}$ and $\beta_{BA}$, respectively. Then we calculate, for country A, the sum of values in parts $\alpha_{BA}$ and $\gamma_{BA}$ as a share of the sum of values in parts $\alpha_{BA}, \alpha_{AB}, \gamma_{BA}$ and $\gamma_{AB}$ together. This is the share of country A in the global intermediate goods trade. Then we calculate, again for country A, the sum of values in parts $\beta_{AB}, \alpha_{BA}$ and $\gamma_{BA}$ as a share of the sum of values in parts $\alpha_{BA}, \alpha_{AB}, \beta_{AB}, \beta_{BA}, \gamma_{BA}$ and $\gamma_{AB}$. This is the share of country A in total value (final demand) delivered in all GVCs.

In its raw form, the backward GVC-RCA indicator for country A is the first share divided by the second share, i.e.,

$$RCA^b_A = \frac{\Sigma \alpha_{BA} + \Sigma \gamma_{BA}}{\Sigma \alpha_{BA} + \Sigma \gamma_{BA} + \Sigma \alpha_{AB} + \Sigma \gamma_{AB}} / \frac{\Sigma \alpha_{BA} + \Sigma \gamma_{BA} + \Sigma \beta_{AB}}{\Sigma \alpha_{BA} + \Sigma \gamma_{BA} + \Sigma \alpha_{AB} + \Sigma \gamma_{AB} + \Sigma \beta_{BA} + \Sigma \beta_{AB}}$$

where a $\Sigma$ indicates summation over the matrix elements of a part. This can also be written as
\[
RCA_A^b = \frac{\sum a_{BA} + \sum y_{BA}}{\sum a_{BA} + 2y_{BA} + 2\beta_{AB}} \cdot \frac{\sum a_{BA} + 2y_{BA} + \sum a_{AB} + \sum y_{AB}}{\sum a_{BA} + 2y_{BA} + \sum a_{AB} + \sum y_{AB} + 2\beta_{AB}}
\]

where the numerator becomes the share of imported intermediate goods in the total value exported by country A as final deliveries, while the denominator is the share of intermediate goods in all value traded globally. This will be > 1 (< 1) if country A relies relatively heavily (little) on foreign inputs for its final demand deliveries to foreign countries.

We re-scale these ‘raw’ RCA indicators by the transformation \((RCA - 1)/(RCA + 1)\), which creates an indicator within the \([-1,1]\) interval and zero as the neutral value. Note also that this indicator is symmetric, i.e., negative and positive values can be assessed on the same scale (this does not hold for the “raw” RCA indicator).

Calculating the forward version of the RCA indicator is very similar to the backward version. The forward version reverses the subscripts on the \(\alpha\) and \(\gamma\) parts, as intermediate inputs used must become intermediate inputs supplied in this indicator. Also, total final demand deliveries in GVCs must be substituted by total value added produced in GVCs. For instance, for country A in our simple example, the raw RCA form of the forward indicators is

\[
RCA_A^f = \frac{\sum a_{AB} + \sum y_{AB}}{\sum a_{AB} + \sum y_{AB} + 2\beta_{AB}} \cdot \frac{\sum a_{AB} + \sum y_{AB} + \sum a_{AB} + \sum y_{AB}}{\sum a_{AB} + \sum y_{AB} + 2\beta_{AB} + 2\beta_{BA}}
\]

Observe that the denominator term for the raw forward and the raw backward RCA indicators are identical, also for all countries (within a given year). Thus, what matters is the nominator terms which are always the share of intermediate goods (imported used or exported delivered) in exported value (see also, e.g., Foster-McGregor, 2019; Antras and Chor, 2022). The transformation of these shares into raw RCA figures and the further normalization into the ultimate indicators mainly serves better interpretability and intertemporal comparability.

In summary, in the backward version, the focus is on a homogenous set of final deliveries, and how intensively this uses a range of different types of foreign inputs, e.g., here we consider how an exported automobile is built from (foreign) steel, plastic, rubber tires, computer chips and glass windows. In the forward perspective, the emphasis is on a homogenous set of produced deliveries, and how this is supplied to a range of different types of (global) final deliveries. For example, we consider how steel is used in a range of different final products such as textile machinery, cars and buildings.

In summary, in the forward perspective, a positive value of the RCA indicator again indicates that the country specializes in providing inputs to GVC value that is delivered as final demand by other countries, and a negative value indicates a specialization providing value that the country itself delivers as final demand. In the backward perspective, a positive RCA indicates that the countries sources relatively much value from other countries for the GVC value that it itself delivers as final demand. In the z-scored and pooled sample of data that we will use, the correlation between the backward and forward version is 0.506, which we consider low enough to keep both separate indicators in the analysis.

Our third and final GVC indicator is the average of two sub-indicators. The definition of these starts by identifying a part of the global economy (as represented by the multi-regional input-output table) that can be characterized as “purely domestic”. For this, we set all elements of matrix \(S^d\) for which the row and column indicate different countries, to zero, and denote the resulting matrix \(\tilde{S}^d\)
(this is obviously the “complement” of matrix $S^d$). The matrix $S^d$ contains value added that is either final demand delivered to the domestic economy, or domestic intermediate deliveries that were used to produce domestic final demand. Note that, unlike the elements of $S^d$, the elements of $\tilde{S}^d$ correspond to value added that never cross a national border.

As before, we consider all value other than that represented by $\tilde{S}^d$ as part of global value chains, because in terms of the multi-regional input-output table, this value crosses at least one national border, either as an intermediate delivery in the production process, or as a delivery of final demand to a foreign economy. Next, we aggregate $\tilde{S}^d$ to a country-by-country format, which will yield a diagonal matrix with positive values on the main diagonal and zeros elsewhere, and denote the diagonal element of country $i$ as $s_i$. Then indicator $b_i$ for country $i$ is a measure for backward GVC integration, and is defined as $b_i = 1 - s_i/FD_i$ where $FD_i$ is total final demand delivered by country $i$ to the domestic economy including itself. $FD_i$ is obtained as a column sum of matrix $S$ as introduced above. Thus $b_i$ is the share of value produced in foreign countries embodied in final demand delivered by country $i$. A higher value of this indicator represents stronger backward integration in GVCs.

In a similar fashion, we construct the indicator $f_i = 1 - s_i/VA_i$, where $VA_i$ is country $i$’s total value added (GDP). $VA_i$ is obtained as a row sum of matrix $S$ as introduced above. This is a forward integration indicator that represents the share of country $i$’s GDP that crosses at least one border, either in the form of exports of final demand, or intermediate deliveries to contribute to final demand deliveries by other countries. After constructing these indicators of backward and forward GVC integration, we find that they are strongly correlated: after $z$-scoring within each year and pooling the six years, the correlation coefficient between them is equal to 0.885. We therefore decided to use the average of ($z$-scored) backward and forward integration as a single GVC indicator, yielding three GVC indicators altogether.

### 3.1. Pooling

Our CCA is performed on a dataset that consists of six years. As explained above, we use the native HS classification that exists in each year. The reason for this is that there are changes in the HS from time to time, and such changes often lead to breaks in the time series for the export data. Concordances exist to handle such breaks, but this does not avoid all breaks. Over the long time period that we use, 1995–2018, the rise of new products and the decline of older ones would not be properly reflected in a common HS classification for the entire period.

As a result of the choice for a native HS scheme for each year, products in the export data are not necessarily comparable between years. This is why we adopt a pooling scheme that assumes no such comparability, but instead leaves maximum flexibility for changes in product (and also country) ordination to be reflected in the results. In other words, our pooling works in such a way that each product-year combination and each country-year combination can appear as a single observation in the biplots. Based on this we can then choose to label and/or connect observations of a single country or product (or groups of products or countries) through time, thus creating a trajectory for the product-(group) or country-(group).

For the six country-level indicators (as explained above), the pooling is straightforward. We just have to stack the six yearly country-by-variable matrices on top of each other. Before we do this, we $z$-score each variable within each year, which makes sure that the entire column is also $z$-scored. This matrix $Y$ of the methodology Section 2 above. The pooling of the RCA trade data (HS products) is slightly more complicated. This matrix takes the role of matrix $X$ as explained in the
The columns of matrix $X$ are the stacked countries, as the rows of matrix $Y$ are (hence matrix $X$ has the same number of columns as matrix $Y$ has rows). Because we want each HS product to be available only in its own year, we define the rows of matrix $X$ as the stacked product lists of the individual years, making sure that the order of years is the same as the order used in the construction of the matrix $Y$. In this way, the RCA matrix will contain blocks around its main diagonal in which the year of the row (HS product RCAs) is identical to the year of the column (country variable observations). These blocks are filled with the respective yearly RCA matrices $X$, while any cells in the pooled RCA matrix outside those blocks are set to zero. This will form a large matrix (in our application, it is 30,968 by 382), but this matrix is only used (after normalization) to create the much smaller and therefore more manageable matrix $C$ from Section 2 (this is a pooled countries-by-pooled countries matrix, hence 382-by-382).

4. Results

4.1. The biplots

Figure 1 shows the biplot of CCA dimension 1 vs CCA dimension 2. Thus, the horizontal and vertical dimensions in this plot are the first two dimensions that were extracted by the CCA. With our six country variables ($z = 6$), we have 15 combinations of canonical axes to choose from for a 2D biplot. Given the ordering of the axes and the resulting decline of inertia with $i$, using the first two columns of the country- and product scores matrices maximizes the share of variance of the trade data captured in the biplot. This is the principle that we follow, in line with most other uses of CCA.

Our first canonical dimension captures about 5.3% of the overall inertia, the second about 2.8%, and the final four dimensions together capture about 7.4%, which implies that the two dimensions that we use in our biplots capture the larger share of the total variance captured by the entire canonical analysis (i.e., all 6 dimensions). These numbers can be benchmarked against the (one dimensional) Hidalgo-Hausmann ECI, which as we will show below, is similar to our first CCA dimension. ECI captures 6.4% of total inertia. Because our CCA dimensions impose the restriction of maximum correlation to the country variables (the ECI has no such restriction), 6.4% of total inertia is the maximum we can obtain for the first CCA index. We may also note that the first and second CCA dimension together capture a larger share of inertia than the ECI.

The country variables are plotted as colored rays from the origin. These lines are obtained for the analysis based on the entire 1995 – 2018 pooled sample, hence they remain constant over time. These lines allow us to interpret the positions of either products or countries (or groups of them) in the biplot, by projecting the products or countries on these lines.

The coordinates to which each ray points at are computed as the diversity-weighted correlation coefficients between the country scores (belonging to the selected CCA dimensions) and the country variable to which the ray refers to. This is called the intra-class correlation coefficient in the ecology literature (Ter Braak, 1986), and is similar to a “loading” in principal components analysis. Thus, the slope of each colored line is indicative of the relative importance of the country variable in each of the two CCA dimensions. A relatively flat (steep) slope indicates that the variable is associated mainly to the first (second) CCA dimension as plotted horizontally (vertically). For visual convenience (i.e., avoid rays that appear too short when the biplot axes cover a range larger than [0,1]), we extend the length of all rays by the same factor, i.e., leaving relative lengths and angles unchanged). Despite this overall scaling, a relatively shorter (longer) ray indicates a relatively lower (higher) correlation in both dimensions, and hence a smaller (larger) role in the overall ordination of the biplot.
The advantage of (and sole reason for) pooling the data for the six individual years in our analysis is that these intra-class correlation coefficients (the variable lines in the biplot) do not vary between years. If, instead of pooling, we would have carried out the CCA for each individual year, there would have been variations in terms of the variable lines for each year. We performed this yearly analysis and found that such variations are minor, i.e., the lines show a large degree of stability over the 6 years. To facilitate interpretation, we show results based on the pooled sample.

Note also both country and product scores are computed from eigenvectors. Although we carefully chose the scale of these eigenvectors (see Section 2), their sign remains arbitrary. Therefore, the direction pointed at by the product and country scores (vis-à-vis country variables) as well as the direction of the country variable rays remain arbitrary in terms of their sign. We address this indeterminacy in the biplots by making sure that the country scores (and the associated product scores) are corrected in their sign in a way that makes sure of a positive correlation between per capita GDP, i.e., the ray corresponding to this variable will always point up and right.

To illustrate the interpretation the position of countries and products in the biplot, we plot one country (USA) and one product (HS30193 – Live carp fish), both plotted with their 2018 coordinates. The relation of products and countries to the variable-lines as represented by the colored rays from the origin is found by a projection country onto the respective variable-lines. This is exemplified for the USA for three variables: GDP per capita, emissions per capita, and the forward-GVC RCA. The USA projects on the positive (i.e., the displayed) part of these lines, indicating that this country has positive values for these three variables. The HS30193 product is projected on the other three variable lines, where it projects on the positive part of those lines. Thus, for example, the ordination shows that live carp fish tends to be exported especially from

---

7 An eigenvector remains a valid solution to an eigenvalue problem also when multiplied by \(-1\).

8 This is also the case with CA, i.e., in the complexity literature, the ECI is said to be positively correlated with per-capita GDP but this is true only because the eigenvector provided by the software algorithm is ultimately multiplied by \(-1\) if it turns out to be negatively correlated by per-capita GDP.
countries with high inequality, high GVC integration, and high backward GVC-RCA. For the USA or HS30193 to be projected on the lines for the three variables for which the projections are not shown, we would have to extend those lines in the opposite direction, i.e., create a “negative” part (not actually drawn in the figure). Such projections point to low scores on those variables.

In Figure 2, we plot the first CCA dimension against the Hidalgo-Hausmann Economic Complexity Index (ECI). As explained in Section 2, the latter is derived as the first dimension obtained from non-canonical CA (we also use the pooled sample of six years for this). The main difference between ECI and our CCA-1 dimension is that the first has been derived without taking into consideration any correlations with the country variables, while the CCA indicator includes these correlations. We can derive the intra-class correlation coefficients with ECI in the same way as we do for CCA, and therefore the colored lines in Figure 2 have the same interpretation as those in Figure 1.

Figure 2. First CCA dimension and Hidalgo-Hausmann ECI

We observe a very tight positive correlation in Figure 2, indicating that, in this case, ECI and the first CCA dimension capture very similar things. This is caused to an important extent by the fact that we include GDP per capita, which is known to strongly correlate to ECI, as a country variable. We also observe that the variable space as indicated by the colored lines is reduced to a single
dimension in Figure 2, with two trios of variables (GDP per capita, emissions and forward GVC-RCA on the one hand, and Gini, backward GVC-RCA and GVC integration on the other hand) stretching at an almost 180° angle.

Thus, ECI is a one-dimensional country indicator that is correlated to these trios of country variables. As there are higher-order eigenvectors in CA as well as in CCA, this one-dimensional indicator could be elaborated to include more dimensions, but in the non-canonical form, this would be hard to interpret, as there are no a priori correlations to country variables that we may expect. In CCA, on the other hand, each higher-order dimension has a clear interpretation that is embedded in the construction of each dimension, by means of the regression part of the algorithm.

Figure 3 shows biplots with the first four of six canonical dimensions, thus showing three additional dimensions as compared to Figure 2. The left panel of Figure 3 shows dimension 1 vs 2 and is therefore similar to Figure 1, although Figure 3 plots all products and all countries for 2018, in both biplots. The biplot on the right in Figure 3 shows CCA dimension 3 against 4.

The added value of CCA with >1 dimensions and a single dimension ECI indicator is illustrated by the much richer impression that results from the variable lines in Figure 3. In the leftmost plot, which captures the largest part of the variance in the underlying trade data, we distinguish two pairs of variables, for which the angles are close to 180°, which means that in terms of the ordination in the biplot, these variables appear as substitutes. One of these pairs is GDP per capita (extending up and right) and inequality as indicated by the Gini coefficient (extending down and left). Thus, the summary of the data that this biplot represents suggests that richer countries tend to be more equal, and vice versa.

We need to stress that the dimensions displayed in the biplot capture only a relatively small part of the variance of the trade data. Therefore, the relationships indicated by the variable lines do not necessarily correspond to overall correlations in the data for the country variables. What the close-to-180° relationship between GDP per capita and inequality suggests is that considering these country variables as substitutes is a useful device to summarize the underlying trade data. Thus, we must not take the angles between the variable lines as indications for causation or even correlation between the variables. Rather, they are guidelines for interpreting the relative position of countries or products, and their movement in this space over time.

The same substitutive relation exists between the GVC integration (combined backward and forward) and the forward GVC-RCA exists. This indicates that, again in terms of the ordination in this biplot, countries that tend to specialize in delivering value in the form of intermediates, also tend to be less integrated into GVCs. Although the variable line for GVC integration is relatively short, indicating that this is a weaker relationship than some of the others, this trade-off between general GVC integration and being specialized as a supplier of intermediates seems to be a major conclusion about the evolution of GVCs as evidenced by our approach, and the detailed trade data that we use.

---

9 Indeed, the greater deal of the literature on economic complexity does not look beyond the first non-trivial eigenvector (i.e., ECI and the corresponding PCI) which is highly correlated with per capita GDP. An exception is van Dam et al., (2021) who recommend the consideration of the higher order CA dimensions, especially for clustering purposes. Their implementation indicates that the second non-trivial dimension of CA ranks countries in a way where exporters of textiles tend to show up on one extreme of the spectrum and oil exporters on the other, leaving the middle part hardly interpretable.
In the rightmost biplot, one interesting fact is that the line for GVC integration is much longer. Thus, if we wanted to have an ordination that includes this variable as a main factor, we would have to use CCA dimensions 3 and 4. However, most of the other lines in this biplot are short, with the exception of the Gini. In the analysis below, in order to keep the number of biplots at a manageable number, we will continue exclusively with CCA dimensions 1 and 2.

The interpretation of the CCA dimensions can also be facilitated by looking at the position of countries (brown dots) and products (blue dots). Here we focus on countries, for some of which we provide labels. In the leftmost biplot, we see, for example, that countries that we know to be specialized in mining (Australia, Kazakhstan, South Africa) tend to be located in the bottom-right. Industrializing Asian countries (Cambodia, Laos, Viet Nam, Indonesia) are located at the center-left. The small EU member Malta is located at the top, and highly-developed Switzerland and the USA are located on the righthand side. In the rightmost biplot, we find a number of large countries (India, Brazil, USA, and to a lesser extent South Africa) on the righthand side, and many European countries on the upper-left side.

4.2. Dynamic trajectories of Lall product groups

For each year in the analysis, we can classify the HS products into 11 categories based on Lall (2000), who classifies export products based on technology level. Our use of the Lall concordance is based on Eurostat’s implementation, but we additionally sub-divided the “resources” category into two parts: one containing mineral resources, and the containing agro-related resources. The latter includes forestry (logging) and fisheries. These resources categories contain raw materials, processed products are included in the one of the other nine categories, which contain three resource-based manufacturing groups, two low-tech manufacturing groups, three medium-tech manufacturing groups, and two high-tech manufacturing groups.

We proceed to present and discuss the dynamic trajectories of the centroids of these Lall categories of products. The centroids are calculated as the weighted average of the coordinates of
the products contained in each group, which makes the centroids the barycenter of the products that it contains. A similar centroid can be calculated for the group of all products (and hence all Lall categories) together (as well as for all countries within a given year). If the group of HS products would have been the same for each year in the analysis, the yearly all-products-centroids would have been exactly at the origin of the biplot (i.e., these centroids would have 0-coordinates). However, because our HS product classification differs (slightly) by year, the yearly all-product-centroids are not at the origin. Because the position of these centroids (and their movement relative to each other) is solely the result of different HS classifications, we filter their movement out of the dynamic trajectories that we will draw. We set the initial (1995) centroid of a Lall category to its actual value, and then filter out the all-product-centroids from subsequent movement (i.e., 1995 becomes the "base year").

Figure 4. Dynamic trajectories for Lall resources and resource-based manufacturing categories

Figure 4 shows the dynamic trajectories for the two resources categories and the two resource-based manufacturing categories in the Lall classification. One should remember that in the
computational sense, the movement that these trajectories display, follows directly from a change in RCAs, in this case changes in the set of countries that are specialized in the products of the Lall category (when plotting country trajectories, it would be changes in the set of products that the country has RCA=1). We also expect that such changes would, by-and-large, correspond to changes in the country variables as indicated by the colored rays, which is indicated by how the trajectories relate to those rays.

This figure, like the ones that will follow, zooms into the part of the biplot around these trajectories, i.e., the scales differ between figures. However, the variable lines are always drawn on the same scale, as long as we look at Lall categories. These four Lall categories appear relatively low in the biplot, they move exclusively on the negative part of the vertical axis in Figure 3 (the left panel). The arrows indicate the direction of time: all trajectories start near the appearance of the labels and then move in the direction that the arrows point. One arrow comprises the movement of one year to the next.

These trajectories move somewhat erratically, but the general tendency along the axis that is formed by the two main GVC indicators, is up, which signals stronger GVC integration and a lesser and exclusive focus on a GVC role that is mostly supplying intermediates (in this case raw materials or processed raw materials). The exception is the 2000-10 move, when, as a result of the financial crisis, movement is generally down (i.e., towards lesser GVC integration). In other words, the main movement is along the axis that is formed by the GVC integration variable and the forward GVC-RCA variable. We also note that the agro-based resources trajectory is located much to the left of the other ones, indicating that these products are mostly found in poorer countries. There is a strong movement of this trajectory to the righthand side of the biplot.

Figure 5. Dynamic trajectories for Lall low-tech manufacturing categories

Next, Figure 5 documents the trajectories for the two low-tech manufacturing categories. One of these is for textiles-related products, the other for other low-tech products (such as simple metal products). These trajectories are higher up in the biplot, indicating larger GVC integration in general. The two low-tech categories are clearly separated on the horizontal axis, indicating that textiles low-tech products are mostly found in poorer countries. In terms of the movement, the direction is opposite: low-tech textiles moves down (lesser general GVC integration and more reliance on a forward supplier role), while the other low-tech moves up, but over a much shorter distance.
Figure 6. Dynamic trajectories for Lall medium-tech manufacturing categories

Figure 6 shows the trajectories for the medium-tech categories. Here we find automotive products (including parts of various kinds), process industries (e.g., chemicals) and engineering industries (e.g., machinery). These trajectories are all located in the part of the biplot that projects positively on the emissions line as well as on the general GVC integration line (i.e., the upper-right part, roughly). The trajectories all move left and slightly up, indicating that relatively poorer countries tend to become specialized in these products, and that these countries become more integrated into GVCs. These trajectories generally cover a long distance (except the process industries one), especially for the period up to 2005. This indicates that these product groups are identified as important bearers of structural changes in trade and GVCs in the early 21st century.

Figure 7. Dynamic trajectories for Lall high-tech manufacturing categories
Finally, Figure 7 shows the dynamic trajectories for the high-tech categories, of which there are two: electrical and electronics, and other high-tech products. These trajectories are in the same part of the biplot as the medium-tech trajectories, and also move in the same general direction. Especially for the electrical and electronics trajectory, the distance covered is large for the early years.

4.2. Dynamic trajectories of countries

We now move to present and discuss the dynamic trajectories of a number of selected countries. Note that the yearly all-products-centroids and the yearly all-countries-centroids are similar to each other in terms of their shape, although their scale differs (because the products are plotted as the barycenter of countries). We will therefore apply the same filtering process that omits the movement of the yearly all-countries-centroids when we plot trajectories for individual countries.

The country selection is somewhat arbitrary, a complete documentation of all 64 countries would take up too much space. We picked a number of countries in different parts of the biplot, and which show some significant movement over time. The selected countries are Australia (AUS), Brazil (BRA), Cambodia (KHM), China (CHE), Estonia (EST), Indonesia (IDN), Korea (KOR), Laos (LAO), Netherlands (NLD), and the United Kingdom (GBR).

Figure 8. Dynamic trajectories for selected countries
The trajectories of the selected countries are presented in Figure 8. One thing that stands out is that although all these countries move, they also stay in their “own” part of the biplot. Thus, Australia (minerals) and Brazil (agro) clearly remain in the resources part of the graph, Korea, the Netherlands and the UK remain in the high/medium-tech part of the biplot, and Cambodia, Indonesia and Laos remain in the low-tech part of the biplot. Estonia and China are in the center, but both move to the right. The movement of individual countries is diverse. Some countries move down, indicating lesser general GVC integration and development of a stronger GVC-RCA in supplying intermediates. Other countries move in the opposite direction. China is an interesting case with a reversal from down to up.

5. Summary and conclusions

The field of economic complexity attempts to relate export specialization performance to economic development. Our application of canonical correspondence analysis aims to do the same but wants to derive more specific relations between the trade specialization data on the one hand, and a set of specific indicators about development on the other hand. The results are promising in the sense that we were able to derive summary indicators (reduced dimensions) of the trade data that are clearly relatable to the economic indicators that we used. Thus, while the economic “complexity” approach yields an indicator that remains multi-interpretable, our CCA application yields clear interpretation of a multi-dimensional set of summary indicators. Together, the indicators that we derive construct an impressionistic “development space” in which we can position countries as well as products, as well as aggregations of those (i.e., product- and country-groups). We feel that this application of CCA can help develop the field of economic complexity further.

In terms of conclusions about development, our results suggest that the ordination (a “development space”) obtained in this way is relatively stable, i.e., both countries and trade products occupy positions in this space that change relatively little over the 30-year period that we consider. In terms of GVCs, this ordination suggests that there is an important trade-off between, on the one hand, being integrated in GVCs in a general sense, and, on the other hand, being specialized in supplying value to GVCs in the form of intermediate goods. There is a tendency of many products and countries to move along this trade-off towards larger integration into global value chains, as is, of course, also documented in the literature on this topic. But there are also countries and (groups of) products move in the opposite way along this trade-off, i.e., in the direction of becoming more specialized as suppliers of intermediate value. We must also note that, even though there is movement, the relative positions of countries (or products) vis-à-vis each other change little.

What does change is that poorer countries (and hence also countries with lower GHG emissions per capita) become more involved in global trade, through the vehicle of global value chains. This is the other main trade-off that our ordination proposes: one between GDP per capita and inequality. The general movement here that poorer and more unequal countries become more involved in trade. This process mainly takes place in medium- and high-tech manufacturing, which are product groups in which rich and develop countries are losing specialization to poorer countries. This is another main form of GVC development, which is not unrelated to the other tendency discussed above. Countries specialized in resources also become more integrated in GVC, but they maintain much less integrated than countries specialized in (medium- and high-tech) manufacturing. Low-tech textiles products remain an “outlier” in the GVC development space that we construct: these products do not share the tendency for stronger GVC integration over time, and remain the domain of a very select group of (relatively poor) countries.
References


Mealy, P., Farmer, J.D. and A. Teytelboym, 2019, Interpreting economic complexity, Science Advances 5, eaau1705, DOI: 10.1126/sciadv.aau1705


Van Dam, A., Dekker, M., Morales-Castilla, I., 2021, Correspondence analysis, spectral clustering and graph embedding: applications to ecology and economic complexity, Sci Rep 11, 8926, DOI: 10.1038/s41598-021-87971-9
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