Contrasting Revealed Comparative Advantages when Trade is (also)in Intermediate Products

Escaith, Hubert

19 October 2020

Contrasting Revealed Comparative Advantages when Trade is (also) in Intermediate Products

Hubert Escaith *

Nous privilégions l'ébauche où l'on sent le tremblement de l'œuvre qui se cherche…
Il n'est pas impossible que le lecteur y soupçonne des prolongements, une ouverture,
des pistes juste indiquées qui ne ferment pas l'œuvre sur elle-même.
Jean Rigaud (1925-2005) «Sur la littérature»

Abstract: The paper reviews and compare a selection of existing and new alternative indicators of Revealed Comparative Advantages, with a special emphasis on trade in intermediate products. The research adopts a statistical approach for both its theoretical and its analytical facets. The formal concepts are those used —inter alia—in statistical inference and information theory. The empirical part applies Exploratory Data Analysis on trade and production data from OECD's Inter-Country Input-Output Tables. International Input-Output data introduce a new dimension in the definition of comparative advantages: upstream or downstream competitiveness. It is shown that One-Way and Two-Way trade indices capture different aspects of trade competitiveness, and are complementary. Comparative advantages being relative by definition, ordinal or dichotomous classifications provide more robust results than the absolute cardinal indices. Even with dichotomous indicators, the classification of best performers remains blurry, fuzziness varying greatly among product categories.

Keywords: international trade, relative comparative advantages, intermediate inputs, indices, exploratory data analysis

JEL codes: C82, F12, F14

Table of Contents

* Contact email: hubert.escaith@outlook.com

Contrasting Revealed Comparative Advantages when Trade is (also) in Intermediate Products

Hubert Escaith, October 2020

1. Introduction

Closely associated with David Ricardo (1772 –1823), comparative advantage is a key concept in explaining the "raison d'être" of trade. Comparative advantages have received renewed attention in economic research. This happened in conjunction with the process of hyper-globalization that characterised the world economy since the early 1990s and the changes in the nature of world trade (global value chains and trade in tasks; emergence of Factory Asia as a major trade hub, etc.). This revival goes beyond trade analysis and has found new applications, particularly in development economics. For example, the measure of revealed comparative advantage is one of the building blocks of the Product-Space concept, used by Hidalgo, Klinger, Baraba´si and Hausmann (2007) to define promising specializations in terms of industrialization potentialities.

Comparative advantages cannot be measured directly and must be inferred from observing the volume, origin and composition of trade flows. In other words, comparative advantages are "revealed" by trade data. During decades, those data were the sole source of information for measuring "Revealed Comparative Advantages" (RCA). Balassa (1965) defined one of the first RCA indices. The Balassa formula remains one of the most widely used today, even if many other alternatives have been proposed since then.

Globalization, with the advent of global value chains (GVC) and the rise of trade in intermediate inputs, has questioned the conceptual basis of comparative advantages. Theoretical models have also improved, and trade analysts are now looking for the microeconomic foundations to what remained in practice a statistical construct. In this line of work, Eaton and Kortum (2002) proposed a model that successfully combined gravity variables and technological factors to define new measures of comparative advantages that would be both measurable and theoretically consistent.

The geographical fragmentation of production along global value chains and the capacity to trade in intermediate products has also changed the way comparative advantages were determined. This mutation of global trade promoted the development of new statistical models combining data on production and trade are combined to measure trade in value-added (WTO and IDE-JETRO, 2011; Koopman, Wang and Wei, 2014). As a result, we have improved on the observability of productivity by industry and by country that are the backbone of the Ricardian model.

New data and new theoretical models offer the possibility to build new indicators, each one pretending to improve on previous ones or to reflect different understandings of how comparative advantages should be measured with data. The trade analyst is now being offered a large palette of alternative indicators, some sharing similar building blocks but having different distributional properties, other amplifying the measurement to include different aspects of trade and production. This paper proposes to guide trade analysts in understanding the logic behind various RCA methodologies; by doing so, it aspires also at answering the following question: what empirical formulation is best suited for the measuring comparative advantages? To this aim, I review and compare a series of existing RCA indices proposed in the literature and modify them to include additional information on trade in intermediate inputs and production.

The examination adopts a statistical approach for both its theoretical and analytical facets. The formal concepts are those used —inter alia—in statistical inference and in information theory. The empirical part borrows its tools from Exploratory Data Analysis. To maintain comparability among the various indicators, I consider only RCAs based on usual trade and production data. Models based on micro-economic foundations or on value-added, which rely on different logic, are reviewed in annex.

The rest of the paper is organized as follows. After this introduction, the following section describes the formal statistical approach used implicitly or explicitly in the measure of most empirical RCA indices. Section III reviews a series of RCA formulations that have been proposed in the literature. Section IV suggests ways to modify these indicators in order to take into consideration trade in intermediate goods. The fifth section is dedicated to the empirical analysis of these indices, while the sixth one applies them to the analysis of changes in countries' RCA between 2005 and 2015. The last section offers some concluding remarks.

2. A few methodological considerations

Most of the RCA indicators that are proposed in the applied trade analysis literature are explicitly or implicitly rooted in a probabilistic approach (Kunimoto, 1977; Bowen, 1983). In this statistical approach, revealed comparative advantages are inferred from the deviation of actual trade flows with their expected value. This "expected" trade pattern is based on an uninformed "prior" (the best rational assessment of the probability of an outcome before collecting new information) where only the marginal distributions of world trade are known (e.g., weight of a country in the world trade and the weight of a given commodity in this total trade).[1]

The Statistician's way of measuring comparative advantages is very similar to a Bayesian approach. The Statistician knows a priori the nature and the origin or destination of the trade flows, thanks to customs data, but ignores the productive specialization of the country of origin (its comparative advantage). Trade data will help getting this information. So, let's put our Statistician hat (it won't last long, I promise).

Denoting country "i" total exports by X^i and total world exports by X^w, let's assume a homogeneous commodity "k" that is randomly traded in a free trade world. Here, "randomly" is meant to say that we do not have detailed observation of the actual trade flows taking place, so we use an uninformed prior assumption. This will be our neutral benchmark to be used when characterising actual trade patterns. The expected prior probability of observing that country "i" will export some product is estimated by the marginal frequency (X^i/X^w):

$$\mu(X^i) = \left(\frac{X^i}{X^w}\right), \quad \mu(X^i) \in]0,1[\qquad [1]$$

[1] The marginal distribution of a variable is the frequency of either the row or column variable in a contingency table (World exports by type of goods, for example). The frequencies are called "marginal" because they can be found by summing the values in a table along rows (X_i, the total exports of country "i") or by columns (X^w_k, the total exports of product "k"), and writing these sums in the margins of the table.

In order to calculate comparative advantages, we require $\mu(X^i)$ to be strictly larger than 0 (country "i" exports some of its production) and lower than 1 (country "i" does not monopolise world trade). Similarly, and without any additional prior knowledge of country "i" production capabilities, the probability to observe that product "k" is exported by any country picked at random (the P(A) in Box 1) is:

$$\mu(X_k^w) = \left(\frac{X_k^w}{X^w}\right), \quad \mu(X_k^w) \in]0,1[\qquad [2]$$

Where X_k^w represents the value of world exports of product "k".

The theoretical literature usually restricts product "k" to being a commodity in order to satisfy the condition of homogeneity. In practice, the analysis is extended to more diversified industrial products, or – as we shall see later when comparing one-way and two-way trade analysis– may even be restricted to these complex industrial products. It is also possible to apply it to trade in services, even if it is less frequent. Unless specified, we will use commodities, goods and products as synonyms.

In absence of special factors affecting "i" ability to export, the probability to observe that country "i" exports commodity "k" (noted here: X_k^i) is given by combining the marginal distributions of X_i and $X^w{}_k$:

$$\mu(X_k^i) = \mu(X^i) \cdot \mu(X_k^w) = \left(\frac{X^i}{X^w}\right) \cdot \left(\frac{X_k^w}{X^w}\right), \quad \mu(X_k^i) \in]0,1[\qquad [3]$$

In other words, with no prior additional information about country "i" production capabilities, we assume statistical independence: the probability of the joint event {country "i" exports product "k"} is equal to the product of the individual probabilities. $\mu(X_k^i)$ is the expected probability of observing exports of product "k" from country "i" in absence of any idiosyncratic factor affecting "i" ability to export "k".

If this hypothetical case (often referred to as the "neutral" situation) is a good representation of actual trade flows, no additional information can be gained by knowing the actual X_k^i export flows (the "microscopic" country properties) because only the knowledge of marginal distributions (the macroscopic World properties) is sufficient. This is also a definition of maximum entropy, a concept used —inter alia— in statistics and in information theory. [2]

Moving from probability [3] to the value of expected gross trade flow, we obtain the statistical expectation of the value of exports on "k" product by country "i" in the neutral situation:

$$E(X_k^i) = \left(\frac{X^i}{X^w}\right) \cdot X_k^w \qquad [4]$$

This formulation is central to the empirical measure of comparative advantages. Most applied RCA indices derive from the following rule: if the observed (X_k^i) is higher than the expected neutral one $E(X_k^i)$, then we conclude that country "i" has special characteristics, other than its sheer economic size, that

[2] In information theory, entropy is maximum when the joint probability of independent sources of information communicates as much information as the individual events separately. The opposite extreme case of minimum entropy would be a situation where each country fully specialises in exporting one good and one good only. In this case, one needs to know the microscopic (i.e., country-level) information; there is no uncertainty once it is known and the entropy is zero.

bestow it with special advantages in exporting the product "k". Indeed, the first family of RCA is based on the ratio between observed and expected trade flows.

$$RCA1_k^i = X_k^i / E(X_k^i); \quad \forall X^i \neq 0 \text{ and } X_k^w \neq 0 \qquad [5]$$

Assuming that all countries export at least one good and that all goods are internationally traded, the calculation creates a list of N*K indicators, where N and K are the total number of countries and products. When RCA_k^i is higher than 1, country "i" has a revealed comparative advantage in exporting "k".

Focusing on exports only is particularly relevant in situations of "one-way" trade, where countries are specialised in the export of some commodities and import those where they have no comparative advantage. Other indices have been proposed that apply to situations of "two-way" trade, where trade takes place in varieties of products and countries can be simultaneously exporters and importers for a given class of products. Two-way trade implies, for each product, a trade balance $(X_k - M_k)$ that results of transactions of inequal value, a majority and a minority flow (Lafay, 1992).

In a situation of "two-way" trade, Aho, Bowen and Pelzman (1980) propose to take net trade into consideration. Taking into consideration that the neutral $E(X_k^i / M_k^i)$ boils down to (X^i / M^i) when imports are measured FOB, it leads to an alternative indicator RCA2: [3]

$$RCA2_k^i = \frac{(X_k^i / M_k^i)}{\left(X^i/_{M^i}\right)} \quad \forall M^i, X^i \text{and } M_k^w > 0 \qquad [6]$$

Where M^i and M_k^i are, respectively, the total imports of country "i" and its imports of product "k".

(X_k^i / M_k^i) is a way of calculating net exports. [4] From a macroeconomic point of view, RCA2 takes into consideration (it "controls for") an unbalanced situation where domestic savings is low or high and country "i" has a structural trade deficit $(X^i / M^i < 1)$ or a surplus $(X^i / M^i > 1)$. It respects also the statistical criteria of Kunimoto (1977).

For Aho et al. (1980), using net exports is the correct way of measuring relative trade performance, but it is meaningful only for manufactured goods (where two-way trade is prevalent). As we shall see, many RCAs used in the literature are based on gross exports; discarding imports is usually explained by the fact that imports are affected by factors unrelated to comparative advantages (trade policy, tariffs, etc.).

3. Empirical Indices of Revealed Comparative Advantage (RCA)

Ballance, Forstner and Murray (1987) distinguish two additional classes of RCA indices, besides the One-Way and Two-Way approaches: the trade-only indices, using only trade data, and the trade-cum-production indicators that use also data on domestic production and consumption. In addition, all RCA

[3] When imports and exports are measured FOB without recording errors and in absence of any significant trade with outer-space, world exports must be equal to world imports. $E(X_k^i / M_k^i)$ simplifies to (X^i / M^i) when considering that $X_k^w = M_k^w$ and $X^w = M^w$. In practice, $X_k^w \neq M_k^w$ because exports and imports are not always recorded similarly by custom offices (differences in valuation method: FOB for exports, CIF for imports; differences in product classification between the exporter and the importer, etc).

[4] Net exports in standard trade literature are usually calculated using the additive formula (X-M). Its multiplicative counterpart (X/M), when it is defined (M>0), has the advantage of taking only positive values. This is an appreciable property for some applications, especially in econometric applications using logarithm.

indices can be interpreted from different perspectives. The traditional way is to consider that the index "quantifies" the comparative advantage enjoyed by a given country for a specific commodity (cardinal approach). The ordinal interpretation means that the RCA results provide a ranking of countries by comparative advantages for a given commodity. The dichotomous interpretation is that RCAs indicate only a demarcation between countries that enjoy comparative advantage for a product, and those that do not. This distinction is important when comparing different indices or when using them in econometric exercises. [5]

a. One-Way Trade RCAs

The pioneering Balassa's RCA index (Balassa, 1965) belongs to this class of indices. It remains very popular today.

- Balassa RCA (BRCA)

This index is calculated as the ratio of product k's share in country "i" exports to its share in world trade. Formally, it reads as:

$$\mathrm{BRCA}_k^i = \left(\frac{X_k^i}{X^i}\right) \Big/ \left(\frac{X_k^w}{X^w}\right) \qquad [7]$$

Intuitively, the index compares country "i" export structure with the World trade situation. A value of the RCA above one in sector "k" for country "i" means that "i" has a revealed comparative advantage in that sector. From a statistical perspective, BRCA measures the ratio between the "observed" exports X_k^i and the "expected" trade flow $\boldsymbol{E}(X_k^i)$ that could be inferred from the relative size of the "i" total exports in World trade.

$$\mathrm{BRCA}_k^i = \frac{X_k^i}{E(X_k^i)} \qquad [8]$$

$$\text{with} \qquad E\left(X_k^i\right) = \left(\frac{X^i}{X^w}\right) \cdot X_k^w$$

Thus, the Balassa index is not only intuitive, but it is also grounded in the probabilistic approach we defined in the previous section: $\boldsymbol{E}(X_k^i)$ corresponds to a situation of maximum entropy under frictionless free trade conditions. An important advantage of this index from a practical perspective is that it is not demanding in terms of data, as only export flows are required.

Despite being widely used, Balassa's RCA suffers from a series of formal weaknesses. Its theoretical foundation has been long debated in the literature since it does not actually reflect the original Ricardian idea of comparative advantages which is based on production and efficiency (Leromain and Orefice, 2013). It was only forty years after Balassa's paper that a seminal article by Eaton and Kortum (2002) revived the quest for a functional analysis of RCAs along formal Ricardian lines, yet incorporating the new results from trade theory (firms' heterogeneity and preference for varieties). We present in Annex this line of research.

BRCA suffers also from a series of practical issues that limit its use for comparative analysis. Hinloopen and Van Marrewijk (2001) find that its distribution is very skewed with a median well below one

[5] After observing a high level of inconsistency among alternative RCA indices, Ballance, Forstner and Murray (1987) recommend incorporating the ordinal and dichotomous perspective in empirical models.

(the neutral value for this index) and a mean well above one. A logarithmic transformation of BRCA is sometimes proposed as an alternative. Deb and Basu (2011) chose this index in their regression analysis because it is close to a normal distribution. LBRCA is defined as long as ($X_k^i > 0$).

$$\mathrm{LBRCA}_k^i \;=\; \log \mathrm{BRCA}_k^i \qquad [9]$$

The logarithmic transformation reduces the statistical bias, but does not correct it entirely. The statistical distribution of the Balassa index is found to differ considerably across countries, making comparisons between countries problematic. [6] Hoen and Oosterhaven (2006) argue that the issue is mainly linked to the multiplicative nature of RCAs and propose an additive measure as alternative.

Moreover, BRCA suffers from systemic biases, in particular it tends to exaggerate the comparative advantages of small countries (Yu, Cai and Leung, 2009). Nevertheless, De Benedictis and Tamberi (2001) find that the advantages of alternative indices that aim at fixing the distributional issues of the Balassa's index are still to be demonstrated.

- Revealed Symmetric Comparative Advantage (RSCA)

BRCA' s skewed distribution violates the assumption of normality in regression analysis, and gives much more weight to values above one, when compared to observations below one. To correct for this bias, Dalum, Laursen and Villumsen (1998) recommend using a symmetric version, obtained by comparing the BRCA with 1, its neutral value. RSCA is simply derived from BRCA:

$$\mathrm{RSCA}_k^i \;=\; (\mathrm{BRCA}_k^i - 1)/(\mathrm{BRCA}_k^i + 1) \qquad [10]$$

The RSCA is similar to a quasi-logarithmic transformation and is often preferred to the alternative logarithmic conversion of BRCA for having a finite inferior limit at -1. Yet, in empirical applications, the sample mean (or neutral) value of BRCA is usually higher than 1, affecting the symmetry of RSCA.

- Additive Comparative Advantage (ARCA)

The unstable sample mean of BRCA index leads to unstable distributions both across countries with respect to commodities, and across commodities with respect to countries. In order to make the distribution of Balassa's index stable with respect to countries, Hoen and Oosterhaven (2006) suggest an Additive Revealed Comparative Advantage (ARCA) index. ARCA uses the difference between the export shares, instead of their ratio as in the BRCA. ARCA is defined as follows:

$$\mathrm{ARCA}_k^i \;= \frac{X_k^i}{X^i} \;-\; \frac{X_k^w}{X^w} \qquad [11]$$

ARCA takes the value of zero when the export share of sector k in country "i" is equal to the world total. It is larger than zero if country "i" has a 'revealed comparative advantage' in sector k, and it is smaller

[6] The BRCA ranges from 0 to ∞, and Hinloopen and Van Marrewijk (2001) show that that the estimated mean obtained through empirical calculation is above the expected theoretical value of 1 for a given country (comparative strength in some sectors balancing comparative weaknesses in others). In addition, Hoen and Oosterhaven (2006) criticise the BRCA because its distribution strongly depends on the number of countries and industries covered, but this dependency on the sample and the level of aggregation is probably inherent to an indicator that is both empirical and comparative.

if it has a 'revealed comparative disadvantage'. Hoen and Oosterhaven (2006) show that the mean of the ARCAs has a value of zero, independent of the number and classification of the sectors or countries. [7]

They discuss the pros and cons of including or excluding the country "i" from the World total. Inclusion keeps the reference group constant. But, in that case the ARCA index becomes biased and the aggregate value at country level differs from 1, being smaller the more specialised and larger the country is. Yu, Cai and Leung (2009) mention also this issue, stating that the ARCA index is not comparable across countries.

While being aware of the potential bias, this should not be an issue if the calculation is done on small countries or on large countries that do not have a dominant situation in the export of some commodities. Or, more formally, when we can safely assume that, for all practical purposes, X^i and X_k^i are small enough for accepting the following approximation:

$$(X_k^w/X^w) \approx (X_k^w - X_k^i)/(X^w - X^i) \qquad [12]$$

- Normalised RCA

Yu, Cai and Leung (2009) propose an alternative that builds on the neutral situation as a starting point. In a comparative-advantage neutral situation where all countries export the same basket of products in proportion of their economic size, country "i" exports of commodity k (noted Ӿ_k^i) would be equal to :

$$\text{Ӿ}_k^i = X^i \cdot \left(\frac{X_k^w}{X^w}\right) \qquad [13]$$

In practices, actual exports differ from the neutral situation and $X_k^i \neq \text{Ӿ}_k^i$. Yu, Cai and Leung (2009) build on this difference to develop their indicator.

$$\mathrm{NRCA}_k^i = (X_k^i - \text{Ӿ}_k^i)/X^w \qquad [14]$$

Substituting Ӿ_k^i in [14], we obtain:

$$\mathrm{NRCA}_k^i = (X_k^i/X^w) - (X^i X_k^w/X^w X^w) \qquad [15]$$

By construction, NRCA is centred on 0, and it can be easily shown that:

$$\sum_i (X_k^i - \text{Ӿ}_k^i) = \sum_k (X_k^i - \text{Ӿ}_k^i) = 0 \qquad [16]$$

In other words, each country or each commodity considered as a whole is comparative-advantage-neutral. NRCA avoids also the "small country" bias present in other approaches. Among the interesting other properties of NRCA, the authors mention that the index is independent of the classification of commodities and countries (the level of product aggregation has no influence).

For Sanidas and Shin (2010), a clear advantage of its "zero-sum" property is to express well the Ricardian notion imbedded in comparative advantage: if a country gains comparative advantage in one sector, then the country loses comparative advantage in other sectors; and if one country gains comparative advantage in a sector, then other countries lose comparative advantage in the sector.

[7]

NRCA scores well on the comparability across space and time criterium: its sum equals zero across space and time, hence so does the mean value. It is bounded within the [-¼; ¼] interval and symmetrical, which loosely approximates the "normality" assumption required by standard econometric exercise. Finally, it does not treat all "0" trade flows equally, which adds to its better treatment of the "small country" bias. A large country with 0 export for one product would receive a higher comparative disadvantage score than a small country. This property is important from a small developing country's perspective, where export diversification at the extensive margin is often limited by objective supply constraints.

b. Two-Way Trade RCAs

All the indices discussed above use export data to reveal comparative advantages. When two-way trade is prevalent, as for most manufactured goods today, Lafay (1992) recalled that it becomes necessary to analyse also the symmetrical ratio of the Balassa RCA, calculated on the import side.

$$\mathrm{BRCAm}_k^i = \left(\frac{M_k^i}{M^i}\right) \Big/ \left(\frac{M_k^w}{M^w}\right) \qquad [17]$$

World imports (when measured FOB) being notionally equal to world exports,

$M_k^w = X_k^w$ and $M^w = X^w$

A priori, Balassa's comparative advantages must meet the condition ($BRCA_k^i > 1 \Rightarrow BRCA^m{}_k^i < 1$) while comparative disadvantage requires ($BRCA_k^i < 1 \Rightarrow BRCA^m{}_k^i > 1$). When results are contradictory, it becomes necessary to look at the trade balance and its composition. The import approach has been criticised, among other things, for being subject to the influence of tariffs and other protectionist measures that influence the volume and composition of imports. This was particularly true when the BRCA index was created in the 1960s. It is less valid today, in particular when analysing non-agricultural imports of developed countries.

A simpler way to take into consideration two-way trade is to consider net exports rather than gross, as suggested by Aho et al. (1980). [8] The following trade balance indicator is often used in the literature:

$$g_k^i = (X_k^i - M_k^i)/(X^i + M^i) \qquad [18]$$

But Lafay (1992) shows that it contains a systematic bias, stemming precisely from the existence of the minority flows in a two-way trade (p.213). He proposes an index based on a GDP weight. In practice, the Lafay index is usually modified to replace GDP by the share of trade (imports plus exports) of product "k" on total trade of "i" $[(X_k^i + M_k^i)/(X^i + M^i)]$ as the scale variable:

$$LRCA_k^i = 100\left[\frac{(X_k^i - M_k^i)}{(X_k^i + M_k^i)} - \frac{(X^i - M^i)}{(X^i + M^i)}\right] \cdot \left[\frac{(X_k^i + M_k^i)}{(X^i + M^i)}\right] \qquad [19]$$

[8] This approach tends, nevertheless, to ignore trade in varieties: Germany may export luxury limousines and import cheaper French cars. The imports of small cars do not reduce Germany's dominance in the luxury car market.

The Lafay index is often used in analytical trade database. Its distribution is centred (mean = 0) for each product k. Gnidchenko and Salnikov (2015) criticise it for being too dependent on product definitions, leading to an aggregation bias due to the heterogeneity of trade classifications.

They propose an index that builds on an approach developed by Bowen (1983) but is based, as the original Lafay formula, on a GDP weight instead of consumption as in Bowen's measure. Their index simultaneously accounts for export and import data and can be expressed as a function of "expected trade turnover" within the Kunimoto (1977) theoretical framework. At the difference of Kunimoto (1977), where the world exports of a commodity are distributed among countries in proportion of their share of total world exports (the neutral situation), Gnidchenko and Salnikov (2015) distributes expected trade of a commodity among countries in proportion of their share of world GDP.

By taking into account the weight of trade in GDP, they wish to put into perspective comparative advantages when a country's trade turnover is small relative to its GDP (typical case for the largest countries) or when the country is not highly integrated in world trade. They call their index the "Net Comparative Advantage Index" (NCAI).

$$NCAI_k^i = \left(\left[\frac{(X_k^i - M_k^i)}{(X_k^i + M_k^i)} \right] \cdot \left[\frac{(X_k^i + M_k^i)}{(GDP^i)} \Big/ \frac{(X_k + M_k)}{(GDP^w)} \right] \right) \qquad [20]$$

With GDP^i and GDP^w being the gross domestic product of country "i" and the world total.

The first part of the right-hand side is the relative net export index (RNX_k^i) and the second part of the formula measures the relative trade openness of country "i" for product "k" (RTO_k^i). RTO_k^i measures also the ratio between the observed exports and imports of the product "k" in country "i" and its expected value considering the relative weight of country "i" GDP. After some manipulations, (RTO_k^i) can be further disaggregated into two components: (RT_k^i), which is the trade intensity of good "k", and (RO_k^i), which reflects the relative openness to trade of the "i" economy. As the authors mention, it "allows us to simultaneously account for economic openness and importance of a trade flow of a certain good for the economy" (p.15).

$$NCAI_k^i = RNX_k^i \cdot RO_k^i \cdot RT_k^i \qquad [21]$$

As other indices based on trade balance for a given product, NCAI reflects also intra-industry trade and comparative advantages emerging from intra-industry specialization. An extension of this index would, ideally, take into consideration the share of imports that constitute the inputs required for producing the exports. I return to this point in the next section.

On the cons side, NCAI may display extreme values and is not centred. Noting that trade intensity is the most volatile part of their index, the authors propose a symmetric version of their index. $SNCAI_k^i$ deals with the issue of extreme values, by using a normalized trade intensity (RT_k^i), a procedure that "impacts the extreme values primarily" (p.16).

$$SNCAI_k^i = RNX_k^i \cdot RO_k^i \cdot (RT_k^i / RT_k^i + 1) \qquad [22]$$

Comparing their results with Leromain and Orefice (2013), Gnidchenko and Salnikov (2015) state that their simpler index has good empirical characteristics and does not need the additional calculations and econometric estimates required by theoretically consistent structural models (see Annex for a review).

4. Accounting for inter-industry Input Output relationships

All the above-mentioned indices build on trade statistics. GDP used by the NCAI in [20] is, in practice, the sole additional indicator that takes into consideration domestic production and income. Since the early 2010s, the dissemination of international input-output tables has provided internationally comparable production and trade data covering inter-industry trade in intermediate inputs. The new stock of information allows to contemplate production and trade-in-intermediate goods models, and suggest new indicators.

a. Comparative advantages and trade in intermediate goods

It is usually argued, in what Amano (1966) calls the "text-book style explanation of comparative advantage", that comparative advantages reflect a country's comparative cost structure. When there are many commodities, (revealed) comparative advantages provide "a scale measuring each trading partner's comparative cost ranking". This explanation of comparative advantage assumes that production costs are domestically defined (labour, in Ricardo's approach). Yet, in today's world trade, a large share of traded goods are intermediate products that are used by the importing industry for its own production. In other words, the pattern of world trade specialisation, at least when processed goods can be produced through global supply chains, may not be entirely predicted by the comparison of pre-trade cost ratios. [9]

Escaith (2019) illustrates this mechanism from the perspective of Efficiency Frontier Analysis using Data Envelopment Analysis (Box 1).

Box 1. Gaining efficiency through trade in intermediate inputs

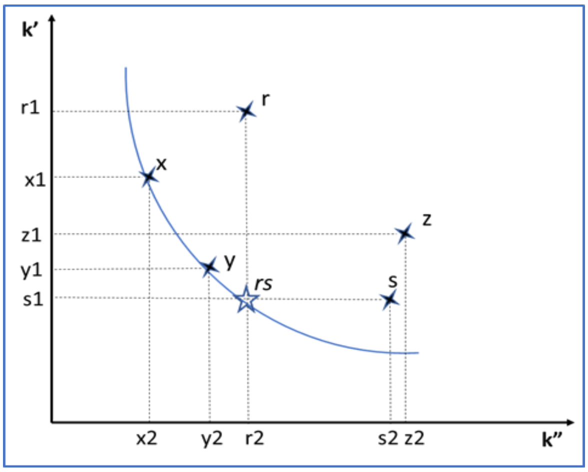
Figure 1 illustrates how two inefficient productors (*r* and *s*) in different countries can join forces and become internationally competitive. All firms use two intermediate inputs to produce a variety of similar goods: the first input (vertical axis) is based on a technology k' intensive in labour and the second one (based on k'' technology) is intensive in high-tech capital. The *r* firm is located in country A while *s* is in country B. Other firms are located in various countries in the Rest of the World.

The isoquant indicates the minimum combination of inputs to produce a given quantity of good q0 and the distance from the isoquant shows the relative efficiency of five production units (r, s, x, y, z) located in different countries. Under frictionless free market, all physical inputs and outputs are priced the same, but firms face different labour wage rates ***w***. Considering that the price of labour is supposed to be inversely proportional to the technology level attained by countries (Balassa–Samuelson effect), Escaith (2019) assumes that the isocost line is curved and blends with an isoquant.

Without trade in intermediate inputs, only two firms (*x, y*) in Figure 1, are on the isoquant and are competitive at world price. Other three firms (*r, s, z*) are away from the curve and inefficient for this production. Yet, *r* inefficiency relates only to its use of input produced using k' technology (a "slack" in Frontier Analysis), while *s* is inefficient for the k" type.

[9] Deardorff (2005) revises the Ricardian Law of Comparative Advantages when trade includes intermediate inputs. Including wage differentials and transport costs, he shows that access to imported inputs provide an additional source of gain from trade.

Figure 1 Gaining efficiency through production sharing



Note: k': use of intermediate input based on labour intensive technology to produce q0; k": use of capital-intensive input; k"': use of natural resource intensive input.
Source: Escaith (2019)

If it is possible to separate the production of intermediate inputs k' and k" in two separate steps, then unbundling the production of q0 in two components allows *r* to specialise in the production of the components intensive in input of k" type, while *s* specialises in the tasks that are labour intensive (technology k'). The joint venture (*r, s*) defines what is known as a "global value chain".

Because slacks are independent of each other by construction of the data envelopment technique defining the efficiency frontier, the unbundling maintains the efficiency of each firm for each zero-slack input (s1 and r2) and creates a new virtual firm *rs* that is cost efficient for the final product and located on the isoquant. Production of the final good q will be physically located in *s*, the country efficient in the labour-intensive inputs (labour being not tradable).

The development of global value chains rendered the use of traditional export data of debateable interest in analysing comparative advantages. When using traditional trade data, one may be able to capture correctly the comparative advantage of countries "*x*" and "*y*" in Box 2, but the competitiveness of the joint venture "*rs*" for the final good produced with technologies k' and k" will entirely be attributed to "*s*". How can we avoid this error and account for the separate contribution of "*r*" and "*s*"? The solution is to use International Input Output tables, like the one depicted in Figure 2.

Figure 2 International Input-Output table

Inputs \ Outputs		Intermediate Use				Final Demand				Total Output
		1	2	…	M=kxn	1	2		M	
Intermediate Inputs	1	Z^{11}	Z^{12}	…	Z^{1m}	Y^{11}	Y^{12}	…	Y^{1m}	X^{1}
	2	Z^{21}	Z^{22}	…	Z^{2m}	Y^{21}	Y^{22}	…	Y^{2m}	X^{2}
	…	…	…	…	…	…	…	…	…	
	M	Z^{m1}	Z^{m2}	…	Z^{mm}	Y^{m1}	Y^{m2}	…	Y^{mm}	X^{m}
Value-added		$(VA^{1})'$	$(VA^{2})'$	…	$(VA^{m})'$					
Total output		$(X^{1})'$	$(X^{2})'$	…	$(X^{m})'$					

Notes: $\boldsymbol{Z}^{ij}$ is an *K*×*K* matrix of intermediate input flows that are produced in country *i* and used in country *j, K* being the number of activity sectors (goods and services) and *N* the number of countries; $\boldsymbol{Y}^{ij}$ is an K×1 vector giving final products produced in country *s* and consumed in country *r*; $\boldsymbol{Q}^{i}$ is also an *K*×1 vector giving gross outputs in country *s*; and $\boldsymbol{V}^{i}$ denotes an *K*×1 vector of direct value added in country *i*.
Source: Adapted from Wang, Wei and Zhu (2013)

Reading the table in line, for each country-sector duplet "ik", (i∈[1,N] and k ∈ [1,K]), the element $z^{ik,jp}$ of matrix $\boldsymbol{Z}^{ij}$ measures the exports of intermediate product "k" by country "i" to the country-sector duplet of destination "jp" (j∈[1,N] and p ∈[1,K]). When i=j, matrix $\boldsymbol{Z}^{ij}$ shows the domestic transactions. In other words, the IIO table provides information on both national and international inter-sector transactions. On the final demand panel, the element $y^{ik,j}$ of vector $\boldsymbol{Y}^{ij}$ measures the exports of final product "k" by country "i" to country "j".

Matrix $\boldsymbol{Z}^{ij}$ can be read in columns, and in this case, element $z^{ik,jp}$ will measure the quantity of inputs "k" purchased from country "i" that the sector "p" in country "j" requires for producing the output x^{jp}. Final demand column vector $\boldsymbol{Y}^{ij}$ indicates the amount of final goods imported by "j" from "i", when i≠j.

There are some important differences in the measure of trade flows between official trade statistics, such as those found in UN-COMTRADE, and those provided by IIO tables. In IIO tables:

1) Trade in services is included. Imports of merchandises are measured FOB.
2) Bilateral trade flows are symmetric: for any given product "k", the value of exports from country "i" to country "j" equals the imports by "j" from "i".
3) Because trade data provide only information by country of origin and destination, the inter-industry disaggregation of bilateral flows results from imputations.
4) The "k" categories are very aggregated and classified by sectors of activities. For example, agricultural products bundle together cereals, meat, fishes, etc. The level of disaggregation varies according to each database. Usually, the wider the geographical coverage, the smaller the level of details. [10]
5) Implicitly, trade takes place in varieties and, at least in theory, products are not easily substitutable. This reflect the fact that columns represent the inputs of a Leontief production function that implies that all inputs enter in fixed (pre-determined) proportions.

Finally, it is easy to differentiate trade in final goods and trade in intermediate inputs. [11] Industrial output can be split in two: part of the product will be used (domestically or exported) as final good and part will be used as intermediate good. These goods are designed to satisfy different purposes. For example, the electronic industry may produce flatscreen used for TV sets (final goods) or monitors (intermediate product) used in producing laptop computers or numerical command machines. The various possible degrees of disaggregation provide additional light on countries' specialization and upstream or downstream specialization in the global value chain.

b. Towards IIO-based RCAs

Based on the review of literature on RCA indices, it appears that prospective IIO-RCAs need to satisfy a series of properties. Our first question, nevertheless, is to decide on the proper approach of trade: one-way or two-way?

Apparently, the world of IIOs describe two-way trade. But if we consider that the Leontief production function implies intermediate products that are not substitutable, each country is expected to specialise into a specific variety. So, what looks two-way is actually one-way, at least for trade in intermediate inputs. For each sector of activity, there is no coexistence of a majority and a minority flow, but an import of particular varieties of intermediate products produced by similar industrial sectors in foreign countries that

[10] For example, WIOD includes 56 goods and services sectors for 43 countries, OECD-WTO TiVA includes 36 sectors for 64 economies and Eora 26 sectors for 190 countries or regions. For comparison, the Harmonised System used for trade in merchandises distinguishes some 5,300 products.

[11] In truth, it is also possible to do it on traditional merchandise trade statistics, using the BEC classification to separate intermediate and final goods.

are used to produce a new variety of processed output. As we shall see, it is also implicitly required that trade is one-way when building the RCA using the statistical approach.

We want to measure the domestic share of the value of its gross exports in order to determine a country's genuine comparative advantage. It is achieved here by deducting all the imports of intermediate products required in the production of these exports. Denoting by χ_k^i the value of k exports by country "i", and assuming that the products exported are produced with the same technology than the products sold locally, I define exports net of imported inputs as:

$$\chi_k^i = \mathrm{X}_k^i - \left[\mathrm{X}_k^i \cdot (M_k^{iQ}/Q_k^i)\right] \qquad [23]$$
$$\text{With } M_k^{iQ} = \sum_{j,p}(z^{ik,jp}), \forall\, j \neq i$$

Where the $z^{ik,jp}$ are the elements of matrix Z^{ij} measuring the imports by industry "k" in country "i" of the intermediate products "p", p ∈ [1,K], produced by the foreign countries j, j∈[1,N]. and p ∈ [1,K]), and required to produce output Q_k^i.

The calculation of $E(\chi_k^i)$ according to the probabilistic approach assumes mutual independence between the random variables M_k^{iQ} and Q_k^i . Because (M_k^{iQ} / Q_k^i) is a parameter of the specific production function of country "i" for the product "k", the independence of M_k^{iQ} and Q_k^i implies the coexistence of a large number of possible techniques to produce "k", some requiring more imports than others. [12]

$$E(\chi_k^i) = E\left(\mathrm{X}_k^i\right) \cdot \left[1 - \frac{E\left(M_k^i\right)}{E(Q_k^i)}\right] \qquad [24]$$

Where $E(X_k^i)$ derives from equation [3]. The same approach applies to $E(M_k^{iQ})$ and $E(Q_k^{iQ})$

$E(M_k^{iQ}) = [(M^{iQ}/M^{wQ}) \cdot (M_k^{wQ}/M^{wQ})] \cdot M^{wQ}$

$E(Q_k^i)] = [(Q^i/Q^w) \cdot (Q_k^w/Q^w)] \cdot Q^w$

Where M^{iQ} and M^{wQ} stand for the total imports of intermediate goods by country "i" and world; Q^i and Q^w measure the total output of "i" and world, all products included.

Equation [24], which gives a central role to production, is similar to the production and consumption-based approach promoted by Bowen (1983). Adapting Bowen's index IT_k^i to our notation gives:

$$BIT_k^i = \frac{[Q_k^i - C_k^i]}{Q_k^w\left(\frac{GDP^i}{GDP^w}\right)}$$

Which simplifies into: [25]

$$BIT_k^i = \left[\frac{X_k^i - M_k^i}{Q_k^w\left(\frac{GDP^i}{GDP^w}\right)}\right]$$

Where C_k^i is the domestic use of product "k" in country "i" $(C_k^i = Q_k^i + M_k^i - X_k^i)$

[12] This is also a hypothesis commonly found in the economists' approaches, for example Eaton and Kortum (2002).

All the one-way trade RCAs that were reviewed in the previous section can also be recalculated using χ_k^i. More generally, the generic equation [5] becomes:

$$\text{IO\_RCA1}_k^i = \chi_k^i \,/\, E(\chi_k^i) \qquad [26]$$

Two-way trade RCAs can also be computed replacing imports M_k^i with M_k^{iQ} in equation [6]

$$\text{IO\_RCA2}_k^i = \frac{\chi_k^i}{M_k^{iQ}} \Big/ \frac{\chi^i}{M^{iQ}} \qquad [27]$$

With $M^{iQ} = \Sigma_k (M_k^{iQ})$

In theory, the IIO approach requires one-way trade in order to be able to calculate equation [24]. But the assumption can be relaxed for practical reasons. Indeed, the IIO approach deals in practice with large aggregates of individual goods and services and not with specific individual products. Unless countries have comparative advantages for all the products varieties produced by a sector of activity, the coexistence of microscopic one-way trade is compatible with the observation of two-way trade at inter-sectoral level. [13]

c. Accounting for double counting

The proper calculation of M_k^{iQ} and M^{iQ} is not as straightforward as it seems. The imports M_k^i and M^i in equation [6] include some intermediate goods that were produced by country "i", exported to third countries and re-imported when purchasing foreign intermediate and final goods. These re-imports must be discounted in order to avoid double-counting. [14] Intuitively, the calculation is based on the following reading of Figure 2, considering a single sector "k" in country A.

Figure 3 Schematic view of domestic intermediate goods exports and reimports

	Intermediate Inputs			Final Demand			Output Q
Country	A	B	C	A	B	C	
A							Q^a
B							Q^b
C							Q^c
Output Q	Q^a	Q^b	Q^c				Q^w

Note: Primary inputs (value-added) are not shown; they are part of domestic inputs.

In order to produce Q^a, A imports M_k^{iQ} intermediate goods from B and C (the light grey cells in Figure 3). Part of this output Q^a is used to produce intermediate goods and another part is used to produce final goods. Out of the intermediate and final goods, some are used domestically and others are exported. When an intermediate product from Q^a is exported as intermediate inputs and used by other countries to produce their own goods (the dark grey cells in Figure 3), some of the Q^a value embodied in these products will be

[13] For Bowen (1983), this is just a convenient way of solving the issue "in practice"; in theory one should expect "macro" indices to be derivable from underlying "micro" trade flows. Thus, when doing this assumption, I call for forbearance under the protection of A. Einstein's famous quote: "In theory, theory and practice are the same. In practice, they are not".

[14] This is upfront double counting and is not directly related to the more complex issue of double counting in Trade in Value-Added measurement, as in Wang, Wei and Zhu (2013).

reimported by country A, either as intermediate good for further production, or as final goods. When incorporated in final goods, it is absorbed and exits the production networks; when incorporated into new intermediate goods, a new production-consumption circuit iteration starts. [15]

In a multi-sector configuration, the total value of foreign inputs M^{iQ} required by "i" to produce all its k= 1 to K outputs is given by:

$$M^{iQ} = \sum_k \left[\sum\nolimits_{j \neq i} (z_k^{ij})\right] \quad [28]$$

Some of these imports include intermediate products that were exported by "i" to other countries "j" then reimported when "i" purchases processed products from "j". Considering for simplicity same proportionality between the different types of utilization countries "j" made of their production Q^j (exported or sold domestically, for final or intermediate use), I assume that the expected share of country "i" exports of intermediate goods ($d\chi^i$) returning home (re-imported) as intermediate or final goods embodied in imports is:

$$E\left(\mathrm{d}\chi^i / M^i\right) = (X_{IG}^w / Q^w) \cdot \left(M^i / M^w\right) \quad [29]$$

Where (X^w_{IG}/Q^w) is the share of world output that is exported as intermediate goods and (M^i/M^w) is the weight of country "i" in total imports. Total imports by "i", net of ($d\chi^i$) the expected reimports of intermediate goods, and noted $Ꞃ^i$ are: [16]

$$Ꞃ^i = M^i \cdot \left[1 - \left(\frac{X_{IG}^w}{Q^w}\right) \cdot \left(\frac{M^i}{M^w}\right)\right] \quad [30]$$

Assuming proportionality, the disaggregation of net imports of intermediate products $Ꞃ^i_{IG}$ is:

$$Ꞃ_{IG}^i = Ꞃ^i \cdot \frac{M_{IG}^i}{M^i} \quad [31]$$

By difference, the net imports of products used for final demand ($Ꞃ^i_{FD}$) is:

$$Ꞃ_{FD}^i = Ꞃ^i - Ꞃ_{IG}^i \quad [32]$$

The same proportionality assumption extends to the net imports of individual "k" products for intermediate or for final use:

[15] A proper accounting of all Qi's contributions to domestic and foreign production and consumption requires measuring trade in value-added rather than in gross commercial value; it would entail undertaking a journey into new concepts and calculations that require drifting away from our present purpose.

[16] For simplicity, I approximate actual re-imports by their expected value. In rigor, it would be possible to calculate the exact value for each sector in each country, at the cost of some cumbersome calculations. In the case of most countries, this value is negligible. This is not the case for large countries deeply involved in GVC trade. Wang, Wei and Zhu (2013) estimate that the share of exported domestic value returning home at 9% for the USA and between 4% and 5% for China and Germany.

$$\Pi^i_{kIG} = \mathrm{M}^i_{kIG} \cdot \left(\frac{\Pi^i_{IG}}{\mathrm{M}^i_{IG}}\right)$$

$$\Pi^i_{kFD} = \mathrm{M}^i_{kFD} \cdot \left(\frac{\Pi^i_{FD}}{\mathrm{M}^i_{FD}}\right) \quad [33]$$

To avoid double-counting, the One-Way and Two-Way RCA families defined by [26] and [27] are calculated substituting Π^i and Π_k^i for M^i and M_k^i.

5. Empirical properties of the RCA indices

As we saw, there are many different approaches for designing empirical RCAs. The probabilistic approach I adopted here, following Kunimoto (1977), provides a rationale to interpret deviations from the expected values. But it gives no indication on what is the best indicator for measuring Ricardian comparative advantages. Bebek (2017) states that there is no rigorous justification in the economic literature as to why one would employ a particular RCA index and not another. In other words, we are confronted to an empirical issue.

In what follows, we will look at the statistical properties of the various RCA candidates, either in their cardinal dimensions to provide a "measure" or in their ordinal capacity to suggest a ranking. Besides their own statistical merits (stability of the first moments, normality of the distribution), we will look at how they compare between themselves. But on the latter criterion, the jury is still out: do you prefer an index which is in-line with the others, or one that provides a different picture? If you are an Econometrician and look for a catch-all index to include in a regression, you will prefer the first option. If you are a Statistician and believe that information is in the variance, you may opt for the second criterion.

a. The data

Being confronted to an empirical issue means looking at the data. For this exercise, I used the latest OECD's ICIO harmonized input-output database, at the basis of the TiVA database. [17] The 2018 edition of the TiVA database provides indicators for 64 economies including all OECD, EU28 and G20 countries, most East and South-east Asian economies and a selection of South American countries. Inter-industrial transactions are disaggregated into 36 unique industrial sectors, covering the period 2005 to 2015. [18]

Not all the 36 sectors producing goods and services in the ICIO tables can be considered as involved in international trade. Table 1 presents the list of the 25 sectors producing "tradable" goods and services for which I calculated the 20 RCAs revised in this paper using the 2015 data, the last year covered by the OECD's ICIO tables. Sector 55T56: Accommodation and food services is included as it may be an important exporter of services under Mode 2 for countries having a strong tourism activity. For each sector, two sets

[17] ICIO data are freely available at https://www.oecd.org/sti/ind/inter-country-input-output-tables.htm

[18] At the difference of the previous OECD-WTO TiVA database, which covered the period 1995-2011 for a smaller group of countries, the 2018 OECD release is based on the 2008 version of the UN System of National Accounts, which has some unfortunate features from the trade in value-added perspective. In particular, it excludes intra-industrial trade in goods when trade takes place within contractually bound international supply chains and inputs at various stage of processing travel the global value chains without changing ownership. SNA2008 does not record the value of trade in this case, as would do a trade statistician, but only the smaller processing fees as manufacturing services. OECD data were used as such, except for aggregating the export-oriented and domestic-oriented sub-tables into a single national one in the case of China and Mexico.

of RCAs were computed, one for exports of intermediate products (sales of intermediate inputs to other industries located in foreign countries) and one for trade in final products (consumer and investment goods and services). GVC trade is particularly involved in inter-industry trade, as the intermediate products that are exported are used as inputs by other industries and reinter the production chain.

Table 1 List of ICIO sectors included in the RCA trade analysis

OECD code	Short label	Long label
01T03	01Agr	Agriculture, forestry and fishing
05T06	02MinF	Mining and extraction of energy producing products
07T08	03MinNF	Mining and quarrying of non-energy producing products
09	04MinSer	Mining support service activities
10T12	05Food	Food products, beverages and tobacco
13T15	06Text	Textiles, wearing apparel, leather and related products
16	07Wood	Wood and products of wood and cork
17T18	08Paper	Paper products and printing
19	09Fuel	Coke and refined petroleum products
20T21	10Chem	Chemicals and pharmaceutical products
22	11Plastic	Rubber and plastic products
23	12NoMet	Other non-metallic mineral products
24	13MetBas	Basic metals
25	14MetFab	Fabricated metal products
26	15Electro	Computer, electronic and optical products
27	16ElecEq	Electrical equipment
28	17Machin	Machinery and equipment, nec
29	18Vehicle	Motor vehicles, trailers and semi-trailers
30	19OthTsprt	Other transport equipment
31T33	20OthMan	Other manufacturing; repair and installation of machinery and equipment
55T56	25Hotel	Accommodation and food services
61	27Telecom	Telecommunications
62T63	28ITserv	IT and other information services
64T66	29Finance	Financial and insurance activities
69T82	31OBuserv	Other business sector services

Note: When required, the calculation of input requirements for these industries includes all sectors, tradable and non-tradable.
Source: Based on OECD, Inter-Country Input-Output (ICIO) Tables, 2018 edition

b. Distributional properties of individual RCAs

A detailed analysis of the statistical properties of 20 RCA indicators calculated on 25 tradable sectors produces lots of data. For a starter, Table 2provides summary statistics on the 18 RCA indicators (log BRCA is omitted due to its similarity with BRCA2).

Table 2 Summary statistics for RCA indexes

	Intermediate Products				Final Products			
Variable	Minimum	Maximum	Mean	Std. deviation	Minimum	Maximum	Mean	Std. deviation
ARCA	-0.12	0.85	0.00	0.07	-0.13	0.60	0.00	0.06
ARCA_IO	-0.12	0.89	0.00	0.07	-0.13	0.59	0.00	0.06
BIT	-3.91	8.67	0.01	0.50	-3.77	3.47	0.01	0.29
BIT_IO	-7.29	8.52	-0.03	0.52	-3.87	2.28	-0.02	0.23
BRCA	0.00	25.70	1.16	1.92	0.00	61.77	1.22	2.66
BRCA_IO	0.00	29.99	1.23	2.09	0.00	64.10	1.30	2.87
BRCA2	-1.00	0.93	-0.23	0.50	-1.00	0.97	-0.20	0.46
BRCA2_IO	-1.00	0.92	-0.28	0.48	-1.00	0.96	-0.28	0.45
LRCA	-15.89	31.05	0.00	3.23	-15.16	21.86	0.00	2.87
LRCA_IO	-16.15	33.95	0.07	3.45	-15.81	25.14	0.08	3.06
NCAI	-20.18	21.29	0.02	1.68	-6.79	72.59	0.17	2.49
NCAI_IO	-59.77	20.18	-0.16	2.34	-6.97	31.50	-0.02	1.56
NRCA	-0.02	0.02	0.00	0.00	-0.01	0.02	0.00	0.00
NRCA_IO	-0.01	0.01	0.00	0.00	-0.01	0.02	0.00	0.00
SNCAI	-4.33	2.81	-0.05	0.40	-2.31	3.81	-0.02	0.45
SNCAI_IO	-3.91	2.27	-0.10	0.39	-2.00	2.22	-0.07	0.37
SNCAI2	-20.18	21.29	0.02	1.68	-0.33	2.01	0.01	0.09
SNCAI2_IO	-59.77	20.18	-0.16	2.34	-0.36	1.00	0.00	0.07

Note: LBRCA excluded, as it derives from the logarithm of BRCA. BRCA equals 0 in a few cases where the IO matrix does not report any export. In this case, a small value is imputed to calculate the log.
1600 observations for each index (Rest of World region excluded).

Source: Based on OECD ICIO 2015 data

Two pieces of data are of interest here: the sample mean (preferably 0 or 1) and the standard deviation. What we look for is an indicator with a normalised mean (before mean-centring the data, evidently) and as much variance as possible (for a Statistician, information is in the variance). On this criterion, LRCA (a two-way RCA indicator) is our preferred one.

Normality is another criterion for assessing the practical relevance of an empirical RCA when econometricians look for a good candidate to include it in their modelling exercises. There are many ways of assessing the normality of a distribution; I use here the Jarque-Bera test.[19] The test is applied to each index, calculated on both intermediate and final products for each of the 25 tradable goods and services sectors. Out of the 1000 results obtained, only a small fraction (less than 10%) tests positive for normality at a significance level of alpha=0.1 for at least one of the calculations on intermediate or on final products. Table 3 shows the indices that produced at least one normal series of results, and the frequency of occurrences.

Some sectors appear to behave more "normally" than others, at least for particular RCA indices. Machinery & equipment and Rubber & plastic products appear in more than 10% of the positive cases in Table 3. At the contrary, Textiles & apparel and Motor vehicles, two sectors frequently analysed in global value chains studies, have very low rate of occurrence (less than 2%). Similarly, Computer, electronic & optical products, one of the most globalized supply chains, do not even appear in the list of sectors that produce a positive normality test for at least one of the RCA indicators.

Table 3 Normality test of alternative RCA indices

Index	Occurences (%)	Index	Occurences (%)
BRCA2	19	SNCAI2	3
BRCA2_IO	18	NCAI_IO	2
SNCAI_IO	11	SNCAI2_I	2
SNCAI	10	BIT_IO	2
LBRCA_IO	9	LRCA_IO	2
LBRCA	8	BIT	2
NRCA_IO	4	NCAI	2
NRCA	4	LRCA	1
SNCAI2	3		

Note: Based on Jarque-Bera normality test of alternative RCA calculated for intermediate and final products.

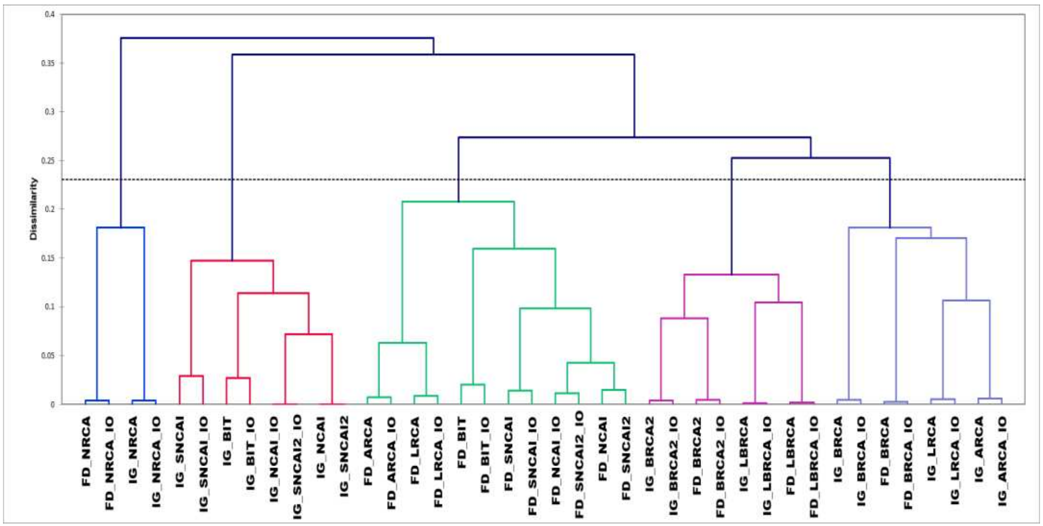
Finally, normality is strongly rejected ($p < 0.0001$) for all indices when all sectors are bundled together. So, our first conclusion is that normality is more the exception than the norm and this must be kept in mind when using RCA indices in econometric modelling. The alternative is to use ordinal indices. By construction, ordinal indices (rank analysis) avoid the asymmetry issue, because all observations are on both sides of the median observation. But rank analysis is less common in econometrics, as it requires using different types of statistical models (e.g., quantile regressions or categorical data analysis).

[19] The Jarque–Bera test is a goodness-of-fit test of whether sample data have the skewness and kurtosis matching a normal distribution. The test is named after Carlos Jarque and Anil K. Bera.

c. Exploring the RCA domain

Let's turn to Exploratory Data Analysis to understand the diversity or similarity of the various RCA indices. I use here agglomerative hierarchical clustering (AHC). Before that, the RCA results are mean-centred and normalised by their standard deviation, in order to have comparable data series. Then, the indicators are paired by increasing dissimilarity: the connections appearing at the bottom of the graph take place between the most similar indicators. Figure 4 shows the result of an AHC on RCA results for all sectors and both intermediate and final goods.

Figure 4 Agglomerative hierarchical clustering of RCAs



Note: Dissimilarity: Pearson dissimilarity on centred data; Agglomeration method: Unweighted pair-group average. The prefix FD stands for Final Demand and IG for Intermediate Goods and Services; RCAs are calculated aggregating all tradable sectors.
Source: Based on OECD ICIO 2015

As could be expected, most "Gross Trade" and "Input-Out-Output" families of indicators start by being paired together when they belong to the same RCA formula and the same type of products (final or intermediate). So, incorporating the GVC dimension in the calculation does not change fundamentally the RCA results. This said, some indicators are more affected by the input-output dimension than others and show more dissimilarity. It is the case for the SNCAI, especially when calculated for the intermediate products and for BIT, for both intermediate and final goods.

There is more to be learned from the AHC analysis. The optimum entropy criterion (smallest number of groups providing significant information on the members of the group) defined five groups (dashed line on Figure 4 above). The NRCA family (a one-way trade indicator) is one cluster all by itself: the NRCA family of indicators will only be regrouped with other RCAs at the very end of the agglomeration process. The logarithmic and pseudo-logarithmic BRCAs (LBRCA and BRCA2) constitute also a closely knitted cluster. Their original source, the Balassa's BRCA, relates more closely to LRCA and ARCA than to its logarithmic avatars; but this is valid only for intermediate products. Indeed, ARCA (one-way trade) and LRCA (two-way trade) calculated on final demand belong to a large family of indicators calculated on final demand: BIT, NCAI and SNCAI (all are two-way trade RCAs). The fifth cluster is made of the same set of two-way-trade indices calculated on intermediate goods: BIT, NCAI and SNCAI.

We can conclude a few things at this stage.

- First, revealed comparative advantages tend to differ between trade in intermediate and trade in final goods
- Second, NRCA index measures something that other RCA indices don't take into consideration, or avoid an issue (like the small country bias) that exists in other indices.
- Third, that one-way trade RCAs share similarities with two-way trade RCAs for final goods but not for trade in intermediate products.
- Fourth, that BIT, NCAI and SNCAI do a good job at differentiating comparative advantages in final and in intermediate products.
- Fifth, that ARCA and LRCA provide an information that is close to the original Balassa's index.
- Finally, that the logarithmic and pseudo-logarithmic derivatives of the original Balassa's index result in very similar results that differ from the original BRCA specification.

In order to check for the robustness of this classification, let's put now a Data Scientist hat. [20] I used fuzzy k-means clustering to create the same number of clusters (5) than above, but adding some fuzziness at the beginning of the classification process in order to allow certain RCA indices located at the periphery of a group to belong at the same time to several different groups (soft clustering). The memberships Table 4 presents for each RCA index the group to which it has been eventually assigned. The latter one is calculated in a final step by choosing the group for which the index's membership probability is maximal.

Table 4 fuzzy k-means clustering of RCAs, 5 clusters

Cluster No	Size	Average silouhette	RCAs							
1	8	0.60	IG_NCAI_IO	IG_SNCAI2_IO	IG_NCAI	IG_SNCAI2	IG_BIT_IO	IG_SNCAI_IO	IG_SNCAI	IG_BIT
2	8	0.38	IG_ARCA_IO	FD_BRCA_IO	FD_BRCA	IG_ARCA	IG_BRCA	IG_BRCA_IO	IG_LRCA	IG_LRCA_IO
3	8	0.70	IG_LBRCA	IG_LBRCA_IO	FD_LBRCA	IG_BRCA2_IO	IG_BRCA2	FD_LBRCA_IO	FD_BRCA2_IO	FD_BRCA2
4	12	0.43	FD_BIT	FD_SNCAI2_IO	FD_BIT_IO	FD_SNCAI2	FD_LRCA_IO	FD_SNCAI	FD_NCAI_IO	FD_LRCA
4		.../...	Cont'd...	FD_NCAI	FD_ARCA	FD_SNCAI_IO	FD_ARCA_IO			
5	4	0.63	FD_NRCA_IO	FD_NRCA	IG_NRCA	IG_NRCA_IO				

Note: Clustering criterion: Cosine dissimilarity; 1600 results from 40 different RCAs in 2015 (RCAs from Rest of World region excluded). The average silhouette indicates the average degree of similarity of each observation with respect to its cluster.
Source: Based on OECD ICIO data

The results using k-means recoup those obtained with agglomerative clustering in Figure 3, but adds additional information. Custer 3, made of the modified Balassa's indices, shows the highest cohesion, with an average silhouette of 0.70. It is followed by the smaller group made of NRCA indices. Cluster 2, joining the original Balassa's BRCA with ARCA and LRCA has the lowest intra-group cohesion, with a silhouette of 0.38.

Increasing by one the number of clusters provides additional information (Table 5). Actually, I prefer this option even if it is not the best on a "pure" (*id est*, uninformed) statistical information criterion.

Table 5 K-means clustering of RCAs, 6 clusters

Cluster No	Size	Average silouhette	RCAs			
1	4	0.78	FD_ARCA_IO	FD_ARCA	FD_LRCA_IO	FD_LRCA

[20] Data scientists is a neologism used to design better-paid statisticians, usually working in the private sector. While using similar tools, their language differs. Statistical inference is called "machine learning" and exploratory data analysis is referred to as "unsupervised machine learning".

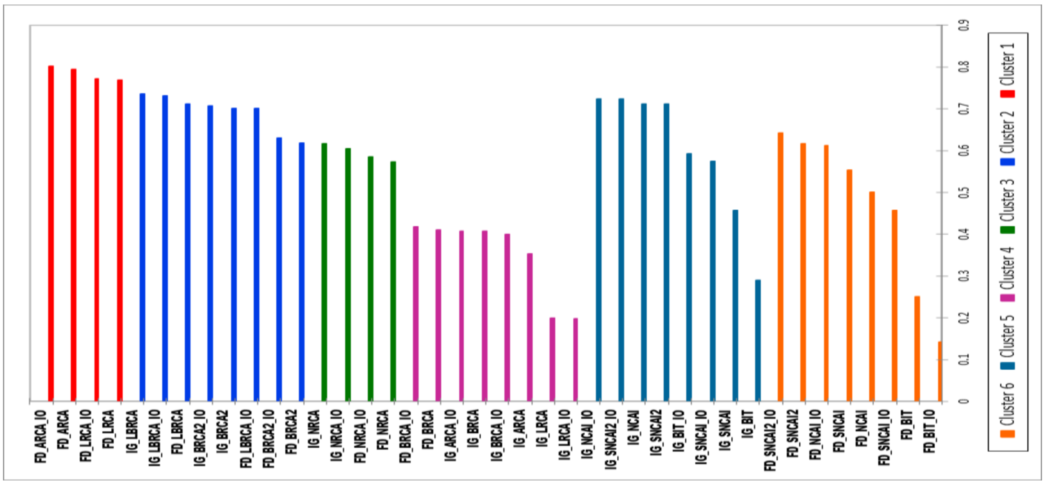
2	8	0.69	IG_LBRCA	IG_LBRCA_IO	FD_LBRCA	IG_BRCA2_IO	IG_BRCA2	FD_LBRCA_IO	FD_BRCA2_IO	FD_BRCA2
3	4	0.59	IG_NRCA	IG_NRCA_IO	FD_NRCA_IO	FD_NRCA				
4	8	0.35	FD_BRCA_IO	FD_BRCA	IG_ARCA_IO	IG_BRCA	IG_BRCA_IO	IG_ARCA	IG_LRCA	IG_LRCA_IO
5	8	0.60	IG_NCAI_IO	IG_SNCAI2_IO	IG_NCAI	IG_SNCAI2	IG_BIT_IO	IG_SNCAI_IO	IG_SNCAI	IG_BIT
6	8	0.47	FD_SNCAI2_IO	FD_SNCAI2	FD_NCAI_IO	FD_SNCAI	FD_NCAI	FD_SNCAI_IO	FD_BIT	FD_BIT_IO

Note: Clustering criterion: Cosine dissimilarity

Source: Based on OECD ICIO 2015 data

In my preferred option, the large cluster 4 in the previous Table 4 is split; it allows ARCA and LRCA indicators for final demand to regroup in a highly cohesive cluster (average silhouette at 0.78). The new cluster #2 inherits from the previous cluster 3 (modified BRCA indices), with a similar high cohesion. NRCA indices conform the new cluster 3, even if their internal cohesion is reduced relative to the other clusters (this can be interpreted as the apparition of a new group that shares some of the characteristics of the NRCA family of indices).The fifth group gathers a series of indicators calculated on intermediate products while the sixth one does the same for trade in final goods. In this new configuration, LRCA indices for intermediate products and BIT indicators for final goods have the lowest cohesion within their cluster (the silhouette plot in Figure 5 indicates the degree of similarity of each observation with respect to its cluster).

Figure 5 Fitness coefficient (silhouette) of k-mean clustering, 6 clusters.



Note: Clustering criterion: Cosine dissimilarity

d. Exploring RCAs' cross-correlations

Most Exploratory Data Analysis methods are based on some measure of distance (similarity or dissimilarity). Correlation is one of these measures. I use Pearson and Spearman correlations to evaluate the similarity of results between two sets of indicators: the RCA indices measured on trade in intermediates and those calculated on trade in final products. Pearson correlation compares absolute values, while Spearman correlation looks at the similarity in rankings. [21] The analysis is done first on aggregated data for 2015, then for each one of the tradable sectors.

Table 6 presents the results obtained when incorporating all sectors in the calculation of the Pearson coefficients. BRCA2 (gross trade and IO formulations) delivers similar values for trade in Final and trade in Intermediate products, with a correlation coefficient of 0.8 (upper shaded area). Except for this case, the

[21] The Spearman correlation coefficient is usually called "rho". It can take values from +1 to -1. A rho of +1 indicates a perfect association of ranks, a rho of zero indicates no association between ranks and a rho of -1 indicates a perfect negative association of ranks.

correlation (matrix diagonal) between the RCA calculated on Final and on Intermediate products is rather low, from 0.6 (BRCA and BRCA_IO) to -0.1 (NCAI_IO).

Very low correlations indicate that several indices (NCAI, SNCAI2, BIT) measures differentiated country specializations in intermediate and in final products. These indices seem therefore to perform better at indicating specificities than convergence, at least on this data set.

When looking at the results obtained at sectoral level (not shown here), this divergence appears very strongly for Textile and apparel, or for Motor vehicles. [22] While the ARCA. BRCA, and NRCA families of indices return Pearson coefficients at 0.9 and above for Textile, all the other ones (except SNCAI) are very low or slightly negative. Interestingly, the IO version of NCAI, SNCAI, SNCAI2 and BIT return a negative correlation, while the calculation on gross trade shows a low, but positive, correlation of 0.3: incorporating the GVC dimension changes the perception of comparative advantages in a non-insignificant way.

The contrast between gross and net is even clearer in the case of Motor vehicles. The NCAI, SNCAI2 and BIT indicators capture the difference between gross exports and exports net of imported inputs. For these indices, the correlation on gross trade is about 0.7 and drops at -0.1 when the input-output dimension is included in the calculation. But this is not a general pattern: for electronics, there is a convergence between intermediate and final goods country results for all RCA indicators.

Table 6 RCA for trade in intermediate and in final products: Pearson correlations

Final Demand / Inter-mediate	BRCA	BRCA_IO	BRCA2	BRCA2_IO	ARCA	ARCA_IO	LRCA	LRCA_IO	NRCA	NRCA_IO	NCAI	NCAI_IO	SNCAI	SNCAI_IO	SNCAI2	SNCAI2_IO	BIT	BIT_IO
BRCA	**0.6**	0.6	0.6	0.6	0.5	0.5	0.5	0.5	0.2	0.2	0.5	0.6	0.5	0.5	0.6	0.6	0.4	0.4
BRCA_IO	0.6	**0.6**	0.6	0.6	0.5	0.5	0.4	0.4	0.2	0.2	0.5	0.6	0.5	0.5	0.6	0.6	0.4	0.4
BRCA2	0.4	0.4	**0.8**	0.8	0.5	0.5	0.5	0.5	0.3	0.3	0.4	0.4	0.6	0.6	0.4	0.5	0.4	0.4
BRCA2_IO	0.5	0.5	0.8	**0.8**	0.5	0.6	0.5	0.5	0.3	0.3	0.4	0.5	0.6	0.6	0.5	0.5	0.4	0.4
ARCA	0.7	0.7	0.6	0.6	**0.6**	0.6	0.5	0.5	0.3	0.2	0.5	0.6	0.5	0.5	0.6	0.5	0.4	0.3
ARCA_IO	0.7	0.7	0.6	0.6	0.6	**0.6**	0.5	0.5	0.2	0.2	0.5	0.6	0.5	0.5	0.5	0.6	0.4	0.3
LRCA	0.6	0.6	0.5	0.5	0.4	0.4	**0.4**	0.4	0.2	0.2	0.3	0.4	0.4	0.4	0.3	0.4	0.2	0.3
LRCA_IO	0.6	0.6	0.4	0.5	0.4	0.4	0.4	**0.4**	0.2	0.2	0.3	0.4	0.4	0.4	0.3	0.4	0.2	0.3
NRCA	0.3	0.3	0.4	0.4	0.3	0.3	0.3	0.3	**0.6**	0.6	0.2	0.3	0.3	0.3	0.2	0.3	0.2	0.2
NRCA_IO	0.3	0.3	0.4	0.4	0.3	0.3	0.3	0.3	0.6	**0.6**	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
NCAI	0.4	0.4	0.3	0.3	0.2	0.2	0.2	0.3	0.1	0.1	**0.0**	0.3	0.3	0.4	0.2	0.4	0.2	0.3
NCAI_IO	0.1	0.2	0.1	0.1	-0.1	0.0	0.0	0.0	0.0	0.0	-0.4	**-0.1**	0.0	0.1	-0.3	0.0	-0.1	0.1
SNCAI	0.4	0.4	0.5	0.5	0.3	0.3	0.4	0.4	0.2	0.2	0.2	0.4	**0.5**	0.5	0.3	0.5	0.4	0.4
SNCAI_IO	0.3	0.3	0.4	0.4	0.2	0.2	0.3	0.3	0.1	0.1	0.0	0.2	0.3	**0.4**	0.1	0.3	0.2	0.3
SNCAI2	0.4	0.4	0.3	0.3	0.2	0.2	0.2	0.3	0.1	0.1	0.0	0.3	0.3	0.4	**0.2**	**0.4**	0.2	0.3
SNCAI2_IO	0.1	0.2	0.1	0.1	-0.1	0.0	0.0	0.0	0.0	0.0	-0.4	-0.1	0.0	0.1	**-0.3**	**0.0**	-0.1	0.1
BIT	0.7	0.7	0.3	0.4	0.3	0.3	0.3	0.3	0.1	0.1	0.2	0.4	0.3	0.4	0.3	0.4	**0.2**	**0.3**
BIT_IO	0.5	0.5	0.3	0.3	0.1	0.1	0.1	0.1	0.1	0.1	-0.1	0.2	0.2	0.3	0.0	0.2	**0.1**	**0.2**

Note: LBRCA not included for its similarity with BRCA2. The matrix is not symmetric and its diagonal differs from 1 because the underlying data for the indicators in row and column are not the same.
Source: Calculations based on OECD ICIO 2015 data

Comparative advantages being, well, comparative, Spearman rho correlations that look at ranks may have better theoretical foundations. As expected, the coefficients obtained (Table 7) are higher than for Pearson. There is much less discrepancy on ranking between the Intermediate and Final types of products.

[22] The comparison at product level does not make sense for all sectors: demand of basic metal for household consumption, for example is uncommon (final demand for this product refers mainly to investment).

As with Pearson, the Balassa's family of indices shows the highest consistency in ranking, above 0.80: if a country ranks high in final products for BRCA, it usually rnaks well on intermediate goods too. There is also some relationship between a low Pearson and a low Spearman, but the lowest rho on the diagonal is 0.4 while Pearson correlations could be negative: Spearman rhos equal or lower than 0.4 are probably not significant in the present context.

Table 7 RCA for trade in intermediate and in final products: Spearman Rho.

Final Demand / Inter-mediate	BRCA	BRCA_IO	ARCA	ARCA_IO	LRCA	LRCA_IO	NRCA	NRCA_IO	NCAI	NCAI_IO	SNCAI	SNCAI_IO	SNCAI2	SNCAI2_IO	BIT	BIT_IO
BRCA	0.8	0.8	0.7	0.7	0.5	0.5	0.6	0.6	0.6	0.5	0.6	0.6	0.6	0.5	0.6	0.5
BRCA_IO	0.8	0.8	0.7	0.7	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.5
ARCA	0.8	0.8	0.7	0.7	0.6	0.6	0.6	0.6	0.6	0.5	0.6	0.6	0.6	0.5	0.6	0.5
ARCA_IO	0.8	0.8	0.7	0.7	0.6	**0.6**	0.6	0.6	0.6	0.5	0.6	0.6	0.6	0.5	0.6	0.5
LRCA	0.5	0.5	0.5	0.5	0.5	0.5	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
LRCA_IO	0.5	0.5	0.5	0.5	0.5	0.5	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
NRCA	0.7	0.7	0.6	0.6	0.5	0.5	0.7	0.7	0.5	0.4	0.5	0.5	0.5	0.4	0.5	0.5
NRCA_IO	0.7	0.7	0.6	0.6	0.5	0.5	0.7	0.7	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
NCAI	0.5	0.5	0.4	0.4	0.5	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.4	0.4
NCAI_IO	0.4	0.4	0.3	0.3	0.4	0.4	0.3	0.3	0.4	0.4	0.4	0.5	0.4	0.4	0.4	0.4
SNCAI	0.5	0.5	0.4	0.4	0.5	0.5	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
SNCAI_IO	0.5	0.5	0.4	0.4	0.4	0.4	0.3	0.3	0.4	0.5	0.5	0.5	0.4	0.5	0.4	0.4
SNCAI2	0.5	0.5	0.4	0.4	0.5	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.4	0.4
SNCAI2_IO	0.4	0.4	0.3	0.3	0.4	0.4	0.3	0.3	0.4	0.4	0.4	0.5	0.4	0.4	0.4	0.4
BIT	0.5	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
BIT_IO	0.4	0.5	0.4	0.4	0.4	0.4	0.3	0.4	0.4	0.5	0.4	0.5	0.4	0.5	0.4	0.4

Note: LBRCA and BRCA2 provide the same ranking than BRCA and are excluded from the table. The matrix is not symmetric and its diagonal differs from 1 because the underlying data for the indicators in row and column are not the same.
Source: Calculations based on OECD ICIO 2015 data

When comparing RCA for final and for intermediate products at sectoral level, there is more homogeneity between the Pearson and the Spearman results. For example, in the case of Electrical equipment, the BRCA, ARCA and NRCA indices correlate at 0.9 and above for both indices, while the NCAI, SNCAI and BIT show similarly low values between 0.1 and 0.3 (see Table 8). Financial services are one of the few cases where Pearson correlation between comparative advantages in the services to firms and the services to households can be highly negative for some indices. But we are probably here on fragile statistical territory, for the difficulty in identifying bilateral flows of services and differentiating between intermediate and final transactions.

Table 8 Comparison of Pearson and Spearman correlations, selected sectors.

Sector	All		Textile		Vehicles		Electronics		Electrical equip.		Finance	
Correlation	**Pears.**	**Spear.**	**Pears.**	**Spear.**	**Pears.**	**Spear.**	**Pears.**	**Spear.**	**Pears.**	**Spear.**	**Pears.**	**Spear.**
BRCA	0.6	0.8	0.9	0.9	0.9	0.9	0.9	1.0	1.0	0.9	1.0	0.8
BRCA_IO	0.6	0.8	0.9	0.9	0.9	0.9	1.0	1.0	1.0	0.9	0.9	0.8
BRCA2	0.8	0.8	0.9	0.9	0.9	0.9	1.0	1.0	0.9	0.9	0.9	0.8
BRCA2_IO	0.8	0.8	0.9	0.9	0.9	0.9	1.0	1.0	0.9	0.9	0.9	0.8
ARCA	0.6	0.7	0.9	0.9	0.9	0.9	0.9	1.0	1.0	0.9	1.0	0.8
ARCA_IO	0.6	0.7	0.9	0.9	0.9	0.9	1.0	1.0	1.0	0.9	0.9	0.8
LRCA	0.4	0.5	0.1	0.4	0.6	0.5	0.8	0.7	0.6	0.4	0.1	0.0
LRCA_IO	0.4	0.5	0.1	0.4	0.6	0.5	0.8	0.6	0.5	0.4	-0.1	0.0
NRCA	0.6	0.7	1.0	0.9	0.9	0.8	0.9	0.9	1.0	0.9	1.0	0.8
NRCA_IO	0.6	0.7	1.0	0.9	0.9	0.8	0.9	0.9	1.0	0.8	1.0	0.7
NCAI	0.0	0.5	0.3	0.4	0.7	0.5	0.7	0.4	0.2	0.2	-0.9	0.0
NCAI_IO	-0.1	0.4	0.0	0.4	-0.2	0.2	0.7	0.4	0.1	0.2	-0.9	0.0
SNCAI	0.5	0.5	0.6	0.4	0.6	0.6	0.6	0.5	0.2	0.3	-0.6	0.1
SNCAI_IO	0.4	0.5	0.5	0.4	0.3	0.4	0.6	0.4	0.2	0.3	-0.7	0.1
SNCAI2	0.2	0.5	0.3	0.4	0.7	0.5	0.7	0.4	0.2	0.2	-1.0	0.0
SNCAI2_IO	0.0	0.4	0.0	0.4	-0.2	0.2	0.7	0.4	0.1	0.2	-0.9	0.0
BIT	0.2	0.5	0.3	0.4	0.7	0.5	0.7	0.4	0.2	0.2	-0.9	0.0
BIT_IO	0.2	0.4	0.0	0.4	-0.1	0.2	0.7	0.4	0.1	0.2	-0.9	0.0

Note: The table shows the correlation of a given index calculated on intermediate products with the same index calculated on final goods (diagonal of the correlation matrices).
Source: Calculations based on OECD ICIO 2015 data

Finally, I looked at country level results for each index and each sector from the dichotomic perspective. The procedure consists in listing for each sector and type of product (intermediate of final use) the top and bottom ten countries. In order to enter into one group or another, a country needs to be classified in the Top-10 or Bottom-10 for at least one RCA index. Once this is done, I proceed to checking if a country that appears in the Top-10 (or Bottom-10) for one index is also classified as such by other indices. The results (Table 9) indicate that there is usually more consistency across indicators in ranking the Top-10 than in ranking the Bottom-10. It is particularly true when the comparison is made with the other type of use (second panel). For example, in the case of Textile & apparel, a country classified in the Top-10 exporter by one RCA index will be also classified in the Top-10 in an average of 15.6 out of the 16 indices analysed (LBRCA – the logarithm of BRCA—and BRCA2 – its pseudo-logarithm– are excluded for their similarity with the ranking provided by BRCA).

Table 9 Top-Bottom comparison of classifications, selected sectors.

	Same use (final or intermediate)			Between types of use		
Number of concordances	Minimum	Maximum	Average	Minimum	Maximum	Average
TEXTILES			*Intermediate Products*			
Top 10	14.6	16	15.6	11.0	13.7	11.7
Bottom 10	4.8	8.2	7.1	2.0	6.1	4.0
			Final Demand			
Top 10	13.9	16.6	16.0	11.0	12.6	11.7
Bottom 10	6.0	9.9	8.4	2.8	4.4	4.0
ELECTRONICS			*Intermediate Products*			
Top 10	11.4	14	13.1	7.9	11.7	10.2
Bottom 10	5.2	9.6	8.0	2.5	7.4	4.3
			Final Demand			
Top 10	12.2	15.5	14.4	8.6	11.5	10.2
Bottom 10	2.3	12.4	10.4	3.2	5.0	4.3
ELECTRICAL EQUIP.			*Intermediate Products*			
Top 10	11.2	14.2	13.2	8.7	11.0	10.0
Bottom 10	4.3	8.7	7.7	1.3	7.2	3.4
			Final Demand			
Top 10	11.3	14.1	13.5	8.4	10.9	10.0
Bottom 10	2.8	10.1	8.5	1.3	5.6	3.4
VEHICLES			*Intermediate Products*			
Top 10	10.8	13.1	12.0	7.2	12.6	9.7
Bottom 10	5.1	10.7	9.0	2.5	8.3	4.5
			Final Demand			
Top 10	15.6	16.7	16.3	8.7	10.3	9.7
Bottom 10	2.9	11.1	9.2	2.5	5.4	4.5

Note: The sectors were selected for being representative of GVC trade, having important trade flows in both intermediate and final products. The table shows the number of times an exporter classified in the Top-10 (resp. Bottom-10) by a given index is also classified in the same Top/Bottom 10 by other RCA indices. The calculation is done twice: for the same type of products (intermediate or final use) or across types of products. The maximum number of similar occurrences is 18 (LBRCA and BRCA2 are excluded because they determine the same ranking than BRCA).
Source: Calculations based on OECD ICIO 2015 data

The higher mark (16 similar classifications out of a maximum of 16) is found for LRCA, NCAI, SNCAI, SNCAI2 and BIT for both their gross and input-output specifications. The minimum (14.6) corresponds to the BRCA and ARCA. The reasoning is the same for the Bottom-10 ranking. There is a clear difference in the stability of rankings between the Top and the Bottom 10 exporters according to the different RCA formulations. The difference is particularly wide for cross-classification. For the sector "Electrical equipment", an exporter classified in the Bottom-10 for intermediate products will appear on average only 3.4 times out of 16 times in the same Bottom-10 for final goods. The minimum (1.3) is found for SNCAI2 and BIT for both their gross and net exports. The maximum (11 similar classifications) is found for exporters classified in the Bottom-10 by BRCA_IO and ARCA_IO.

In general, there is more consistency:

- In the classification derived from the BRCA and ARCA families. This may be due to the fact that most RCA indices are derived more or less directly from Balassa's BRCA and generate similar rankings.
- In the classification of exporters in the Top-10 than in the Bottom-10 groups. A practical consequence would be to give more weight to the former when characterising the comparative advantages of an exporter.

A tentative conclusion is that instead of looking only at cardinalities or ordinalities in RCA indices, a more robust classification of comparative advantages would be based on dichotomies (yes/no) based on quantiles: belonging to the top 10% or 20% of the exporter for a "well behaved" empirical RCA index is a strong indication of having a comparative advantage for this particular product.

e. Dichotomous classification and the convergence of RCA indices

The analysis of dichotomous or qualitative indicators, usually inserted in econometric models as "dummy" variables, is relatively well covered by Statistics, in particular under the family of descriptive or predictive classification algorithms. Here, I am interested in knowing whether there is convergence or divergence between indicators in gross and in input-output terms when classifying countries in the Top-10 performers. The lower the number of discrepancies between indicators for a given industry, the more robust the Top-10 classification for this sector. Or, at the contrary, the more interesting it will be to contrast gross and input-output approaches when calculating RCAs as it brings additional information (variance). To interpret the results, it may help imagining that each RCA index is a member in a jury tasked with nominating the best 10 performers in a competition, all others being discarded into a "failed" category.

Table 10 Top-10 concordance and discrepancy between Gross and Input-Output RCA indices, selected sectors

Type of traded goods	Number of countries classified Top-10 [a]	Number of Discrepancies [b]	BRCA2	ARCA	LRCA	NRCA	NCAI	SNCAI	SNCAI2	BIT
Textiles, wearing apparel, leather and related products										
- Intermediate	13	0	0	0	0	0	0	0	0	0
-Final	14	12	2	2	0	2	2	0	2	2
Electrical equipment										
- Intermediate	18	24	2	2	2	2	4	4	4	4
-Final	19	20	2	2	2	6	2	2	2	2
Computer, electronic and optical products										
- Intermediate	17	32	2	2	4	0	6	6	6	6
-Final	16	24	2	2	2	2	4	4	4	4
Motor vehicles, trailers and semi-trailers										
- Intermediate	18	30	0	0	4	2	6	6	6	6
-Final	12	4	2	2	0	0	0	0	0	0
Total column:	...	***146***	***12***	***12***	***14***	***14***	***24***	***22***	***24***	***24***

Note: a/ The total is higher than 10 because a country needs to be classified at least by one RCA to enter the group. b/ Number of times the Gross and Input-Output specifications for an RCA index differed when classifying a country in the Top-10. BRCA and LBRCA are excluded because they determine the same ranking than BRCA2.
Source: Calculations based on OECD ICIO 2015 data

There is a relationship between the number of RCA candidates nominated for the Top-10 group and the number of discrepancies, even if it is not a strict one. A large number of discrepancies means that the indicators did not always converge on the same diagnostic (more on that in the next section). All the indicators showing more discrepancies between the Top-10 classifications generated by Gross and the IO specifications belong to the family of Two-Way trade RCAs. This conforms the previous results suggesting higher variance for these indicators.

The differences are also found across industries, and between trade in intermediate products and trade in final goods. For example, no discrepancy at all when the RCAs identify their Top-10 candidates for trade in intermediate goods in the textile industry, while the indicators show 12 cases of divergence for consumer goods (mainly apparel). [23] Computer, electronic & optical products (56 discrepancies in total) and Electrical equipment (44) are the sectors where more discrepancies were found. At the difference of Textile & apparel industry, the RCA jury was more uncertain when it dealt with trade in intermediate goods (the exporters specializing in the upstream part of the GVC) than in final products (the downstream segment).

Almost no discrepancies are observed when it comes to exporting vehicles in their final stage of production. The conditions for entering the competition are unambiguous and demanding on this market. Several countries collected all the top marks from the RCA jury: Czech Republic, Germany, Hungary, Japan, Korea, Mexico and Slovakia. The competition is both more open and diverse when it comes to exporting parts and components for motor vehicles. Only three competitors collected the top marks for all indicators: Germany, Hungary and Japan. Actually, there was a large number of discrepancies (30) between Gross and Input-Output RCAs. Most of them (24) were due to Two-Way trade indicators.

To conclude this section on the dichotomous approach to RCAs, I conducted a Descriptive Discriminant Analysis (DDA) to look at the contribution of RCA indicators in classifying a country in the Top-10 (in other words, my question is: are some members of the RCA jury better than others at identifying the winners?). DDA uses a classification function to “assign” each country to the group of Top-10 or not, and compare the result (posterior classification) with the observed one. For each of the four industries in Table 10, I applied a DDA on intermediate and final goods classification, separating each time the contribution of “gross” and “input-output” indicators. In addition to the RCA indices, I included also as covariates the relative net export index (RNX_k^i), the trade intensity of good “k” (RT_k^i) together with the variants SRT and SRT2; and the relative openness to trade of the “i” economy (RO_k^i) (see equation [21]).

Annex 2 provides more detailed information on the process and the results. I focus here on the most salient ones, deriving a few general conclusions:

- Firstly, high relative trade openness (RO) for a product does not lead to higher comparative advantages, especially when the use of imported inputs are taken into consideration.
- Secondly, the dichotomous approach to RCA is better done on two groups only (Best performers vs. All the rest) than on three groups (Top-Middle-Bottom) because the classification in the lowest group is not robust for most products.
- Thirdly, the classification in the Best Performer group (I used here the Top-10, but it could have been the first decile or quintile) varies from one RCA index to another one, each index measuring particular aspects. Even if there is a general convergence, there is also some divergence, especially when comparing One-Way trade and Two-Way trade indicators. This divergence is often amplified when the indicators are specified using the input-output relationships.

[23] The discrepancy occurred for Portugal (3 divergences), Indonesia (3), Bulgaria (2), Peru (2), Cambodia (1) and Romania (1).

- Finally, the classification based simply on ranking by individual RCAs can be enriched when considering them altogether plus some additional trade indicators, as I did in the DDA exercise (but one can use other multi-criteria data analysis techniques). A few countries that were not considered in a first instance by any individual RCA index were reclassified in the group of top performers. A personal inference from these results (an opinion to be confronted to hard data) is that the number of discrepancies between individual RCA indices and the number of reclassifications may indicate a more open competition for entering and upgrading in a given world trade market.

6. RCA indices and the evolution of comparative advantages 2005-2015

In this last review of the behaviour of alternative RCA indices, we will look at the evolution of countries' comparative advantages through time. I calculated the same set of RCAs on 2005, using the same OECD ICIO tables. Based on the conclusions from previous section, I focus on the Top-10 classification of countries for each product, which seems to provide more consistent results across indicators. Then, I compare the 2005 results with the 2015 ones.

There are two ways of looking at the results. One is more methodological and looks at stability of country classification of any given RCA index. The other one is analytical and aims at differentiating the sectors where rankings remain stable from those where changes occurred.

a. Looking at countries

Table 11 reviews for each index the number of times a country classified in the Top-10 in 2015 was also classified as such in 2005. Two indicators are calculated: the first one looks at stability for the same index, and ranges from 0 to 10 (10 means that all Top-10 in 2015 were also ranked in the Top-10 in 2005, but not always with the same raking). The other indicator reveals how often a country ranked in the Top-10 for one RCA index was also ranked Top-10 by other indices in 2005. This measure ranges from 0 to 160 (16 indicators, 10 ranking options for each indicator). For each indicator, its standard deviation is provided to inform on the variance of the mean across sectors.

Table 11 Coincidence being Top 10 in both 2015 and 2015, by RCA index

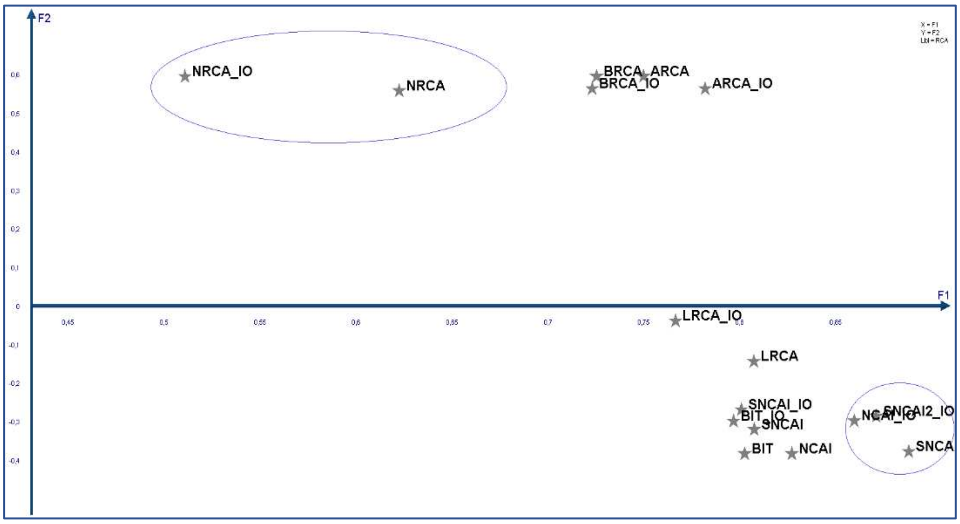
Index	BRCA	BRCA_IO	ARCA	ARCA_IO	LRCA	LRCA_IO	NRCA	NRCA_IO	NCAI	NCAI_IO	SNCAI	SNCAI_IO	SNCAI2	SNCAI2_IO	BIT	BIT_IO
Average same index	7.3	7.4	7.3	7.1	7.1	7.7	7.6	6.9	6.9	6.7	6.8	6.9	6.8	6.9	6.8	7.3
- Std. Dev.	1.4	1.4	1.4	1.4	1.3	1.2	1.3	1.3	1.4	1.3	1.2	1.3	1.4	1.3	1.4	1.4
Average All indices	89.3	98.2	99.4	97.4	99.2	92.2	92.5	99.7	99.4	96.2	98.6	100.6	100.4	100.4	100.6	89.3
- Std. Dev.	22.1	21.0	22.4	19.6	19.6	22.8	23.0	19.5	20.7	21.8	20.6	19.3	20.8	19.3	20.3	22.1

Note: Include trade in intermediate and final products. Log(BRCA) and BRCA2 not included as they duplicate the BRCA ranking.

The classifications as Top-10 by BRCA and BIT_IO for both 2005 and 2015 are the least stable when considering the average classification by all the RCA indices (89.3 out of 160 or 56% of the cases), even if they are within the average (7.3 out of 10) for the same index. Yet, it is difficult to extract workable information from the aggregated results. Looking for clues, let's apply Principal Component Analysis (PCA) to the disaggregated data.

When looking only at number of coincidences in 2005 and 2015 for the same RCA indicator calculated at sectoral level for trade in intermediate and final products (corresponding to the first average results in Table 11), we find a distinction (albeit not a large one) between the "one-way trade" and the "two-way trade" indicators. It is clear on Figure 6: The vertical axis "explains" only 19% of the total variance, but shows the clear dichotomy existing between the One-Way Trade indices (above the horizontal line) and the Two-Way Trade indices (below the horizontal line). [24] This should not be surprising: while One-Way indices vary only in function of exports, Two-Way have two sources of variations: imports and exports (BIT includes also GDP variations). If one wants to capture both aspects, Figure 6 suggests to choose NRCA and SNCAI formulations (gross or net for foreign inputs) in order to have the most contrasting options. This choice is valid only for the current data, and may change for other data sets.

Figure 6 PCA on same RCA coincidence being Top 10 in both 2015 and 2015



Note: Based on results for sectoral trade in intermediate and final products (50 observations for each RCA index). Log(BRCA) and BRCA2 not included as they duplicate the BRCA ranking.

a. Looking at sectors

The other way of looking at the results is to identify the sectors where the comparative advantages remained stable during the ten years 2005-2015 from those that registered variations in their Top-10 contenders. Table 12 below shows the number of time countries appearing in the Top-10 RCAs in 2015 were similarly classified in 2005. As before, there are two ways of qualifying stability: no change for the same RCA index (a country that was present in 2015 is also in the 2005 Top-10) or stability within the wider set of RCAs (a country was present for at least a RCA in 2015 appears also in one of the RCA indices in 2005).

The sectors showing RCA stability between 2005 and 2015 are mainly industries relying on natural resources (extractive activities, food and wood products) or heavy industries (chemicals, basic metals). Two manufacturing activities that are associated with different stage of GVC industrialisation are also in this group: Textile and apparel (an entry position with relatively simple GVC) and Motor vehicles (requiring

[24] In order to interpret the first component on the horizontal axis, one has to look at the observations (individual sectors) and not at the RCAs. Stable sectors (where Top-10 composition does not change much between 2005 and 2015) are concentrated on the right-hand side of the graph, while least stables one are on the left (see Table 12).

complex GVC arrangements). International finance for final consumers is the sole services sector showing stability.

On the least stable sectors, we find several services sectors. A statistical effect cannot be excluded, as these products are mainly produced for the domestic market and exports are marginal, with important year to year variations. Surprisingly, "Other non-metallic mineral products" (The sector transforming mineral raw materials into other non-metallic mineral products for use, among others, by the construction industry, the food and beverages sector, or households in the form of consumer durables) is also classified in this category for both intermediate and final demand. When looking at country composition, this is due to the raise of countries like Bulgaria, Malaysia or Vietnam in the Top-10 ranking for this industry.

Table 12 Stability in the composition of Top-10 most competitive exporters, 2005-2015

Top-15 most stable sectors				Top-15 least stable sectors			
Average coincidence [a]				*Average coincidence [a]*			
Same RCA	*All RCA (%)*	*Exports for :*	*Sector*	*Same RCA*	*All RCA (%)*	*Exports for :*	*Sector*
8.9	83.8	**FD**	Mining and extraction of energy producing products	4.5	36.3	**IG**	Mining support service activities
8.8	79.1	**FD**	Chemicals and pharmaceutical products	4.9	43.2	**FD**	Mining support service activities
8.6	85.2	**IG**	Mining and quarrying of non-energy producing products	4.9	37.9	**IG**	Telecommunications
8.6	75.8	**IG**	Textiles, wearing apparel, leather and related products	5.6	50.8	**FD**	IT and other information services
8.4	80.5	**IG**	Mining and extraction of energy producing products	5.9	50.8	**IG**	Coke and refined petroleum products
8.4	77.2	**FD**	Mining and quarrying of non-energy producing products	5.9	53.5	**IG**	Financial and insurance activities
8.4	68.8	**IG**	Paper products and printing	6.1	54.2	**FD**	Rubber and plastic products
8.2	62.7	**FD**	Food products, beverages and tobacco	6.1	48.2	**FD**	Telecommunications
8.1	73.6	**FD**	Textiles, wearing apparel, leather and related products	6.1	55.7	**IG**	IT and other information services
8.1	62.9	**IG**	Wood and products of wood and cork	6.2	54.5	**IG**	Other non-metallic mineral products
8.1	81.0	**FD**	Financial and insurance activities	6.3	42.7	**IG**	Rubber and plastic products
8.1	70.4	**FD**	Other manufacturing; repair and installation of machinery and equipment	6.3	50.1	**FD**	Other non-metallic mineral products
7.9	73.3	**FD**	Motor vehicles, trailers and semi-trailers	6.3	53.8	**IG**	Other manufacturing; repair and installation of machinery and equipment
7.8	59.5	**IG**	Basic metals	6.4	54.2	**FD**	Electrical equipment
7.8	63.9	**IG**	Machinery and equipment, nec	6.5	48.1	**FD**	Other business sector services

Note: a/ number of times a country classified in the Top-10 in 2015 was also classified as such in 2005; same RCAT: by the same index, All RCA: by same or other indices in 2005. Sectors are ranked by increasing or decreasing coincidence for the same RCA index.

Before we part company, I would like to add another word of caution when comparing, as I did, the results of RCA indices at two separate points in time. All the basic data used for the calculations are in nominal USD. If one believes in the law of one price, it should not be a big issue unless the products, as it is the case here, are highly aggregated: all similar individual products are sold and bought at the same price on the international market. Yet, many things can go wrong: firms may have different pricing schedules according to the countries of final destination. Even if firms do not price to market, prices may diverge with exchange rates when exporters and importers belong to a large currency area, such as the Euro zone.

The issue is amplified when the "product class " is a large set aggregating many individual goods, as is the case here. Through time, there may be large variations in the relative price of individual products belonging to the same aggregate (for example, the price of textile vs. the price of apparel). The last point is even more relevant when a Two-Way trade RCA index is used, because the price of the export flow relative to the unit cost of imports may change significantly, affecting the results.

7. Conclusions

A single unified statistical concept – deviation of observed trade from the expected one– for measuring revealed comparative advantage a key that opened the door to many different practical formulations. The empirical application showed that One-Way and Two-Way trade indices capture different aspects of trade competitiveness, all indices having their own merits. The availability of new database providing harmonized trade and production data increases the depth of analysis by being able to factor-in the use of imported inputs in the production of exports.

The new indicators proposed in the paper extend the analytical power of traditional indicators by focusing on the domestic contribution to competitiveness and neutralising the effect of imported inputs. Taking into consideration imported inputs was particularly relevant for the index SNCAI, proposed by Gnidchenko and Salnikov (2015), especially when calculated for the intermediate products and for BIT (Bowen, 1983) for both intermediate and final goods. Incorporating the GVC dimension changes the perception of comparative advantages for several industries; the contrast between traditional measure (gross exports) and net (domestic share only) was particularly clear in the case of Motor vehicles.

Incorporating the input-output dimension allows also to analyse separately the exports of intermediate products used by the importing industries to produce a new output and the exports of final goods and services aimed at satisfying final demand in the importing country. This introduces a new dimension in the definition of competitive advantages: upstream or downstream competitivity. A country can be competitive in exporting one category of products but not the other one. A typical example is the Textile and Apparel industry, Textile, mainly used as input, is upstream and capital intensive, while Apparel is downstream (close to final demand) and labour intensive.

The quest for a perfect indicator failed. Each category of indicators provides relevant information but suffers from statistical or conceptual biases. The review of the empirical properties of each one of them shed some light on the best use that can be done of them. In a few words, the findings are: don't rely on a single indicator and stay clear of absolute (cardinal) values.

First, One-Way and Two-Way trade indices capture different aspects of trade competitiveness. There is no good reason to privilege one over the other. Contrasting traditional indicators based on gross trade statistics with those incorporating the input-output dimension may help understanding the source of trade competitiveness. This is in particular important when separating trade in intermediate inputs from trade in final products: some RCAs perform better than others in spotting specificities.

Second, comparative advantages are, well, comparative. It means the trade analysts will be on a firmer theoretical ground using ordinal classification than the absolute cardinal result provided by any index you decide to rely on. The results obtained here show that the classification is more robust for the cases where strong competitiveness is observed than for the lowest ranked cases. Better than ordinal classification, thus, they should opt for dichotomy based on quantile distribution: for example, top 20% vs. rest of traders.

Additional caution must be exerted when doing historical comparisons or using absolute RCAs for normative purpose. The data and the rankings may reflect changes in relative prices and in exchange rates. The bias can be somewhat reduced if the product category is disaggregated and One-Way trade indicators are used. When the exporters belong to a deeply integrated non-dollar currency area, extra regional trade flows should be preferred as they reflect better international prices.

The input-output approach suggested in the paper is limited by the coverage of existing databases and their product aggregation. The methodology may, nevertheless, be applied to disaggregated trade data such as COMTRADE by assuming that all exported products belonging to a given industrial sector share the same Leontief input-output technology. Relevant input-output data may be taken directly from national tables, which usually present a much more detained disaggregation than the international ones.

8. Bibliography

Amano, A. (1966) 'Intermediate Goods and the Theory of Comparative Advantage: A Two-Country, Three-Commodity Case' Weltwirtschaftliches Archiv, 96, 340-345

Anderson, J (1979), "A theoretical foundation for the gravity equation", American Economic Review 69(1), pp. 106-116.

Anderson, J and E van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle", American Economic Review 93(1), pp. 170-192.

Andrey A. Gnidchenko, A. and V. Salnikov (2015) 'Net Comparative Advantage Index: Overcoming The Drawbacks of The Existing Indices' Working Papers WP BRP 119/EC/2015, Moscow

Balassa B. (1965). Trade Liberalization and Revealed Comparative Advantage. The Manchester School, 33(2), 99-123

Ballance, R., H. Forstner and T. Murray (1987) 'Consistency Tests of Alternative Measures of Comparative Advantage', The Review of Economics and Statistics, Vol. 69, No. 1, pp. 157-161

Bebek, U.G (2017) 'RCA: Choosing the Right Measure', Department of Economics, University of Birmingham

Borin, A and M. Mancini (2019) 'Measuring What Matters in Global Value Chains and Value-Added Trade' Policy Research Paper 8804, World Bank

Bowen, H.P (1985) 'On measuring comparative advantage: A reply and extension' Weltwirtschaftliches Archiv 121, 351–354

Carrère, C., M. Mrazova and P. Neary (2020) 'Gravity Without Apology: The Science of Elasticities, Distance, and Trade' CESifo Working Paper No. 8160

Choi, N. and S. Park (2016) 'Comparative Advantage of Value Added in Exports: The Role of Offshoring and Transaction Costs',KIEP Working Paper 16-09

Dalum, B., K. Laursen and G. Villumsen (1998) 'Structural change in OECD export specialisation patterns: de-specialisation and stickiness' International Review of Applied Economics, 12, 447–467

De Benedictis, L. and M. Tamberi (2001) 'A note on the Balassa Index of Revealed Comparative Advantage' Mimeo SSRN: https://ssrn.com/abstract=289602

Deardorff, A.V. (2005) 'Ricardian Comparative Advantage with Intermediate Inputs', North American Journal of Economics and Finance, 16, March p. 11.34

Deb, K. and P. Basu (2011) 'Indices of Revealed Comparative Advantage and Their Consistency with the Heckscher-Ohlin Theory' Foreign Trade Review, 46(3):3-28.

Deb, K. and W. Hauk (2015) 'RCA Indices, Multinational Production and the Ricardian Trade Model' International Economics and Economic Policy 14, 1–25

Eaton, J. and S. Kortum (2002) 'Technology, Geography, and Trade' Econometrica 70 (5), pp.1741-1779

Gnidchenko, A. and V. Salnikov (2015) 'Net Comparative Advantage Index: Overcoming the Drawbacks of the Existing Indices' WP BRP 119/EC/2015, National Research University Higher School of Economics, Moscow

Hidalgo, C. A., Klinger, B., Baraba´si, A.-L. & Hausmann, R. (2007) 'The Product Space Conditions the Development of Nations' Science 317, pp. 482–487

Hinloopen, J. and C. Van Marrewijk (2001) 'On the Empirical Distribution of the Balassa Index', Weltwirtschaftliches Archiv 137(1):1-35

Hummels, D., J. Ishii and K. Yi (2001) 'The Nature and Growth of Vertical Specialisation in World Trade' Journal of International Economics, Vol. 54, pp. 75-96

Jones, L., M. Demirkaya and E. Bethmann (2019) 'Global Value Chain Analysis: Concepts and Approaches' Journal of International Commerce and Economics, April pp 1-29.

Koopman, R., Z. Wang, and S.-J. Wei (2014) 'Tracing Value Added and Double Counting in Gross Exports' American Economic Review 104(2), pp. 459-494

Kunimoto, K. (1977) 'Typology of Trade Intensity Indices Hitotsubashi Journal of Economics, Vol. 17, pp. 15-32

Lafay, G. (1992) 'The measurement of revealed comparative advantages' In: Dagenais, M. and P. Muet (eds.) "International Trade Modelling", London: Chapman & Hall

Leamer, E. (1983) 'Let's Take the Con Out of Econometrics' The American Economic Review Vol. 73, No. 1 pp. 31-43

Leromain E., G. Orefice (2013) 'New Revealed Comparative Advantage Index: Dataset and Empirical Distribution' CEPII Working Paper No 2013-20

Low, P. (2013) 'The role of services in global value chains' in D.K Elms and P. Low (eds) "Global Value Chains in a Changing World", Geneva, WTO/Fung Global Institute, pp. 61–81.

Miroudot, S. and C. Cadestin (2017) 'SErvices in Global Value Chains: Trade patterns and gains from specialisation", OECD Trade Policy Paper No. 208.

Miroudot, S. and M. Ye (2018) 'A simple and accurate method to calculate domestic and foreign value-added in gross exports' MPRA Paper 89907

Quast, B. and V. Kummrit (2015) 'Decompr: Global Value Chain Decomposition in R', CTEI PAPERS CTEI-2015-01, Graduate Institute Geneva

Sanidas, E. and Y. Shin (2010) 'Comparison of Revealed Comparative Advantage Indices with Application to Trade Tendencies of East Asian Countries' 9th Korea and the World Economy Conference, Incheon.

Shiozawa, Y. (2017) 'The New Theory of International Values: an Overview'' in Shiozawa, Y., Oka, T. and Tabuchi, O. (Eds) 'A New Construction of Ricardian Theory of International Values: Analytical and Historical Approach', Springer, Singapore, pp. 3-73

WTO and IDE-JETRO (2011) 'Trade patterns and global value chains in East Asia: From trade in goods to trade in tasks', Geneva.

Yamazawa, I. (1970) 'Intensity Analysis of World Trade Flow', Hitotsubashi Journal of Economics, 10(2) pp.61-90

9. Annexes

Annex 1. Other methodological approaches

1) Theoretically-consistent RCA

Most of the RCA indices that have been reviewed in the paper try to overcome some or all of the empirical weaknesses of the Balassa (1965) revealed comparative advantage have been based on observed ex-post trade flows. As occurred with the gravity model, an empirical approach which received micro-foundations (Anderson and van Wincoop, 2001), there were also attempts at developing new theoretically-consistent indicators of comparative advantages. It is interesting to compare the statistical method adopted in the main text with these alternative attempts.

The issue is complex because the standard micro-economic approach is expected to include both demand and supply functions, their convergence to equilibrium being driven by price. Indeed, most RCA indices implicitly or explicitly rely on homothetic demand functions and the "law of one-price". As Bowen (1985) concluded, the identical country/homothetic preference assumption is an imperfect but practical option in absence of a better alternative.

Taking aside the issue of heterogeneity on the demand-side, CES demand functions allow to focus on the production side and on firms' heterogeneity. This is by itself a formidable task, even from a purely empirical perspective. An option is to split the problem and separate the analysis of production and the analysis of trade, as in Eaton and Kortum (2002). In a nicely written paper, the authors develop and quantify a Ricardian model of international trade that incorporates differences in technology and in trade costs.

Encasing their formal analysis into a Ricardian approach, the authors examine how technology and geography (read: "trade costs") determine patterns of specialisation in a continuum of products. In the process, they look at the impact of trade in intermediate inputs, one of the main new characteristics of the early 21st century trade. At the difference of the empirical RCAs, the construction of Ricardian comparative advantage starts with a model of technology, prices and trade flows. Prices are defined on the basis of production and trade costs. Production costs, in turn, are related to labour inputs and technology. Eaton and Kortum (2002) assume perfect competition, so that buyers in a given country "j" will shop around for the best deal and buy from country "i" the cheapest product satisfying their requirements.

Without entering here into the technical details, one of the key empirical parameters is the representation of technologies and efficiencies. While technologies are exogeneous, each particular country has a country-specific probability distribution of these technological options that it can efficiently put at work. The likelihood that country "i" supplies a particular good k to country "j" is that the price offered by "i" on the "j" market (including trade costs) is the lowest. The convenient statistical hypothesis here is that the efficiency distribution follows some kind of extreme value distribution called "Fréchet". [25] Fréchet is built on two parameters, one called "T_i" defines country's "i" state of technology; a higher "T_i" implies that high efficiency technologies are more probable. The other parameter is "θ", common to all countries, and reflect the variation within the Fréchet distribution. A larger "θ" implies less variability.

The price parameter is also crucial in Eaton and Kortum (2002) model. It conveys information about the state of technology and the costs of inputs in the world as well as the trade costs affecting bilateral trade. In absence of trade costs, the law of one price holds for each good; this price being defined by the most efficient producer. We find here the same results obtained in Neo-Ricardian models "à la Shiozawa (2017)". Because the model includes also trade in intermediate goods, the existence of trade costs affects also the price of inputs in each country as well as final production cost (sum of labour and intermediate inputs). Trade shares respond to costs at the extensive margin: When a supplier becomes more expensive for some of its trade partners due to production or trade costs, it exports a narrower range of goods. [26]

An important parameter for empirical application is "θ", common to all countries, which reflects the technological variation within the Fréchet distribution. The lower "θ", the more heterogenous productivity across countries. The authors exclude therefore agriculture and mining activities from their analysis, to concentrate on manufacture industries that are expected to show more homogeneity in technology. Estimating "θ" requires micro data and is very dependent on the model specification and the regression method applied. Testing various options for data and methodology, their estimates range from a 3.60 obtained with 2SLS method using wage to 12.86 using 2SLS on prices. Their preferred estimate of "θ = 8.28" is obtained using the method of moments, which "lies very much in the middle of the range of estimates we obtain from our alternative approaches" (page 1765).

Costinot, Donaldson and Komunjer (2012) build on Eaton and Kortum (2002) to estimate comparative advantages, but eventually select "θ = 6.53" as their preferred estimate on the basis of openness corrected exports. Their estimation uses producer prices conditional to R&D expenditures and exports, using data from 21 countries and 13 manufacturing sectors (here again, natural resources-based activities like agriculture and mining are excluded). Even within the same sample, estimates vary from "θ = 4.62" (EU members only) to "θ = 8.06" (producer price data above median quality). Then Costinot et al. (2012) proceed to estimate "revealed" measures of productivity for country "i" and industry "k" based on observed bilateral trade flows. The estimate is retrieved using a fixed-effect regression.

$$\ln X_k^{i,j} = \delta_{i,j} + \delta_{j,k} + \delta_{i,k} + \varepsilon_{i,j,k} \qquad [34]$$

[25] This type of distribution is often used to model the maxima of random variables in large samples.

[26] Eaton and Kortum (2002) page 1750 point that this adjustment at the extensive margin is a major difference with the Armington models where the adjustments are done at the intensive margin.

Where $\delta_{i,j}$ is country-pair fixed effect, $\delta_{j,k}$ is an importer-industry fixed effect and $\varepsilon_{i,j,k}$ the error term. The technological differences $z_{i,k}$ are captured through the exporter-fixed effect $\delta_{i,k}$ and deducted from the theoretical model and the parameter "θ" in the following way:

$$z_{i,k} = e^{(\frac{\delta_{i,k}}{\theta})} \quad [35]$$

According to Costinot et al. (2012), $z_{i,k}$ is a good proxy of the comparative advantage because it can be considered as the part of the trade flow that is only due to the intrinsic productivity level of a given industry "k" in a country "i".

Leromain and Orefice (2013) adapt Costinot et al. (2012) to build a set of "synthetic" empirical measures for Ricardian comparative advantages. To characterise the empirical distributional properties of their results, they compare the "synthetic" (also referred to as "structural") RCA (SRCA) with Balassa's formula for 20 developed and developing countries over the 1995-2010 period at two different product disaggregation levels, excluding agriculture and mineral fuels that depend on natural resources endowments. They show that the SRCA has the desired symmetric thin tail distribution and is not subject to the BRCA's size effect. While the mean value of BRCA indices fluctuate over the 1995-2010 period, the structural results are more stable and the mean value of SRCA can be considered being stationary.

On the minus side, the index is theoretically grounded for the manufacturing sector only. Its relevance to the monopolistic competition that characterises complex industrial products is limited by the choice of CES demand functions (Carrère et al., 2020). On the empirical side, its application remains dependent on the choice of the technical parameter "θ" for intra-industry heterogeneity. This choice remains largely arbitrary as it is conditioned by the micro-datasets and the estimation method used to calibrate the parameter. The construction of the structural SRCA is cumbersome and requires implementing a large number of regressions (one for each country-sector duple). The task becomes almost impossible when applying the methodology to trade in goods at high level of disaggregation due to the large number of producer-product fixed effects to include in the model specification. Leromain and Orefice (2013) propose an alternative strategy to reduce the number of fixed effects, at the cost of running another set of regressions similar to equation [34].

2) Value-Added-based structural measures

The trade in value-added approach is in-between the above mentioned micro-economic approach and the macro approach adopted by the trade approach to RCAs we revised in the paper. As we mentioned page 11, the availability of international input-output tables such as Figure 2 provides new opportunities to measure and analyse comparative advantages from the trade and production venue. In the paper, we chose to follow the traditional trade statistics approach and write down our input-output based RCAs in gross terms. These new RCA indicators use international input-output data to re-parameterize trade indicators using information on inter-industrial trade in intermediate products.

But the modelling of RCAs can also be done entirely in value-added terms, substituting the gross value of inputs in the $\mathbf{Z}$ matrices of Figure 2 by their value-added content by country and sector of origin. A very general definition of value-added exports is the value-added that is produced in a country but is used in another country, either for final demand or for producing exports. The domestic value-added embodied in exports includes the contribution of the exporting sector itself, but also the supply of value-added from of other sectors of the national economy that contributed to the domestic value chain. This is a simple definition, and the specialised literature refine it by differentiating the value-added that is absorbed by the importing country from the value-added that is reexported in intermediate or final goods, or that returns home through direct or indirect re-imports.

This strand of empirical research is directly associated to the analysis of trade along Global Value Chains (GVCs), also known as "Trade in Value-Added". It is closely associated with new dimensions in trade statistics, following the concept of Vertical Specialization (Hummels, Ishii and Yi, 2001). The first application using official data was published in 2011 by WTO and IDE-JETRO, with an application on Eastern Asia. It was also the guiding methodology used by the OECD-WTO Trade in Value-Added (TiVA) database in 2012.

Since the availability of consistent world-wide databases, a growing literature has looked at the wedge between traditional trade statistics and trade in value-added.[27] The basic idea behind trade in value-added is that individual economies participate in global value chains by importing foreign inputs to produce the goods and services they export (backward GVC participation) and also by exporting domestically produced inputs to partners in charge of downstream production stages (forward GVC participation). The value-added decomposition of trade starts with the so-called Leontief model. Following the notation used in this paper, the model starts with the following identity:

$$\boldsymbol{Q} \equiv \boldsymbol{A} \cdot \boldsymbol{Q} + \boldsymbol{FD} \qquad [36]$$

where:

$\boldsymbol{Q}$: is an n.k*1 vector of the output of k industries within an economy of n countries.

$\boldsymbol{A}$: is the n.k*n.k matrix of technical coefficients describing the interrelationships between industries; with, for each product "k" $a^{ij} = \boldsymbol{Z}^{ij} / \boldsymbol{Q}^{j}$ the ratio of inputs from country "i" used for the production of the output of country "j".

$\boldsymbol{FD}$: is an n.k*1 vector of final demand for goods and services.

(Note: Matrix and vectors are in bold characters)

The contribution of exports to the countries' GDP is equal to:

$$\boldsymbol{VA} = \boldsymbol{v} \cdot (\boldsymbol{I} - \boldsymbol{A})^{-1} \cdot \boldsymbol{e} \qquad [37]$$

where:

$\boldsymbol{v}$: is a 1 x n.k vector components m_j (ratio of value-added to output in industry j)

$\boldsymbol{I}$: is an n.k x n.k identity matrix.

$\boldsymbol{e}$: is a n.k x 1 vector of gross exports by industry.

Equation [37] provides the basis for measuring the domestic value-added content of traded products, but the actual computation is more complex. A key component in most calculations is the Value-Added Multiplier $\mathbf{VB} = \mathbf{V}\ (\mathbf{I}\text{-}\mathbf{A})^{-1}$ where $\mathbf{V}$ is a diagonal matrix, the diagonal representing the direct value-added contribution of each industry in each country. $\mathbf{VB}$ is a n.k x n.k matrix. Two types of domestic contents can be calculated: the domestic content of national exports (similar in its idea to the approach used in equation [26]) and the domestic content found in other countries exports, which corresponds to exports of domestic inputs that were processed and embodied into the exports of third countries.

A contentious issue remains the proper calculation of "double counting". Gross exports include some 'double counting' in the sense that the same value-added (already defined as domestic or foreign) is counted twice or more. Therefore, a simple Leontief decomposition as in Equation [37] will not properly decompose the gross value of exports into the sum of the domestic and foreign value-added (Miroudot and Ye, 2018). This "Leontief decomposition" approach has been further refined by Koopman, Wang and Wei (2014) who decompose GVC trade into several trade in value-added indicators. Wang, Wei and Zhu (2013) extend the information contained in inter-country input-output tables to decompose GVC trade and derive additional indicators. Borin and Mancini (2019) present an overview of the various measures suggested in the literature and propose a solution to the double-account issues.

Input-output tables are balanced by construction, and total supply equals total demand. Rather than measuring the domestic content in exports (a supply-side perspective), it is also possible to identify the national and sectoral origin of the value-added embodied in the domestic and imported products absorbed to satisfy final demand (Value Added embodied in trade). Thus, there are two possible perspective: the "source-based" measure the value-added where it originates (the supply side perspective) while the

[27] See Jones, Demirkaya and Bethmann (2019) for a comprehensive review of the applications of this concept to trade analysis in the business and economics literature.

alternative, called the "sink-based" approach measures it from the perspective of the country that ultimately absorbs it in its final demand (Meng, Fang and Yamano, 2012).

In their comprehensive review of the different methodologies used to measure trade in value-added and the various accounting issues faced, Borin and Mancini (2019) conclude that the source-based approach is better suited to analyse exports and production linkages, while the sink-based perspective illustrates the role of a country's final demand in activating world-wide production and trade flows, such as, for instance, in an analysis of bilateral trade balances.

I could add to this that the source-based approach is closer to the trade analyst's preoccupation. Trade in value-added is also called "trade in tasks" and relates to the contributions each country's firms do to an international supply chain by contributing their own intermediate inputs. This supply-side approach can be easily related to the business and microeconomic perspectives: when optimizing its supply chain, a Lead Firm will subcontract tasks (R&D, specific parts and components, assembly) to the most efficient contractors. The decision to sub-contract tasks to firms in one country rather than in another would be based on various considerations, including production and trade costs, suppliers' reputation and country's business environment, etc. The sink-based measure is also of interest for trade analysts because it relates to the issue of preferential treatment in regional trade agreements, when preferences are conditional to the origin of the value-added embodied in the finished goods.

From a practical perspective, a researcher may simply use the trade in value-added indicators built-in the OECD TiVA database, which includes various source-based and sink-based measures. WTO's "Trade in value-added and global value chains: statistical profiles" are more synthetical and provide aggregated information on the value-added content in an economy's exports, its participation in global value chains and the contribution of services to the value-added content of exports (https://www.wto.org/english/res_e/statis_e/miwi_e/countryprofiles_e.htm). Another option is the GVC Indicator Database (UIBE, Beijing, China) at http://rigvc.uibe.edu.cn/english/D_E/database_database/index.htm. This research database covers the most widely used Trade in Value Added and GVC indicators, calculated on several databases. The UNCTAD-Eora portal (https://worldmrio.com/unctadgvc/)presents also a series of key GVC indicators derived from the Eora data used in this paper. A middle way, for those willing to build their own indicators without dealing with the intricate matrix calculus is to use a dedicated program, such as Quast and Kummrit (2015), which is based on the "R" language and calculate the Leontief and the Wang, Wei and Zhu (2013) decompositions.

Because of the increasing fragmentation of production in the 1990s and 2000s, it is to be expected a divergence between indicators based on gross trade statistics with those that account only for the domestic contribution to the final value of the traded product. Choi and Spark (2016) conclude their analysis of trade in value-added stating that RCAs have greater influence on determining the patterns of trade in a world with global value chains.

Another important contribution of the value-added approach is to provide a much better idea of RCAs in trade in services. The role of services is probably the most important information revealed by the measure of Trade in Value-Added (Low, 2013). The value-added method reveals the real contribution of services in trade because many services are not exported directly through cross-border trade but are embodied in exported goods. Additionally, because services are produced with fewer foreign inputs, their domestic content is generally higher than manufacturing goods once the value of foreign intermediate inputs is removed. Interestingly, international competitiveness of goods is enhanced by the incorporation of services, from the most upstream R&D to the most downstream after-sale services. For example, Miroudot and Cadestin (2017) show that many countries that are specialised in exports of manufacturing goods rely for their comparative advantages on a large range of services activities.

Annex 2. Results from the Discriminant Data Analysis

The main interest here is to see if some countries are classified differently when using the more comprehensive DDA compared to their ranking based only on RCAs. A large number of divergences would indicate that the frontier between high-RCA exporters and the other exporters is fuzzy and that there is more than one way for identifying competitiveness. My prior here is that the more complex the production process and the product varieties, the fuzzier the classification.

The first information a DDA provides is the mean of each "explicative" variable by class. The results obtained for the RCA indices is almost tautologic: The Top-10 countries are always in the highest range of the RCA indicator. Similarly, net export index (RNX) and the trade intensity of the good "k" (RT and its variants) are in general higher for the countries classified in the Top-10 group for this industry. At the contrary, the relative openness (RO) does not seem to influence the classification: Top-10 countries are usually more open to trade, but the difference is marginal. It is even 0 in the case of final textile & apparel products when RO is calculated in its Input-Output specification.

DDA uses a classification function to "assign" each country to the group of Top-10 or not, and compare the result (posterior classification) with the observed one. The classification is performed by comparing the probability of belonging to each of the classes (two, in the present case). Thus, three types of information can be extracted from the DDA results in the present case: [28] the overall robustness of RCA classification for a given industry (percentage of stable classifications); the countries that need to be reclassified on the basis of their score, and those that have some characteristics of the other group, but not enough to justify a reclassification.

Table 13 identifies for each of the four selected industries the countries' classification in the Top-10 and the cases that were reclassified after the DDA. We observe that this reclassification concerns only cases that were not initially identified in the Top-10, all initial Top-10 cases were correctly reclassified as such for both intermediate and final goods. I show some countries not retained for the Top-10 group when the probability of belonging to this group is above 0.1 according to the discriminant analysis.

To illustrate the results in the table, let's consider the case of Textile. The DDA based on gross indicators and on Input-Output ones provided identical results. Were reclassified to Top-10 the following countries: For exports of intermediate goods: Croatia, Hong Kong and Romania; for exports of final goods: Lithuania. Spain, not classified in the Top-10 for intermediate goods, shared 21% of the characteristics measured on gross trade of final goods (more formally: the probability of Spain belonging to the Top-10 group for these products was 0.21) and the probability rose to 29% when the specification of the "explanatory" variables was done on Input-Output specification. Lithuania shared 18% of the Top-10 score for intermediate goods when the specification of the "explanatory" variables was done on Input-Output specification. On final goods, Colombia, despite not been classified in the Top-10, shared 28% of the characteristics of this group and the Philippines 16%.

Textile is the simplest case and the classification process is more uncertain for other sectors, in particular Electrical equipment, as can be seen in the table below. For this industry, the end-result for final product when applying the DDA on Gross explanatory variables delivers an inflated "Top-10" group of 24 cases! Additionally, 6 of the countries that were not retained had probabilities of belonging to the top group ranging between 0.50 and 0.12. Israel, with a probability of 0.495, was really close to becoming the 25th "Top-10" member. The industry producing electrical equipment for final use is also the sole case where the competition for entering into the group of best performers is larger than for the production of intermediate products.

Table 13 Ex-post classification in the Top-10 group by Discriminant Analysis, selected sectors

Textiles, wearing apparel, leather and related products

Intermediate						Final					
Gross			INPUT-OUTPUT			Gross			INPUT-OUTPUT		
ISO3	Posterior	Pr(1)	ISO3	Posterior	Pr(1)	ISO3	Posterior	Pr(1)	ISO3	Posterior	Pr(1)
PRT	1	1.00	VNM	1	1.00	BGR	1	1.00	TUN	1	1.00
TWN	1	1.00	PRT	1	1.00	TUN	1	1.00	BGR	1	1.00
IND	1	1.00	KHM	1	1.00	KHM	1	1.00	MAR	1	1.00
KHM	1	1.00	TWN	1	1.00	MAR	1	1.00	TUR	1	1.00
VNM	1	1.00	CHN	1	1.00	CHN	1	1.00	IDN	1	1.00
CHN	1	1.00	IND	1	1.00	VNM	1	1.00	VNM	1	1.00
ITA	1	1.00	ITA	1	1.00	IDN	1	1.00	CHN	1	1.00
TUR	1	1.00	TUR	1	1.00	TUR	1	1.00	IND	1	1.00
TUN	1	1.00	TUN	1	1.00	IND	1	1.00	PRT	1	1.00
THA	1	1.00	THA	1	1.00	PRT	1	1.00	KHM	1	1.00
BGR	1	1.00	IDN	1	1.00	THA	1	1.00	ITA	1	1.00
IDN	1	1.00	BGR	1	1.00	ROU	1	1.00	ROU	1	1.00
MAR	1	0.98	MAR	1	0.99	ITA	1	1.00	THA	1	1.00
HRV	1	0.96	ROU	1	0.95	PER	1	0.98	PER	1	0.97
ROU	1	0.91	HRV	1	0.89	LTU	1	0.91	LTU	1	0.71
HKG	1	0.85	HKG	1	0.71	COL	0	0.28	COL	0	0.10
ESP	0	0.21	ESP	0	0.29	PHL	0	0.16			
LTU	0	0.03	LTU	0	0.18						

Electrical equipment

Intermediate						Final					
Gross			INPUT-OUTPUT			Gross			INPUT-OUTPUT		
ISO3	Posterior	Pr(1)	ISO3	Posterior	Pr(1)	ISO3	Posterior	Pr(1)	ISO3	Posterior	Pr(1)
HKG	1	1.00	HKG	1	1.00	TUN	1	1.00	MYS	1	1.00
TUN	1	1.00	TUN	1	1.00	MYS	1	1.00	TUN	1	1.00
CHN	1	1.00	ROU	1	1.00	MAR	1	1.00	MAR	1	1.00
ROU	1	1.00	CHN	1	1.00	CHN	1	1.00	CHN	1	1.00
SVN	1	1.00	SVN	1	1.00	KOR	1	1.00	KOR	1	1.00
MAR	1	1.00	TUR	1	1.00	HKG	1	1.00	SVN	1	1.00
TUR	1	1.00	MAR	1	1.00	ROU	1	0.99	ROU	1	1.00
KOR	1	1.00	JPN	1	1.00	SVN	1	0.99	HKG	1	0.99
JPN	1	0.99	KOR	1	1.00	JPN	1	0.98	CZE	1	0.99
SGP	1	0.99	MEX	1	0.99	CZE	1	0.98	JPN	1	0.99
MEX	1	0.99	SGP	1	0.99	POL	1	0.96	DEU	1	0.97
CZE	1	0.99	CZE	1	0.99	SGP	1	0.95	POL	1	0.96
POL	1	0.98	DEU	1	0.98	DEU	1	0.93	SGP	1	0.95
AUT	1	0.94	POL	1	0.98	TUR	1	0.92	AUT	1	0.91
DEU	1	0.94	AUT	1	0.97	MEX	1	0.82	TUR	1	0.91
HRV	1	0.88	IND	1	0.91	AUT	1	0.82	FIN	1	0.86
IND	1	0.87	HRV	1	0.91	SVK	1	0.80	ITA	1	0.71
ITA	1	0.64	ITA	1	0.75	FIN	1	0.78	SVK	1	0.71
BGR	1	0.91	BGR	1	0.66	VNM	1	0.74	MEX	1	0.69
PHL	1	0.56	PHL	1	0.57	ITA	1	0.73	VNM	1	0.66
PRT	0	0.49	DNK	0	0.37	THA	1	0.69	PHL	1	0.64
CHE	0	0.35	FIN	0	0.30	PHL	1	0.67	THA	0	0.40
DNK	0	0.27	CHE	0	0.24	HUN	1	0.66	HUN	0	0.38
FIN	0	0.26	PRT	0	0.21	EST	1	0.60	EST	0	0.37
CRI	0	0.21	SWE	0	0.17	ISR	0	0.50	ISR	0	0.36
SWE	0	0.21	FRA	0	0.15	ESP	0	0.31	BGR	0	0.35
FRA	0	0.15	ESP	0	0.12	BGR	0	0.31	ESP	0	0.32
EST	0	0.12	CRI	0	0.10	PRT	0	0.28	COL	0	0.22
ESP	0	0.11				COL	0	0.27	SWE	0	0.18
						SWE	0	0.18	DNK	0	0.14
						CHE	0	0.18	PRT	0	0.14
						DNK	0	0.12	CHE	0	0.13

Computer, electronic and optical products

Intermediate						Final					
Gross			INPUT-OUTPUT			Gross			INPUT-OUTPUT		
ISO3	Posterior	Pr(1)	ISO3	Posterior	Pr(1)	ISO3	Posterior	Pr(1)	ISO3	Posterior	Pr(1)
SGP	1	1.00	SGP	1	1.00	SGP	1	1.00	SGP	1	1.00
PHL	1	1.00	PHL	1	1.00	PHL	1	1.00	PHL	1	1.00
KOR	1	1.00	CHN	1	1.00	TWN	1	1.00	IRL	1	1.00
CHN	1	1.00	MYS	1	1.00	CHN	1	1.00	CHN	1	1.00
KHM	1	1.00	KHM	1	1.00	EST	1	1.00	TWN	1	1.00
MLT	1	1.00	KOR	1	1.00	IRL	1	1.00	EST	1	1.00
MEX	1	1.00	MLT	1	1.00	MLT	1	1.00	MYS	1	1.00
MYS	1	1.00	TWN	1	1.00	KOR	1	1.00	MLT	1	1.00
TWN	1	1.00	CHE	1	1.00	MYS	1	1.00	CHE	1	1.00
CHE	1	1.00	ISR	1	1.00	HUN	1	1.00	KOR	1	1.00
ISR	1	1.00	JPN	1	1.00	ISR	1	1.00	ISR	1	1.00
IRL	1	0.99	MEX	1	1.00	SVK	1	1.00	HUN	1	1.00
JPN	1	0.99	IRL	1	0.99	CHE	1	1.00	SVK	1	1.00
THA	1	0.99	THA	1	0.99	NLD	1	1.00	THA	1	1.00
SWE	1	0.95	SWE	1	0.96	MEX	1	1.00	MEX	1	0.99
NLD	1	0.95	DNK	1	0.95	THA	1	1.00	CZE	1	0.99
DNK	1	0.86	HKG	1	0.58	VNM	1	1.00	FIN	1	0.95
HKG	1	0.58	NLD	1	0.54	CZE	1	0.97	VNM	1	0.89
AUT	0	0.32	AUT	0	0.41	FIN	1	0.80	NLD	1	0.74
FIN	0	0.22	DEU	0	0.39	TUN	0	0.12	TUN	1	0.65
KAZ	0	0.21	FIN	0	0.34						
LVA	0	0.17	KAZ	0	0.17						
EST	0	0.13	LVA	0	0.13						
DEU	0	0.12									

Motor vehicles, trailers and semi-trailers

Intermediate						Final					
Gross			INPUT-OUTPUT			Gross			INPUT-OUTPUT		
ISO3	Posterior	Pr(1)	ISO3	Posterior	Pr(1)	ISO3	Posterior	Pr(1)	ISO3	Posterior	Pr(1)
TUR	1	1.00	JPN	1	1.00	JPN	1	1.00	JPN	1	1.00
JPN	1	1.00	HUN	1	1.00	HUN	1	1.00	SVK	1	1.00
SVK	1	1.00	TUR	1	1.00	SVK	1	1.00	HUN	1	1.00
KOR	1	1.00	DEU	1	1.00	MEX	1	1.00	DEU	1	1.00
HUN	1	1.00	KOR	1	1.00	CAN	1	1.00	MEX	1	1.00
DEU	1	1.00	SVK	1	1.00	DEU	1	1.00	CZE	1	1.00
MEX	1	1.00	CZE	1	1.00	CZE	1	1.00	CAN	1	1.00
THA	1	1.00	THA	1	1.00	THA	1	1.00	THA	1	1.00
CZE	1	1.00	MEX	1	0.99	KOR	1	1.00	KOR	1	1.00
VNM	1	1.00	VNM	1	0.99	SWE	1	1.00	SWE	1	0.99
SVN	1	1.00	FRA	1	0.98	SVN	1	0.97	ESP	1	0.93
POL	1	0.99	ROU	1	0.98	ESP	1	0.95	ROU	1	0.91
FRA	1	0.99	SVN	1	0.98	ZAF	1	0.90	ZAF	1	0.87
ROU	1	0.98	POL	1	0.98	ROU	1	0.81	SVN	1	0.85
AUT	1	0.97	IND	1	0.94	POL	1	0.56	POL	1	0.69
ESP	1	0.95	AUT	1	0.93	AUT	0	0.22	AUT	0	0.27
PRT	1	0.95	ESP	1	0.90	VNM	0	0.14	TUR	0	0.20
MAR	1	0.94	MAR	1	0.87	PRT	0	0.14	ARG	0	0.16
IND	1	0.85	SWE	1	0.83				VNM	0	0.13
TUN	1	0.70	TUN	1	0.80				PRT	0	0.12
TWN	1	0.62	PRT	1	0.67						
SWE	1	0.52	ARG	1	0.66						
ARG	0	0.11	USA	1	0.53						
			GBR	0	0.46						
			TWN	0	0.43						
			BRA	0	0.40						
			ITA	0	0.37						
			ZAF	0	0.14						

Note: Shaded cases: countries reclassified in the "Top-10 RCA" by the discriminant function. Cut-off at probability of belonging to Top-10 not smaller than 0.10

Annex 3. List of RCA indices and their calculation

The calculation for the 20 RCA indices uses the conventions in R programming. RCA and T.Indic are tables (data frame), each RCA$x and T.Indic$x is a variable (a column in the table).

1. RCA$BRCA <- (T.Indic$Xki/T.Indic$Xi)/(T.Indic$Xwk/Xw) #Balassa RCA BRCA= (Xki/Xi)/(Xwk/Xw)
2. RCA$BRCA_IO <- (T.Indic$IO_Xki/T.Indic$IO_Xi)/(T.Indic$Xwk/Xw) #IO version of BRCA assuming (IO_Xwk/IO_Xw)≈(Xwk/Xw)
3. RCA$LBRCA <- Log(RCA$BRCA) #Log of BRCA
4. RCA$LBRCA_IO <- Log(RCA$BRCA_IO) #Log of BRCA_IO
5. RCA$BRCA2 <- (RCA$BRCA - 1)/(RCA$BRCA + 1) # BRCA normalised according to theoretical mean=1
6. RCA$BRCA2_IO <- (RCA$BRCA_IO - mean(RCA$BRCA_IO))/(RCA$BRCA_IO + mean(RCA$BRCA_IO)) # BRCA2 normalisation centred on observed mean
7. RCA$ARCA <- (T.Indic$Xki/T.Indic$Xi) - (T.Indic$Xwk/Xw) #Hoen and Ooserhaven additive RCA ARCA= (Xki/Xi)-(Xwk/Xw)
8. RCA$ARCA_IO <- (T.Indic$IO_Xki/T.Indic$IO_Xi) - (T.Indic$Xwk/Xw) #IO Hoen and Ooserhaven assuming (IO_Xwk/IO_Xw)=(Xwk/Xw)
9. RCA$LRCA <- 100*((((T.Indic$Xki-T.Indic$Mki)/(T.Indic$Xki+T.Indic$Mki)) - (T.Indic$Xi.Indic$Mi)/(T.Indic$Xi+T.Indic$Mi))) * ((T.Indic$Xki+T.Indic$Mki)/(T.Indic$Xi+T.Indic$Mi)))
10. RCA$LRCA_IO <- 100*((((T.Indic$IO_Xki-T.Indic$IO_Mki)/(T.Indic$IO_Xki+T.Indic$IO_Mki)) - ((T.Indic$IO_Xi-T.Indic$IO_Mi)/(T.Indic$IO_Xi+T.Indic$IO_Mi)))*((T.Indic$IO_Xki+T.Indic$IO_Mki)/(T.Indic$IO_Xi+T.Indic$IO_Mi)))
11. RCA$NRCA <- ((T.Indic$Xki) - (T.Indic$Xi*T.Indic$Xwk/Xw))/Xw #Normalised RCA NRCAki = (Xki / Xw)- (Xi Xwk/ Xw Xw)
12. RCA$NRCA_IO <- ((T.Indic$IO_Xki) - (T.Indic$IO_Xi*T.Indic$Xwk/(Xw)))/Xw #Same Assuming (IO_Xwk/IO_Xw)=(Xwk/Xw)
13. RCA$NCAI <- T.Indic$RNX * T.Indic$RO * T.Indic$RT
14. RCA$NCAI_IO <- T.Indic$RNX_IO * T.Indic$RO_IO * T.Indic$RT_IO
15. RCA$SNCAI <- T.Indic$RNX * T.Indic$RO * T.Indic$SRT # Original formula
16. RCA$SNCAI_IO <- T.Indic$RNX_IO * T.Indic$RO_IO * T.Indic$SRT_IO #Original modified for IO relationship
17. RCA$SNCAI2 <- T.Indic$RNX * T.Indic$RO * T.Indic$SRT2 #Modified for extreme values
18. RCA$SNCAI2_IO <- T.Indic$RNX_IO * T.Indic$RO_IO * T.Indic$SRT2_IO #Modified for extreme values, IO version
19. RCA$BIT <- (T.Indic$Xki - T.Indic$Mki)/((T.Indic$GDP/GDP_w)*T.Indic$Qk) # Bowen index (Xki-- Mki / E(Qki))
20. RCA$BIT_IO <- (T.Indic$IO_Xki - T.Indic$IO_Mki)/((T.Indic$GDP/GDP_w) * (T.Indic$Qk*(1- (T.Indic$Xi/X_w))) # Bowen index IO adjusted

Where:

T.Indic$Mw_byki # Imports of inputs k by sector k in country i

T.Indic$Mki # Total Imports of product k by country i (all i sectors included)

T.Indic$Qki #Production of ki (total: domestic a nd export of intermediate and final goods purposes)

T.Indic$Mxw_byki <- (T.Indic$Mw_byki*(T.Indic$Xki/T.Indic$Qki)) * (1-T.Indic$Xi/X_w) # Foreign imputs required by Xki net of reimports (proportionality assumption: Xi/X_w is the share of returning intermediates).

T.Indic$IO_Xki<- T.Indic$Xki - T.Indic$Mxw_byki #net export of ki, gross less net intermediate imports of foreign intermediates

T.Indic$IO_Mki <- T.Indic$Mki*(1- (T.Indic$Xi/X_w)) # Proportionality assumption.

T.Indic$IO_Mi <- T.Indic$Mi*(1- (T.Indic$Xi/X_w)) # Mi: Net Imports of inputs required for production of total output Qi Proportionality assumption

T.Indic$IO_Xi <- T.Indic$Xi - (T.Indic$IO_Mi*T.Indic$Xi/T.Indic$Qi) #Ne#Net of imported inputs required for the production of total export Xi

T.Indic$GDP <- Qki.df$GDPi # GDP country i including ROW

T.Indic$RO <- ((T.Indic$Xi+T.Indic$Mi)/T.Indic$GDP) / ((Xw+Xw)/GDP_w) # Xw=Mw

T.Indic$RO_IO <- ((T.Indic$IO_Xi+T.Indic$IO_Mi)/T.Indic$GDP) / ((Xw+Xw)/GDP_w) # Same for IO analysis assuming (IO_Xwk/IO_Xw)=(Xwk/Xw)

T.Indic$RT <- ((T.Indic$Xki+T.Indic$Mki)/(T.Indic$Xi+T.Indic$Mi)) / (T.Indic$Xwk/Xw) # Simplified (T.Indic$Xwk + T.Indic$Mwk)/(Xw+Mw) as Xwk=Mwk

```
T.Indic$RT_IO <- ((T.Indic$IO_Xki+T.Indic$IO_Mki)/(T.Indic$IO_Xi+T.Indic$IO_Mi)) / (T.Indic$Xwk/Xw)
#Same for IO analysis Assuming (IO_Xwk/IO_Xw)=(Xwk/Xw)
T.Indic$SRT <- T.Indic$RT / (T.Indic$RT + 1) # Original formula
T.Indic$SRT2 <- T.Indic$RT- min(T.Indic$RT) / (T.Indic$RT + max(T.Indic$RT)) # Not the original formula
RT/RT+1 as max(RT)not eq 1
T.Indic$SRT_IO <- T.Indic$RT_IO / (T.Indic$RT_IO + 1) # Original formula
T.Indic$SRT2_IO <- T.Indic$RT_IO - min(T.Indic$RT_IO) / (T.Indic$RT_IO + max(T.Indic$RT_IO)) # Not the
original formula RT/RT+1 as sample value for max(RT) is not equal to  1
T.Indic$RNX <-(T.Indic$Xki-T.Indic$Mki)/(T.Indic$Xki+T.Indic$Mki)
T.Indic$RNX_IO <-(T.Indic$IO_Xki-T.Indic$IO_Mki)/(T.Indic$IO_Xki+T.Indic$IO_Mki) #Same for IO
```