



# Time stationarity, shape and ordinal ranking bias of RCA indexes: a new set of measures

Rémi Stellan<sup>1</sup> · Jair N. Ojeda-Joya<sup>2</sup> · Jenny P. Danna-Buitrago<sup>3</sup>

Accepted: 1 September 2023  
© The Author(s) 2023

## Abstract

A new set of tools to measure the time stationarity, shape and ordinal ranking bias of Revealed Comparative Advantage (RCA) indexes is suggested. The aim is to help select the most consistent RCA index for a given set of countries, products and time periods, which is especially relevant for economic policy based on comparative advantages. The GMM estimation of an AR(1) process based on RCA indexes provides three measures of time stationarity that are more rational than the measures available in the literature. Furthermore, we revise the statistics that are commonly used to capture shape. Finally, with respect to ordinal ranking bias, we modify the use of Spearman's rank correlation coefficient in Leromain and Orefice (Int Econ 139:48–70, 2014) and generalize the non-parametric measure of Stellan and Danna-Buitrago (J Appl Econ 22(1):349–379, 2019). We discuss different methods of ranking RCA indexes according to all proposed measures. An application to 33 RCA indexes, 67 trade areas and three product classifications shows that, on average, the most accurate RCA indexes are Contribution-to-the-Trade-Balance indexes, Revealed Competitiveness indexes and regression-based indexes.

**Keywords** RCA index · Empirical properties of RCA indexes · GMM estimation · Statistical distribution · Non-parametric measure

**JEL Classification** C10 · C23 · C65 · F14 · F15

---

✉ Rémi Stellan  
rstellan@javeriana.edu.co

<sup>1</sup> Department of Business Administration, Pontificia Universidad Javeriana, AK. 7 #40B-36 Bogotá, Colombia

<sup>2</sup> Banco de la República (Central Bank of Colombia), Cra. 7 #14-78 Bogotá, Colombia

<sup>3</sup> Faculty of Economics, Management and Accounting, Los Libertadores University Institute, Cra. 16 #63A-68 Bogotá, Colombia

## 1 Introduction

Since Balassa (1965), comparative advantages have been measured through Revealed Comparative Advantage (RCA) indexes. RCA indexes are calculated using trade data under the assumption that trade flows can “reveal” comparative advantages. To the question “how can trade data be used to calculate an RCA index?”, Balassa (1965) suggests a simple and intuitive answer: if a product has greater weight in some country’s total exports than in a given trade area, then this divergence can be interpreted as the consequence of comparative advantage. In this regard, the RCA index from Balassa (1965), hereafter referred to as the B index, is calculated as the share of that product in the country’s total exports divided by the share of the product in total exports from all countries in the trade area. Thus, the B index reveals comparative advantages (disadvantages) if its value is greater (less) than 1.

The widespread use of the B index would convey the idea that comparative advantages should be measured according to that index. Recent examples of the use of the B index include: Barattieri (2014), Nath et al. (2015), Nath and Goswami (2018), Bagci (2016), Carrère and Strauss-Kahn (2017), Kathuria (2018), Lectard and Rougier (2018), Bahar et al. (2019). In addition, various international organizations seem to support the B index, such as the World Bank in its *World Integrated Trade Solutions* database, the World Trade Organization in its *Practical guide to trade policy analysis* (Bachetta et al., 2012), and the United Nations Conference on Trade and Development in its *RCA radar plots*.

However, doubts about the B index have been expressed in the literature. These doubts primarily concern the “theoretical” properties of the B index, namely the nature of the variables used by the B index and the method used to incorporate these variables into a formula. For instance, the B index is based on export data but ignores import data. However, using both export and import data would increase the precision of measurements of comparative advantages by better capturing both the demand and supply dimensions (Vollrath, 1991). Furthermore, the B index is subject to a *size bias*: a country can be associated with high values of the B index and, consequently, strong comparative advantages even if the country represents a relatively small share of exports (i.e., a small “size”) in a given trade area (Leromain & Orefice, 2014). Similarly, the B index reveals comparative advantages if its value is greater than 1 and comparative disadvantages if its value belongs to  $[0, 1)$ . Because there is an upper-bound for comparative disadvantages but not comparative advantages, the B index does not measure comparative advantages and disadvantages in the same way (Laursen, 2015). Lastly, the B index is not *additive*; that is, “EU’s comparative advantage in producing vehicles can be measured by the sum of its membership countries’ comparative advantage in producing vehicles; and China’s comparative advantage in the various labor-intensive products as a whole can be measured by the sum of China’s comparative advantage in each specific labor-intensive product” (Yu et al. 2009, p. 273).

In addition to doubts about the theoretical properties of the B index, concerns have arisen regarding its “empirical” properties, namely the properties of the

values of the B index when applied to a given set of countries, products and periods. Three main shortcomings have been reported. First, the B index proves to be *unstable over time* even though comparative advantages are supposed to be sticky over time. Second, the shape of its statistical distribution is *asymmetric* despite the fact that symmetry better reflects the fact that comparative disadvantages counterbalance comparative advantages, and exhibits *fat tails* even though observations suggest that strong comparative (dis)advantages are relatively rare. Third, the B index is subject to an *ordinal ranking bias* as the values of the B index may not rank countries in a consistent way (Yeats, 1985; Hoen & Oosterhaven, 2006; Yu et al., 2009; Leromain & Orefice, 2014; Liu & Gao, 2019).

This paper focuses on these three empirical properties—time stationarity, symmetric shape with thin tails, and minimization of ordinal ranking bias—and suggests a new set of measures to evaluate the ability of the B index and its alternatives to fit these properties. These new measures represent revisions of the measures already used in the literature, with the aim of increasing their reliability. Ultimately, these improved measures enhance the selection of the most appropriate RCA index to apply to a given set of countries, products and periods, based on a discussion of the empirical properties of RCA indexes together with a consideration of their theoretical properties. This is especially relevant for the design of economic policies on the basis of comparative advantages (on this point, see Gallardo 2005). We discuss different methods to rank RCA indexes according to all proposed measures, and we apply these methods to a data set of more than 595 million values of 33 RCA indexes encompassing different trade areas and product classifications.

The rest of the paper is organized as follows. Section 2 briefly reviews the most representative RCA indexes in the literature. Sections 3, 4 and 5 present our new tools to evaluate the time stationarity, shape and ordinal ranking bias of RCA indexes, respectively. Section 6 describes our ranking of RCA indexes according to all measures, and Sect. 7 presents our empirical assessment of RCA indexes. Concluding remarks are given in Sect. 8.

## 2 An overview of RCA indexes

Let  $J$  be a set of countries that form a trade area,  $K$  a set of products (or sectors) and  $T$  a set of time periods. An RCA index is a (many-to-one) function from  $J \times K \times T$  to  $\mathbb{R}$  or a subset of  $\mathbb{R}$  such that every triplet  $(i, k, t) \in J \times K \times T$  is associated with a unique real number, which we denote as  $RCA_{ikt}$ . There exists a comparative-advantage “neutral” value  $\phi$  such that the inequality  $RCA_{ikt} > \phi$  reveals comparative advantages for country  $i$  in  $J$  with respect to product  $k$  in time period  $t$ . The inequality  $RCA_{ikt} < \phi$  reveals comparative disadvantages, and the equality  $RCA_{ikt} = \phi$  reveals the absence of comparative advantages and disadvantages. In this regard:

**Table 1** The B index and alternatives (1)

Index	Formula	Range	$\phi$	Proposed by
B	$\frac{X_{it}/X_{i,t}}{X_{kt}/X_{..t}}$	$\mathbb{R}_+$	1	Balassa (1965)
AB	$\frac{X_{ikt}}{X_{i,t}} - \frac{X_{.kt}}{X_{..t}}$	$[-1, 1]$	0	Hoen and Oosterhaven (2006)
SB	$\frac{B_{ikt} - 1}{B_{ikt} + 1}$	$[-1, 1]$	0	Dalum et al. (1998), Laursen (2015)
WB <sup>K</sup>	$\frac{B_{ikt}}{\frac{1}{\#K} \sum_{l \in K} B_{ilt}}$	$\mathbb{R}_+$	$\frac{1}{\#K} \sum_{l \in K} B_{ilt}$	Proudman and Redding (1998, 2000)
WB <sup>J</sup>	$\frac{B_{ikt}}{\frac{1}{\#J} \sum_{j \in J} B_{jkt}}$	$\mathbb{R}_+$	$\frac{1}{\#J} \sum_{j \in J} B_{jkt}$	Amador et al. (2011)
B2	$\frac{X_{ikt} - M_{ikt}}{X_{ikt} + M_{ikt}}$	$[-1, 1]$	0	Balassa (1986)
B2D	$(B_{2it} / \frac{X_{i,t} - M_{i,t}}{X_{i,t} + M_{i,t}} - 1) \times s_{it}$	$\mathbb{R}$	0	Donges and Riedel (1977)
B2G	$B_{2ikt} \times \frac{(X_{ikt} + M_{ikt})/Y_{it}}{(X_{kt} + M_{kt})/Y_{..t}}$	$\mathbb{R}$	0	Gnidchenko and Salnikov (2015)
M	$\frac{X_{ikt}}{X_{i,t}} - \frac{M_{ikt}}{M_{i,t}}$	$\mathbb{R}$	0	Michaely (1962)

$\phi$  is the neutral value of an RCA index. Regarding the B2D index,  $s_{it} = \text{sign}(X_{i,t} - M_{i,t}) = -1$  if  $X_{i,t} - M_{i,t} < 0$ , else 1. Regarding the B2G index,  $Y_{it}$  is the GDP of country  $i$  in  $t$ , and  $Y_{..t} = \sum_{j \in J} Y_{jt}$

- If  $RCA_{i_1kt} > RCA_{i_2kt} > \phi$ , country  $i_1$  has a higher comparative advantage than country  $i_2$  for product  $k$  in period  $t$ . If  $\phi > RCA_{i_1kt} > RCA_{i_2kt}$ ,  $i_1$  has a lower comparative disadvantage than  $i_2$  for  $k$  in  $t$ .
- If  $RCA_{ik_1t} > RCA_{ik_2t} > \phi$ ,  $i$  has a higher comparative advantage for product  $k_1$  than for product  $k_2$  in  $t$ . If  $\phi > RCA_{ik_1t} > RCA_{ik_2t}$ ,  $i$  has a lower comparative disadvantage for  $k_1$  than for  $k_2$  in  $t$ .
- If  $RCA_{ikt_1} > RCA_{ikt_2} > \phi$ ,  $i$  has a higher comparative advantage for  $k$  in period  $t_1$  than in period  $t_2$ ; and if  $\phi > RCA_{ikt_1} > RCA_{ikt_2}$ ,  $i$  has a lower comparative disadvantage for  $k$  in  $t_1$  than in  $t_2$  (Danna-Buitrago, 2017).

We define  $B_{ikt}$  as the B index calculated for  $(i, k, t)$ :  $B_{ikt} = (X_{ikt}/X_{i,t})/(X_{kt}/X_{..t})$  where  $X_{ikt} \in \mathbb{R}_+$  denotes the exports of  $k$  from  $i$  to the other countries among  $J$  in  $t$ ;  $X_{i,t} = \sum_{l \in K} X_{ilt}$  denotes the exports from  $i$  to the other countries among  $J$  for all products among  $K$  (in  $t$ );  $X_{.kt} = \sum_{j \in J} X_{jkt}$  denotes the exports of  $k$  from every country to another among  $J$ ; and  $X_{..t} = \sum_{j \in J} \sum_{l \in K} X_{jlt}$  denotes the exports from every country to another among  $J$  for all products among  $K$ .  $B_{ikt} > 1$  implies that  $k$  has more weight in the exports of  $i$  in  $t$  relative to a “typical” country among  $J$  (French, 2017) and therefore should reveal comparative advantages for  $(i, k, t)$ . On the contrary,

**Table 2** The B index and alternatives (2)

Index	Formula	Range	$\phi$	Proposed by
RC	$SB_{ikt} = \frac{(M_{kt}/M_{it})/(M_{kt}/M_{..t}) - 1}{(M_{kt}/M_{it})/(M_{kt}/M_{..t}) + 1}$	[-2,2]	0	Algieri et al. (2022), Danna-Buitrago and Stellian (2022)
RC <sup>r</sup>	RC computed with trade flows weighted by $w_{kr}/w_{ki}$	[-2,2]	0	Danna-Buitrago and Stellian (2022)
RC <sup>Y</sup>	$RC_{ikt} \times m^{-Y_{it}}/\bar{Y}_{t+1}$	[-2m,2m]	0	
RC <sup>y</sup>	$RC_{ikt} \times m^{-y_{it}}/\bar{y}_{t+1}$	[-2m,2m]	0	
RC <sup>Y,r</sup>	RC <sup>Y</sup> computed with trade flows weighted by $w_{kr}/w_{ki}$	[-2m,2m]	0	
RC <sup>y,r</sup>	RC <sup>y</sup> computed with trade flows weighted by $w_{kr}/w_{ki}$	[-2m,2m]	0	

$\phi$  is the neutral value of an RCA index.  $w_{kt} = (X_{kt} + M_{kt})/(X_{..t} + M_{..t})$  is the share of  $k$  in the total trade of  $J$  in  $t$ .  $Y_{it}$  and  $y_{it}$  are the GDP and GDP *per capita* of country  $i$  in  $t$ , respectively.  $\bar{Y}_t$  and  $\bar{y}_t$  are the averages of  $Y_{it}$  and  $y_{it}$  throughout  $J$ , respectively

**Table 3** The B index and alternatives (3)

Index	Formula	Range	$\phi$	Proposed by
Z	$\frac{z_{ikt}}{z_{i,t}} / \frac{\bar{z}_{kt}}{\bar{z}_{..t}}$	$\mathbb{R}_+$	1	Leromain and Orefice (2014)
AZ	$z_{ikt}/z_{i,t} - \bar{z}_{kt}/\bar{z}_{..t}$	$\mathbb{R}$	0	This paper
NY	$\left( X_{ikt} - \frac{X_{kt}}{X_{..t}} X_{it} \right) / X_{..t}$	$\mathbb{R}$	0	Yu et al. (2009)
CTB <sup>W</sup>	$(X_{ikt} - M_{ikt} - w_{it}(X_{it} - M_{it})) / (X_{..t} + M_{..t})$	$\mathbb{R}$	0	Lafay (1987)
CTB <sup>Y</sup>	$(X_{ikt} - M_{ikt} - w_{it}(X_{it} - M_{it})) / Y_{it}$	$\mathbb{R}$	0	Lafay (1992)
CTB <sup>Y,r</sup>	CTB <sup>Y</sup> computed with trade flows weighted by $w_{kr}/w_{ki}$	$\mathbb{R}$	0	De Saint Vaulry (2008)
CTB <sup>W,r</sup>	CTB <sup>W</sup> computed with trade flows weighted by $w_{kr}/w_{ki}$	$\mathbb{R}$	0	Stellian and Danna-Buitrago (2022)
CTB <sup>W,Y</sup>	$CTB_{ikt}^W \times m^{-Y_{it}}/\bar{Y}_{t+1}$	$\mathbb{R}$	0	
CTB <sup>W,y</sup>	$CTB_{ikt}^W \times m^{-y_{it}}/\bar{y}_{t+1}$	$\mathbb{R}$	0	
CTB <sup>W,Y,r</sup>	CTB <sup>W,Y</sup> computed with trade flows weighted by $w_{kr}/w_{ki}$	$\mathbb{R}$	0	
CTB <sup>W,y,r</sup>	CTB <sup>W,y</sup> computed with trade flows weighted by $w_{kr}/w_{ki}$	$\mathbb{R}$	0	

$\phi$  is the neutral value of an RCA index. For more details about the variables included in the Z index, please see the text.  $w_{kt} = (X_{kt} + M_{kt})/(X_{..t} + M_{..t})$  is the share of  $k$  in the total trade of  $J$  in  $t$ .  $Y_{it}$  and  $y_{it}$  are GDP and GDP *per capita* of country  $i$  in  $t$ , respectively.  $\bar{Y}_t$  and  $\bar{y}_t$  are the across-country averages of  $Y_{it}$  and  $y_{it}$ , respectively

$0 \leq B_{ikt} < 1$  reveals comparative disadvantages, and  $B_{ikt} = 1$  is the comparative-advantage neutral value.

As summarized in Tables 1, 2 and 3, various alternatives to the B index have been proposed in the literature. Some of these alternatives stem from modifications

of the B index. The “additive” version of the B index (Hoen & Oosterhaven, 2006), referred to as the AB index, calculates the difference between  $X_{ikt}/X_{i,t}$  and  $X_{kt}/X_{..t}$  instead of the ratio, namely  $AB_{ikt} = X_{ikt}/X_{i,t} - X_{kt}/X_{..t}$ . The “symmetric” version of the B index (Dalum et al., 1998; Laursen, 2015), which we denote as SB, is an approximation of the log of the B index around the neutral value of the B index (i.e., 1):<sup>1</sup>  $SB_{ikt} = (B_{ikt} - 1)/(B_{ikt} + 1)$ . Furthermore, in “weighted” versions of the B index,  $B_{ikt}$  is normalized by the average B index either across products ( $K$ ) for a given country (Proudman & Redding, 1998, 2000) or across countries ( $J$ ) for a given product (Amador et al., 2011). We write as  $WB^K$  and  $WB^J$  these weighted versions of the B index:  $WB^K_{ikt} = B_{ikt} / \frac{1}{\#K} \sum_{l \in K} B_{ilt}$  and  $WB^J_{ikt} = B_{ikt} / \frac{1}{\#J} \sum_{j \in J} B_{jkt}$ .

Other RCA indexes are based not only on exports but also on imports. For this purpose, let us denote as  $M_{ikt}$  the imports of  $k$  by  $i$  in  $t$ , and as  $M_{i,t}$ ,  $M_{kt}$  and  $M_{..t}$  as the counterparts of  $X_{i,t}$ ,  $X_{kt}$  and  $X_{..t}$ , respectively. The RCA index suggested by Balassa (1986) and referred to as the B2 index consists of the trade balance associated with  $(i, k, t)$ , that is,  $X_{ikt} - M_{ikt}$ , normalized by the total trade associated with  $(i, k, t)$ , namely  $X_{ikt} + M_{ikt}$ . Donges and Riedel (1977) and Gnidchenko and Salnikov (2015) subsequently proposed modifications of B2. Donges and Riedel (1977) calculate the ratio of  $B2_{ikt}$  to the same type of ratio calculated at the level of  $K$ —that is,  $(X_{i,t} - M_{i,t})/(X_{i,t} + M_{i,t})$ —before subtracting 1 from that ratio and ultimately multiplying by 1 or -1 depending on the sign of the total trade balance  $X_{i,t} - M_{i,t}$  associated with  $(i, t)$ . Gnidchenko and Salnikov (2015) take into account GDP in addition to exports and imports. Their RCA index, which we denote as B2G, weights  $B2_{ikt}$  by the *relative openness rate*, defined as the ratio of trade  $X_{ikt} + M_{ikt}$  to the GDP of  $i$  in  $t$  divided by the same ratio at the level of  $J$ .

Michaely (1962) proposes an RCA index that, like the B2, B2D and B2G indexes, takes into account both exports and imports. This RCA index calculates the difference between  $X_{ikt}/X_{i,t}$  and  $M_{ikt}/M_{i,t}$  instead of the difference between  $X_{ikt}/X_{i,t}$  and  $X_{i,t}/X_{..t}$ , contrary to the AB index. Consequently, the share of  $k$  in  $i$ 's imports substitutes for the share of  $k$  in  $J$ 's exports. Another solution is suggested by Algieri et al. (2022) and Danna-Buitrago and Stellian (2022) and is the difference between the SB index and the same kind of metric in relation to imports, that is, the log-approximation of  $(M_{ikt}/M_{i,t})/(M_{kt}/M_{..t})$ . This RCA index is denoted as RC and is rather similar to the RCA index labelled “Revealed Competitiveness” suggested by Vollrath (1991). Vollrath (1991) proposes removing the trade flows involving  $k$  and/or  $i$  when exports are added up across products and/or countries. For example,  $X_{kt} - X_{ikt}$  substitutes for  $X_{kt}$ ; therefore,  $X_{ikt}/X_{i,t}$  becomes  $X_{ikt}/(X_{i,t} - X_{ikt})$ . This would enable a clear distinction between  $k$  and the other products, and between  $i$  and the other countries. Danna-Buitrago and Stellian (2022) explain that preserving these flows is necessary because otherwise the values of the corresponding RCA index cannot be

<sup>1</sup>  $SB_{ikt} < \ln(B_{ikt})$  if  $B_{ikt} > 1$ ; that is, the approximation gives a smaller value than the log if applied to a number higher than 1.  $SB_{ikt} > \ln(B_{ikt})$  if  $0 < B_{ikt} < 1$ ; that is, the approximation gives a higher value than the log if applied to a number smaller than 1 (and greater than zero). Lastly, the approximation is more accurate if the B index is closer to 1. Note that the approximation of the log must substitute for the log because the log of B is undefined if  $B_{ikt} = 0$ , which occurs when  $i$  does not export  $k$  in  $t$  (Laursen, 2015).

computed if a country is the sole exporter or importer of some product<sup>2</sup>. In addition, Vollrath (1991) applies the log to the corresponding export-based and import-based ratios instead of the approximation of the log. Again, as argued by Danna-Buitrago and Stellan (2022), using the log prevents the calculation of an RCA index throughout a given set of countries, products and time periods.<sup>3</sup>

Further improvements to the RC index have subsequently been suggested, thus making RC the baseline RCA index of a whole class of RCA indexes. First, the RC index can be adjusted by multiplying it by a coefficient whose value is based on the respective GDPs of all countries in  $J$ . The purpose of this coefficient is to better measure comparative advantages as GDP differentials between countries is a proxy for differences in the availability of economic resources for creating comparative advantages or enhancing existing comparative advantages. The coefficient is given by  $m^{-Y_{it}/\bar{Y}_t+1} \in [0, m)$  where  $m$  is strictly positive and  $\bar{Y}_t$  denotes the across-country average GDP in  $t$ . The corresponding RCA index is denoted  $RC^Y$  if the aforementioned coefficient is based on GDP, and  $RC^y$  if GDP *per capita* is used instead (Danna-Buitrago & Stellan, 2022; Stellan & Danna-Buitrago, 2022).

Second, following the RCA indexes in terms of contribution to the trade balance—which are presented thereafter—every trade flow can be adjusted so that the share of  $k$  in total trade is the same in each period and corresponds to the share of a given reference period  $r$ . For this purpose, every export/import flow associated with  $(k, t)$ — $X_{ikt}$  and  $M_{ikt} \forall i \in J$ —is multiplied by  $w_{kr}/w_{kt}$  where  $w_{kt} = (X_{.kt} + M_{.kt})/(X_{.t} + M_{.t})$  is the share of  $k$  in the total trade of  $J$  in  $t$ .<sup>4</sup> The corresponding RCA index is denoted as  $RC^r$ . Third, GDP adjustment and trade-flow adjustment can take place together. The corresponding RCA indexes are denoted as  $RC^{Y,r}$  and  $RC^{y,r}$ , respectively.

Another RCA index is suggested by Leromain and Orefice (2014). This RCA index, which we denote as  $Z$ , is based on the variable  $z_{ikt}$ , which is a proxy for the Ricardian fundamental productivity level of  $i$  with respect to  $k$  in  $t$ . In line with the micro-founded Ricardian model of Costinot et al. (2012),  $z_{ikt}$  verifies  $z_{ikt} = \exp(\delta_{ikt}/\theta)$  where  $\delta_{ikt}$  is derived from the OLS estimation of  $\ln(x_{ijkt}) = \delta_{jit} + \delta_{ikt} + \delta_{jkt} + \epsilon_{ijkt}$ . According to this equation, the log of  $x_{ijkt}$ , which is the trade flow of  $k$  from  $i$  to another country  $j$  in  $t$ , is decomposed additively into an exporter-importer fixed effect ( $\delta_{jit}$ ), an exporter-product fixed effect ( $\delta_{ikt}$ ) and an importer-product fixed effect ( $\delta_{jkt}$ );  $\epsilon_{ijkt}$  is the residual term. The parameter  $\theta$  captures productivity dispersion across varieties of the same  $k$ .  $z_{ikt}$  is divided by  $i$ 's average

<sup>2</sup> If  $i$  is the sole exporter of  $k$ , then the denominator of  $X_{ikt}/(X_{it} - X_{ikt})$  would be equal to zero. Being the sole exporter implies that  $i$  has comparative advantages for  $k$  (in  $t$ ), but the value of the index remains undefined because of a numerical exception, namely division by zero. The same occurs if  $i$  is the sole importer of  $k$ , as the denominator of  $M_{ikt}/(M_{it} - M_{ikt})$  would be equal to zero (Danna-Buitrago & Stellan, 2022).

<sup>3</sup> If  $X_{ikt} = 0$  or  $M_{ikt} = 0$ , namely  $i$  does not export or import some product, then the approximation of the log can be calculated, but the log faces a numerical exception. Various other RCA indexes may also face numerical exceptions, but these exceptions can be solved. See the online appendix (click here).

<sup>4</sup> Because the exports from a given country are the imports from other countries,  $X_{.kt} = M_{.kt}$  and  $X_{.t} = M_{.t}$ . Consequently,  $w_{kt}$  can be calculated as  $X_{.kt}/X_{.t}$  or  $M_{.kt}/M_{.t}$  instead of  $(X_{.kt} + M_{.kt})/(X_{.t} + M_{.t})$ .

productivity across products, and the same ratio is calculated at the level of  $J$ . These two ratios are analogous to  $X_{ikt}/X_{i,t}$  and  $X_{,kt}/X_{,t}$  in the B index. Instead of dividing the former ratio by the latter ratio, we can also compute the difference, similar to the AB index with respect to the B index. AZ denotes this additive version of Z. The AZ index may show more interesting empirical properties than the Z index or other RCA indexes. Consequently, it is useful to include the AZ index among the set of RCA indexes that should be evaluated for a given space  $J \times K \times T$ .

Lastly, the RCA index from Yu et al. (2009) and the RCA indexes in terms of *Contribution to the Trade Balance* (CTB) should be mentioned. Yu et al. (2009) assume that  $i$  would have neither comparative advantages nor comparative disadvantages if the total exports of  $i$  are distributed according to the share of each product in the total exports of all countries in  $J$ . In this respect,  $(X_{,kt}/X_{,t})X_{i,t}$  is the comparative-advantage neutral value of exports. Actual exports higher (lower) than this neutral value should reveal a comparative advantage (disadvantage) for  $i$  with respect to  $k$  in  $t$ . The RCA index from Yu et al. (2009), which we denote as NY, consists of calculating the difference between actual exports and the comparative-advantage neutral value of exports; this difference is then normalized by the total exports of  $J$ .

The CTB indexes are similar to the NY index. The first difference is that CTB indexes do not calculate a comparative-advantage neutral value of exports but a comparative-advantage neutral value of the trade balance associated with  $(i, k, t)$ . In addition, this value is calculated by weighting the trade balance of  $i$  (instead of the sole exports of  $i$ ) by  $w_{kt}$ , which has been defined previously as the share of  $k$  in the total trade of  $J$  and is calculated as  $(X_{,kt} + M_{,kt})/(X_{,t} + M_{,t})$ . Consequently  $w_{kt}(X_{i,t} - M_{i,t})$  is the comparative-advantage neutral value of  $X_{ikt} - M_{ikt}$ . The difference between the actual trade balance and its neutral value can be normalized either by the total trade of  $J$  (CTB<sup>W</sup>) or by the GDP of  $i$  (CTB<sup>Y</sup>). In addition, every trade flow can be adjusted so that the share of  $k$  in total trade is the same in each period and corresponds to the share of a given reference period  $r$  (as explained before in relation to the class of RCA indexes arising from RC). The corresponding CTB index, also normalized by GDP, is denoted as CTB<sup>Y,r</sup>.

As in the case of RC-inspired indexes, a whole class of CTB indexes can be conceptualized starting from the difference between the actual trade balance and the theoretical trade balance. First, once this difference is computed on the basis of adjusted trade flows, it can be normalized by total trade calculated from the adjusted trade flows instead of a country's GDP. The corresponding CTB index is denoted as CTB<sup>W,r</sup>. Second, the standard CTB index with normalization by (unadjusted) total trade can be weighted by the coefficient  $m^{-Y_{it}/\bar{Y}_t+1}$  or  $m^{-y_{it}/\bar{y}_t+1}$ . The corresponding CTB indexes are denoted as CTB<sup>W,Y</sup> and CTB<sup>W,y</sup>, respectively. The combination of normalization by total trade and the GDP-related coefficient can be further augmented by the use of adjusted trade flows (Stellian & Danna-Buitrago, 2022). The corresponding CTB indexes are denoted as CTB<sup>W,Y,r</sup> and CTB<sup>W,y,r</sup>, respectively.<sup>5</sup>

<sup>5</sup> All the CTB indexes mentioned previously have in common the calculation of the difference between the actual trade balance and the comparative-advantage neutral trade balance. It is possible to compute CTB indexes based on the ratio of the actual trade balance to the comparative-advantage neutral trade balance. However, (Stellian & Danna-Buitrago, 2022) show that these "multiplicative" CTB indexes



### 3 Time stationarity

Comparative advantages arise from structural factors like technology, factor endowments and institutions, so comparative advantages tend to change only in the medium/long term. However, because an RCA index is calculated on the basis of trade data, short-term fluctuations in trade flows might induce variations in an RCA index even though comparative advantages might not change. Consequently, an RCA index that varies substantially over time might not adequately reveal comparative advantages. Ultimately, an RCA index is preferable to another if its time stationarity is higher, as an RCA index with higher time stationarity is better able to reveal comparative advantages despite the short-term fluctuations inherent to trade flows.

Our review of the literature reveals that time stationarity has been evaluated in the following ways:

- Unconditional standard deviation through time of the values of an RCA index calculated across products and possibly countries (Leromain & Orefice, 2014; Danna-Buitrago, 2017; Stellian & Danna-Buitrago, 2019).
- The OLS estimation of  $\beta_i$  in  $RCA_{ikt_1} = \alpha_i + \beta_i RCA_{ikt_0} + \varepsilon_{ik}$  (Dalum et al., 1998; Laursen, 2015; Danna-Buitrago & Stellian, 2022). According to this equation, time stationarity between initial period  $t_0$  and final period  $t_1$  is higher if  $\beta_i$  is closer to 1 and  $\alpha_i$  is closer to 0. This measure of time stationarity concerns the values of an RCA index calculated across products for a single country (hence the index  $i$  in  $\beta_i$  and  $\alpha_i$ ) on the basis of  $\#K$  observations, each one corresponding to a product  $k$ .
- The OLS estimation of  $\beta$  in  $RCA_{ikt_1} = \alpha + \beta RCA_{ikt_0} + \gamma_i + \varepsilon_{ik}$  (Stellian & Danna-Buitrago, 2019, 2022). According to this equation, time stationarity between  $t_0$  and  $t_1$  is higher if  $\beta$  is closer to 1 and  $\alpha$  is closer to 0. This method of measuring time stationarity concerns the values of an RCA index calculated across products *and* countries instead of a single country (hence  $\#J \times \#K$  observations, each one corresponding to a country-product pair  $(i, k)$ ). Country heterogeneity is captured by the fixed effect  $\gamma_i$ .
- $\chi^2$ -tests to determine whether the values of an RCA index calculated across products for a single country in  $t_0$  and the same kind of set in  $t_1$  are significantly different (Hoen & Oosterhaven, 2006). These tests can be extended to the values of an RCA index calculated across products and countries instead of considering a single country.
- The Harris-Tzavalis unit-root test based on the OLS estimation of  $RCA_{ikt} = \rho RCA_{ikt-1} + \gamma_{ik} + \varepsilon_{ikt}$  (Stellian & Danna-Buitrago, 2019), which has the same interpretation as the unit-root test presented in this paper (see below).
- The OLS estimation of  $\beta_{ik}$  in  $RCA_{ikt} = \alpha_{ik} + \beta_{ik}t + \varepsilon_{ikt}$  (Yu et al., 2010). According to this equation, time stationarity is higher if  $\beta_{ik}$  is closer to zero.

---

Footnote 5 (continued)

were systemically outperformed by their “additive” counterparts over different configurations of  $J$  involving more than 4.75 million values for each CTB index. This is the reason why we choose not to include multiplicative CTB indexes in the present paper.

This method of measuring time stationarity concerns the values of an RCA index calculated for a single product and a single country but can be extended to multiple products and countries, possibly with fixed/random effects associated with countries and/or products.

- Calculation of *transition probabilities* that measure the likelihood that the value of an RCA index belonging to a given interval in  $t_0$  remains in the same interval in  $t_1$ . If these transition probabilities are closer to 1, then time stationarity is higher (De Benedictis & Tamberi, 2004).

We propose a more robust method in which the starting point is the panel data set arising from the collection of RCA indexes calculated for every element of  $J \times K \times T$ . In this set, each individual corresponds to a country-product pair  $(i, k) \in J \times K$ ; hence the cross-sectional dimension comprises  $\#J \times \#K$  individuals. The evaluation of the time stationarity of an RCA index is based on the Arellano-Bond GMM estimation of the following AR(1) process:

$$RCA_{ikt} = \beta t + \rho RCA_{ikt-1} + \gamma_{ik} + \varepsilon_{ikt} \quad (1)$$

In this equation,  $\gamma_{ik}$  is a specific intercept for each country-product pair, which is useful to control for heterogeneity in the estimation.  $\varepsilon_{ikt}$  is the residual term. Using Eq. 1, we can implement the unit-root test according to which an RCA index shows time stationarity if the null hypothesis  $\rho = 1$  is rejected and thus the alternative hypothesis  $|\rho| < 1$  is accepted. Indeed, if  $|\rho| < 1$ , then the value of the RCA index associated with each country-product pair tends to fluctuate around the mean calculated as  $(\alpha + \beta t + \gamma_{ik})/(1 - \rho)$ . We can reject an RCA index as a suitable measure of comparative advantages for  $J \times K \times T$  if  $\rho = 1$  is accepted. Thereafter, for the RCA indexes for which  $|\rho| < 1$  is accepted, it is possible to apply three measures of time stationarity:

1. The distance between their respective estimations of  $\beta$  and 0. A smaller distance implies higher time stationarity.
2. The distance between their respective estimations of  $\rho$  and 0. By construction, a smaller distance implies higher time stationarity.
3. The conditional standard deviation of an RCA index, calculated as  $\sigma_\varepsilon / \sqrt{1 - \rho^2}$ , where  $\sigma_\varepsilon$  denotes the standard deviation of the residuals. A lower conditional standard deviation implies lower deviations of  $RCA_{ikt}$  around the mean, giving rise to higher time stationarity.

Equation 1 has four valuable features compared with already available tools. First, instead of an unconditional standard deviation, Eq. 1 rests the calculation of standard deviation on an AR(1) process. Second, the GMM estimation according to the Arellano-Bond method is more robust than the OLS estimation, especially for “large  $N$  - small  $T$ ” panel data sets like ours. GMM estimation provides consistent estimators in the case of panel models with the presence of endogeneity or lagged dependent variables, whereas OLS estimation does not. Under the usual moment condition in which instrumental variables and errors are orthogonal, the two-step

**Table 4** Estimation of Eq. 1: An illustration

	CTB <sup>W,y</sup>	NB	Z	RC <sup>Y,r</sup>
$\beta$	0.269e-7*** (- 2.04)	0.977e-7 (- 1.61)	- 0.0003*** (- 3.39)	- 1.458e-04*** (- 0.72)
$\rho$	0.7833*** (21.53)	0.8037*** (30.32)	0.4311*** (14.88)	0.8067*** (88.02)
$\sigma_\varepsilon$	0.4380e-4	0.5771e-4	0.8930e-1	0.3230
$\sigma_\varepsilon/\sqrt{1-\rho^2}$	0.7045e-4	0.9697e-4	0.9896e-1	0.5583

Source: Authors' calculations

\*\*\* $p \leq 0.01$ , \*\* $p \leq 0.05$ , \* $p \leq 0.10$

GMM provides the most efficient estimator. Even if regressors are strongly exogenous, two-step GMM is more efficient than pooled OLS in the case that lagged or squared independent variables are added to the estimation (Cameron & Trivedi, 2005). Third, Eq. 1 takes into account all periods rather than only the initial period ( $t_0$ ) and the final period ( $t_1$ ). Fourth, Eq. 1 comprises the linear time trend as suggested by Yu et al. (2010).

**Remark 1** The GMM estimation by Arellano and Bond (1991) assumes that errors are not serially correlated. These authors devised an autocorrelation test using the residuals of the estimation. They also recommend applying a Sargan test for over-identifying restrictions, which is analogous to the test proposed by Hansen (1991) within a general GMM regression. However, the former test is more important than the latter since the absence of serial correlation is an assumption for the GMM estimation. By contrast, the test for over-identifying restriction is only an early warning for the possibility of misspecification problems in the estimation.

We use a few options to guarantee a more reliable estimation of Eq. 1. First, the variance-covariance matrix is computed to be robust to the presence of heteroskedasticity. Second, the GMM estimation is optimal since it uses two steps to compute the coefficients. Third, we restrict the number of lags to be used as instrumental variables to improve the asymptotic behavior of the estimation.

**Example 1** Equation 1 is estimated for the CTB<sup>W,y</sup>, NB, Z and RC<sup>Y,r</sup> indexes according to the 259 items in the 3-digit Standard International Trade Classification (SITC) for each year from 1995 to 2020. The trade area comprises Germany and its main trading partners,<sup>6</sup> of which there are 34. Hence, the panel data comprises 26 time

<sup>6</sup> As further explained, we define the main trading partners of a given "reference" country as the smallest set of countries whose (i) respective shares are the highest and (ii) whose cumulative share is greater than 90%. According to UNCTADstat yearly trade data from 1995 to 2020, the main trading partners of Germany are Australia, Austria, Belgium, Brazil, Canada, China, Taiwan, Czechia, Denmark, Finland, France, Hungary, India, Ireland, Italy, Japan, South Korea Malaysia, Mexico, the Netherlands, Norway, Poland, Portugal, Romania, Russia, Singapore, Slovakia, South Africa, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States of America.

**Table 5** Ranking of RCA indexes in Table 4 according to time stationarity: an illustration

	CTB <sup>W,y</sup>	NB	Z	RC <sup>y,r</sup>
$\beta$	1	1	4	3
$\rho$	2	3	1	4
$\sigma_\varepsilon / \sqrt{1 - \rho^2}$	1	2	3	4
Mean rank	1.33	2	2.67	3.67

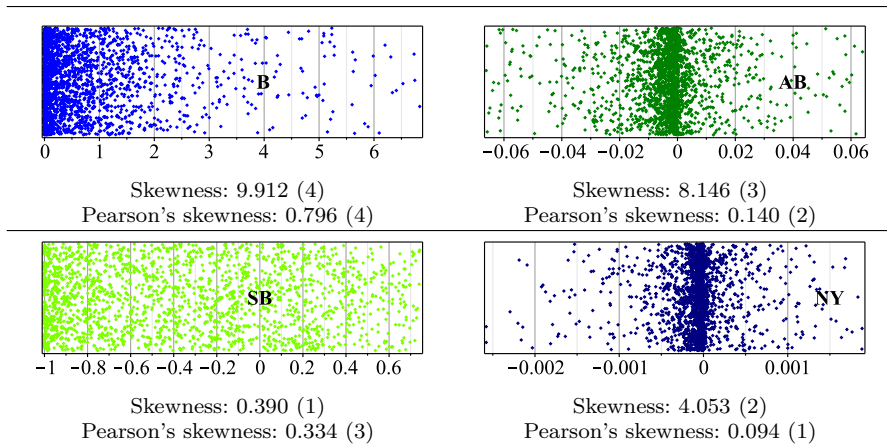
Source: Authors' calculations

periods and  $35 \times 259 = 9065$  country-product pairs for each RCA index. Estimates are reported in Table 4. For each RCA index, the 95% confidence interval of  $\rho$  does not include 1, allowing us to reject the null hypothesis of unit root ( $\rho = 1$ ). Consequently, all RCA indexes can be considered stable over time, and the next step is to determine the extent of time stationarity for each RCA index. According to Table 4, CTB<sup>W,y</sup> is the most stable RCA index over time from the vantage point of  $\beta$  because CTB<sup>W,y</sup> minimizes the distance of  $\beta$  from zero. However, the p-value associated with the estimate of  $\beta$  with NB is greater than 10%, and hence we cannot reject  $\beta = 0$  with NB. In addition, from the vantage point of  $\rho$ , Z is the most stable RCA index over time because Z minimizes the distance of  $\rho$  from zero. Ultimately, CTB<sup>W,y</sup> minimizes the conditional standard deviation and therefore would be the most stable RCA index over time from that third vantage point.

Table 5 ranks the four RCA indexes according to each criterion and then computes the mean rank. Rank 1 is attributed to CTB<sup>W,y</sup> and NB because the corresponding estimates of  $\beta$  are roughly equal and very close to zero. Eventually, CTB<sup>W,y</sup> achieves the lowest mean and therefore the best rank on average for the three criteria. In this example, CTB<sup>W,y</sup> is the most stable RCA index over time.

## 4 Shape

The distribution of an RCA index across countries and products should be as *symmetric* as possible in a given period. If the distribution is symmetric, the values that are greater than the mean offset the values that are smaller than the mean. Assuming that the mean is close to the neutral value of the RCA index under consideration, symmetry implies that the values that reveal comparative advantages tend to counterbalance the values that reveal comparative disadvantages. This is consistent with the relative nature of comparative advantages: a country that has comparative advantages for a product has comparative disadvantages for at least one other product, and at least one other country has comparative disadvantages for the former product. Ultimately, an RCA index is preferable to others if its asymmetry is lower. In addition, the distribution of an RCA index should contain a few outliers and therefore have *thin tails*. Such a distribution is necessary to reflect the fact that, generally speaking, a country has high comparative advantages and disadvantages for a few products. Consequently, fat tails should be avoided because they imply that outliers are more frequent than they should be.



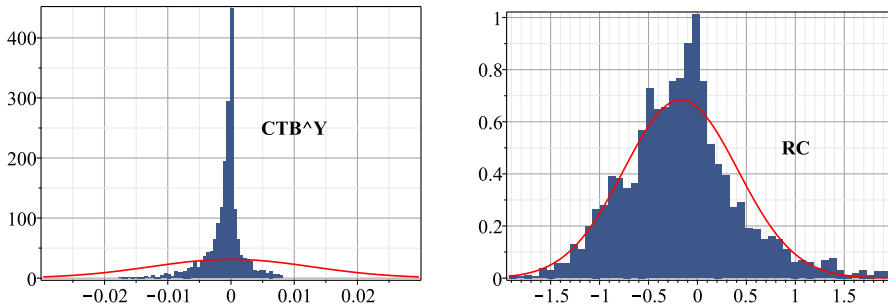
**Fig. 1** Scatter plots and asymmetry statistics of B, AB, SB and NY; trade area comprising Japan and its main trading partners, 2020. *Source:* Authors' calculations. Note: All points are assigned a random  $y$ -value to allow a better visual representation

Leromain and Orefice (2014) compare the respective asymmetries of the B and Z indexes based on skewness and mean-minus-median. Both statistics provide evidence of a right-skewed distribution if positive, or a left-skewed distribution if negative. Asymmetry is smaller if the respective distances of skewness and mean-minus-median from zero are smaller. However, it is difficult to compare the mean-minus-median of heterogeneous RCA indexes. If two RCA indexes have the same scale, then the aforesaid comparison is possible. For instance, the neutral value is 1 for both the B and Z indexes, and both indexes reveal comparative advantages within the range  $(1, +\infty)$  and comparative disadvantages within the range  $[0, 1)$ . Consequently, if the B index has a higher mean-minus-median than the Z index, it is possible to say that the former has a higher lack of symmetry. Nonetheless, if a third RCA index enters the comparison and has another scale (the AB index for example), mean-minus-median should be interpreted with caution.<sup>7</sup>

Ultimately, to compare the asymmetry of heterogeneous RCA indexes, we replace the mean-minus-median with the coefficient known as Pearson's second coefficient of skewness, which is calculated as  $3(\text{mean} - \text{median})/\sigma$ . Similar to skewness, the normalization by standard deviation gives a measure in a dimensionless unit and therefore enables comparisons between RCA indexes with different scales.<sup>8</sup> This coefficient provides evidence of a right-skewed distribution if positive, and a left-skewed distribution if negative.

<sup>7</sup> Leromain and Orefice (2014) also compare percentiles, but for the same reason as before, this kind of comparison might not prove useful for RCA indexes of different scales.

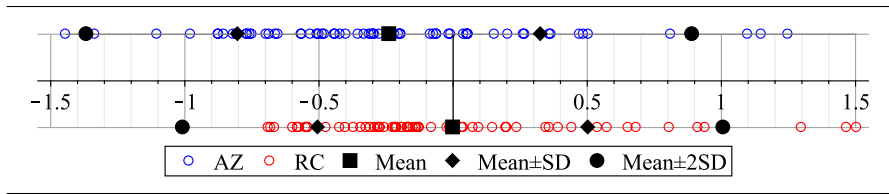
<sup>8</sup> Pearson's first coefficient of skewness is  $(\text{mean} - \text{mode})/\sigma$ , but this coefficient requires that the mode is unique. If not, the multiplicity of modes leaves undetermined which mode should enter the calculation.



**Fig. 2** Distributions of the  $CTB^Y$  index (left) and the RC index (right) for the trade area comprising Ghana and its main trading partners, 2020. *Source:* Authors' calculations. Note: The area of each bar is proportional to the number of data points in the corresponding bin. Moreover, the histogram is rescaled so that the total area of the bars is equal to 1. The red curve represents the probability density function of a normal distribution whose mean and standard deviation are taken from the distribution of the corresponding RCA index

**Example 2** Assume that the B, AB, SB and NY indexes are computed for the year 2020 according to the 65 items of the 2-digit SITC for the 30-country trade area comprising Japan and its main trading partners. The sets of values of each RCA index are represented in the scatter plots in Fig. 1. Each set comprises 1950 values, one per item and per country. The skewness and Pearson's skewness associated with each set are reported. The numbers in parentheses are the respective ranks of the RCA indexes for each coefficient. For example, B ranks fourth for each coefficient because B gives rise to the longest distance between each coefficient and zero. Consequently, the B index is the least symmetric RCA index in this example. SB ranks first in terms of skewness because SB minimizes the distance of skewness from zero (0.390), whereas NY ranks first in terms of Pearson's skewness because NY minimizes the distance of Pearson's skewness from zero. If we compute the mean rank, NY is the most symmetric RCA index in this example (mean rank equal to 1.5), followed by SB (2), AB (2.5) and B (4).

Kurtosis is the conventional statistic for measuring tail fatness. According to the standard interpretation, a higher kurtosis implies fatter tails. The reference point is 3, which is the kurtosis of a normal distribution. A kurtosis higher than 3 implies fatter tails than if the distribution were normal. However, Fig. 2 shows that this might not be true, similar to Westfall (2014) who gives several counterexamples showing that kurtosis cannot always be considered a reliable measure of the concentration of a distribution about its mean ("peakedness"). Figure 2 presents the distributions of the values computed for the  $CTB^Y$  and RC indexes for 2020 according to the 65 items of the 2-digit SITC for the 32-country trade area comprising Ghana and its main trading partners. We compare the distributions of these RCA indexes with normal distributions with the same mean and standard deviation. On the one hand, the left graph shows that the values of the  $CTB^Y$  index tend to remain below the probability density function of the corresponding normal distribution, and hence the distribution of the  $CTB^Y$  index shows thinner tails than in the normal case. On the other



**Fig. 3** Measure of outliers: Mexico in the trade area comprising this country and its main trading partners, 2020. *Source:* Authors’ calculations

hand, the right graph shows a fat right tail for the RC index, as values higher than 1 are more frequent than if the distribution were normal. Nonetheless, the kurtosis associated with the  $CTB^Y$  index is higher than the kurtosis associated with the RC index, namely 280.09 versus 3.83. According to the conventional interpretation of kurtosis, both distributions would have fat tails, but this is not the case for the  $CTB^Y$  index, even if its distribution has higher kurtosis than that of the distribution of the RC index.

To avoid possible misinterpretations of kurtosis, we suggest another measure of tail fatness. Let  $\mu_{it}$  and  $\sigma_{it}$  be the mean value and standard deviation of an RCA index calculated for the  $\#K$  values associated with a given country  $i$  and period  $t$  (one value per  $k$ ), and let  $rca_{it}^p = [\mu_{it} - p \cdot \sigma_{it}, \mu_{it} + p \cdot \sigma_{it}]$  be the range given by  $p \in \mathbb{N}_*$  standard deviation(s) of the mean. We define an *outlier of order  $p$*  as a value of an RCA index that does not belong to that range. A higher number of outliers implies fatter tails. Let  $out_{it}^p$  be the number of outliers of order  $p$  with respect to  $(i, t)$ :

$$out_{it}^p = \#\{k : RCA_{ikt} \notin rca_{it}^p\} \tag{2}$$

Our measure of tail fatness is the average of  $out_{it}^1$  and  $out_{it}^2$  throughout  $T$  for each country in  $J$ . These averages are written as  $\overline{out}_i^1$  and  $\overline{out}_i^2$ :

$$\overline{out}_i^p = \frac{1}{\#T} \sum_{t \in T} out_{it}^p \quad p = 1, 2 \tag{3}$$

Smaller values of  $\overline{out}_i^1$  and  $\overline{out}_i^2$  imply thinner tails.  $\overline{out}_i^p$  is an unambiguous measure of tail fatness. Its consistency arises from the fact that it is calculated on the basis of the standard deviation of each distribution. Lastly, it is possible to compare two RCA indexes on the basis of their respective values of  $\overline{out}_i^p$ .

**Example 3** Figure 3 represents the values of the AZ and RC indexes in 2020 according to the 65 items of the 2-digit SITC for Mexico in the 13-country trade area comprising this country and its main trading partners. The graph shows the location of the mean, the interval  $[\mu - \sigma, \mu + \sigma]$  and the interval  $[\mu - 2\sigma, \mu + 2\sigma]$  for both RCA indexes. Therefore, it is possible to observe that, for example, one value of AZ is inferior to  $\mu - 2\sigma$  and three other values are superior to  $\mu + 2\sigma$ . Consequently, there are four second-order outliers. According to the same logic, both RCA indexes

present nineteen first-order outliers, whereas RC has fewer second-order outliers than AZ (three instead of four). Ultimately, RC gives better measures of comparative advantages than AZ in this example because RC gives rise to fewer outliers than AZ.

## 5 Ordinal ranking bias

Assume that country  $i$  is ranked as the  $x^{\text{th}}$  country according to the B index for some product  $k$  (in time period  $t$ ). There is an *ordinal ranking bias* if for another product  $h \neq k$  the B index is higher but  $i$  has a lower rank than  $x$ , or is lower but  $i$  has a higher rank than  $x$ . Ordinal ranking bias should be avoided as much as possible to avoid misleading interpretations of RCA indexes.

Leromain and Orefice (2014) measure ordinal ranking bias by means of Spearman's rank correlation coefficients. Assume that in a three-country trade area product  $k$  has the fourth highest RCA index among the  $\#K$  indexes (one per product) associated with  $i$ . In addition, the RCA index associated with  $i$  for product  $k$  is the highest RCA index compared with the RCA indexes associated with the other countries. We thus have a first pair of integers (4, 1) where the first integer is the ranking of RCA indexes in the corresponding country's comparative advantage distribution and the second integer is the ranking of RCA indexes across countries. Assume that for the two other countries the respective pairs are (14, 2) and (7, 3). Spearman's rank correlation coefficient measures the extent to which two vectors of discrete data place the distinct items in the same order. In the case of RCA indexes, the coefficient captures the magnitude of match/mismatch between countries' respective distributions of RCA indexes and the ranking of countries according to RCA indexes. In our example, the Spearman's rank correlation coefficient is calculated for the rank vectors (4, 14, 7) and (1, 2, 3) and equals 0.292 when computed as Pearson's correlation coefficient (Kvam & Vidakovic, 2007). A coefficient closer to 1 implies a higher consistency between product classification and country classification, which leads to a lower ordinal ranking bias.

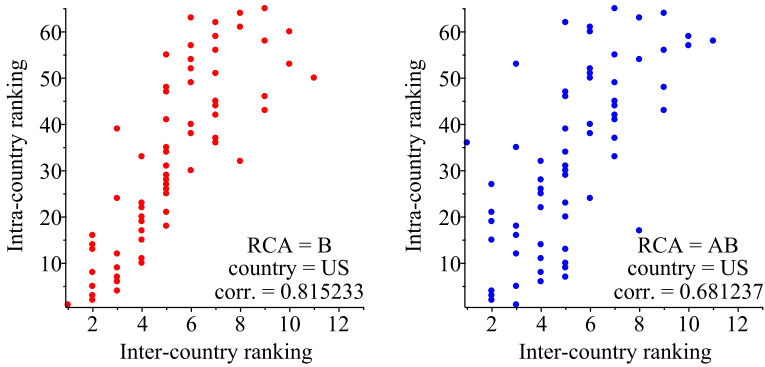
Leromain and Orefice (2014) calculates Spearman's rank correlation coefficient for each product under consideration (and for a given period); hence the number of these coefficients can be large. To obtain a synthetic measure, we calculate a unique Spearman's rank correlation coefficient across all products (and periods). Formally, for each  $(i, k, t)$ , the computation of Spearman's rank correlation coefficient requires the calculation of  $i$ 's rank relative to the other countries with respect to  $k$  in  $t$ . This rank is written as  $\hat{K}_{ikt} \in \{1, 2, \dots, \#J\}$ :

$$\hat{K}_{ikt} = \#\{x \in J : \text{RCA}_{xkt} > \text{RCA}_{ikt}\} + 1 \quad (4)$$

$i$  is ranked as the  $x^{\text{th}}$  country if the value of the RCA index is higher for  $x - 1$  countries in relation to  $(k, t)$ . Similarly, the computation of Spearman's rank correlation coefficient requires the calculation of  $k$ 's rank relative to the other products with respect to  $i$  in  $t$ . This rank is written as  $\check{K}_{ikt} \in \{1, 2, \dots, \#K\}$ :

$$\check{K}_{ikt} = \#\{x \in K : \text{RCA}_{ixt} > \text{RCA}_{ikt}\} + 1 \quad (5)$$





**Fig. 4** Spearman’s rank correlation coefficient between intra- and inter-country product rankings, United States in the Mexico-related trade area, 2020. *Source:* Authors’ calculations

$k$  is ranked as the  $x^{\text{th}}$  product if the value of the RCA index associated with  $(i, t)$  is higher for  $x - 1$  products in relation to  $(i, t)$ . If two countries/products share the same value of the RCA index, they have the same rank. For example, for 4 countries, assume that  $RCA_{i_1kt} = 1.2$ ,  $RCA_{i_2kt} = RCA_{i_3kt} = 1$  and  $RCA_{i_4kt} = 0$ .  $i_1$  is ranked first with respect to  $k$  in  $t$ , both  $i_2$  and  $i_3$  are ranked second, and  $i_4$  is ranked fourth. No country is ranked third (the RCA index values of three countries are higher than the value of  $i_4$ , and therefore  $i_4$  is ranked fourth). The same logic applies to products.

Ultimately, the Spearman’s rank correlation coefficient is calculated as Pearson’s correlation between the paired ranks from  $\hat{K}$  and  $\check{K}$  over  $\#K \times \#T$  for each country.

**Example 4** We compute  $\hat{K}_{ikt}$  and  $\check{K}_{ikt}$  for the trade area comprising Mexico and its main trading partners for a single year, 2020. Figure 4 describes each pair  $(\hat{K}_{ikt}, \check{K}_{ikt})$  with  $i = \text{United States}$  (and  $t = 2020$ ), namely one of Mexico’s main trading partners, in the case of the B index and the AB index. The inter-country ranking (i.e.  $\hat{K}_{ikt}$ ) is represented on the horizontal axis, and the intra-country ranking (i.e.  $\check{K}_{ikt}$ ) is represented on the vertical axis. Because the Mexico-related trade area comprises 13 countries, the inter-country ranking ranges from 1 to 13; because we use the 65 items in the 2-digit SITC as in the previous example, the intra-country ranking ranges from 1 to 65. Spearman’s rank correlation coefficient is greater for the B index than for the AB index. In this example, B is better able than AB to avoid ordinal ranking bias.

We complement the computation of Spearman’s rank correlation coefficient by the generalization of the measure of ordinal ranking bias by Stellian and Danna-Buitrago (2019). Indeed, Spearman’s rank correlation coefficient might not adequately capture ordinal ranking bias, especially if the number of countries in  $J$  is small, as this would imply a few possible ranks for each country and therefore might distort the coefficients. The non-parametric measure suggested by Stellian and Danna-Buitrago (2019) was initially designed for a four-country trade area, namely the Pacific Alliance (Chile, Colombia, Mexico and Peru). Here we provide formulas that

enable the application of this measure to any trade area [see also the formalization aspects of ordinal ranking bias in Danna-Buitrago and Stellian (2022) and Stellian and Danna-Buitrago (2022)]. For this purpose, we implement a four-step procedure. First, for each  $(i, t)$ , we distribute  $K$  into  $J$  subsets  $K_{it}^1, K_{it}^2, \dots, K_{it}^J$ . If  $k \in K_{it}^j$  then  $k$  implies that  $i$  is ranked as the  $j^{\text{th}}$  country according to the value of its RCA index compared with the values associated with the other countries and the same product. This means that a set of  $\#j - 1$  of countries other than  $i$  have a higher rank than  $i$ . These countries verify  $RCA_{xkt} > RCA_{ikt} \exists k, t$  and  $x \neq i$ . Ultimately:

$$K_{it}^j = \{k : \#\{x : RCA_{xkt} > RCA_{ikt}\} = j - 1\} \tag{6}$$

Second, we calculate the mean value of the RCA index that leads  $i$  to be ranked as the  $j^{\text{th}}$  country (in  $t$ ). We denote this mean as  $RCA_{it}^j$ :

$$RCA_{it}^j = \frac{1}{\#K_{it}^j} \sum_{k \in K_{it}^j} RCA_{ikt} \tag{7}$$

This second step is the use of a function from  $\mathbb{R}^{\#K_{it}^j}$  to  $\mathbb{R}$  such that the set  $\{RCA_{ikt} : k \in K_{it}^j\}$  is associated with a unique value that is representative of the values in that set. We choose the function that computes the mean, but other descriptive statistics such as the median or percentiles, among others, are potential candidates for  $RCA_{it}^j$ . Further research should explore this subject.

For the third step, we introduce the following definition:

**Definition 1** An RCA index implies an ordinal ranking bias if a country  $i$ , a time period  $t$ , a product  $k$  and two ranks  $j_1, j_2$  exist such that  $k \in K_{it}^{j_1}, j_1 > j_2$  and  $RCA_{ikt} > RCA_{it}^{j_2}$ , or  $k_1 \in K_{it}^{j_1}, j_1 < j_2$  and  $RCA_{ik_1t} \leq RCA_{it}^{j_2}$ .

Assume that  $i$  has rank  $j_1$  with respect to  $k$  in  $t$ , namely  $k \Rightarrow \#\{x : RCA_{xkt} \geq RCA_{ikt}\} = j_1 - 1$ . There is an ordinal ranking bias if the value of the RCA index associated with  $(i, k, t)$  is higher than the mean value that leads  $i$  to have a higher rank than  $j_1$  in  $t$ . Conversely, there is an ordinal ranking bias if the value of the RCA index associated with  $(i, k, t)$  is lower than the mean value that leads  $i$  to have a lower rank than  $j_1$  in  $t$ . Definition 1 implies the following two properties. First, an RCA index is exempt from ordinal ranking bias if  $RCA_{ikt} < RCA_{it}^{j_2} \leq \#J \forall t, i, j_1 > j_2$  and  $k \in K_{it}^{j_1}$ , and  $RCA_{ikt} > RCA_{it}^{j_2} \forall t, i, j_1 > j_2 \geq 1$  and  $k \in K_{it}^{j_1}$ . Second, it is possible that  $K_{it}^{j_1} = \emptyset$ , namely  $i$  does not have rank  $j_1$  in  $t$ . If so, rank  $j_1$  does not imply ordinal ranking biases in relation to  $(i, t)$ . Similarly, rank  $j_1$  does not imply ordinal ranking biases in relation to  $(i, t)$  and  $j_2 \neq j_1$ .

The third step consists of counting the number of ordinal ranking biases for every  $(i, k, t)$ . We denote this number as  $\text{orb}_{i,k,t} \in \mathbb{N}$ :

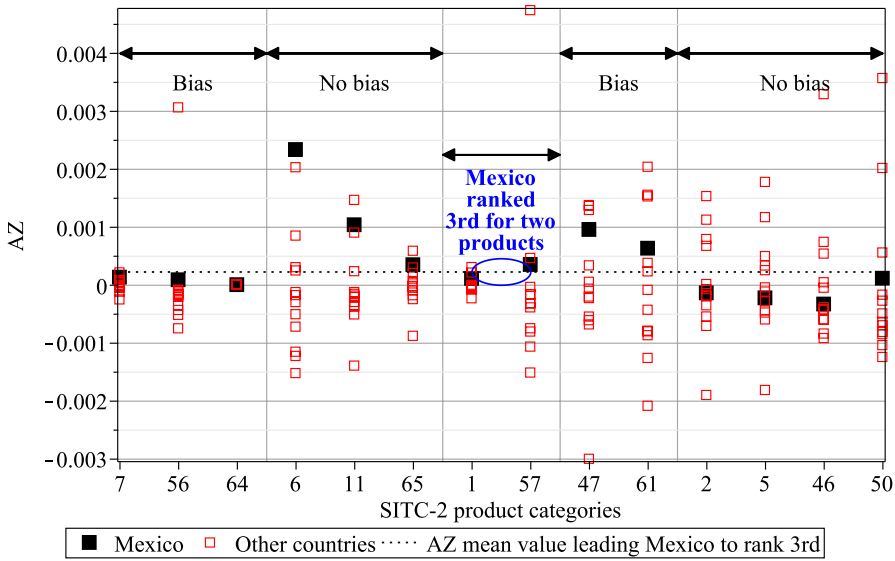


Fig. 5 Ordinal ranking bias, Mexico and its main trading partners, 2020. Source: Authors' calculations

$$\begin{aligned}
 orb_{ikt} = & \sum_{j=1}^{\#J} \# \left\{ k \in \bigcup_{x=j+1}^{\#J} K_{it}^x : RCA_{ikt} > RCA_{it}^j \right\} \\
 & + \sum_{j=1}^{\#J} \# \left\{ k \in \bigcup_{x=1}^{j-1} K_{it}^x : RCA_{ikt} < RCA_{it}^j \right\}
 \end{aligned}
 \tag{8}$$

As a last step, we compute the average value of  $orb_{i,k,t}$  across products and time. This average is written as  $\overline{orb}_i$ :

$$\overline{orb}_i = \frac{\sum_{k \in K} \sum_{t \in T} orb_{i,k,t}}{\#K \times \#T}
 \tag{9}$$

A smaller value of  $\overline{orb}_i$  implies a smaller ordinal ranking bias. For replication purposes, an algorithm to compute  $\overline{orb}_i$  is available in an online repository ([click here](#)).

**Example 5** Figure 5 describes the values taken by the AZ index in 2020 for the countries in the trade area comprising Mexico and its main trading partners for a subset of 2-digit SITC items. Figure 5 shows that Mexico is ranked third for two items, that is, the first and fifty-seventh items (“Live animals” and “Furniture and parts thereof”, respectively). Indeed, for both items, there are only two values greater than the value associated with Mexico. This third rank is associated with  $AZ = 0.10905e-3$  for the first item and  $0.34806e-3$  for the second item. Consequently, an average value of  $0.22855e-3$  leads Mexico to rank third. This leads to five ordinal ranking biases in this example:

- On the left of the graph, items 7, 56 and 64 imply an ordinal ranking bias. Indeed, for these three items, Mexico is ranked first or second, but the corresponding value of AZ is less than  $0.22855e-3$ . Mexico is expected to rank second or first if the value of AZ is greater than the mean value that ranks Mexico third. This occurs for items 6, 11 and 65, which therefore do not imply ordinal ranking bias, but not for items 7, 56, and 64.
- On the right, items 47 and 61 also imply an ordinal ranking bias. Indeed, for these two items, Mexico is ranked fourth, but the corresponding value of AZ is greater than  $0.22855e-3$ . Mexico is expected to rank fourth if the value of AZ is smaller than the mean value that leads Mexico to rank third. This occurs for items 2, 5, 46 and 50, which therefore do not imply ordinal ranking bias, but not for items 47 and 61.

## 6 Ranking of RCA indexes according to all measures

In summary, sections (3), (4) and (5) give rise to nine measures for evaluating the empirical properties of an RCA index for a given universe  $J \times K \times T$ . The first empirical property is time stationarity, and its three measures are  $\beta$  (trend),  $\rho$  (persistence) and the conditional standard deviation ( $\sigma$ ) associated with the AR(1) process in which the RCA index computed for  $(i, k, t)$  is the dependent variable and the RCA index computed for  $(i, k, t - 1)$  is the independent variable (see Eq. 1). The process is estimated once for all values of  $J \times K \times T$ , so each RCA index is associated with a single triplet  $(\beta, \rho, \sigma)$ . This gives rise to three rankings, each one associated with a measure in that triplet. Provided that a value of  $\beta$  closer to zero implies higher time stationarity, rank  $x$  is given to an RCA index in terms of  $\beta$  if  $x - 1$  RCA indexes achieve a lower distance of  $\beta$  from 0. RCA indexes can be ranked in the same manner using  $\rho$  and  $\sigma$ .

The second empirical property is shape, which can be divided into asymmetry and tail fatness. The two measures of asymmetry are skewness and Pearson's second coefficient of skewness. These shape statistics are calculated for all countries and products in each period; namely these asymmetry metrics are calculated  $\#T$  times with all values in  $J \times K$ . Provided that asymmetry is lower if skewness is closer to zero, RCA indexes can be ranked  $\#T$  times according to the distance of their respective values of skewness from zero, one rank per period, and then the across-period average rank is calculated. Another solution is to rank RCA indexes a single time according to the distance of their respective values of across-period average skewness from zero. The same kind of ranking applies to Pearson's second coefficient of skewness.

The two measures of tail fatness are the numbers of outliers based on one standard deviation of the mean— $\overline{\text{out}}_i^1$ —and two standard deviations of the mean— $\overline{\text{out}}_i^2$ —given by Equations (2) and (3).  $\overline{\text{out}}_i^1$  and  $\overline{\text{out}}_i^2$  are calculated for each country in  $J$ . Provided that a smaller value of  $\overline{\text{out}}_i^p$  implies lower extreme value frequency,

RCA indexes are ranked according to their ability to minimize  $\overline{\text{out}}_i^p$  for each country before computing the across-country average rank.

Lastly, the two measures of ordinal ranking bias are Spearman’s rank correlation coefficient (see Eqs. 4 and 5) and the non-parametric measure given by Eqs. (6–9). The two metrics are calculated for each country in  $J$ . Values of Spearman’s rank correlation coefficient closer to 1 imply a lower ordinal ranking bias. RCA indexes are ranked for each country according to the distance of their respective values of this coefficient from 1 before computing the across-country average rank. Ultimately, RCA indexes are ranked for each country according to their ability to minimize the non-parametric measure of ordinal ranking bias before computing the across-country average rank.

To obtain more synthetic rankings, a solution is to compute the mean rank of each RCA index for each of the three empirical properties. The underlying assumption is that all measures associated with a given empirical property have the same weight, which is theoretically plausible. In this regard, let  $R_{u,v} \in \mathbb{N}_*$  be the rank of RCA index  $u$  with respect to measure  $v \in V$ . The nonuple  $\langle R_{u,v} \in \mathbb{N}_* \rangle_{v \in V}$  is converted into the triplet  $(\bar{R}_{u,1}, \bar{R}_{u,2}, \bar{R}_{u,3}) = (1/3 \sum_{v \in V_1} R_{u,v}, 1/4 \sum_{v \in V_2} R_{u,v}, 1/2 \sum_{v \in V_3} R_{u,v})$  where:

- $V_1 = \{\beta, \rho, \sigma\}$ ;
- $V_2 = \{\text{Skewness, Pearson's 2}^{\text{nd}} \text{ skewness, } \overline{\text{out}}^1, \overline{\text{out}}^2\}$ ; and
- $V_3 = \{\text{Correlation coefficient between } \hat{\mathcal{K}} \text{ and } \check{\mathcal{K}}, \overline{\text{orb}}\}$

This triplet enables the identification of the most accurate RCA indexes in relation to time stationarity, shape and ordinal ranking bias for a given universe  $J \times K \times T$ . Note that an RCA index is not necessarily the best or among the best for each empirical property. Going further, one could calculate a “global” rank as a unique rank arising from the values of  $(\bar{R}_{u,1}, \bar{R}_{u,2}, \bar{R}_{u,3})$ . The simplest global rank is the simple arithmetic mean of  $(\bar{R}_{u,1}, \bar{R}_{u,2}, \bar{R}_{u,3})$ . Let  $\bar{R}_u^0$  be this global rank. Therefore:

$$\bar{R}_u^0 = \frac{1}{3}(\bar{R}_{u,1} + \bar{R}_{u,2} + \bar{R}_{u,3}) = \frac{1}{3} \left( \frac{1}{3} \sum_{v \in V_1} R_{u,v} + \frac{1}{4} \sum_{v \in V_2} R_{u,v} + \frac{1}{2} \sum_{v \in V_3} R_{u,v} \right) \quad (10)$$

This method of calculation assumes that the mean ranks associated with each of the three empirical properties have the same weight in determining the global rank. However, one empirical property might not be as important as the two others for a specific analysis. In this regard, using the simple arithmetic mean is not necessarily relevant. Consequently, two other solutions are suggested. The first solution is to maintain the ranking of RCA indexes for each empirical property separately instead of converting  $(\bar{R}_{u,1}, \bar{R}_{u,2}, \bar{R}_{u,3})$  into a (single) global rank. The second solution consists of the two following weighted arithmetic means denoted as  $\bar{R}_u^1$  and  $\bar{R}_u^2$ :

$$\begin{cases} \bar{R}_u^1 = \frac{1}{4}(\bar{R}_{u,1} + \bar{R}_{u,2}) + \frac{1}{2}\bar{R}_{u,3} \\ [9pt]\bar{R}_u^2 = \frac{1}{6}\bar{R}_{u,1} + \frac{1}{3}\bar{R}_{u,2} + \frac{1}{2}\bar{R}_{u,3} \end{cases} \quad (11)$$

Regarding  $\bar{R}_u^1$ , ordinal ranking bias alone determines 50% of the global rank, whereas the other empirical properties—time stationarity and shape—determine equally the other 50% of the global rank.  $\bar{R}_u^1$  implies that ordinal ranking bias is twice as important as time stationarity or shape. On the one hand, ordinal ranking bias concerns the informational content of RCA indexes. The measurement of ordinal ranking bias evaluates the ability of an RCA index to provide reliable information about the strengths and weaknesses of a country relative to others in international trade. In this regard, the measurement of ordinal ranking bias evaluates the ability of an RCA index to provide useful guidance for economic policy. On the other hand, time stationarity and shape are empirical properties that are more concerned with *how closely* an RCA index resembles the ideal of an RCA index that is stable over time, symmetric and without fat tails. Contrary to ordinal ranking bias, time stationarity and shape are not concerned with *what* an RCA index tells us for economic policy purposes. Consequently, calculating  $\bar{R}_u^1$  amounts to a policy-oriented standpoint according to which the informational content of RCA indexes is more important than the compatibility of RCA indexes with the ideal of an RCA index that is stable over time, symmetric and without fat tails.

With respect to  $\bar{R}_u^2$ , the respective weights of  $\bar{R}_{u,1}$ ,  $\bar{R}_{u,2}$  and  $\bar{R}_{u,3}$  imply not only that ordinal ranking bias is twice as important as time stationarity and shape—similar to  $\bar{R}_u^1$  and the corresponding policy-oriented standpoint—but also that shape is twice as important as time stationarity. Indeed, trade patterns are changing more rapidly now than before because of factors such as geopolitics-induced trade tensions (e.g. rising prices of various commodities), shifting consumer preferences, new developments in digital technology and the COVID-19 pandemic. Provided that trade patterns arise from comparative advantages, more rapidly changing trade patterns imply that comparative advantages are less sticky over time than before. Ultimately, this shift reduces the importance of time stationarity relative to shape when considering how closely an RCA index resembles the ideal RCA index. Simultaneously, the informational content of RCA indexes still determines 50% of the global rank.

**Example 6** Thirty-three RCA indexes are ranked in the case of the 44-country trade area comprising Italy and its main trading partners. The corresponding product classification consists of the 26 sectors arising from the EORA database (agriculture, fishing, mining and quarrying, food and beverages, etc.). The RCA indexes are those described in Sect. 2 (see Tables 1, 2 and 3). The RC and CTB indexes based on adjusted trade flows are calculated twice: first using the first available year as the reference year for adjusting trade flows (1990) and then using the last available year (2017).

**Table 6** Ranking of RCA indexes: 44-country trade area comprising Italy and its main trading partners, EORA26 product classification, 1990–2017

RCA	$\beta$	$\rho$	$\sigma$	TS	Sk	P	$\overline{\text{out}}^1$	$\overline{\text{out}}^2$	SH	S	$\overline{\text{orb}}$	ORB
B	11	32	31	24.67	33	32	13.55	12.95	22.88	12.86	16.75	14.81
AB	24	15	15	18.00	18	15	9.34	20.73	15.77	23.27	25.39	24.33
SB	2	17	19	12.67	4	26	26.59	10.41	16.75	13.00	17.36	15.18
WB <sup>K</sup>	12	31	32	25.00	32	27	13.75	15.61	22.09	10.05	12.41	11.23
WB <sup>K</sup>	16	29	28	24.33	29	33	13.45	12.89	22.09	12.05	12.16	12.10
B2	13	28	18	19.67	1	16	23.45	8.75	12.30	21.00	10.95	15.98
B2D	1	33	33	22.33	22	21	23.45	8.75	18.80	17.84	16.70	17.27
B2G	17	23	30	23.33	24	17	12.16	16.05	17.30	20.89	8.05	14.47
M	22	16	16	18.00	17	14	10.20	22.68	15.97	23.52	25.61	24.57
RC	18	7	4	9.67	3	19	22.48	10.16	13.66	12.82	15.55	14.18
RC <sup>f</sup>	7	20	17	14.67	2	18	23.18	10.39	13.39	12.45	15.91	14.18
RC <sup>l</sup>	3	18	20	13.67	6	20	21.89	10.41	14.57	12.68	15.82	14.25
RC <sup>Y</sup>	3	18	20	13.67	15	28	22.48	10.16	18.91	13.05	14.55	13.80
RC <sup>Y,f</sup>	8	25	25	19.33	11	29	23.18	10.39	18.39	12.39	14.36	13.38
RC <sup>Y,l</sup>	5	26	26	19.00	16	30	21.89	10.41	19.57	13.20	14.39	13.80
RC <sup>y</sup>	5	26	26	19.00	10	24	22.48	10.16	16.66	17.14	16.86	17.00
RC <sup>y,f</sup>	10	21	22	17.67	7	23	23.18	10.39	15.89	16.64	17.39	17.01
RC <sup>y,l</sup>	9	24	23	18.67	13	25	21.89	10.41	17.57	17.25	16.82	17.03
NB	18	7	4	9.67	27	8	9.34	20.73	16.27	18.86	18.70	18.78
Z	14	30	29	24.33	31	31	16.52	9.36	21.97	11.02	4.00	7.51
AZ	15	22	24	20.33	14	22	26.80	8.39	17.80	4.09	5.07	4.58
CTB <sup>W</sup>	20	4	2	8.67	28	3	8.34	19.14	14.62	21.11	20.32	20.72
CTB <sup>W,f</sup>	21	14	11	15.33	30	2	10.43	21.57	16.00	20.77	19.82	20.30
CTB <sup>W,l</sup>	28	10	12	16.67	25	7	9.57	16.89	14.61	18.57	18.61	18.59
CTB <sup>Y</sup>	27	5	8	13.33	19	12	8.34	19.14	14.62	21.89	22.27	22.08
CTB <sup>Y,f</sup>	26	6	6	12.67	20	11	10.43	21.57	15.75	21.55	21.39	21.47
CTB <sup>Y,l</sup>	23	1	7	10.33	9	13	9.57	16.89	12.11	19.07	20.41	19.74
CTB <sup>W,Y</sup>	29	3	1	11.00	8	4	8.34	19.14	9.87	21.27	22.09	21.68
CTB <sup>W,Y,f</sup>	25	11	9	15.00	12	1	10.43	21.57	11.25	21.18	22.48	21.83
CTB <sup>W,Y,l</sup>	30	13	10	17.67	5	10	9.57	16.89	10.36	18.61	20.00	19.31
CTB <sup>W,y</sup>	33	2	3	12.67	23	6	8.34	19.14	14.12	20.93	19.57	20.25
CTB <sup>W,y,f</sup>	31	9	13	17.67	26	5	10.43	21.57	15.75	20.23	19.73	19.98
CTB <sup>W,y,l</sup>	32	12	14	19.33	21	9	9.57	16.89	14.11	18.86	19.48	19.17

Source: Authors' calculations

TS time stationarity, Sk skewness, P Pearson's skewness, SH Shape, S Spearman's rank correlation coefficient, ORB ordinal ranking bias. With respect to RC, r in RC<sup>r</sup>, RC<sup>Y,r</sup> and RC<sup>y,r</sup> equals f to refer to "first-year adjustment of trade flows" (r = f = 1990) and equals l to refer to "last-year adjustment of trade flows" (r = l = 2017). The same notation applies to CTB indexes

**Table 7** Top 5 CTB indexes according to different ranking methods, trade area comprising Italy and its main trading partners, EORA26 product classification, 1990–2017

	TS ( $\bar{R}_{u,1}$ )	SH ( $\bar{R}_{u,2}$ )	ORB ( $\bar{R}_{u,3}$ )	$\bar{\bar{R}}_u^0$	$\bar{\bar{R}}_u^1$	$\bar{\bar{R}}_u^2$
1	CTB <sup>W</sup> 8.67	CTB <sup>W,Y</sup> 9.87	AZ 4.58	RC	12.50	AZ 11.82
2	NB+RC 9.67	CTB <sup>W,Y,I</sup> 10.36	Z 7.51	CTB <sup>Y,I</sup>	14.06	RC 12.92
3		CTB <sup>W,Y,f</sup> 11.25	WB <sup>K</sup> 11.23	RC <sup>f</sup>	14.08	RC <sup>f</sup> 14.11
4	CTB <sup>Y,I</sup> 10.33	CTB <sup>Y,I</sup> 12.11	WB <sup>f</sup> 12.10	RC <sup>f</sup>	14.16	RC <sup>f</sup> 14.19
5	CTB <sup>W,Y</sup> 11.00	B2 12.30	RC <sup>Y,f</sup> 13.38	CTB <sup>W,Y</sup>	14.18	SB 14.95
						Z 15.14

Source: Authors’ calculations

TS time stationarity, SH shape, ORB ordinal ranking bias.  $\bar{\bar{R}}_u^0$  is the simple arithmetic mean of the ranks associated with TS, SH and ORB.  $\bar{\bar{R}}_u^1$  is the policy-oriented weighted mean with two times more weight for ORB than for TS and SH.  $\bar{\bar{R}}_u^2$  is the policy-oriented weighted mean with two times more weight for ORB than for TS and SH, and with two times more weight for SH than for TS

Table 6 shows the ranking of these RCA indexes (variable  $R_{u,v}$ ) after calculating the nine measures evaluating their empirical relevance as illustrated in Examples 1–5. Variables  $\bar{R}_{u,1}$ ,  $\bar{R}_{u,2}$  and  $\bar{R}_{u,3}$ , which are the mean ranks associated with each of three empirical properties—time stationarity (TS), shape (SH) and ordinal ranking bias (ORB), respectively—are also reported. Table 7 shows the five RCA indexes with the highest ranks (i.e. the RCA indexes with the lowest values of  $\bar{R}$  or  $\bar{\bar{R}}$ ) according to the four different methods discussed previously. In this example, CTB indexes have a clear advantage regarding time stationarity and shape but not ordinal ranking bias. Z, AZ and the weighted versions of B are better suited to avoiding ordinal ranking bias. When the three empirical properties are combined as simple or weighted arithmetic means, RC is the best option ( $\bar{\bar{R}}^0$ ) or the second best option ( $\bar{\bar{R}}^1$  and  $\bar{\bar{R}}^2$ ) after AZ. The trade-adjusted versions of RC are ranked third and fourth, respectively. There is only one CTB index represented in these three global rankings, and both Z and SB appear a single time with rank five.

The next section provides a general empirical assessment of RCA indexes on the basis of  $\langle R_{u,v} \in \mathbb{N}_* \rangle_{v \in V}$ , then  $(\bar{R}_{u,1}, \bar{R}_{u,2}, \bar{R}_{u,3})$  and ultimately the global ranks  $\bar{\bar{R}}_u^0$ ,  $\bar{\bar{R}}_u^1$  and  $\bar{\bar{R}}_u^2$ .

## 7 An empirical assessment

We compute the RCA indexes in Sect. 2, apply the set of measures in Sect. 3–5 and then rank RCA indexes as discussed in Sect. 6 for different trade areas (i.e. different values of  $J$ ) and different product classifications (i.e. different values of  $K$ ); the time span ( $T$ ) will be the longest available depending on the database employed to collect the trade data. The objective of working with a sample of trade areas instead of a single trade area is to obtain more general results about the empirical relevance of RCA indexes. According to the same logic, working with different product



**Table 8** Main trading partners: the example of Nepal

	Country	Share of Nepal's total trade	Cumulative share (%)
1	India	61.28%	61.28
2	China	10.19%	71.47
3	United States of America	3.02%	74.49
4	United Arab Emirates	2.55%	77.04
5	Singapore	1.71%	78.75
6	Indonesia	1.66%	80.41
7	Germany	1.57%	81.98
8	Thailand	1.27%	83.24
9	Japan	1.13%	84.38
10	France	1.13%	85.50
11	Argentina	1.09%	86.60
12	Canada	1.09%	87.69
13	Hong Kong	1.04%	88.72
14	Malaysia	1.02%	89.74
15	South Korea	0.95%	90.69
	Other countries	Less than 0.95% each	100

Source: Authors' calculations based on yearly trade data from 1995 to 2020 in *UNCTADstat*

classifications provides an opportunity to analyze the influence of product classification on the empirical relevance of RCA indexes.

To build a sample of trade areas, we follow the path suggested by Stellan and Danna-Buitrago (2022). We conceptualize a trade area by gathering a given country—the “reference” country—and its “main trading partners”. Using trade data supplied by *COMTRADE* and then refined by *UNCTADstat*, all countries are ranked according to their respective shares in the reference country's total trade over the 26-year time span of that database, namely 1995–2020. The set of main trading partners is the smallest set of countries with the highest shares and that together represent at least 90% of the reference country's total trade.<sup>9</sup> Table 8 illustrates the conceptualization of the kind of trade area with Nepal as the reference country.<sup>10</sup> Nepal has fifteen main trading partners, so the corresponding trade area comprises sixteen countries. From 1995 to 2020, 61.28% of Nepal's total trade with the rest of the world consists of exports to India or imports from this country. No other country has a higher share, so India is the first main trading partner of Nepal. China and the

<sup>9</sup> Formally,  $\tilde{x}_{ij} = \sum_{k \in K} \sum_{t \in T} (x_{ijkt} + x_{jik t})$  denotes the total trade between countries  $i$  and  $j$  over time span  $T$  and product classification  $K$ . Country  $j$  is ranked as the  $p$ -th trading partner of country  $i$  among a set  $P$  of countries if  $\#\{j' : \tilde{x}_{ij'} > \tilde{x}_{ij} \text{ and } j, j' \in P\} = p - 1$ . This rank is written as  $r_{ij}$ . Ultimately, the subset  $J \subset P$  of countries that give rise to a trade area on the basis of reference country  $i$  is defined as  $\arg \min_{J \subset P} \sum_{j \in J} r_{ij} \text{ s.t. } \sum_{j \in J} \tilde{x}_{ij} \geq 90\%$ .

<sup>10</sup> Trade areas are available as supplementary material in an online repository ([click here](#)).

**Table 9** List of reference countries for the conceptualization of trade areas

Albania (22)	China (41)	Guatemala (25)	Mexico (13)	Romania (29)
Argentina (36)	Taiwan (25)	Honduras (20)	Morocco (34)	Senegal (39)
Australia (25)	Colombia (28)	India (44)	Nepal (16)	South Africa (42)
Austria (27)	Costa Rica (26)	Indonesia (26)	Netherlands (36)	Spain (43)
Bangladesh (31)	Denmark (31)	Israel (29)	New Zealand (30)	Sweden (30)
Belarus (21)	Dom. Rep. (23)	Italy (44)	Nicaragua (20)	Switzerland (31)
Belgium (28)	Ecuador (23)	Jamaica (27)	Nigeria (29)	Thailand (32)
Bolivia (20)	Egypt (42)	Japan (30)	Norway (23)	Turkey (48)
Botswana (13)	El Salvador (20)	Kenya (37)	Pakistan (36)	Ukraine (44)
Brazil (41)	Estonia (24)	South Korea (35)	Panama (32)	UK (34)
Bulgaria (33)	Finland (32)	Latvia (24)	Paraguay (23)	USA (35)
Cameroon (35)	France (39)	Lithuania (22)	Peru (25)	
Canada (17)	Germany (35)	Luxembourg (18)	Philippines (19)	
Chile (25)	Ghana (32)	Malaysia (23)	Portugal (29)	

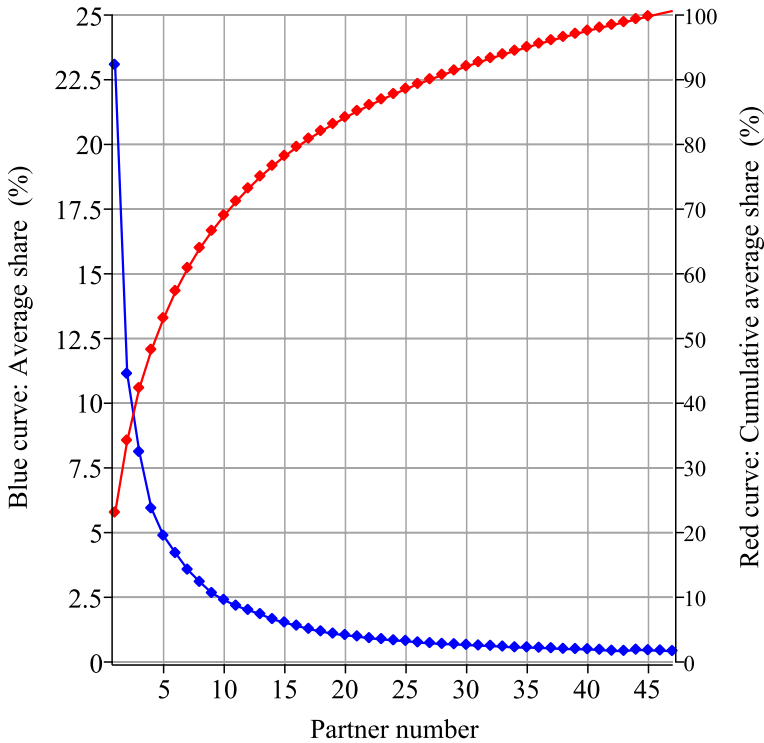
The numbers in parentheses are the respective numbers of countries in each trade area (the reference country is included in each number)

**Table 10** Number of trade areas each country belongs to in the sample of trade areas

CN	67	CH	48	ID	25	IR	13	KZ	7	BO	4	MD	3	NA	2	GN	1	GY	1
US	66	RU	44	CZ	24	PR	13	LT	7	BY	4	PY	3	SN	2	GQ	1	HT	1
DE	65	TW	41	AT	21	UA	13	MA	7	CI	4	SY	3	TG	2	GW	1	JO	1
FR	63	SE	36	DK	21	GV	11	SK	7	DO	4	TT	3	TM	2	KH	1	PG	1
GB	61	TR	36	VN	20	NG	11	EC	6	EE	4	UY	3	ZM	2	LA	1	RW	1
JP	61	HK	32	FI	18	RO	11	IQ	6	HR	4	BD	2	ZW	2	LK	1	SO	1
IT	60	TH	32	IE	18	IL	10	QA	6	LV	4	BF	2	GM	1	LS	1	SZ	1
NL	60	AE	31	HU	17	PA	10	BG	5	LY	4	CD	2	GA	1	LU	1	TD	1
ES	57	SG	31	NO	17	AL	9	GT	5	PK	4	CG	2	AF	1	MH	1	TZ	1
KR	56	AU	28	ZA	17	GR	9	HN	5	SI	4	CM	2	BB	1	ML	1	UG	1
BE	53	MY	28	VE	16	PE	9	NI	5	TN	4	GH	2	BH	1	MM	1	UZ	1
BR	53	PL	28	CL	15	PH	9	OM	5	AZ	3	LB	2	BW	1	MZ	1		
CA	52	SA	28	AR	14	CR	7	SV	5	GE	3	MK	2	CY	1	NP	1		
IN	51	MX	27	CO	14	KW	7	AO	4	KE	3	MR	2	FJ	1	NZ	1		

Source: Authors' calculations. Note: Countries are identified by their respective Alpha-2 ISO codes: CN for China, US for the United States, DE for Germany, etc

United States of America represent 10.19% and 3.02% of Nepal's total trade, respectively, and thus are the second and third main trading partners of Nepal, respectively. Together, India, China and the US represent 74.49% of Nepal's total trade. Twelve other countries must be taken into account so that this cumulative share tends toward (without being less than) 90%: the United Arab Emirates, Singapore, Indonesia, ... and South Korea.



**Fig. 6** Trading partner's average share of reference country's 1995–2020 total trade *Source:* Authors' calculations

This kind of trade area is conceptualized for 67 reference countries from each main region of the world and with diverse levels of development. Table 9 describes the set of reference countries. The minimum number of countries in a trade area is 13 (Mexico as the reference country with a set of main trading partners comprising 12 countries), the maximum is 48 (Turkey), and the mean is 29.41; 137 countries are represented in at least one trade area either as a reference country or as a main trading partner. Table 10 describes these countries and the number of trade areas they belong to. This number ranges from 1 (e.g. Uzbekistan) to 67 (China, which belongs to all trade areas under consideration). On average, a country belongs to 13.91 trade areas. Ultimately, Fig. 6 shows the average trade share that makes a country the  $j$ -th main trading partner of a reference country. Specifically, on average the first main trading partner represents approximately 23% of a reference country's total trade, the second main trading partner approximately 11%, and so on for the other main trading partners. On average, half of total trade is associated with the first five main trading partners.

Regarding  $K$ , we use three different product classifications. The first two classifications are the 2-digit SITC (65 items) and the 3-digit SITC, respectively. SITC is employed by *UNCTADstat* to report trade flows, and we use this classification

with two different levels of disaggregation to compute RCA indexes according to this database. These trade flows are “gross” trade flows, that is, exports and imports as reported by custom officials (and adjusted or complemented if necessary). However, gross trade flows may not provide sufficient information about international trade because of the international fragmentation of the production process and the subsequent global value chains. Indeed, “the export of a computer, for example, is in a fragmented world no longer reflecting the production of that computer from start to finish. The country involved might only contribute a (small) fragment of the production process, or in other words, add only a part of the total value added of the final product” (Brakman and Van Marrewijk 2017 p. 62).

For that reason, as the third product classification, we use the “value-added” trade flows provided by the *UNCTAD-EORAGlobal Value Chain (GVC) database* (Casella et al., 2019). In contrast to the SITC classifications, working with this database requires a simplified *26-sector classification* to enable the computation of RCA indexes for all trade areas under consideration. Indeed, the EORA sector classification is not the same depending on the reporting country. Value-added trade flows are reported according to these 26 sectors for various countries (e.g. Algeria, Egypt, Mali or Pakistan, among others) whereas for other countries these flows are reported according to a higher number of more disaggregated sectors, ranging from 49 (Canada) to 512 (UK). Consequently, to combine different countries into a trade area, sectors must be aggregated for countries reporting more than 26 sectors.<sup>11</sup>

To sum up, the use of value-added trade flows (EORA) instead of gross trade flows (SITC) implies a lower level of product/sector disaggregation (26 *versus* 65 or 259). We work with the aforementioned three classifications to account for different trade-offs between information accuracy (“gross” *versus* “value-added” trade flows) and the level of disaggregation among products/sectors.

For the RCA indexes that take into account GDP and GDP *per capita* (B2G, part of CTB indexes and part of RC indexes), we use the GDP data provided by *UNCTA Dstat*. As in Stellian and Danna-Buitrago (2022), the value of  $m$  to adjust an RCA index depending on GDP differentials in a trade area is set to 2; that is, GDP differentials at most double the level of comparative advantages or comparative disadvantages. For the period  $r$  that serves as the reference period to adjust trade flows for the calculation of some CTB indexes and some RC indexes, we use two different years: the first available year and the last available year. The corresponding values

<sup>11</sup> The Eora website contains a [concordance table](#) documenting how to implement the sectoral aggregation. For each country  $i$ , the concordance table contains a  $S_i \times 26$  matrix where  $S_i > 26$  is the number of disaggregated sectors associated with  $i$ . In that matrix, the element in row  $a$  and column  $b$  is a coefficient ranging from 0 to 1 that indicates which part of the value-added trade flow in the  $a$ -th disaggregated sector contributes to the  $b$ -th aggregated sector (the sum of coefficients for each row therefore equals 1). Let  $C_i$  denote that matrix and  $V_j$  the  $S_i \times \#J$  matrix whose element in row  $a$  and column  $j$  is the value-added trade flow of the  $a$ -th disaggregated sector from country  $i$  to the  $j$ -th country (among a set  $J$  of countries). Ultimately, sectoral aggregation arises from the standard matrix product  $C_i' \cdot V_j$ , which is the  $26 \times \#J$  matrix whose element in row  $b$  and column  $j$  is the value-added trade flow of the  $b$ -th aggregated sector from country  $i$  to the  $j$ -th country.

of  $r$  are written  $f$  (for “first”) and  $l$  (for “last”). For the SITC classifications, the first and last available years are 1995 and 2020, respectively. For the Eora-26 classification, the first and last available years are 1990 and 2017, respectively. Using the first available year for the adjustment of trade flows can be labelled a “forward-looking adjustment”, and using the last available year can be labelled a “backward-looking adjustment” (Stellian & Danna-Buitrago, 2017). Ultimately, regarding the Z and AZ indexes, the value of  $\theta$  is set to 6.534 as suggested by Costinot et al. (2012).

Note that an RCA index may face numeric exceptions. The online appendix explains how these numeric exceptions can be solved. Eventually, a full set of RCA indexes is computed. This set comprises:

- 33 RCA indexes: the nine RCA indexes in Table 1 (the standard B index and eight other RCA indexes related to the standard B index); nine RC indexes (the six RC indexes in Table 2, three of which are computed with two different values of reference period  $r$ ); Z and AZ; NY; and 12 CTB indexes (the eight CTB indexes in Table 3, four of which are computed with two different values of reference period  $r$ ).
- 67 different values of  $J$ . Each value is a trade area centered on the main trading partners of a given country among a representative sample of 67 countries (see Table 9).
- 2 values of  $K$  associated with  $T = \{1995, 1996, \dots, 2020\}$ , namely 2-digit SITC (65 items) and 3-digit SITC (259 items); and a third value of  $K$ , EORA-GVC (26 items), associated with  $T = \{1990, 1991, \dots, 2017\}$ .

Given the number of countries in each possible value of  $J$  (see Table 9), the number of items in each possible value of  $K$  (26, 65 or 259) and the number of years in each possible value of  $T$  (26 or 28), our empirical assessment is based on more 595 million RCA index values.

Table 11 reports the ten best RCA indexes according to the mean values of  $\bar{R}_{u,1}$ ,  $\bar{R}_{u,2}$  and  $\bar{R}_{u,3}$ , and then  $\bar{R}_u^0$ ,  $\bar{R}_u^1$  and  $\bar{R}_u^2$ . These mean values are computed for the whole sample of 67 trade areas and are reported separately for each of the three product classifications<sup>12</sup> (2-digit SITC, 3-digit SITC and EORA-26 GVC). The corresponding calculations are described in Sect. 6 and illustrated in Tables 6 and 7.

For the three product classifications, the class of CTB indexes has a clear advantage regarding time stationarity and shape. A CTB index is always ranked first, and at least 6 CTB indexes belong to the top 10. There is an even sharper contrast regarding shape: the top 10 is fully occupied by CTB indexes in relation to the 3-digit SITC, and all positions except one in the top 10 are occupied by CTB indexes in relation to the EORA-26 GVC. NB is in the second position in the top 10 regarding time stationarity for each product classification. However, NB does not remain in the top 10 in terms of shape in relation to the 3-digit SITC and EORA-2 GVC and is last

<sup>12</sup> The results of the auto-correlation tests (see Remark 1) are available upon request. For most RCA indexes, these tests do not find any serial correlation in the errors. We regard the remaining cases as inconclusive due to the low number of time periods in the database.

**Table 11** Top 10 RCA indexes, mean values for the sample of trade areas

	TS ( $\bar{R}_{it,1}$ )	SH ( $\bar{R}_{it,2}$ )	ORB ( $\bar{R}_{it,3}$ )	$\bar{R}_{it}^0$	$\bar{R}_{it}^1$	$\bar{R}_{it}^2$
<i>2-digit SITC</i>						
1	CTB <sup>W,y</sup>	CTB <sup>W</sup>	RC	CTB <sup>W</sup>	CTB <sup>W</sup>	RC
2	NB	CTB <sup>W,l</sup>	RC <sup>l</sup>	CTB <sup>W,y</sup>	CTB <sup>W,y</sup>	CTB <sup>W</sup>
3	CTB <sup>W,y</sup>	CTB <sup>W,y,l</sup>	RC <sup>l</sup>	CTB <sup>W,y</sup>	CTB <sup>W,y</sup>	CTB <sup>W,y</sup>
4	CTB <sup>W</sup>	CTB <sup>W,y</sup>	AZ	RC	RC	RC <sup>l</sup>
5	CTB <sup>y</sup>	CTB <sup>W,y</sup>	WB <sup>K</sup>	RC <sup>l</sup>	RC <sup>l</sup>	CTB <sup>W,y</sup>
6	CTB <sup>y,l</sup>	CTB <sup>W,f</sup>	B2G	RC <sup>l</sup>	RC <sup>l</sup>	RC <sup>l</sup>
7	CTB <sup>y,f</sup>	RC <sup>l</sup>	RC <sup>y</sup>	AZ	AZ	AZ
8	M	RC	RC <sup>y</sup>	NB	NB	RC <sup>y</sup>
9	AB	CTB <sup>W,y,l</sup>	CTB <sup>W</sup>	B2	B2	B2
10	AZ	NB	CTB <sup>W,y</sup>	SB	SB	RC <sup>y</sup>
<i>3-digit SITC</i>						
1	CTB <sup>W,y</sup>	CTB <sup>W,l</sup>	Z	CTB <sup>W</sup>	CTB <sup>W</sup>	RC <sup>l</sup>
2	NB	CTB <sup>W</sup>	RC <sup>l</sup>	RC <sup>l</sup>	RC <sup>l</sup>	RC
3	CTB <sup>W,y</sup>	CTB <sup>W,y</sup>	RC	CTB <sup>W,y</sup>	CTB <sup>W,y</sup>	RC <sup>l</sup>
4	CTB <sup>W</sup>	CTB <sup>W,y,l</sup>	RC <sup>l</sup>	RC	RC	Z
5	CTB <sup>y,l</sup>	CTB <sup>W,y,l</sup>	B2G	Z	Z	B2G
6	CTB <sup>y</sup>	CTB <sup>W,f</sup>	AZ	RC <sup>l</sup>	RC <sup>l</sup>	CTB <sup>W</sup>
7	M	CTB <sup>W,y,f</sup>	RC <sup>y,l</sup>	RC <sup>l</sup>	CTB <sup>W,y</sup>	AZ
8	CTB <sup>y,f</sup>	CTB <sup>W,y</sup>	RC <sup>y</sup>	RC <sup>y</sup>	CTB <sup>W,y</sup>	CTB <sup>W,y</sup>
9	AB	CTB <sup>y</sup>	RC <sup>y,l</sup>	RC <sup>y</sup>	CTB <sup>W,y</sup>	RC <sup>y,l</sup>
10	Z	CTB <sup>y,l</sup>	RC <sup>y</sup>	NB	NB	RC <sup>y</sup>

Table 11 (continued)

	TS ( $\bar{R}_{u,1}$ )	SH ( $\bar{R}_{u,2}$ )	ORB ( $\bar{R}_{u,3}$ )	$\bar{R}_u^0$	$\bar{R}_u^1$	$\bar{R}_u^2$						
<i>EORA-26 GVC</i>												
1	CTB <sup>W</sup>	7.25	CTB <sup>W,Y,f</sup>	10.92	AZ	6.10	RC	12.34	RC	9.25	AZ	11.73
2	NB	8.61	CTB <sup>W,Y</sup>	11.82	Z	9.48	CTB <sup>W</sup>	13.61	CTB <sup>W</sup>	10.21	RC	13.18
3	RC	8.61	CTB <sup>W,y</sup>	12.90	WB <sup>K</sup>	11.11	AZ	13.66	AZ	10.24	RC <sup>f</sup>	13.93
4	CTB <sup>W,Y</sup>	10.20	CTB <sup>W</sup>	13.56	WB <sup>f</sup>	12.98	RC <sup>f</sup>	14.13	RC <sup>f</sup>	10.60	RC <sup>f</sup>	14.36
5	CTB <sup>Y,f</sup>	11.93	CTB <sup>W,y,f</sup>	13.61	RC <sup>f</sup>	13.18	NB	14.26	NB	10.69	B2	14.87
6	CTB <sup>W,y</sup>	12.09	B2	13.65	RC	13.65	RC <sup>f</sup>	14.59	RC <sup>f</sup>	10.94	SB	14.91
7	CTB <sup>Y,l</sup>	12.55	CTB <sup>W,Y,l</sup>	13.86	RC <sup>f</sup>	13.93	CTB <sup>W,Y</sup>	14.63	CTB <sup>W,Y</sup>	10.97	CTB <sup>W</sup>	15.74
8	RC <sup>Y</sup>	14.41	CTB <sup>W,y,f</sup>	14.10	B	14.06	CTB <sup>W,y</sup>	15.03	CTB <sup>W,y</sup>	11.27	Z	15.76
9	RC <sup>f</sup>	14.41	CTB <sup>W,l</sup>	14.11	B2G	14.24	SB	15.03	SB	11.27	RC <sup>Y</sup>	15.77
10	CTB <sup>W,f</sup>	14.65	CTB <sup>W,f</sup>	14.26	SB	14.29	B2	15.49	B2	11.62	NB	15.93

Source: Authors' calculations

in the top 10 in relation to the 2-digit SITC. Consequently, our empirical assessment suggests that CTB indexes are most able to generate empirical measurements that fit the ideal of an RCA index that is stable over time, symmetric and without fat tails.

In addition, as argued by Stellan and Danna-Buitrago (2022), CTB indexes present *theoretical* robustness, namely robustness before any empirical assessment. Indeed, the formulas defining CTB indexes present the following features independently of the universe  $J \times K \times T$  under consideration:

1. Compatibility with the Kunimoto-Vollrath principle, according to which the magnitude of the trade balance or another trade-related variable is compared to its theoretical—“expected”—value to reveal comparative advantages;
2. Calculation on the basis of the overall structure of exports and imports to adapt an RCA index to the relative nature of comparative advantages and both the supply- and demand-side dimensions of comparative advantages;
3. Symmetry around its neutral point to ensure that comparative advantages and comparative disadvantages are measured in a homogeneous way;<sup>13</sup>
4. Use of GDP data to ensure a more precise measurement of comparative advantages;
5. In some cases, additivity across products and even countries to make measurements of comparative advantages independent of product/country classifications.

Ultimately, CTB indexes warrant attention because they resemble most the ideal RCA index—stable over time, minimizing both asymmetry and tail fatness—and because of their theoretical robustness. However, with respect to *what* an RCA index tells us for economic policy purposes, Z, AZ and the class of RCA indexes based on RC tend to be more reliable RCA indexes because they are more able than CTB indexes to avoid ordinal ranking bias. For each product classification:

- Z and AZ belong to the top 10, or at least AZ belongs to the top 10 and therefore places regression-based indexes among the RCA indexes most able to avoid ordinal ranking bias. In addition, the first rank is occupied by Z in relation to the 3-digit SITC, while AZ and Z rank first and second, respectively, in relation to the EORA-26 GVC.
- At least RC,  $RC^l$ , and  $RC^f$  belong to the top 10. These three indexes even occupy the three first ranks in the case of the 2-digit SITC. The other RCA indexes of the same class are present among the ten best RCA indexes:  $RC^Y$  and  $RC^Y$  in relation to the 2-digit SITC and 3-digit SITC and  $RC^{Y,l}$  and  $RC^{Y,l}$  in relation to the 3-digit SITC.

<sup>13</sup> This use of the concept of symmetry is different from symmetry in the distribution of an RCA indexes throughout a given value of  $J \times K$ . Rather, at the theoretical level, an RCA index is symmetric if the values revealing comparative disadvantages lie within the interval  $[\phi - a, \phi)$  or  $(-\infty, \phi)$  while the values determining comparative advantages lie within the interval  $(\phi, \phi + a)$  or  $(\phi, +\infty)$ , where  $a > 0$  and  $\phi$  is the neutral value of an RCA index.



No other RCA index appears so frequently in each top 10, which makes the RC index, its further improvements and the regression-based RCA indexes most able to avoid ordinal ranking bias and therefore the most useful RCA indexes for economic policy purposes. In addition, as explained by Danna-Buitrago and Stellan (2022) and Stellan and Danna-Buitrago (2022), the theoretical robustness of these RCA indexes is close to the theoretical robustness of CTB indexes. Ultimately, CTB indexes are less able than regression-based indexes or RC-inspired indexes to avoid ordinal ranking bias, but their informational content should not be ignored because CTB indexes are most compatible with the ideal of an RCA index that is stable over time with the most desirable shape.

When the ranks arising from time stationarity, shape and ordinal ranking bias are combined into a global rank, each product classification gives rise to a specific situation. In the case of the 2-digit SITC, the highest positions are occupied by the same three CTB indexes— $CTB^W$ ,  $CTB^{W,y}$  and  $CTB^{W,Y}$ —and three RC indexes—RC,  $RC^l$  and  $RC^f$ . In summary, with the 2-digit SITC and under different weights given to time stationarity, shape and ordinal ranking bias, *the best RCA indexes are the CTB indexes with normalization by total trade or normalization by both total trade and the GDP-related coefficient  $m^{-y_{it}/\bar{y}_t+1}$  or  $m^{-y_{it}/\bar{y}_t+1}$ , and the standard RC index with or without adjustment of trade flows*. Note that, interestingly, none of these RCA indexes combine the GDP-related coefficient and the use of adjusted trade flows. This kind of combination confers greater theoretical robustness but is absent from the top 10. This suggests a trade-off between theoretical robustness and empirical accuracy. Ultimately, the last positions are occupied by AZ and B2, as well two RCA indexes among SB, NB,  $RC^Y$  and  $RC^y$ .

In the case of the 3-digit SITC, compared with the 2-digit SITC, RC,  $RC^l$  and  $RC^f$  remain classified among the six first positions. Nonetheless, these positions are no longer shared with  $CTB^W$ ,  $CTB^{W,y}$  and  $CTB^{W,Y}$  exclusively. On the contrary, Z is now classified as fifth regarding  $\bar{R}_u^0$  and  $\bar{R}_u^1$  and fourth in terms of  $\bar{R}_u^2$ ; B2G occupies the fifth position when the global rank is calculated as  $\bar{R}_u^2$ .  $CTB^W$  still occupies the best position with respect to  $\bar{R}_u^0$  and  $\bar{R}_u^1$  but falls to sixth regarding  $\bar{R}_u^2$ .  $CTB^{W,y}$  occupies the third position according to  $\bar{R}_u^0$  and  $\bar{R}_u^1$  but falls to eighth regarding  $\bar{R}_u^2$ .  $CTB^{W,Y}$  occupies the seventh position in terms of  $\bar{R}_u^0$  and  $\bar{R}_u^1$  but no longer appears among the top 10 regarding  $\bar{R}_u^2$ . In summary, with the 3-digit SITC and under different weights given to time stationarity, shape and ordinal ranking bias, *CTB indexes with normalization by both total trade and the GDP-related coefficient are less relevant from an empirical standpoint, and Z gains empirical relevance*. Interestingly, other CTB indexes appear among the lowest positions of the top 10, namely  $CTB^{W,l}$  and  $CTB^{W,f}$ .

In the case of the EORA-26 GVC, the same classification arises from  $\bar{R}_u^0$  and  $\bar{R}_u^1$ , where RC,  $CTB^W$  and AZ occupy the three first positions, followed by  $RC^l$ , NB and  $RC^f$ . If the weights underlying a global rank are those of  $\bar{R}_u^2$ , with less importance given to time stationarity than shape, AZ occupies the first position followed by RC with or without adjustment of trade flows.

As a general result, we can say that the standard RC index with or without adjustment of trade flows tend to provide relevant measures of comparative advantages

independently of product classification; CTB indexes without adjustment of trade flows and normalization by total trade combined or not with GDP adjustment are more reliable RCA indexes in relation to the 3-digit SITC; and regression-based indexes can usefully complement RC and CTB indexes. All these RCA indexes present greater theoretical robustness than the B index and the other RCA indexes based on exports only. Indeed, the main theoretical robustness of regression-based RCA indexes is that they are supported by a Ricardian model of international trade, whereas CTB indexes and RC indexes are RCA indexes arises from the Kunimoto-Vollrath principle, are calculated on the basis of the overall structure of exports and imports while being symmetric around their neutral point, and can be adapted to GDP differentials in a given trade area. In addition, some CTB indexes are additive across countries and/or products. Therefore, more theoretical robustness seems to confer greater empirical relevance to an RCA index. However, as suggested before, this positive relationship between theoretical robustness and empirical accuracy has some limits because no global rank includes among the ten best RCA indexes the RCA indexes that combine adjusted trade flows and a GDP-related coefficient.

## 8 Conclusion

The conventional RCA index from Balassa (1965) has been questioned multiple times but remains widely used along with many alternative RCA indexes. This puzzling situation suggests that there is still no clear consensus on which RCA index should be used to analyze the comparative advantages of a given set of countries, products and periods. This paper contributes to filling this gap by suggesting a consistent set of measures to evaluate the empirical properties of RCA indexes based on their time stationarity, shape and ordinal ranking bias. This set comprises nine measures that together rank different RCA indexes according to stability over time, the presence of a symmetric and thin-tailed distribution, and the absence of ordinal ranking bias. These measures combine GMM estimation, descriptive statistics and non-parametric measures.

We compute 33 RCA indexes for 67 different trade areas comprising different sets of countries centered on a reference country and its main trading partners. These computations were repeated for three different product classifications and, in relation to these classifications, two sets of time periods. Eventually, a database of more 595 million values of RCA indexes is used as the input to rank RCA indexes according to their empirical accuracy. Different methods of rankings can support a discussion in which the empirical properties of RCA indexes are balanced against their theoretical properties. Contribution-to-the-Trade-Balance (CTB) indexes, the "Revealed Competitiveness" index and its further improvements, and regression-based indexes tend to provide the most relevant measures of comparative advantages from an empirical standpoint. In addition, these RCA indexes tend to have the greatest theoretical robustness.

To conclude, three future lines of research are worth mentioning. The first line of research is the use of alternative dynamic panel data methods to measure the time stationarity of RCA indexes, with the aim of working with more flexible

serial correlation assumptions for the residuals. The second line of research is other methods of evaluating ordinal ranking bias. Specifically, there is an ordinal ranking bias if a country has lower (higher) rank than  $x$  with a value of an RCA index greater (lower) than the mean value leading the same country to have rank  $x$ . Instead of the mean value, other representative values like the median or quartiles might be used to measure ordinal ranking bias. The third line of research is to analyze the average ranking of RCA indexes for different subsamples in which reference countries and/or their respective main trading partners present some specific characteristics instead of analyzing the average ranking of RCA indexes throughout the sample of 67 trade areas. This kind of exercise would strengthen the analysis of RCA indexes from the vantage point of their empirical relevance. Ultimately, further examining the empirical properties of RCA indexes will help academics and practitioners select the most adequate RCA index without ignoring the theoretical foundations of RCA indexes.

**Acknowledgements** We would like to thank Sergi Basco, Bruno Casella, Daniel Moran and Gianluca Orefice for useful comments and suggestions. The usual disclaimer applies.

**Author contributions** All data used in this manuscript are available online or upon request. All authors (i) made substantial contributions to the final manuscript; and (ii) read and approved the final manuscript.

**Funding** Open Access funding provided by Colombia Consortium. This work was supported by *Pontificia Universidad Javeriana*, Bogotá, Colombia (Grant No. 9486).

## Declarations

**Conflict of interest** The authors have no competing interests to declare.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Algieri, B., Aquino, A., & Succurro, M. (2022). Trade specialisation and changing patterns of comparative advantages in manufactured goods. *Italian Economic Journal*. <https://doi.org/10.1007/s40797-022-00185-4>
- Amador, J., Cabral, S., & Maria, J. R. (2011). A simple cross-country index of trade specialization. *Open Economies Review*, 22(3), 447–461.
- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2), 277–297.
- Bachetta, M., Beverelli, C., Cadot, O., Fugazza, M., Grether, J.-M., Helble, M., Nicita, A., & Piermartini, R. (2012). *A practical guide to trade policy analysis*. WTO.
- Bagci, K. (2016). Measuring within diversification. *International Trade Journal*, 30(3), 223–236.

- Bahar, D., Rosenow, S., Stein, E., & Wagner, R. (2019). Export take-offs and acceleration: Unpacking cross-sector linkages in the evolution of comparative advantage. *World Development*, 117, 48–60.
- Balassa, B. A. (1965). Trade liberalization and revealed comparative advantage. *The Manchester School of Economic and Social Studies*, 33(2), 92–123.
- Balassa, B. A. (1986). Comparative advantages in manufactured goods: A reappraisal. *Review of Economics and Statistics*, 68(2), 315–319.
- Barattieri, A. (2014). Comparative advantage, service trade, and global imbalances. *Journal of International Economics*, 92(1), 1–13.
- Brakman, S., & Van Marrewijk, C. (2017). A closer look at revealed comparative advantage: Gross-versus value-added trade flows. *Papers in Regional Science*, 96(1), 61–92.
- Cameron, C., & Trivedi, P. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.
- Carrère, C., & Strauss-Kahn, V. (2017). Export survival and the dynamics of experience. *Review of World Economics*, 153(2), 271–300.
- Casella, B., Bolwijn, R., Moran, D., & Kanemoto, K. (2019). Improving the analysis of global value chains: The UNCTAD-Eora database. *Transnational Corporations*, 26(3), 115–142.
- Costinot, A., Donaldson, D., & Komunjer, I. (2012). What goods do countries trade? A quantitative exploration of Ricardo's ideas. *Review of Economic Studies*, 79, 581–068.
- Dalum, B., Laursen, K., & Villumsen, G. (1998). Structural change in OECD export specialisation patterns: De-specialisation and 'stickiness'. *International Review of Applied Economics*, 12(3), 423–443.
- Danna-Buitrago, J. P. (2017). La Alianza del Pacífico+4 y la especialización regional de Colombia: Una aproximación desde las ventajas comparativas. *Cuadernos de Administración*, 55, 39–52.
- Danna-Buitrago, J. P., & Stellian, R. (2022). A new class of revealed comparative advantage indexes. *Open Economies Review*, 33, 477–503.
- De Benedictis, L., & Tamberi, M. (2004). Overall specialization empirics: Techniques and applications. *Open Economies Review*, 15(4), 323–346.
- De Saint Vaulry, A. (2008). Base de données CHELEM commerce international du CEPII. Technical Report 9, Paris, Centre d'études Prospectives et d'Informations Internationales.
- Donges, J., & Riedel, J. (1977). The expansion of manufactured exports in developing countries: An empirical assessment of supply and demand issues. *Weltwirtschaftliches Archiv*, 113(1), 58–87.
- French, S. (2017). Revealed comparative advantage: What is it good for? *Journal of International Economics*, 106, 83–103.
- Gallardo, J. L. (2005). Comparative advantage, economic growth and free trade. *Revista de Economía Contemporánea*, 9, 313–335.
- Gnidchenko, A., & Salnikov, V. (2015). Net comparative advantage index: Overcoming the drawbacks of the existing indices. Working paper BRP 119/EC/2015, National Research University, Higher School of Economics.
- Hansen, L. (1991). Large sample properties of generalized method of moments estimators. *Econometrica*, 59, 1029–1054.
- Hoen, A. R., & Oosterhaven, J. (2006). On the measurement of comparative advantage. *The Annals of Regional Science*, 40(3), 677–691.
- Kathuria, L. M. (2018). Comparative advantages in clothing exports: India faces threat from competing nations. *Competitiveness Review*, 28(5), 518–540.
- Kvam, P. H., & Vidakovic, B. (2007). *Nonparametric statistics with applications to science and engineering*. Wiley.
- Lafay, G. (1987). Avantage comparatif et compétitivité. *Économie Prospective Internationale*, 29, 39–52.
- Lafay, G. (1992). The measurement of revealed comparative advantages. In M. G. Dagenais & P.-A. Muet (Eds.), *International trade modelling* (pp. 209–234). Chapman & Hall.
- Laursen, K. (2015). Revealed comparative advantage and the alternatives as measures of international specialization. *Eurasian Business Review*, 5(1), 99–115.
- Lectard, P., & Rougier, E. (2018). Can developing countries gain from defying comparative advantage? Distance to comparative advantage, export diversification and sophistication, and the dynamics of specialization. *World Development*, 102, 90–110.
- Leromain, E., & Orefice, G. (2014). New revealed comparative advantage index: Dataset and empirical distribution. *International Economics*, 139, 48–70.
- Liu, B., & Gao, J. (2019). Understanding the non-gaussian distribution of revealed comparative advantage index and its alter. *International Economics*, 158, 1–11.

- Michaely, M. (1962). *Concentration in international trade*. North-Holland.
- Nath, H. K., & Goswami, B. (2018). India's comparative advantages in services trade. *Eurasian Economic Review*, 8(2), 323–342.
- Nath, H. K., Liu, L., & Tochkov, K. (2015). Comparative advantages in U.S. bilateral services trade with China and India. *Journal of Asian Economics*, 38, 79–92.
- Proudman, J., & Redding, S. (1998). *Openness and Growth*. Bank of England.
- Proudman, J., & Redding, S. (2000). Evolving patterns of international trade. *Review of International Economics*, 8(3), 373–396.
- Stellian, R., & Danna-Buitrago, J. P. (2017). Competitividad de los productos agropecuarios colombianos en el marco del tratado de libre comercio con Estados Unidos: análisis de las ventajas comparativas. *Revista CEPAL*, 122, 139–163.
- Stellian, R., & Danna-Buitrago, J. P. (2019). Revealed comparative advantages and regional specialization: Evidence from Colombia in the Pacific Alliance. *Journal of Applied Economics*, 22(1), 349–379.
- Stellian, R., & Danna-Buitrago, J. P. (2022). Revealed comparative advantage and contribution-to-the-trade-balance indexes. *International Economics*, 170, 129–155.
- Vollrath, T. L. (1991). A theoretical evaluation of alternative trade intensity measures of revealed comparative advantage. *Weltwirtschaftliches Archiv*, 127(2), 265–280.
- Westfall, P. H. (2014). Kurtosis as peakedness, 1905–2014: R.I.P. *Am Stat*, 68(3), 191–195.
- Yeats, A. J. (1985). On the appropriate interpretation of the revealed comparative advantage index: Implications of a methodology based on industry sector analysis. *Weltwirtschaftliches Archiv*, 121(1), 61–73.
- Yu, R., Cai, J., & Leung, P. S. (2009). The normalized revealed comparative advantage index. *Annals of Regional Science*, 43(1), 267–282.
- Yu, R., Cai, J., Loke, M. K., & Leung, P. S. (2010). Assessing the comparative advantage of Hawaii's agricultural exports to the US mainland market. *Annals of Regional Science*, 45(2), 473–485.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.