CENTRE D'ÉTUDES PROSPECTIVES ET D'INFORMATIONS INTERNATIONALES

No 2013 – 20
June

DOCUMENT DE TRAVAIL

New Revealed Comparative Advantage Index: dataset and empirical distribution

---

Elsa Leromain, Gianluca Orefice

TABLE OF CONTENTS

New Revealed Comparative Advantage Index: dataset and empirical distribution

Elsa Leromain, Gianluca Orefice

Highlights

- This paper presents a database containing a new Revealed Comparative Advantage (RCA) index, based on an econometric estimation procedure recently suggested by Costinot et al. (2012)
- The new RCA index shows better better statistical properties than Balassa Index

Abstract

Balassa Index (Balassa 1965) is widely used in the literature to measure country-sector Revealed Comparative Advantage (RCA). However, being computed on observed trade flows, it mixes up all the factors influencing trade flows. In particular, Balassa Index cannot isolate exporter-sector (*ex ante*) specific factors which are the source of comparative advantage in the spirit of the traditional trade model. Furthermore, Balassa Index suffers some empirical distribution weaknesses, mainly time instability and poor ordinal ranking property (Yeats 1985; Hinloopen and Van Marrewijk 2001). A recent paper by Costinot et al. (2012) provides a micro-founded version of the Ricardian model and suggests a new measure for comparative advantage. We build up on this paper, and present a dataset providing a new econometric based measure for Ricardian RCA.

JEL Classification: F11, F14

Keywords: Revealed Comparative Advantage, Ricardian model, Exports.

NEW REVEALED COMPARATIVE ADVANTAGE INDEX: DATASET AND EMPIRICAL DISTRIBUTION

Elsa Leromain, Gianluca Orefice

POINTS CLEFS

- Ce papier présente une base de données contenant un nouvel indice d'avantage comparatif construit grâce à une méthode économétrique développée dans un article récent de Costinot et al.(2012).
- Ce nouvel indice s'avère avoir de meilleures propriétés statistiques que l'indice de Balassa.

RÉSUMÉ

L'indice de Balassa (Balassa 1965) est le plus utilisé dans la littérature pour mesurer les avantages comparatifs révélés (ACR) pour un pays et un secteur donnés. Cependant, sa construction s'appuyant sur des flux de commerce observés, cet indice n'est pas en mesure d'isoler les facteurs propres à l'exportateur ou au secteur qui sont à la source des avantages comparatifs dans les modèles traditionnels de commerce. De plus, l'indice de Balassa présente une distribution empirique conceptuellement problématique, puisque, notamment, elle n'est pas constante d'une année sur l'autre (Yeats 1985 ; Hinloopen and Van Marrewijk 2001). Un récent article de Costinot et al.(2012) développe une nouvelle version, micro-fondée, du modèle ricardien et suggère une nouvelle méthode de calcul des avantages comparatifs à partir d'une estimation économétrique. La base de données que nous présentons dans ce papier contient une nouvelle mesure des ACR inspirée de cette méthode.

Classification JEL : F11, F14

Mots clés : Avantage comparatif révélé, Modèle ricardien, Exportations

NEW REVEALED COMPARATIVE ADVANTAGE INDEX: DATASET AND EMPIRICAL DISTRIBUTION[1]

Elsa Leromain* Gianluca Orefice†

1. INTRODUCTION

Theoretical foundation and empirical measures of comparative advantage have long been analysed by trade economists. In particular, Ricardian comparative advantage has long been perceived as a useful pedagogical tool: a country should produce (and export) relatively more in those industries in which it is relatively more productive. However, Ricardian comparative advantage received little attention in empirical studies. The main reason behind this lack in empirical tests of Ricardian model is the absence of a clear theoretical micro-foundation and theoretically-consistent measure of comparative advantage.

In the last few years, since the seminal paper by Eaton and Kortum (2002),[2] we have seen a renewed interest in empirical works on the sources of comparative advantages - Chor (2010), Kerr (2009), Levchenko and Zhang (2011). But nothing has been done (to our knowledge) in improving synthetic empirical measures for Ricardian comparative advantage. In this paper we aim to fill this lack by providing a new dataset on Ricardian comparative advantage measures.

A recent paper by Costinot et al. (2012) provides a theoretical micro-foundation for the Ricardian model of trade. They build a structural Ricarian model with multiple countries and industries, one factor of production (labor), allowing for intra-industry heterogeneity (Eaton and Kortum 2002). In the process, they also propose a theoretically-consistent empirical measure for comparative advantage able to fit the Ricardian ideas of comparative advantage in a proper way.

The contribution by Costinot et al. (2012) revitalized the importance of technological differences (i.e. productivity) in analysing the patterns of trade, and therefore renewed the need for a proper measure of comparative advantage. Indeed, the Balassa Index of Revealed Comparative

[1]We are grateful to Arnaud Costinot, Matthieu Crozet, Lionel Fontagné and Sébastien Jean for helpful comments and suggestions. The usual disclaimers apply.
*Université Paris 1. Email: elsa-leromain@hotmail.fr
†CEPII. Email: gianluca.orefice@cepii.fr
[2]Eaton and Kortum (2002) introduced intra-industry productivity heterogeneity in a standard Ricardian model and show that, under an appropriate parametrization for the underlying distribution of productivity, productivity differences drive comparative advantages

Advantage (RCA) - as proposed by Bela Belassa (1965)[3] - has been widely used to approximate countries' sectorial specialization[4] but suffers both theoretical foundation and empirical distribution weaknesses.

The theoretical foundation of the Balassa Index has long been debated in the literature since it does not really match the original Ricardian idea of comparative advantage (Bowen 1983; Vollrath 1991)[5]. Ricardian comparative advantage, indeed, is based on the intrinsic (ex-ante) nature of the country in being relatively more efficient in the production of a certain good. Unfortunately, Balassa index fails in fitting this idea since it is based on the actual (ex-post) realization of bilateral sector's trade flows, mixing up exporter with importer and sector specific factors affecting trade.[6]

Balassa Index has also been criticized for its poor empirical distribution characteristics (Hinloopen and Van Marrewijk 2001; De Benedictis and Tamberi 2004): (i) it does not have a stable distribution over time (which is a crucial property in view of the ex-ante nature of Ricardian comparative advantage) and (ii) it provides poor ordinal ranking property (UNIDO 1982; Yeats 1985).

Several attempts have been made in the literature to overcome the former empirical weakness of the pure Balassa index: Lafay index (Lafay 1992) combines together trade and production variables, symmetric revealed comparative advantage indices (Dalum et al. 1998) and weighted RCA measures (Proudman and Redding 2000). However all these new indices only partially solve for the previous statistic distribution and cross-country comparison problems. Moreover, being based on ex-post trade flows, they still miss the *ex ante* spirit of Ricardian comparative advantage concept.

The new theoretically-consistent measure of Ricardian RCA proposed by Costinot et el. (2012) is able to isolate the exporter-specific factors driving trade flows, and thus it fits better the original idea of Ricardian comparative advantage.[7] Relying on an econometric technique, the new

[3]In the spirit of Balassa Index, a country's revealed comparative advantage in the trade of a certain industry is assessed by the share of that industry in the country's total exports relative to the industry's share in total world exports of manufactures.

[4]Amighini et al. (2011); Amiti (1999); Ferto and Hubbard (2003); Richardson and Zhang (1999).

[5]Bowen (1983) analysed the theoretical basis of Balassa Index and its interpretation. In particular he found that the interpretation of Balassa Index (trade intensity above one as a signal of comparative advantage) relies on the assumption that a certain country exports every commodity, this assumption being unrealistic, the traditional interpretation of Balassa Index is invalidated

[6]We are aware that the pure Ricardian comparative advantage is a relative concept, which derives from the comparison with the sector efficiency (technology) in a benchmark country. But as we are interested in a synthetic measure of comparative advantage, we consider the rest of the world as benchmark country. Moreover, Balassa index does not take into account the two-way trade flows (imports in the same industry). The index we present in this paper also control for this issue.

[7]Although the comparative advantage measure proposed by Costinot et al. (2012) directly derives from a Ricardian style model, it could also be ideally extended to a wider meaning of comparative advantage (i.e. factor intensity or economics of scale driven comparative advantage). Indeed, the main property of the index proposed

RCA index is therefore clean for partner country and sector specific factors that may affect ex-post trade flows and thus the traditional Balassa Index (such as import demand shocks, income effect, and tradability of goods). We picked up the idea from Costinot et al. (2012) and created a new dataset of Revealed Comparative Advantage (RCA)[8]. In this paper, we describe the construction of this new dataset and we compare the distribution characteristics and the ordinal ranking properties of the new RCA index with the traditional Balassa Index.

We also propose some improvements with respect to the seminal paper by Costinot et al. (2012). First, we cover a higher product disaggregation; while Costinot et al. (2012) provides index for 13 ISIC industries, our database contains RCA measures at both chapter and product level (HS-2 and HS-4 digit classification). We also extend the sample of partner countries and the time span used by Costinot et al. (2012). We used BACI trade flows data for 20 exporting and 76 importing countries along the period 1995-2010, while Costinot et al. (2012) rely on a sample of 21 exporting and importing countries in 1997. The bigger sample of partner countries guarantees more robust RCA estimations, while the longer time period allows us to analyse the time stability of RCA index distribution. Our dataset covers manufacturing sectors only, for which we have homogeneous and highly disaggregated product classification.[9]All in all we show that the new RCA index based on Costinot et al. (2012) has a symmetric and stationary distribution with better ordinal ranking property characteristics, compared to Balassa Index.

The rest of the paper is organized as follows. Section 2 presents the new RCA measure proposed by Costinot et al. (2012) and describes how we extended their approach to a higher product disaggregation. Section 3 presents the dataset we created based on the former measure. Section 4 shows the statistic distribution characteristics of the new index with the aim to compare it with Balassa Index. Final section concludes.

2. A new measure of Revealed Comparative Advantage

A measure of revealed comparative advantage, in the spirit of the Ricardian model of trade, points to capture the innate productivity of a country in a given industry or product relatively to the other countries. The idea of Balassa index is to compare the performance of a country in one industry to the performance of a reference group of countries using export flows. In doing so, Balassa Index mixes up comparative advantage driven with other determinants of trade flows in approximating the RCA. Indeed, good export performance can be due to several factors that are not directly linked to comparative advantage (formal or informal trade barriers, historical trade relationships, internal demand shock in a country, difference in preferences, etc.). India can easily export tea in the United Kingdom, more than what China could do, because Indian exporters have more information on the United Kingdom market as a consequence of former colony rela-

by Costinot et al. (2012) is to be an exporter-industry specific measure cleaned from all other factors affecting the pattern of exports.

[8]Available in CEPII web site $http://www.cepii.fr/CEPII/en/bdd_modele/bdd.asp$

[9]Bilateral trade data on services are not as detailed as needed for the correct calculation of our RCA measure.

tionship. But this link has nothing to do with relative productivity (as in the Ricardian model). Hence export flows, in order to be good proxies for exporters-sector technological advantages (i.e. innate productivity) as in Ricardo, must be cleaned from all country-pair specific factors (such as factor endowment and trade barriers between countries: physical distance, existence of colonial ties, use of a common language etc.) and partner-sector specific determinants of trade (variation in policy barriers, demand shock, bias in the tastes of consumers, etc.).

Relying on the framework presented in Costinot et al. (2012), we control for the former bias by providing a new econometric based index of comparative advantage. The framework proposed by Costinot et al. (2012) is a Ricardian model with one factor of production (labor) and K industries[10] in a perfect competition setting. The key assumption is that the fundamental productivity of country i in an industry k, named $z_{i,k}$, is randomly drawn from a Fréchet distribution as in Eaton and Kortum (2002).

Using this model, trade flows can be defined as follows:[11]

$$ln\left(x_{i,j,k}\right) = \delta_{i,j} + \delta_{j,k} + \theta ln\left(z_{i,k}\right) + \varepsilon_{i,j,k} \tag{1}$$

where i, j and k indicate respectively exporter, importer and industry (chapter HS-2 and product HS-4 in our estimations), $\delta_{i,j}$ are country-pair fixed effects, $\delta_{j,k}$ are importer-industry fixed effects. Finally, the third term $z_{i,k}$ approximates for the fundamental productivity level of country i in sector k (i.e. technological coefficient in Ricardo model). They assume that technological differences are exporter-industry specific and depend on two parameters: the fundamental productivity $z_{i,k}$ that is exporter-industry specific and a measure of productivity dispersion θ which is country invariant. $z_{i,k}$ is the parameter of interest and it is essential to built a proper index of Ricardian comparative advantage. It captures factors related to cross-country variation of productivity such as climate, infrastructure, institution that affect all producers in a given country and industry. The parameter θ derives from the Fréchet distribution of productivity and represents the intra-industry productivity heterogeneity. It reflects the idea that technological know-how changes across products. As in Costinot et al. (2012) we assume θ as common to all countries and industries.

The realization of random $z_{i,k}$ is ex-ante unknown, but it can be retrieved approximating the technological differences by an exporter-industry fixed effect in the empirical counterpart of equation (1) as follows:

$$ln\left(x_{i,j,k}\right) = \delta_{i,j} + \delta_{j,k} + \delta_{i,k} + \varepsilon_{i,j,k}{}^{12} \tag{2}$$

[10]In their framework an industry k is defined using ISIC Rev 3.1 codes, cf. Table 1

[11]See Costinot et al. (2012) for further details in the theoretical model.

[12]Althought the zero trade flow problem does not arise in HS-2 level regression (small number of zero flows) it arises at HS-4 level. Finally, we decided to exclude the zero flows from our sample considering that zero flows are indeed direct signals of comparative disadvantage.

From the OLS estimation of equation (2) we capture the measure of technological differences through the exporter-industry fixed effect $\delta_{i,k}$. For the value of θ we use the preferred estimate in Costinot et al. (2012) comfortably in line with previous estimates in the literature ($\theta = 6.53$)[13]. Hence, we are able to recover the parameter $z_{i,k}$ from (1) as follows:

$$z_{i,k} = e^{\delta_{i,k}/\theta}. \tag{3}$$

$z_{i,k}$ is a good proxy for comparative advantage because it can be considered as the part of the trade flows that is only due to the intrinsic productivity level of a given industry k in a country i. In fact, it is cleaned from all other determinants of export performance (kept by importer-exporter fixed effect and importer-industry fixed effect). Having values of $z_{i,k}$ we could continue in following Costinot et al. (2012) and compute pairwise index of comparative advantages: keeping one sector-country fixed, all comparative advantage measures would be a pairwise comparison with the sector-country benchmark (this choice implies an exponential increase in the number of observations considered all the possible combinations in the exporter-sector benchmark). However, we aim in providing a dataset of synthetic measures of comparative advantages which do not depend on a specific country-sector benchmark, to this end we decide to normalize $z_{i,k}$ according to a reference group of countries. We use all exporter countries in our dataset as benchmark group.

Doing so, we depart from Costinot et al. (2012) idea of a pair-wise index[14]. We compute a weighted index as follows:

$$RCA_{i,k} = \frac{z_{ik} z_{..}}{z_{i.} z_{.k}} \tag{4}$$

where $z_{..}$ is the average of all z_{ik} coefficients across all industries and countries, $z_{i.}$ is the average of z_{ik} for the country i across all sectors and $z_{.k}$ is the average of z_{ik} for the sector k across all exporters.

Given the formula in equation (4) a country i has a comparative advantage in sector k if $RCA_{i,k}$ is greater than 1. When the RCA index takes values higher than one, it means that, given the worldwide average level of productivity $z_{..}$ (which $de facto$ plays as a sample weight), country i in sector k has a productivity level higher than the expected one - $z_{i.} z_{.k}$. Indeed, the expected productivity level in country i and sector k is the product between the average country's productivity $z_{i.}$ and the sector average productivity $z_{.k}$.

Using this framework, we are able to directly compute RCA index at chapter (HS 2-digit) level. However, our aim is also to compute RCA measures at product level (HS 4-digit). Indeed, we

[13] See Simonovska and Waugh (2011); Donaldson (2010).

[14] The following normalization has been used to consider the all sample of exporters countries as a benchmark. However having the values of z_{ik} it is still possible to build a bilateral index (in the style of pure Ricardo model) by choosing a country-sector of reference and express our RCA index as comparison with that country-sector.

believe that having a RCA index at a higher level of product disaggregation is essential to assess a country's performance in terms of productivity.

The only difference so far with respect Costinot et al. (2012) concerns the sector classification: we use chapter HS 2-digit while Costinot et al. (2012) refers to ISIC rev 3.1 industries. Thus we are implicitly assuming that the parameter θ is the same in the two sector classifications (HS 2-digit and ISIC). Ideally, we would like to have a proxy of a proper intra-chapter (HS 2-digit) heterogeneity. Nevertheless, as pointed out by Costinot et al. (2012), the estimation of parameter θ is challenging and requires firm-level data; and as it is not the main focus of our paper, we simply rely on θ estimation used by Costonit et al. (2012) as a proxy for intra-chapter (HS 2-digit) heterogeneity.[15]

Having a RCA index at product level (HS 4-digit) is a bit more complicated. The assumptions that we made to extend our methodology from industry (ISIC) to chapter (HS 2-digit) are likely to apply to product as well. Hence, we could estimate equation (2) to retrieve RCA coefficients at product level. However, for computational reasons[16], we can not run regression (2) defining k as HS 4-digit. Thus, we propose an alternative strategy to reduce the number of exporter-product fixed effects in the regression. We further assume that the product level productivity index $z_{i,k}$ (where k stands now for HS 4-digit) can be decomposed into two parts: one common to all the products in the chapter (HS-2) - $z_{i,K}$ (where K stands for chapter HS 2-digit); and the other being product-specific (HS-4) given the chapter HS-2, namely $z_{i,k|K}$. Thus, the final $z_{i,k}$ at HS-4 level can be computed as follows:

$$z_{i,k} = z_{i,K} * z_{i,k|K} \tag{5}$$

Ideally, the product specific RCA is composed by: (i) a chapter specific component common to all products in the same chapter, and (ii) by a product-specific component which differentiates the specific product with the rest of products in the same chapter in terms of productivity.

Each product requires a set of specific skills. Even if a country owns a powerful devise that makes it very productive in making "vehicles other than railway or tramway rolling stock" (chapter 87), this specific devise might be more useful for "tractors" (product 8701) than for "baby carriages" (product 8715) production. Hence, we need a product specific productivity component in order to have proper HS-4 measure of RCA - $z_{i,k|K}$. The second component of equation (5) is thus obtained by estimating equation (2) (where k stands for product HS 4-digit) using the trade flows of all products in the chapter K. Then, we use the same formula as in (3) to compute $z_{i,k|K}$. For the aggregation of the two components, as in (5), to be consistent it is necessary to normalize $z_{i,k|K}$ around one by the geometric average of all $z_{i,k|K}$ coefficients in the

[15]Moreover, ISIC rev. 3.1 classification used by Costinot et al (2012) is not far from our HS 2-digit chapter classification, so our approximation on θ does not seem to be problematic.

[16]The number of fixed effects increases significantly when using flows in HS 4-digit and STATA software does not support such huge number of dummies.

chapter K. This assures that the average of the RCA indexes for country i - product k in the chapter K -$z_{i,k}$- is equal to the RCA index for country i in chapter K -$z_{i,K}$ - computed directly at HS2 digit.

3. Dataset Description

The new database we propose contains thus revealed comparative advantage measures (RCA) obtained by using the methodology suggested by Costinot et al. (2012) and adjusted as described in the previous section. The index is then provided for 20 countries over the period 1995-2010[17] at two different product disaggregation levels (HS-2 and HS-4). This results in two databases, respectively named, $RCA\ HS-2$, $RCA\ HS-4$. Each database contains 5 variables:

- *country* is the country of interest (i.e. the exporter country);
- *isocode* is the ISO 3166-1 numeric code of the country;
- *year* is the year dimension;
- $hs2$ or $hs4$ indicates the sector of interest (respectively HS-2 and HS-4);
- *RCA* is the index of revealed comparative advantage;

As mentioned earlier to build the index we had to estimate the coefficients in equation (2). Our estimation only requires data on trade flows, the dependent variable. We use trade flow data from the international trade database at the product level (BACI) constructed by the CEPII (HS revision 1992) which provides trade flows for more than 200 countries from 1989 to 2010. Because of data availability at HS-4 disaggregation, we compute RCAs on the period 1995-2010. Moreover, we compute the index for a restricted sample of countries because of estimation constraints that arise for high level of product disaggregation. We performed the estimation on a set of 20 exporting countries and 76 destination countries[18] that we believe representative in terms of trade (our sample of 76 importer countries represent the 96% of total imports in 2010). Hence, our database contains revealed comparative advantage indexes for 20 countries, mainly G20 countries, which are the leaders in manufacturing exports. We adjusted the G-20 group[19] by excluding Saudi Arabia (negligible as we got rid of the oil sector) and including Spain and the Netherlands (which are important players in world trade).

Concerning the product coverage, we stuck to the manufacturing sector because the RCA index suggested by Costinot et al. (2012) is theoretically grounded for the manufacturing sector.

[17] Equation (2) has been estimated year by year from 1995 to 2010.

[18] We aim at having a sample of partner countries that allows a reasonable number of fixed effects in the regression but is still representative. To select the set of destination countries we ranked the countries in descending order according to the value of imports they received in 2010, we kept the top 76

[19] The final list of countries if the following : Argentina, Australia, Brazil, Canada, France, Germany, Indonesia, Italy, Japan, South Korea, Mexico, the Netherlands, Russia, India, South Africa, Spain, Turkey, United Kingdom, the United States

Table 1 – Summary description of the RCA datasets, industry, HS-2 and HS-4 disaggregation

	HS-2	HS-4
Number of countries	20	20
Number of partner countries	76	76
Number of chapter/product	70	1018
Number of years	16	16
Number of non-missing observations	22,069	295,015

Source: RCA database, CEPII

Hence, we dropped chapters related to agriculture (chapter 1 to 24) as well as the chapter concerning art objects (chapter 97). Moreover, we decided to drop the chapter concerning mineral fuels (chapter 27) because it is not relevant for RCA as it mainly depends on natural endowments of countries. The 2-digit level database contains 70 chapters (HS-2). The 4-digit level database contains 1,018 HS-4.

For some exporter-sector combinations (mainly at HS-4 level) the number of trade partners and years with no missing trade flows is small or however insufficient to estimate $\delta_{i,k}$ coefficient in equation (2). This is the reason for missing values that the user could find in our database; however, given the reason for such missing values they could be interpreted as comparative disadvantage (few export flows indeed). In table (1) is reported a brief description of the numbers of observations, exporter and importer countries, and number of sectors for each of the two datasets.

As an example of the information provided by our dataset, Table 2 shows the RCA index at industry level for the sample of the exporting country we cover in the dataset (ISIC aggregation as in Costinot et al. 2012 to the sake of clarity of the table). We find France having comparative advantage in the Food sector and, with a lesser extend in Textile and Wood sector. Germany and Japan are the top-ranked countries in the Machinery sector, while China is definitely the worldwide leader in the Textile industry.

This is only an example of potential applications for this dataset. Also changes over time of comparative advantage might be analysed and other econometric applications are possible. In these cases the user needs to be informed about the empirical distribution characteristics of the new RCA index; this is what we do in the following section.

Table 2 – Industries RCA in 2010

country	Food	Textile	Wood	Paper	Chemicals	Plastic	Minerals	Metals	Machinery	Electrical	Transport	Misc. Manuf.
Argentina	1.99	0.79	1.01	0.81	0.84	0.98	0.80	0.82	0.75	0.86	0.89	1.13
Australia	1.24	0.84	0.93	1.08	0.95	0.88	1.01	1.01	1.01	0.99	0.95	1.00
Brazil	1.53	0.81	1.36	0.81	0.85	0.96	1.21	1.11	0.87	0.85	0.91	0.96
Canada	1.07	0.88	1.21	1.07	0.85	0.90	1.14	1.03	1.00	1.00	0.95	0.98
China	0.85	1.44	1.13	1.21	0.99	0.99	1.05	1.18	1.10	1.13	1.16	NA
France	1.05	1.02	1.01	1.00	0.98	0.99	1.01	0.97	1.01	0.99	1.00	0.98
Germany	0.93	0.94	1.11	1.05	1.00	1.01	1.08	1.00	1.05	1.01	1.03	0.94
India	1.03	1.26	0.95	0.94	1.03	1.09	0.90	1.01	1.04	0.94	0.93	0.95
Indonesia	1.07	1.16	1.27	0.96	NA	0.86	1.12	1.06	0.74	0.91	0.94	0.82
Italy	0.95	1.16	1.15	0.98	0.95	0.92	1.03	1.10	1.09	0.95	1.01	0.90
Japan	0.63	0.81	NA	1.00	0.90	0.94	0.81	0.90	1.12	1.11	1.06	1.01
Korea	0.72	1.04	0.75	0.93	1.12	1.07	0.90	0.97	0.98	1.15	1.13	1.03
Mexico	1.00	0.86	0.78	1.07	0.98	1.04	0.82	0.98	0.90	1.13	1.14	1.02
Netherlands	1.19	0.88	1.02	1.01	1.19	1.08	1.06	0.97	1.00	0.99	0.96	0.89
Russia	0.97	0.72	1.20	0.89	1.24	1.29	1.18	1.10	0.86	0.86	0.87	1.10
South Africa	1.23	0.85	0.89	0.94	0.86	1.12	1.00	0.95	0.90	1.01	0.90	1.10
Spain	1.08	1.14	1.10	0.97	1.01	1.01	1.09	1.02	0.94	0.90	0.97	0.94
Turkey	1.05	1.38	1.09	0.79	0.84	0.97	0.98	1.03	1.05	0.91	1.01	0.94
UK	1.03	1.00	0.90	1.19	0.95	0.99	0.98	1.01	1.01	1.00	0.98	0.96
US	1.02	0.86	1.13	1.07	1.06	1.01	1.02	0.98	1.07	1.06	0.99	0.92

Source: RCA database, CEPII

4. Empirical Distribution and ordinal rank property of RCA index

A lot of studies on commercial and industrial policy extensively relied on the concept of revealed comparative advantage, often measured by Balassa Index and used by cross-country and cross-industry comparison. However the statistical properties of Balassa Index distribution have been criticized (Hinloopen and Van Marrewijk 2001) and its power in cross country (industry) comparison questioned (Yeats 1985). Hence, this section describes the statistical distribution properties of the RCA index compared with the traditional Balassa index of revealed comparative advantage. In particular we focus on: (i) basic distribution's characteristics (namely, shape and stationarity), and (ii) the ordinal country-sector ranking property (rank correlations). The results clearly show differences in the basic distribution characteristics between the RCA and Balassa index, and better country-sector ranking performances of RCA index with respect to the traditional Balassa index.[20]

4.1. Shape and time stationarity

In this section we investigate the shape of the distribution and the time stationarity of RCA and Balassa indexes (BI). While computing basic distribution characteristics of the RCA index has only a descriptive purpose[21], the time stability of the distribution is an important feature in assessing whether the new index is a proper measure of Ricardian comparative advantage. Indeed, in the spirit of Ricardo, technological coefficients are country-sector specific and mainly sticky along time, since changes in technological coefficients are only due to structural technological changes. Thus, a proper measure of comparative advantage should not vary a lot along time. Indeed, one of the most relevant critique in the literature concerns the scarce time stationarity of the Balassa Index (Hinloopen and Van Marrewijk 2001; De Benedictis and Tamberi 2004). We show that the new RCA index based on Costinot et al. (2012) has higher time stationarity than Balassa Index.

Table 3 shows basic distribution characteristics of RCA index as compared with Balassa index. Although they have similar mean values (almost 1), Balassa index has a higher dispersion, namely six times than RCA distribution. Moreover, according with both skewness index and simple comparison of percentile measures, Balassa index is more skewed than RCA index; meaning that Balassa index has a higher lack of symmetry than RCA index.

The former difference in the symmetry of distributions can be also shown by simple density function graphs in Figure 1. The density function of RCA index (continuous line) is symmetric around one (being one the threshold for having comparative advantage in a certain sector) and

[20] Although our RCA measure and Balassa index differ in statistical properties, they are positively correlated. At HS-2 level the correlation index is 0.647, while at HS4 level correlation index is 0.510. Table A1 shows RCA-Balassa correlation index by country in 2010. Appendix figures A1 and A2 show the positive (and statistically significant) correlation between RCA and Balassa index at HS2 and HS4 level.

[21] However statistic distribution characteristics might be interesting for potential future econometric analysis using the RCA index.

Table 3 – Empirical distribution characteristics of RCA and Balassa index - across HS-4 and countries (time average)

Percentile:	RCA index	Balassa Index
1	0,575	0,007
5	0,692	0,029
10	0,755	0,059
25	0,858	0,181
50	0,970	0,502
75	1,094	1,139
90	1,246	2,183
95	1,370	3,338
99	1,665	6,880
Mean	0,991	0,927
Std. Dev.	0,211	1,276
Variance	0,044	1,629
Skewness	1,051	3,197
Kurtosis	6,534	16,193

Source: RCA database, CEPII

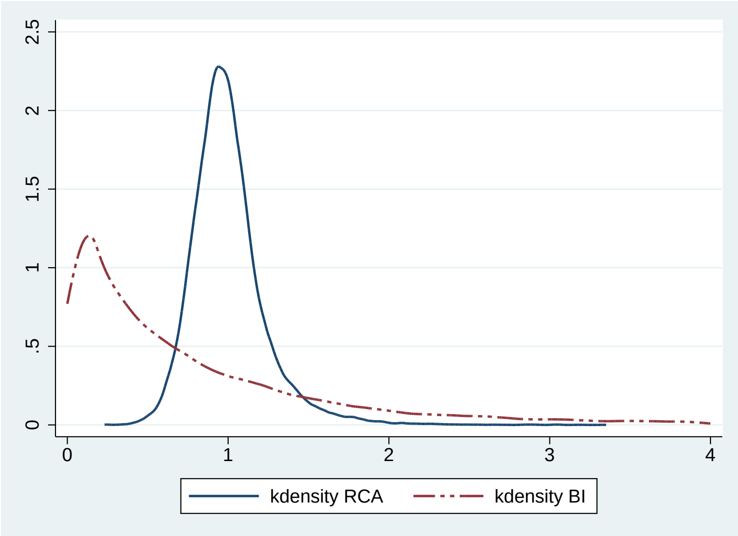
mostly close to a normal distribution. Differently, the density function of a traditional Balassa index (dashed line in Figure 1) has a zero left bound but it is upward unbounded with a long right tail. This first evidence shows an important property of the RCA index as compared with a Balassa measure. The RCA index has a symmetric thin tailed distribution, while BI has a strongly asymmetric distribution with fat right tail.

Asymmetric distribution and fat right tail are indeed consequences of BI formula and its related *size bias*, and make complicated any comparison across sectors (and countries) of high values of BI. The value of Balassa index, in fact, depends on the share of country exports on world exports, in other words it depends on the size of the country, and this may imply some confusion in the cross-country interpretation of index.[22] Let's consider the following example. Suppose that two countries share the total world export market of a certain good, their world sectorial export shares will be equal to 50 per cent. But the two countries are different in size, and one country represents a small share of world total export, while the other country is big and represents a large share of world export flows. A simple comparison between the big and the small country (as in the previous example) using the Balassa Index, would suggest the small country having higher comparative advantage than the big country, even if it has nothing to do with the productivity of the two countries - *size bias*. More importantly, it might imply huge values of Balassa Index for very small countries (long right tail distribution of Balassa Index). This may cause some confusion in interpreting (and or comparing) high values of BI. The RCA index, being cleaned from size effect ($de\ facto$ kept by country pair fixed effect in equation 2)

[22]Indeed Balassa index $\frac{x_{ik}}{X_{i,all}} \frac{X_{world,all}}{X_{world,k}}$ can be written as $\frac{x_{ik}}{X_{world,k}} \frac{X_{world,all}}{X_{i,all}}$ where x_{ik} is country i's export in sector k, $X_{world,k}$ is the world level of export in sector k, $X_{world,all}$ is the overall level of world's exports and $X_{i,all}$ is the world's level of exports in country k.

and being symmetric small tailed distributed, does not suffer this problem, and cross-country interpretation is then easier.

Figure 1 – Density distribution of RCA index



Source: RCA database, CEPII Note: to the sake of clarity in the picture we drop the long right tail of BI distribution at 4

Table 4 shows, by country, mean and median values of RCA and Balassa indexes. In both cases, the mean is higher than the median, providing further evidence of the asymmetry of the two distributions skewed to the right (even at country level). Moreover, the difference between mean and median for Balassa index distribution is systematically higher than RCA distribution. All in all, Balassa index distribution is more asymmetric than RCA distribution. This may represent a problem in using Balassa Index as explanatory variable in econometric based analysis, on the contrary the RCA index, being symmetric and close to a normal distribution can be included as explanatory variable in econometric based analysis. [23]

Finally we move to the time stability of distributions. Table 5 reports the 10, 25, 50, 75 and 90 percentile values of the two distribution (Balassa and RCA index) along the period 1995-2010 and the overall period percentage change. The RCA index is more stable along time: for each percentile of the distribution, the 2010-1995 change (in percentage points) is systematically higher in the Balassa than in the RCA index distribution.[24] The same table reports the mean values of RCA and Balassa index by year (and 1995-2010 change), the mean value of Balassa index fluctuates substantially over time. It is not really surprising given the high skewness and kurtosis of Balassa index distribution. It also reflect the fact that Balassa index, being upward

[23] However some caution is recommended. Indeed, RCA comes from fixed effects regression having export flows as dependent variable; thus using RCA measure to explain trade flows may be endogenous.

[24] Further evidence of the scarce time stationarity of Balassa Index is provided by using Markov transition matrix as in De Benedictis and Tamberi (2004).

Table 4 – Symmetry of RCA and Balassa Index distributions

country	Median RCA	Median BI	Mean RCA	Mean BI	Mean-Median RCA	Mean-Median BI
Argentina	0,953	0,273	1,015	1,266	0,062	0,993
Australia	0,967	0,320	1,013	1,482	0,046	1,162
Brazil	0,934	0,361	1,005	1,129	0,071	0,768
Canada	0,950	0,374	0,986	0,862	0,036	0,489
China	0,996	0,701	1,038	1,435	0,042	0,734
France	1,001	0,775	1,001	0,977	0,000	0,201
Germany	0,988	0,880	0,988	0,981	-0,001	0,102
India	0,988	0,563	1,042	1,994	0,054	1,431
Indonesia	0,947	0,334	1,018	1,475	0,072	1,141
Italy	1,006	0,854	1,011	1,222	0,005	0,369
Japan	0,930	0,441	0,951	0,690	0,021	0,249
Korea	0,956	0,377	0,966	0,731	0,010	0,354
Mexico	0,951	0,298	0,963	0,633	0,012	0,335
Netherlands	0,973	0,650	0,987	1,066	0,015	0,416
Russia	0,923	0,301	1,017	1,382	0,093	1,081
South Africa	0,941	0,217	1,010	1,621	0,069	1,403
Spain	1,011	0,777	1,011	1,155	0,000	0,378
Turkey	0,979	0,369	1,016	1,539	0,037	1,170
UK	1,002	0,783	1,003	1,000	0,000	0,217
US	0,987	0,771	0,987	0,878	0,001	0,107

Source: RCA database, CEPII

unbalanced, may assume very high values (say outlying values). On the contrary, RCA mean does not fluctuate that much, meaning that it does not suffer the presence of outlying values (notice that this property does not derive by its construction - see equation 4). It follows that the mean value of Balassa index is a poor indicator of structural comparative advantage (see Hinloopen and Van Marrewijk 2001), while the mean value of RCA distribution, being stationary, can be well used as a measure of structural comparative advantage.

4.2. Ordinal Ranking properties: RCA-Balassa comparison

One of the main problems in using traditional Balassa index for economic analysis is its poor ordinal ranking property (Yeats 1985). Indeed it may be the case that for a given sector, the majority of country specific indexes of comparative advantage (namely Balassa index) are concentrated slightly above (or below) one; in this situation the top-rank country in the sector may have a relatively low comparative advantage index with respect its own specialization in other sectors. On the other hand, it may also be the case that, in another sector export flows are highly concentrated in few countries; in this case the country with the lowest comparative advantage index may still have a very high Balassa index. As a consequence, the numeric values of Balassa index not necessarily provide the right ordinal ranking of a country's comparative

Table 5 – Empirical distribution of RCA and Balassa index based on yearly export flows

	RCA index						Balassa Index					
Percentile:	10	25	50	75	90	Mean	10	25	50	75	90	Mean
1995	0,732	0,843	0,972	1,117	1,305	1,006	0,038	0,145	0,492	1,201	2,471	1,276
1996	0,734	0,845	0,974	1,120	1,293	1,002	0,046	0,157	0,509	1,209	2,506	1,263
1997	0,707	0,822	0,963	1,119	1,304	0,993	0,046	0,159	0,507	1,205	2,484	1,259
1998	0,746	0,852	0,976	1,113	1,287	1,004	0,049	0,162	0,511	1,214	2,479	1,269
1999	0,733	0,852	0,982	1,125	1,314	1,011	0,049	0,171	0,516	1,225	2,540	1,307
2000	0,738	0,852	0,979	1,123	1,313	1,012	0,045	0,163	0,511	1,226	2,509	1,267
2001	0,739	0,849	0,979	1,123	1,302	1,010	0,046	0,167	0,512	1,201	2,528	1,256
2002	0,736	0,850	0,974	1,112	1,286	1,001	0,048	0,165	0,510	1,214	2,464	1,236
2003	0,733	0,847	0,975	1,113	1,284	1,000	0,044	0,162	0,502	1,192	2,461	1,226
2004	0,736	0,853	0,979	1,120	1,312	1,014	0,044	0,161	0,498	1,191	2,435	1,201
2005	0,725	0,847	0,979	1,125	1,324	1,012	0,040	0,152	0,485	1,169	2,407	1,168
2006	0,715	0,836	0,970	1,121	1,309	1,003	0,038	0,149	0,485	1,168	2,371	1,152
2007	0,719	0,842	0,980	1,125	1,307	1,004	0,038	0,147	0,481	1,163	2,376	1,149
2008	0,712	0,846	0,993	1,154	1,348	1,022	0,037	0,144	0,473	1,160	2,352	1,126
2009	0,700	0,824	0,964	1,110	1,297	0,994	0,038	0,141	0,475	1,161	2,327	1,152
2010	0,713	0,842	0,979	1,127	1,342	1,019	0,033	0,134	0,465	1,160	2,371	1,128
Change (95-10) in % points	-2,5	-0,2	0,6	0,9	2,8	1,2	-13,5	-8,2	-5,5	-3,4	-4,1	11,6

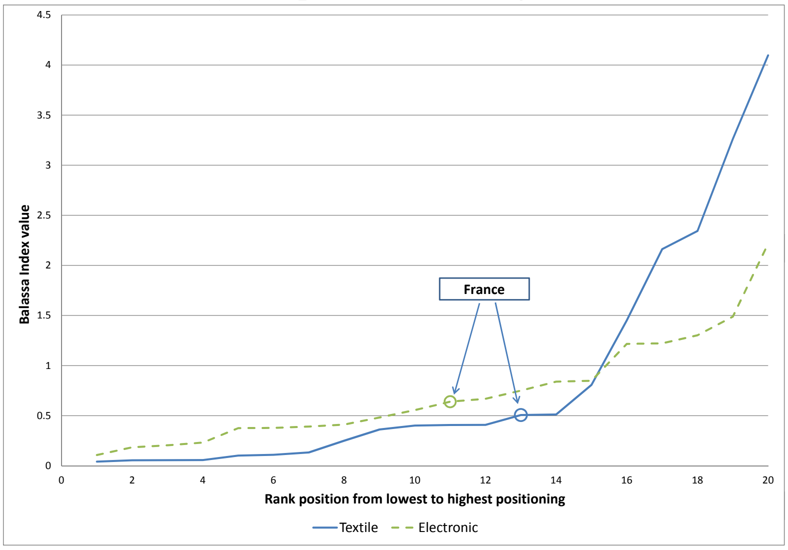
Source: RCA database, CEPII

advantage when the underlying distribution of index values are different across industries (see UNIDO 1982).

The intuition of the previous ordinal ranking bias in Balassa Index is shown in figure 2. We ranked the exporter countries on the base of their Balassa Index values (in 2010) in textile and electronic sector and reported the industry specific rank in the horizontal axis and the associated Balassa Index value on the vertical axis. Thus the two curves show the distribution of the countries' Balassa measures for the two industries. As an illustrative example let's consider the case of France. It has a higher positioning in the textile than in electronic sector even if the Balassa Index associated to textile is lower than that in electronic sector. This is what literature refers as bias in the revealed comparative advantage ordering (Yeats 1985; UNIDO 1982) - bias in the *ordinal ranking property*.

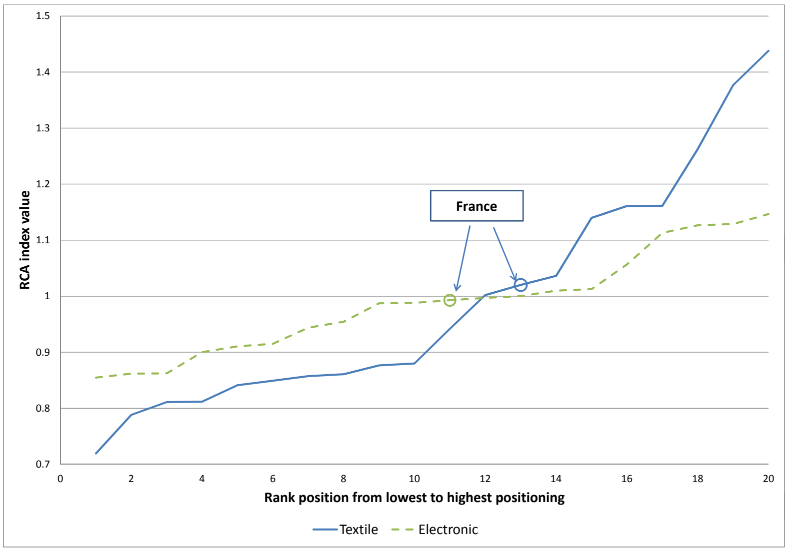
Thus in this section we compare the RCA and the Balassa index in terms of their ordinal ranking properties. To this end we firstly replicate the graph in figure (2) but using our measures of RCA, see figure (3). Differently to what shown for Balassa measure, now France is better ranked in textile with also a higher RCA value in textile than in Electronic sector. So there is no bias in the ordering by using the new RCA index. Only a first piece of evidence in favour of RCA index concerning its ordering property.

Figure 2 – True Industry Comparative Advantage Reversal in Balassa Index



Source: Authors calculation on BACI (CEPII) data

Figure 3 – True Industry Comparative Advantage Reversal in RCA index



Source: Authors calculation on BACI and RCA (CEPII) data

To be more rigorous on this point, we compare the ordinal ranking properties of RCA and Balassa index in table 6 and 7 respectively. For each HS2 sector (column 1 in tables 6 and 7) we report the top-rank country and its comparative advantage measure (columns 2 and 3); then in column 4 we report the rank of such sector in the within country comparative advantage measures (across sectors). Column 5 reports the comparative advantage value for the top-rank sector within country. Column 6 assumes value zero when the top-rank country for a given sector and the top-rank sector for such country coincide. Finally, Spearman Rank Correlation index in column 7 shows the correlation between the ordinal ranking of revealed comparative advantage indices within each sector (across countries) and the position of these indices in the corresponding country's comparative advantage distribution (across industries).

For the sake of clarity, concerning RCA index (table 6), Turkey is the top RCA country in the sector 25 with a RCA index equal to 1.286; however sector 25 is for Turkey the 9th sector in the within country ranking of comparative advantage values. Indeed, the top Turkey's value of comparative advantage is 1.54 (column 5), implying a 0.25 difference with the comparative advantage value in sector 25 (column 6). Spearman Rank Correlation index in column 7 (0.938) suggests that there is a good ordering property of RCA in sector 25 (meaning that top RCA index countries in the sector 25 present in their respective country specific rankings good positioning for sector 25).

To evaluate the ordering property of the two indices we use the Spearman correlation index (or rank correlation): in 60 out of 70 sectors, rank correlation values for RCA is higher than for Balassa index. The former feature shows that the RCA approach performs better that the Balassa Index in terms of strict ordinal ranking property and it provides a more statistically significant rankings of sectors according to revealed comparative advantage. The explanation for such better ordering performance of RCA index is related to its higher homogeneity in sector specific RCA distribution (across countries) compared with Balassa index distribution (see table A2).

Table 6 – National and Industry RCA index discrepancies

HS2	Top RCA country by sector	RCA index for the top country (by sector)	Position in the country's RCA ranking	Top country's RCA value	Difference (3)-(5)	Spearman rank correlation
(1)	(2)	(3)	(4)	(5)	(6)	(7)
25	Turkey	1.286399	9	1.54031	-.2539111	0,938
26	Brazil	1.776277	1	1.776277	0	0,915
28	Russia	1.385714	9	2.321885	-.9361712	0,948
29	India	1.205548	12	1.653853	-.4483044	0,972
30	Australia	1.252129	7	1.666545	-.4144163	0,946
31	Russia	2.321885	1	2.321885	0	0,935
32	Argentina	1.235652	8	1.826324	-.5906726	0,968
33	France	1.358667	2	1.461363	-.1026959	0,990
34	Netherlands	1.159827	7	1.294232	-.1344057	0,966
35	Argentina	1.403165	4	1.826324	-.4231597	0,951
36	Russia	1.320552	13	2.321885	-1.001333	0,946
37	Japan	1.629987	1	1.629987	0	0,970
38	Netherlands	1.161873	6	1.294232	-.1323594	0,927
39	Korea	1.190535	10	1.528697	-.3381615	0,930
40	Indonesia	1.320917	11	1.901992	-.5810748	0,933
41	Argentina	1.810128	2	1.826324	-.0161961	0,969
42	China	1.474481	4	1.79142	-.3169383	0,878
43	Argentina	1.826324	1	1.826324	0	0,933
44	Russia	1.508309	6	2.321885	-.8135767	0,975
45	Spain	1.347449	2	1.446859	-.0994101	0,942
46	Indonesia	1.901992	1	1.901992	0	0,816
47	Canada	1.548649	1	1.548649	0	0,924
48	Russia	1.232086	20	2.321885	-1.089799	0,940
49	UK	1.330552	1	1.330552	0	0,969
50	India	1.653853	1	1.653853	0	0,942
51	Argentina	1.732894	3	1.826324	-.0934299	0,946
52	India	1.523878	3	1.653853	-.1299751	0,975
53	India	1.564187	2	1.653853	-.0896653	0,969
54	Indonesia	1.410638	9	1.901992	-.4913542	0,949
55	Indonesia	1.476321	6	1.901992	-.4256713	0,933
56	Korea	1.170644	15	1.528697	-.3580531	0,960
57	Turkey	1.54031	1	1.54031	0	0,939
58	Turkey	1.454855	2	1.54031	-.0854548	0,956
59	Korea	1.239133	6	1.528697	-.2895635	0,962
60	Korea	1.528697	1	1.528697	0	0,952
61	Turkey	1.444538	3	1.54031	-.0957716	0,933
62	Indonesia	1.429069	7	1.901992	-.4729238	0,967
63	India	1.364578	6	1.653853	-.2892747	0,943
64	Indonesia	1.573512	3	1.901992	-.32848	0,931
65	China	1.422169	9	1.79142	-.3692502	0,966
66	China	1.767657	2	1.79142	-.0237626	0,965
67	China	1.721706	3	1.79142	-.0697137	0,884
68	Brazil	1.195323	11	1.776277	-.5809547	0,888
69	Spain	1.446859	1	1.446859	0	0,945
70	Turkey	1.266827	11	1.54031	-.2734829	0,870
71	South Africa	1.301541	7	1.652338	-.3507972	0,953
72	Russia	1.67936	3	2.321885	-.6425254	0,980
73	Argentina	1.25817	7	1.826324	-.5681547	0,968
74	Russia	1.311569	14	2.321885	-1.010316	0,976
75	Russia	1.859686	2	2.321885	-.4621997	0,946
76	Russia	1.427264	8	2.321885	-.894621	0,915
78	Australia	1.664493	3	1.666545	-.0020524	0,897
79	Australia	1.523862	4	1.666545	-.142683	0,931
80	Indonesia	1.816115	2	1.901992	-.0858773	0,958
81	Russia	1.542331	4	2.321885	-.7795541	0,972
82	Brazil	1.13377	17	1.776277	-.6425078	0,880
83	Italy	1.167716	12	1.313768	-.1460526	0,969
84	Japan	1.221677	9	1.629987	-.4083096	0,940
85	Korea	1.234884	7	1.528697	-.2938125	0,933
86	South Africa	1.316152	6	1.652338	-.3361858	0,926
87	Japan	1.502927	2	1.629987	-.1270595	0,956
88	US	1.651538	1	1.651538	0	0,964
89	Russia	1.310251	15	2.321885	-1.011634	0,937
90	Japan	1.296426	5	1.629987	-.3335608	0,970
91	China	1.423979	7	1.79142	-.3674408	0,963
92	Indonesia	1.518199	4	1.901992	-.3837934	0,908
93	Russia	1.347734	10	2.321885	-.9741514	0,950
94	Indonesia	1.264598	13	1.901992	-.637394	0,962
95	China	1.42381	8	1.79142	-.3676099	0,959
96	Japan	1.293943	6	1.629987	-.3360431	0,962

Source: RCA database, CEPII

Table 7 – National and Industry Balassa Index index discrepancies

HS2	Top BI country by sector	BI index for the top country (by sector)	Position in the country's BI ranking	Top country's BI value	Difference (3)-(5)	Spearman rank correlation
(1)	(2)	(3)	(4)	(5)	(6)	(7)
25	Turkey	5.995914	4	11.06007	-5.064159	0,931
26	Australia	25.56693	1	25.56693	0	0,945
28	Australia	6.350285	6	25.56693	-19.21664	0,924
29	Netherlands	2.276666	4	3.822494	-1.545828	0,816
30	UK	2.897759	1	2.897759	0	0,965
31	Russia	15.46576	2	17.4136	-1.947845	0,911
32	Netherlands	1.869282	9	3.822494	-1.953212	0,829
33	France	4.382196	1	4.382196	0	0,950
34	Netherlands	2.038254	6	3.822494	-1.78424	0,918
35	Argentina	4.990786	5	19.81263	-14.82185	0,905
36	China	3.187779	13	9.120917	-5.933139	0,933
37	Japan	2.49461	2	2.573856	-.0792463	0,974
38	Argentina	2.027803	12	19.81263	-17.78483	0,843
39	Netherlands	2.044887	5	3.822494	-1.777608	0,890
40	Indonesia	4.613765	8	36.24853	-31.63476	0,879
41	Argentina	19.81263	1	19.81263	0	0,969
42	India	5.56766	7	12.86299	-7.295334	0,874
43	Argentina	8.662229	2	19.81263	-11.1504	0,943
44	Indonesia	6.955878	3	36.24853	-29.29265	0,962
45	Spain	18.75652	1	18.75652	0	0,861
46	China	9.093101	2	9.120917	-.0278168	0,727
47	Brazil	8.047337	2	13.83692	-5.78958	0,961
48	Canada	3.17503	6	6.326344	-3.151313	0,931
49	UK	2.677098	2	2.897759	-.2206612	0,943
50	India	9.026817	3	12.86299	-3.836177	0,961
51	Australia	20.66965	2	25.56693	-4.897278	0,930
52	India	9.538302	2	12.86299	-3.324692	0,923
53	India	8.782625	4	12.86299	-4.080369	0,948
54	Indonesia	3.709504	12	36.24853	-32.53902	0,862
55	Indonesia	5.395952	6	36.24853	-30.85258	0,893
56	Italy	2.17314	10	4.399635	-2.226495	0,921
57	India	12.86299	1	12.86299	0	0,922
58	Turkey	4.600917	8	11.06007	-6.459157	0,811
59	Korea	2.520327	4	5.463738	-2.943411	0,845
60	Korea	5.005698	2	5.463738	-.4580402	0,897
61	Turkey	11.06007	1	11.06007	0	0,926
62	Turkey	5.795514	5	11.06007	-5.26456	0,923
63	Turkey	8.955243	2	11.06007	-2.104831	0,872
64	Indonesia	4.623774	7	36.24853	-31.62475	0,922
65	China	4.805635	7	9.120917	-4.315282	0,766
66	China	8.852024	3	9.120917	-.2688932	0,901
67	China	9.120917	1	9.120917	0	0,803
68	Italy	2.656797	7	4.399635	-1.742838	0,786
69	Spain	4.334494	2	18.75652	-14.42203	0,927
70	Turkey	2.155069	17	11.06007	-8.905005	0,788
71	South Africa	17.96588	1	17.96588	0	0,928
72	Russia	6.40905	5	17.4136	-11.00455	0,931
73	Argentina	2.751005	6	19.81263	-17.06163	0,744
74	Russia	6.025295	6	17.4136	-11.38831	0,854
75	Russia	17.4136	1	17.4136	0	0,919
76	Russia	7.840696	3	17.4136	-9.572906	0,939
78	Australia	19.01382	3	25.56693	-6.553104	0,880
79	Australia	7.929682	5	25.56693	-17.63725	0,915
80	Indonesia	36.24853	1	36.24853	0	0,888
81	Russia	7.48174	4	17.4136	-9.931863	0,930
82	China	1.638718	24	9.120917	-7.4822	0,846
83	Italy	1.883854	14	4.399635	-2.515781	0,920
84	Italy	1.260009	26	4.399635	-3.139626	0,825
85	Mexico	1.963531	2	2.021656	-.0581257	0,925
86	Russia	2.388403	15	17.4136	-15.0252	0,899
87	Spain	1.995204	6	18.75652	-16.76132	0,911
88	France	3.264653	2	4.382196	-1.117543	0,922
89	Korea	5.463738	1	5.463738	0	0,881
90	US	1.681298	4	2.496875	-.8155775	0,941
91	China	3.675922	12	9.120917	-5.444996	0,900
92	Indonesia	6.567279	4	36.24853	-29.68125	0,940
93	Russia	3.226315	10	17.4136	-14.18729	0,953
94	Italy	2.556877	8	4.399635	-1.842758	0,954
95	China	5.613165	4	9.120917	-3.507752	0,890
96	China	2.50104	15	9.120917	-6.619878	0,931

Source: RCA database, CEPII

5. Concluding remarks

This paper intended to present a new database on new Ricardian comparative advantage measure proposed by Costinot et al. (2012). In doing this we also presented some empirical distribution features of the new index as comparison with the traditional Balassa Index.

The new measure proposed by Costinot et al (2012) conceptually fits the ex-ante and country-sector specific nature of Ricardian comparative advantage better than Balassa Index. In fact, being based on ex-post export flows computation, Balassa Index mixes up exporter with importers and sector specific factors driving export flows. The new RCA measure presented here is the results of fixed effects estimation regression explaining bilateral trade flows. The final measure is based on the coefficient associated with exporter-sector specific fixed effects and thus it is clean from importers-sector specific factors driving trade flows (importer demand shocks, sectorial productivity shocks etc.).

We extend the measure proposed by Costinot et al. (2012) in two main directions. First we provide RCA index with a higher level of sector disaggregation, HS-2 and HS-4 digit classification. Second we use a bigger set of partner countries and provide the RCA index for the period 1995-2010.

The new measure of RCA proposed here shows better statistical properties than Balassa index: (i) symmetric distribution, (ii) time stability and (iii) satisfactory order ranking properties. The symmetry of the distribution concurs to reduce the size bias usually arising in Balassa index. The stability of the distribution is a further element making the RCA index a better proxy for Ricardian comparative advantage than Balassa Index (technological coefficient in Ricardo model are hardly varying along time). Finally, the RCA provides good ordinal (and cardinal) measure of country's revealed comparative advantage.

Symmetric and small tailed RCA's distribution, along with time stability and good ordinal ranking properties make across sectors and countries comparison more reasonable and stable than comparison using Balassa Index. This property is particularly useful for applied research and policy evaluation studies aiming to compare the specialization pattern of different countries.

Several improvements are possible on this measure. Higher sector disaggregation level, up to HS 6-digit, is possible by using the Abowd et al. (2002) algorithm which avoids the computational limitations related with the huge amount of fixed effects in the estimation of equation (2); or different normalization procedure with respect the one we used in (4). However, this paper represents the first step in the direction of new econometrically based revealed comparative advantage measures.

6. References

Abowd, J.M., R. Creecy and F. Kramarz (2002) "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data", Cornel University Working Paper.

Amighini, A., M. Leone and R. Rabellotti (2011) "Persistence versus Change in the International Specialization Pattern of Italy: how much does the 'Direct Effect' Matter?", in *Regional Studies* 45, pp. 381-401.

Amiti, M. (1999) "Specialization Patterns in Europe", in *Weltwirtschaftliches Archiv*, 135, pp. 573-593.

Balassa, B. (1965), "Trade liberalization and Revealed Comparative Advantage", *The Manchester School of Economic and Social Studies* 33: 92-123.

Bowen, H. (1983) "On the Theoretical Interpretation of Indices of Trade Intensity and Revealed Comparative Advantage", in *Weltwirtschaftliches Archiv*, 119, pp. 464-472.

Costinot, A., D. Donaldson and I.Komunjer (2012) "What goods do countries trade? A quantitative exploration of Ricardo's Ideas", *Review of Economic Studies*, 79, 581-608.

Chor, D. (2010) "Unpacking sources of comparative advantage: A quantitative approach", in *Journal of International Economics*, 82, pp.152-167.

Dalum, B., K. Laursen and G. Villumsen (1998) "Structural Change in OECD export specialization pattern: de-specialization and 'stickiness'", in *Internaional Review of Applied Economics*, vol. 12, pp. 423-443.

De Benedictis, L. and M. Tamberi (2004) "Overall Specialization Empirics: Techniques and Applications" in *Open Economies Review*, vol. 15(4), pages 323-346.

Donaldson D. (2010) "Railroads of the Raj: Estimating the impact of Transportation Infrastructure", NBER working paper 16487.

Eaton J. and S. Kortum (2002) "Technology, Geography, and Trade",in *Econometrica*, 70, 1741–1779.

Ferto, I. and J. Hubbard (2003) "Revealed Comparative Advantage and Competitiveness in Hungarian Agri-Food Sectors", in *The World Economy*, 26(2), pp. 247-259.

Hinloopen, J. And C. Van Marrewijk (2001) "On the empirical distribution of the Balassa Index", in *Weltwirtschaftliches Archiv*, pp. 1-35.

Kerr, W.R. (2009) "Heterogeneous Technology Diffusion and Ricardian Trade Patterns", Unpublished Manuscript, Harvard Business School.

Lafay, G. (1992) "The measurement of revealed comparative advantage", in Dagenais, M.G. and Muet P.A. (eds.) *InternationalTradeModelling*. London

Levchenko, A. and J. Zhang (2012) "The Evolution of Comparative Advantage: measurement and Welfare implications", in *Economic Policy*, 27:72, 567-602.

Proudman, J. and S. Redding (2000) "Evolving pattern of international trade", in *Review of International Economics*, vol. 8, pp. 373-396.

Richardson, J.D. and C. Zhang (1999) "Revealing Comparative Advantage: chaotic or coherent patterns across time and sector and US trading partner?", NBER working paper 7212.

Simonovska, I. and M.E. Waugh (2011) "The Elasticity of Trade: Estimates and Evidence" , NBER working paper 16796.

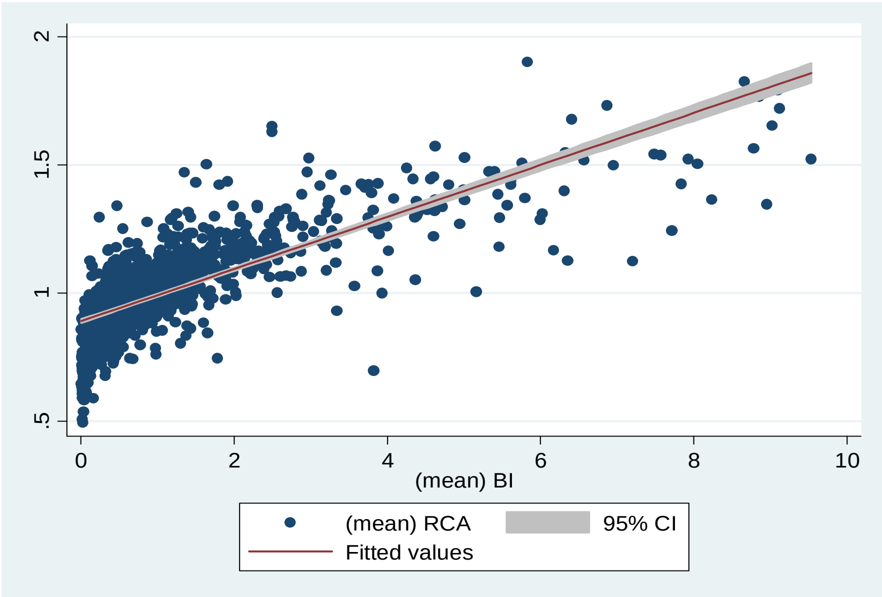
UNIDO (1982) "Changing Pattern of Trade in World Industry: an Empirical Study on Revealed Comparative Advantage". New York.

Vollrath, T. (1991) "A Theoretical Evaluation of Alternative Trade Intensity Measures of Revealed Comparative Advantage", in *Weltwirtschaftliches Archiv*, 127, pp. 265-280.

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, pp. 61-73.

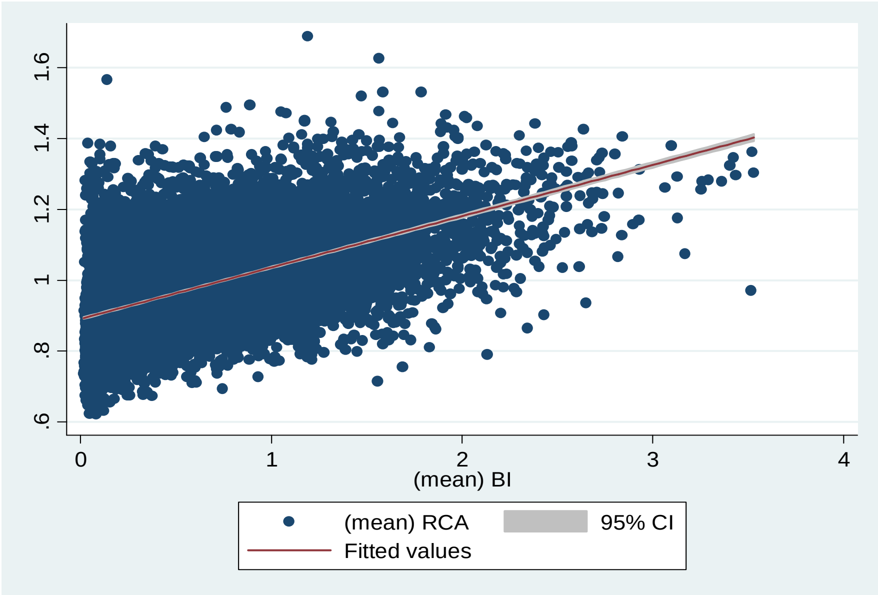
7. APPENDIX

Figure A1 - RCA and Balassa index correlation - across countries and sectors HS2



Source: RCA database, CEPII

Figure A2 - RCA and Balassa index correlation - across countries and sectors HS4



Source: RCA database, CEPII

Table A1 - RCA - Balassa index correlation by country (year 2010)

	HS2	HS4
Argentina	0.6481	0.5153
Australia	0.856	0.5139
Brazil	0.7624	0.4198
Canada	0.425	0.5135
China	0.9305	0.7724
France	0.744	0.5011
Germany	0.5637	0.4488
India	0.8314	0.6639
Indonesia	0.5981	0.5845
Italy	0.7558	0.6375
Japan	0.8072	0.5316
Korea	0.7192	0.554
Mexico	0.6148	0.4352
Netherlands	0.7231	0.5527
Russia	0.7691	0.6049
South Africa	0.6525	0.5809
Spain	0.3398	0.3901
Turkey	0.8394	0.5866
UK	0.5822	0.3905
US	0.6992	0.5253

Source: RCA database, CEPII

Table A2 - Sector specific RCA and Balassa distribution characteristics

	RCA index				Balassa Index			
Sector	Mean	Standard Deviation	Min	Max	Mean	Standard Deviation	Min	Max
25	1.032	0.174	0.718	1.346	1.643	1.774	0.304	8.368
26	1.290	0.536	0.683	2.388	3.182	6.303	0.014	25.378
28	1.005	0.176	0.735	1.609	1.415	1.361	0.303	6.109
29	1.028	0.134	0.779	1.263	0.991	0.591	0.092	2.371
30	0.959	0.199	0.593	1.217	0.969	0.949	0.052	3.293
31	1.043	0.409	0.675	2.568	1.608	3.910	0.051	17.254
32	0.990	0.107	0.803	1.193	0.961	0.553	0.296	1.956
33	0.977	0.153	0.693	1.256	1.123	1.101	0.247	4.720
34	0.997	0.082	0.854	1.154	0.991	0.577	0.098	2.133
35	0.970	0.175	0.690	1.398	1.179	1.403	0.163	6.267
36	1.070	0.169	0.814	1.461	0.852	0.604	0.147	2.319
37	1.028	0.243	0.609	1.644	0.798	0.942	0.017	3.503
38	0.963	0.090	0.782	1.090	1.050	0.986	0.208	4.529
39	1.011	0.066	0.909	1.183	0.888	0.441	0.184	1.880
40	1.060	0.169	0.855	1.482	1.246	1.382	0.084	6.835
41	1.018	0.303	0.619	1.882	2.236	3.793	0.085	16.765
42	1.014	0.173	0.715	1.382	0.725	0.934	0.025	2.814
43	1.031	0.327	0.548	2.026	1.286	1.439	0.005	4.543
44	1.070	0.296	0.648	1.773	1.482	1.966	0.021	7.625
45	0.914	0.133	0.687	1.225	1.343	3.998	0.012	18.142
46	1.148	0.338	0.882	2.051	0.587	1.203	0.009	4.024
47	0.924	0.313	0.489	1.573	1.733	2.628	0.005	10.513
48	1.023	0.143	0.855	1.371	1.108	0.780	0.307	3.227
49	0.986	0.134	0.729	1.338	0.824	0.769	0.094	3.427
50	0.885	0.212	0.512	1.399	0.732	1.271	0.000	4.626
51	0.974	0.308	0.626	1.821	2.029	3.860	0.046	15.818
52	1.080	0.224	0.771	1.620	1.328	2.155	0.026	9.658
53	1.164	0.260	0.810	1.846	1.127	1.966	0.010	8.783
54	1.038	0.194	0.788	1.436	1.082	1.049	0.052	3.309
55	1.041	0.237	0.729	1.656	1.211	1.693	0.044	6.864
56	0.998	0.081	0.827	1.151	0.954	0.553	0.106	2.413
57	1.046	0.255	0.790	1.736	1.466	2.746	0.044	10.248
58	0.998	0.190	0.583	1.402	0.978	1.123	0.075	5.039
59	0.983	0.100	0.704	1.150	0.854	0.431	0.129	1.599
60	0.973	0.190	0.643	1.473	0.923	1.576	0.022	6.312
61	1.069	0.232	0.700	1.535	0.973	1.539	0.012	6.085
62	1.060	0.227	0.759	1.475	0.903	1.182	0.019	3.717
63	1.044	0.152	0.844	1.402	0.893	1.448	0.076	4.924
64	1.101	0.295	0.666	1.752	0.824	1.149	0.008	4.085
65	1.006	0.134	0.718	1.390	0.507	0.730	0.038	3.384
66	0.971	0.238	0.677	1.737	0.342	0.946	0.004	4.321
67	1.010	0.252	0.687	1.633	0.689	1.436	0.005	5.225
68	1.035	0.102	0.855	1.277	1.075	0.806	0.126	2.941
69	1.024	0.156	0.810	1.404	0.928	0.944	0.101	3.369
70	1.007	0.110	0.767	1.210	0.820	0.455	0.223	1.671
71	0.956	0.130	0.681	1.226	2.056	2.826	0.155	12.368
72	1.012	0.241	0.769	1.799	1.392	1.215	0.334	5.440
73	1.029	0.085	0.898	1.281	0.970	0.507	0.240	1.934
74	0.981	0.148	0.689	1.467	1.408	1.257	0.080	5.774
75	0.946	0.319	0.561	1.777	2.436	4.615	0.006	19.206
76	0.989	0.142	0.765	1.415	1.492	1.297	0.321	5.753
78	0.900	0.254	0.646	1.520	1.752	2.699	0.016	11.741
79	0.971	0.229	0.690	1.543	1.818	2.056	0.040	6.469
80	1.045	0.308	0.733	2.213	2.533	9.612	0.004	43.346
81	0.953	0.237	0.441	1.524	0.967	1.204	0.026	5.530
82	1.010	0.109	0.739	1.209	0.704	0.367	0.147	1.401
83	0.995	0.091	0.791	1.142	0.803	0.500	0.129	1.776
84	1.002	0.085	0.847	1.188	0.740	0.367	0.199	1.284
85	1.016	0.113	0.864	1.290	0.709	0.549	0.093	1.947
86	0.943	0.150	0.702	1.312	0.903	0.757	0.054	2.345
87	0.996	0.155	0.802	1.446	1.045	0.710	0.098	2.616
88	0.905	0.280	0.482	1.517	0.879	1.266	0.034	5.411
89	1.011	0.197	0.756	1.443	0.843	1.163	0.037	5.218
90	0.985	0.142	0.717	1.206	0.723	0.576	0.098	1.961
91	0.941	0.161	0.679	1.276	0.580	0.667	0.025	2.154
92	1.049	0.257	0.727	1.820	0.940	2.064	0.013	9.477
93	0.924	0.228	0.408	1.291	1.021	0.924	0.087	3.066
94	1.025	0.103	0.889	1.299	0.747	0.645	0.091	2.374
95	1.024	0.157	0.747	1.387	0.473	0.721	0.024	3.362
96	1.010	0.142	0.705	1.333	0.706	0.467	0.077	2.164

Source: RCA database, CEPII