Int Econ Econ Policy (2017) 14:1–25
DOI 10.1007/s10368-015-0317-z

ORIGINAL PAPER

RCA indices, multinational production and the Ricardian trade model

Kaveri Deb[1] **· William R. Hauk Jr.**[2]

Published online: 9 July 2015

Abstract The practice of using Revealed Comparative Advantage (RCA) Indices to determine the flow of goods trade among countries is well established. But an important issue that demands attention is whether the RCA indices reflect the essentials of comparative advantage theory. Deb and Basu Foreign Trade Rev 46(3):3–28, (2011) examined the consistency of alternative RCA indices with the Heckscher-Ohlin theory of comparative advantage, leaving scope for re-examination of the indices in the context of the Ricardian comparative advantage theory, which insists on relative factor productivity differences among countries contrary to Heckscher-Ohlin's relative factor endowment differences. The other issue which has been overlooked in much of the existing literature is the importance of value-added trade. With the growing importance of global production chains, RCA indices based on gross export values may not portray an accurate picture of the underlying comparative advantage of countries. In this context, adjusting the RCA indices to incorporate domestic value-added in exports seems to be quite relevant. This paper explores the consistency of RCA indices based on domestic value-added in exports with the Ricardian theory of comparative advantage using a panel data approach. A brief review on the structures of alternative RCA indices is also provided. The Log-of-Balassa index is found to be the best performer in this empirical examination, although the deficiencies of the index for cross-country or cross-commodity comparison must be acknowledged. The index of Yu et al. Ann Reg Sci 43(1):267–282, (2009) does possess the latter feature but in our study its

✉ William R. Hauk, Jr.
hauk@moore.sc.edu

Kaveri Deb
kaveri@iitg.ernet.in

[1] Department of Humanities and Social Sciences, Indian Institute of Technology Guwahati, Guwahati 781039 Assam, India

[2] Department of Economics, Moore School of Business, University of South Carolina, 1014 Greene Street, Columbia, SC 29208, USA

performance is quite poor and hence its consistency with the Ricardian theory of comparative advantage is questionable.

Keywords Revealed comparative advantage index · Ricardian theory · Exports · Value-added exports

JEL Classification F14 · C12

1 Introduction

In the literature on international trade theory a considerable volume of work has been devoted on the determination of factors that drive the flow of trade between countries. The theories in this area explore the reasons behind a particular country exporting one good while importing another from its trading partner. In contrast to Adam Smith's theory on absolute cost advantages in trade, David Ricardo recognized in 1817 that in a two country – two commodity framework, even if one country has an absolute cost advantage in both the commodities, it is possible for it to engage in gainful trade with the other country if each country specializes in the production of the product in which they have a comparative cost advantage. This idea came to be widely known as the Ricardian comparative advantage theory of trade. In the Ricardian model, comparative advantage is the outcome of the differences in technology or labour productivity between countries, assuming labour to be the only factor of production. Later Eli Heckscher (1919) and Bertil Ohlin (1933) focused on the relationship between the composition of a country's factor endowments and the flow of commodity trade, and the Heckscher-Ohlin model developed thereby identified comparative advantage as the outcome of differences in relative factor endowments between countries.

Thus, while both of these two theories predict that trade is driven primarily by the principle of comparative advantage, they make different assumptions about what lies behind comparative advantage. The Ricardian model focused on differences in labour productivity as the main driver of comparative advantage, without reference to the relative factor endowments of countries. On the other hand, Heckscher and Ohlin, as noted above, focused on differences in factor endowments (such as capital, labor, and natural resources) across countries and, for tractability purposes, assumed that factor productivity was roughly the same across countries. Subsequent research has shown (see Bowen et al. 1987; Trefler 1995) that this last assumption has severely hampered the empirical performance of the Heckscher-Ohlin model, thus highlighting the importance of technological differences. Trefler (1995) shows the predictions of the model to improve with the incorporation of productivity differences across countries. It seems thus, both differences in technology and factor endowments across sectors and countries are important determinants of the pattern of trade between countries. Hence it is quite logical to claim that a complete description of comparative advantage needs to account for both the Ricardian and Heckscher-Ohlin theories.

While theories describing the idea of comparative advantage have been developed, an important question that arises in this context is how to apply this idea in determining the comparative advantage of countries in real world. The issue was addressed with the development of the Revealed Comparative Advantage (RCA) Index of Balassa (1965), which is an index based on gross export values. The index has been extensively used in the literature for determining the export potential of various countries in various products. But as subsequent discussions would reveal, the index has been subjected to various criticisms, and its applicability towards cross-country or industry comparisons has been questioned. These critics recommended several alternative RCA indices to address one or more shortcomings of the Balassa's index. However, empirically examining and comparing the consistency of various RCA indices with the theories of comparative advantage has rarely been attempted.[1]

Deb and Basu (2011) made an attempt in this regard by empirically analysing the consistency of the existing RCA indices based on gross export values with the Heckscher-Ohlin model. However, their analysis left behind a few gaps which have been addressed in this paper. First, their paper explores only the consistency of RCA indices with the Heckscher-Ohlin model. Second, being a cross sectional analysis, their results are based on limited number of observations. Third, and most importantly, in the event of growing importance of global production chains, an analysis of RCA indices that are based solely on gross export values is questionable. With global value added supply chains gaining more and more prominence, reconstructing the RCA indices using value-added in export values rather than gross export values and examining their performance, is perhaps a better way to proceed. This paper thus tries to address the three points above. We therefore perform a panel data analysis in order to check the consistency of the RCA indices based on value-added in exports, with the Ricardian theory on comparative advantage, with the primary objective of identifying an index which is most compatible with the theory. We also simultaneously study the features of alternative RCA indices. An index which performs well in the empirical analysis and has desirable structural features could be reliably used for determining the comparative advantage of countries in various products. It is to be noted, in this paper we restrict ourselves to export based indices only due to their extensive use in the literature.[2]

Section 2 of this paper contains a brief review of the alternative RCA indices to be studied in this paper, and generalizes the importance of adjusting RCA indices to take into account the growth of multinational production chains. Section 3 describes the methodology applied and the sources of data. Section 4 empirically analyses the indices based on value-added in exports, and thereby identifies the RCA index which is most consistent with the Ricardian theory. Section 5 performs a robustness check on the results generated in section 4. Section 6 concludes the paper by citing the difficulty in selecting one particular index which has all the features of an ideal index and at the same time would be empirically consistent with the theory on comparative advantage. Any researcher's choice of a particular index would ultimately be determined by the

[1] In a theoretical study, Hillman (1980) explored the relation between Balassa's RCA index values and comparative advantage as indicated by the pre-trade relative prices and failed to establish their one to one correspondence under all circumstances. His idea has been extended by Bebek (2011) in the context of other export based RCA indices.

[2] An earlier version of this article looked at the possibility of incorporating imports and exports into our RCA indices. Theoretically, using net exports rather than gross exports as the basis of a measure of comparative advantage might be desirable in sectors with large amounts of intra-industry trade and re-exports. However, converting all of our indices was difficult given that some of our indices are in ratio form, while others are in deviation form, forcing us to adopt the approach of Vollrath (1991) regarding the incorporation of imports in the indices. Furthermore, our studies in this area yielded little real impact on our primary results.

objectives of the researcher and could therefore result in favoring one feature over the other in order to meet the needs of the policy makers.

2 The evolution of RCA indices and value-added trade

The RCA index of Balassa (1965) has been extensively applied throughout the literature on international trade. Stated in ratio form, the index measures the level by which exports of a particular commodity by a country relative to its total exports exceed the exports of the same commodity by all countries relative to their total exports. The index takes the following form:

$$\frac{X_a^i/X_t^i}{X_a^w/X_t^w} \tag{1}$$

Where X refers to exports, i denotes a specific country, and w stands for the world totals (or the sum of all countries on which data are available). Subscripts a and t refer to a particular commodity and to all traded commodities, respectively. Country i has a comparative advantage (disadvantage) in product a if the index takes a value greater (lesser) than unity.

The index of Balassa was later criticized in the literature for various reasons and several alternatives were suggested at different points to address these issues. For instance, Dalum et al. (1998) and Laursen (1998) noted that the Balassa index is asymmetric. Countries that have a revealed comparative disadvantage in the production of a good will have an index value between 0 and 1, while countries that have a comparative advantage in the production of a good will have a value between 1 and infinity. Consequently, they suggested a quasi-logarithmic transformation of the Balassa's index, called the Revealed Symmetric Comparative Advantage (RSCA) index, with the aim of making Balassa's inherently asymmetric distribution symmetric, and thereby generating a normal distribution for the resulting index.[3] The index takes the following form:

$$\frac{\frac{X_a^i/X_t^i}{X_a^w/X_t^w} - 1}{\frac{X_a^i/X_t^i}{X_a^w/X_t^w} + 1} \tag{2}$$

The index is symmetric about the value zero, the comparative advantage neutral point. In an attempt to establish the superiority of their index over the index of Balassa, Laursen estimated separate Galtonian regressions by incorporating each index.[4] Using

[3] This move was primarily aimed at ensuring the reliability of the corresponding regression estimates.

[4] The concept of Galtonian regression was first adopted by Hart and Prais (1956) from Galton's (1889) methodology while studying the heights of fathers and sons and analyzing size wise concentration. Since then, the Galtonian regression method has been extensively employed in the literature to examine technological specialization patterns (Cantwell 1989), convergence of productivity levels over time (Hart 1995) and changes in the structure of trade specialization using RCA indices (Dalum et al. 1998; Laursen 1998; Frantzen 2008; Sharma and Dietrich 2007; Sanidas and Shin 2010, 2011).

Jarque-Bera test for normality of estimated residuals, Laursen proved that the RSCA index outperforms the Balassa's index.

Balassa's index is characterized by unstable distributions both across countries (with respect to commodities) and across commodities (with respect to countries) due to its unstable mean. In order to make the distribution of Balassa's index stable with respect to countries, Hoen and Oosterhaven (2006) suggested an Additive Revealed Comparative Advantage (ARCA) index.[5] Their index, stated in deviation form, measures the extent to which the exports of a commodity by a country relative to its total exports, exceed the exports of the same commodity by all countries relative to their total exports. The index takes the following form:

$$\frac{X_a^i}{X_t^i} - \frac{X_a^w}{X_t^w} \tag{3}$$

The index is symmetric about the value zero. However as pointed out by Yu et al. (2009) while this particular index could be compared across commodities for a country, its comparability across countries is doubtful. The arithmetic mean of the index values across countries for a commodity would never be zero and would vary from commodity to commodity, which in effect implies an unstable distribution across countries.

In another paper by Vollrath (1991), the Balassa RCA index was modified by considering its logarithmic transformation. This new index also has the potential to address the asymmetrical distribution of Balassa's index. However, Vollrath defined the rest of the world and the rest of the commodities exported as the reference group of countries and commodities respectively. To avoid having the reference group of countries or commodities change from country to country or commodity to commodity, all countries and all commodities exported will be considered as the reference groups in this paper. Thus the suggested RCA index could be represented in the following form:

$$ln\frac{X_a^i/X_t^i}{X_a^w/X_t^w} \tag{4}$$

The index is symmetrically distributed about the value zero, which is the comparative advantage neutral point. Being a monotonic transformation of the Balassa's index, the current index tends to preserve most of the features of the Balassa's index in terms of ranking of sectors or countries and difficulties in comparability across sectors or countries. However, the greatest advantage of the index arises from its logarithmic

[5] A distribution with an unstable mean across sectors with respect to a country, has two consequences for comparability – (1) If a country gains a comparative advantage in one sector, it is impossible to say for certain that it has lost a comparative advantage in some other sector. Hence cross-sectoral comparison becomes difficult. (2) The same sectoral value of the index may have a different meaning for different countries, which leads to difficulty in comparing country index values for that particular sector. Hence, ranking countries by RCA for a sector would not be reliable. Similar problems would exist in case of a distribution with an unstable mean across countries with respect to a sector.

form, which to a certain extent ensures that the estimated residuals from any regression with the index as the dependent variable could be normally distributed. As a result, the distributional feature needed to ensure the reliability of the t and the F statistics will be present.

In an attempt to overcome many of the deficiencies of Balassa's index and other indices, Yu et al. (2009) proposed a Normalized Revealed Comparative Advantage (NRCA) index, where comparative advantage of a country in a particular commodity is measured by the deviation of the actual exports of that commodity by that country from the expected exports of the same commodity in a world of no comparative advantages. The actual and expected exports are further normalized by world total exports. The index takes the following form:

$$\frac{X_a^i}{X_t^w} - \frac{X_t^i}{X_t^w}\frac{X_a^w}{X_t^w} \tag{5}$$

Not only is the index symmetric, it also has a stable distribution with a stable mean, as would be evident from the calculated arithmetic mean of index values across sectors for a country or across countries for a sector. As discussed above, this feature ensures that there is no difficulty in comparing the index values across sectors or countries. Its deviation form and the normalization by total world exports also make the index less susceptible to the size of sector or country compared to other indices (Deb and Basu 2011). The proponents of the index also proclaim its usage for temporal comparability due to the stability of its mean over time.

However, the index falls short with respect to one particular feature. Normality of the distribution is not guaranteed by the symmetry of the index's distribution. The same is true in case of ARCA index. In fact, Hoen and Oosterhaven (2006) show that the ARCA index does not fit into the specifications of a normal distribution. As a result, just like Balassa's index, some caution must be employed when analyzing the results of regression for these two indices. From these discussions, it follows that none of the indices discussed is perfect in every respect. In fact, Sanidas and Shin (2010) after analyzing the statistical properties of several RCA indices also concluded that none of the indices could be inferred as perfect. Although the NRCA index, unlike the indices of Balassa, RSCA and ARCA, can be reliably compared across time, sectors or countries, it could generate non-normally distributed residuals, which may be problematic for certain parametric estimation techniques.

Deb and Basu (2011) empirically tested the consistency of these RCA indices with the Heckscher-Ohlin theory of comparative advantage, and found both the RSCA index and Log-of-Balassa's index generating favourable results. However as already pointed out, their analysis had limited scope due to reliance only on the Heckscher-Ohlin theory, a smaller sample of data, and perhaps most relevantly, overlooking the importance of value-added trade. Before addressing these issues in the current paper, we briefly discuss the incidence of value-added trade and its contribution towards redefining the RCA indices in the following subsection.

2.1 Vertical specialization and value added trade: implications for RCA indices

Because of the changing nature of trade flows throughout the world over the last four to five decades, not only in terms of quantity but also quality, concerns have been raised over the relevance of trade indices based on gross exports or imports. Hummels et al. (1998) consider the internationalization of production to be the primary reason for this change. As Feenstra (1998) notes, trade in intermediate goods, such as parts and components, now constitute the largest share of world trade and very often final consumer goods sold in one country are made up of components produced in several other countries. Meng et al. (2012) tried to map the evolution of trade in goods by different end use categories over the period 1995 to 2010.[6] Their analysis reflected the fact that trade in intermediate goods is the main driver of growth in world goods trade, followed by household consumption and capital goods.[7] The phenomenon of a higher volume of intermediate goods being produced for later production stages in different countries and then exported to other countries for further processing describes the development of a global value chain which has alternatively been termed as "vertical specialization" or "fragmentation of production" or "outsourcing" or "global supply chains" (Meng et al. 2012).

The development of global value chains has made traditional gross export measures of doubtful use in observing patterns of trade. As a result, more recent literature has used measures based on domestic value-added trade as a means of reflecting the potential of a country in exporting a particular product. Thus, several papers in this literature have developed measures of comparative advantage based on value-added to exports figures, arguing that the comparative advantage exhibited by such measures differ widely than the typical RCA measures based on gross export figures (Koopman et al. 2014; Meng et al. 2012).

Hummels et al. (2001) provided an empirical measurement of a country's participation in vertically specialized trade. They define the vertical specialization (VS) of country k in sector i as,

$$VS_{ki} = \left(\frac{imported\ intermediates}{gross\ output\ of\ i} \right) \times (Exports\ of\ i\ by\ k) \quad (6)$$

Thus according to their measurement VS is simply foreign value-added in exports. The first term in the above expression denotes the share of imported input in gross output. Multiplying this share by the exported amount provides a dollar value of the imported input content of exports. The authors also allow for the fact that imported inputs could be used both directly and indirectly in the

[6] End use categorization classifies goods as intermediate goods, final consumption goods and capital goods.

[7] The recent change in global merchandise trade could be explained by several factors. The first explanation is, a higher rate of reduction in tariffs and other trade costs on intermediate products compared to final products over the last 20 years, has contributed to greater movement of parts and components all over the world (Miroudot et al. 2009). Second, the growth of foreign direct investment (FDI) boosted the trade in intermediate products (Miroudot et al. 2009). Third, induced by the first and second factors, intra-firm trade has increased and has contributed to increased intermediate goods trade (Yi 2003). Finally, domestic market oriented reforms by countries like China have enabled other countries and multinational enterprises to involve China in their global production network (Meng et al. 2012, 4).

production of an export good. Thus the production of a domestic input which is directly used to produce exports might require imported inputs. Further, a domestic input directly used to produce exports might require another domestic input, which necessitates imported inputs in its production. Similarly, other stages could be identified where the imported input is embodied in exports at the fourth, fifth or sixth stage. Thus considering N sectors, the vertical specialization or foreign content in gross exports of country k could be expressed as:

$$u\left(A^M X + A^M A^D X + A^M A^D A^D X + A^M A^D A^D A^D X + \ldots\ldots\right) = u\,A^M\left(I - A^D\right)^{-1} X \quad (7)$$

Where u is the (1x N) vector of 1s,A^M is the (N x N) imported input–output coefficient matrix (where each element $a_{ij}{}^M$ represent the amount of imported input from sector i used to produce one unit of sector j's output), I is the identity matrix, A^D is (N x N) domestic input–output coefficient matrix (where each element $a_{ij}{}^D$ represent the amount of domestically produced input from sector i used to produce one unit of sector j's output), and X is the (Nx1) vector of exports. In this scenario, $(I - A^D)^{-1}$ captures the total domestic output requirement from each sector i to produce exports from sector j.

Koopman et al. (2012) used Hummels et al.'s (1998) conceptual framework to show that the total output in any sector equals the sum of direct value-added in that particular sector and the cost of intermediate inputs from all domestically produced and imported products. With this set up, they derived the expression of domestic content or domestic value-added in exports as:

$$A_v\left(I - A^D\right)^{-1} X \quad (8)$$

In addition to the matrices defined above, A_v denotes the (1XN) row vector of domestic value-added coefficients where each $a^v{}_j$ represents sector j's ratio of value-added to gross output.

The methodology of Hummels et al. was applied by the National Research Council (2006) to determine the foreign content in the exports of US. However, subsequent articles have questioned the validity of two basic assumptions made by Hummels et al., which limits the applicability of their measure in the real world. These two assumptions are:

(1) Imported inputs used in the manufacture of export goods or goods meant for domestic final consumption must be used in similar intensity. However, in the presence of differing technologies used in the export and domestic consumption sectors (such as might be found in a country that makes considerable use of export processing zones) this assumption would not hold.
(2) All imports have 100 % foreign content. This assumption would be violated if more than one country exports intermediate goods (Koopman et al. 2014). In that case, a country could import intermediate goods, add value and export them to be

processed further by the country from whom it had previously imported the intermediate goods. Or a country could import intermediate goods which contain value that had been previously added domestically at an earlier stage in the production process.

To address the first assumption Koopman et al. (2012) compute the shares of foreign and domestic content in a country's exports in the presence of differing input–output matrices for the domestic and export sectors, and establish Hummels et al.'s measure as a special case of their advocated general measure. They treat export processing zones as a category distinct from normal exports, and allow for differences in the use of imported and domestic intermediates for export processing, and goods meant for normal exports and final domestic sales. They also show that their expressions for foreign content and domestic content shares in a country's total exports would reduce to expressions similar to Hummels et al. if one assumes that similar technologies are used in the exporting and domestic consumption sectors. In support of their argument, the authors constructed the relevant input–output tables and used their new formula to estimate the share of foreign value-added in China's manufactured exports for the years 1997 and 2002. They found it to be about 50 % of the gross export value, which is almost double the amount found in Hummels et al. (2001).

Daudin et al. (2011) tried to overcome the limitations of the second assumption by computing not only the import content of exports (VS as per Hummels et al.'s definition) but also the share of exports used as inputs in further exports (which Hummels et al. define as VS1 but refrain from its computation). The authors also definedVS1* as that part of VS1, which comes back to the country of origin after being processed abroad. Daudin et al. define vertical trade of a country as being twice the sum of VS and VS1* and total value-added trade as being total standard trade minus the sum. The authors found through their calculations that the industrial and geographic patterns of value-added trade differ widely from standard trade.

Johnson and Noguera (2012) also made an effective attempt at addressing the second assumption of Hummels et al. They formulated the value added export ratios or the VAX ratios to compute and analyse the value-added content of trade at the bilateral level, using input–output and bilateral trade data. The authors allowed for the fact that a country's imports are not 100 % foreign-sourced. Within that two country framework, they also reduced their VAX ratio to the corresponding formula of Hummels et al. (2001) by assuming away the presence of intermediate goods trade between the two countries. The authors, using data from GTAP database version 7.1, found through their calculations that US exports to Canada are 40 % smaller in value-added terms and the US-China deficit to be 30–40 % smaller than when measured using gross exports.

All of the above mentioned literature recognized the importance of value-added measures of trade as compared to gross export measures of trade in the rapidly changing world trade scenario by demonstrating the deviations of value-added exports from gross exports. As noted above, Koopman et al. (2014) and Meng et al. (2012) recomputed RCA indices after taking into account the importance of value-added trade. Koopman et al. (2014) decomposed each country's gross exports into various value-added components and tried to formally link value-added measures of trade with official trade statistics. In the process, their constructed framework incorporated almost all the measures of value-added and vertical trade suggested in the literature. The authors also provided a comparative analysis of Balassa's RCA indices computed on the basis of gross exports and domestic value-added in gross exports, for several countries and several sectors. Their analysis revealed contradictory pictures for some sectors, e.g., India has a comparative advantage in sectors such as finished metal products and business services on the basis of gross exports data, while having a comparative disadvantage in the same sectors on the basis of domestic value-added in exports data. For each sector, the authors also noticed substantial changes in the country rankings. Similar changes were also noticed by Meng et al. (2012) after they made a distinction between value-added in final goods and value-added in intermediate goods, and used international input–output tables and relevant bilateral data to reconstruct RCA indices similar to Balassa, but only made up of value-added components.

As already discussed, due to the recent growth of global production chains, the importance of adjusting RCA indices to incorporate domestic value-added in exports becomes necessary in order to determine the true potential of a country in exporting any product. Though Koopman et al. (2014) and Meng et al. (2012) re-analyse the RCA index values after incorporating domestic value-added in exports, they limit themselves to Balassa's index only. In this paper, we apply domestic value-added in exports to all the five RCA indices considered and try to examine the consistency of those indices with the Ricardian model. It is in this context that our paper is not a mere replication of the existing papers. This paper thus conducts a panel data analysis of the consistency of the five RCA indices based on domestic value-added in exports with the Ricardian theory on comparative advantage.

Before moving on to the analysis, it is necessary to address certain relevant issues with respect to the Ricardian trade model, multinational production and intra-industry trade. Trade in parts and components, which is the essence of value-added trade, brings to the fore the concept of intra-industry trade. As observed by Pittiglio (2014), the primary purpose of both the theoretical and empirical literature on intra-industry trade has been to analyse the exchange of different varieties of products without any reference to intermediate goods. Krugman (1979, 1980), Lancaster (1980), Helpman (1981), Eaton and Kierzkowski (1984), Balassa and Bauwens (1988) considered intra-industry trade to exist in horizontally differentiated products between countries with similar factor endowments. Such trade was explained by economies of scale in this literature. On the other hand, Falvey (1981), Falvey and Kierzkowski (1987) insisted on vertical product differentiation, which could take place among countries with dissimilar factor endowments, and could therefore be explained by the Heckscher-Ohlin relative factor endowment model. Unlike horizontal intra-industry trade, which relied on monopolistic competition and increasing returns to scale, the proponents of vertical intra-industry trade based their models on perfectly competitive markets and constant returns to scale.

However, Grubel and Lloyd (1975) argued that the simultaneous exchange of goods within the same industry between two countries need not be restricted to final products only, but also to exchange of final goods for intermediate inputs or intermediate goods for other intermediate goods (as cited by Pittiglio p. 469). In the more recent literature, Jones et al. (2002), Ando (2006) and Türkcan (2011) have argued that international fragmentation of production may lead to the inter-country exchange of intermediate goods within the same industrial classification. According to Pittiglio (2014) this exchange of intermediate goods may take three forms – the exchange of horizontally differentiated intermediate goods, the exchange of vertically differentiated goods distinguished by quality, and vertical specialization which involves value-addition by the associated countries. Arndt (1997), Feenstra and Hanson (1997), Deardorff (2001), and Jones and Kierzkowski (2001) suggest that vertical specialization involving value addition incorporates the ideals of Heckscher-Ohlin model, whereby production fragmentation is undertaken to exploit the factor cost differences across countries. Türkcan and Ates (2011), by dividing total intra-industry trade in auto parts between USA and 29 OECD countries into horizontal and vertical components, found that a substantial portion of this trade is vertical. The extent of this vertical intra-industry trade, which in effect is an indicator of the international fragmentation of production, is positively related to average market size, differences in market size, differences in factor endowments and outward foreign direct investment. The fact that multinational corporations emerge to take advantage of cross country differences in factor rewards has also been recognized by Helpman (1985). As observed by Kinoshita and Campos (2006), the location of foreign direct investments associated with the international fragmentation of production is closely linked with the comparative advantages of the host country, which the home country in turn has to take into account to increase the profitability of investments.

In this context adjusting the RCA indices to account for domestic content in exports would help in determining the inherent advantages a country has to offer. The fragmentation of production process is primarily undertaken to reap the benefits of cheaper factor costs or differences in technology. In this scenario, vertical specialization can be explained by the theories on vertical intra-industry trade, but not by the theories on horizontal intra-industry trade. Hence the assumptions underlying the inter-industry trade theories of Ricardo or Heckscher-Ohlin are not contradictory to those underlying the vertical intra-industry trade models. Even if the importance of horizontal intra-industry trade is realized, Davis (1995) shows that increasing returns to scale is not essential for intra-industry trade to occur and some of the characteristics of intra-industry trade highlight the relevance of Ricardian determinants of trade. When intra-industry trade is defined to be the exchange of goods with similar factor intensity, trade becomes sensitive to even minor technical differences.[8] Thus,

[8] Goods with similar factor intensity are characterized by excellent substitution possibilities. Moreover, as inherent in the intra-industry trade literature, with number of goods being substantially large relative to the number of factors, it is possible to expand the production of some goods at the expense of others without raising the marginal opportunity costs. Both these characteristics of intra-industry trade underline the essence of Ricardian determinants of trade such that with linear transformation curves, trade could take place on the basis of technical differences (Davis 1995).

even when the markets are perfectly competitive and returns to scale are constant, Davis shows that intra-industry trade could be explained by the traditional Ricardian theory, i.e., technical differences could account for intra-industry trade. The disaggregation of intra-industry trade in intermediate goods into horizontal differentiation, vertical differentiation and vertical specialization have further reinforced the importance of Ricardian or Heckscher-Ohlin determinants of comparative advantage in the context of intra-industry trade and the prevalence of classical assumptions of perfect competition and constant returns to scale. Thus, inter and intra industry trade may not be two completely different strands of literature, and integrating the two concepts through the adjustment of RCA indices for domestic value-added in exports is an important step in the literature. Further as noted by Ando (2006), the exchange of products under vertical specialization need not be restricted to products within the same industrial classification. Value addition could result in intermediate products or in final goods with different industrial classifications. Under such circumstances, vertical specialization would go beyond the boundary of intra-industry trade and would more be at par with the concept of inter-industry trade. Hence, the adjustment of RCA indices for domestic value-added in trade need not necessarily imply trying to determine inter-industry trade through the instruments of intra-industry trade.

3 Methodology and data

The phenomenon of global value chains has been extensively studied in the literature and Hummels et al.'s idea has been suitably modified to deal with real life complications. The focus of this paper is more about analyzing the changes in performance of the RCA indices in the presence of value-added trade. Hence, for our purposes, it is not so important for us to know the destination of a country's exports and the ultimate use of those exports in the destination country. In our setup we rely on the established methodology of Hummels et al. (2001) while calculating the domestic value-added in a country's exports and in the process, ignore the presence of processing trade and end use of a country's exports. We assume that the technology applied to produce goods meant for domestic consumption and goods meant for exports are the same. Instead of relying on gross exports, we thus adjust the RCA indices for each sector to take into account how much of a particular sector's domestic output is actually embedded in the country's exports. In effect, for each country we enquire about how much domestic content from a sector is embedded in the exports of that sector and other sectors.[9] We therefore make use of the following formula for

[9] This formulation however differs to some extent from the domestic value-added concept of Koopman et al. (2012) represented by expression (8). They basically enquire about how much of the domestic content from various sectors could be found in a country's final exports of a product. But in order to determine the comparative advantage of a country in a sector, one should enquire about that sector's domestic content in exports of various products from a country.

calculating the amount of domestic value-added in exports (*VAX*) and use this new expression to reconstruct the RCA indices:

$$VAX = \hat{A}_V \left(I - A^D\right)^{-1} X \tag{9}$$

Where $\hat{A}_v$ is an $(N\ X\ N)$ diagonal matrix corresponding to the row vector A_v defined in expression (8). All other variables have the same meaning as in expression (8). This expression produces an $(N\ X\ 1)$ vector that estimates for each country the amount of domestic value-added from each of the *N* sectors in its exports of all sectors.

To test the consistency of the RCA indices with the Ricardian theory of comparative advantage, we calculate the ratio of average labour productivity in a particular sector relative to all sectors in a particular country, to the average labour productivity of that sector relative to all sectors in all countries for which we have data.[10] Mathematically this ratio is:

$$\frac{APL_{j,i}}{APL_i} \Big/ \frac{APL_{j,w}}{APL_w} \tag{10}$$

Where $APL_{j,i}$ is the average product of labor in sector j in country *i*, APL_i is the average product of labor across all sectors in country *i* and the expressions with subscript *w* indicate the corresponding world averages. According to the Ricardian theory of comparative advantage, a country will export a product in which it incurs lower opportunity cost of production compared to its trading partners. One way to estimate the relative cost advantages of countries would be through the measurement of differences in average productivity ratios. Hence, if the expression (10) takes on a value greater than unity, then in line with the Ricardian model, the country should have a comparative advantage in sector *i*. On the other hand, the country should have a comparative disadvantage in sector *i* if the expression takes on a value less than unity.

Based on the above argument, any RCA index would be in line with the Ricardian predictions if the RCA index values and the average productivity ratio moved in the same direction. In other words, a positive association between the RCA index values and average productivity ratio would show that the RCA index values are consistent with the predictions of the Ricardian theory.

For our purpose, we therefore formulate the following null hypothesis:

H_0: There exists no relation between the RCA index values and the average productivity ratio.

[10] In principle, what matters for the determination of comparative advantage is the marginal product of labour (MPL), rather than the APL. However, data limitations prevent us from directly observing the marginal product, even though the APL and MPL will not generally be the same. However, for our purposes, it is sufficient that these measures should be highly correlated with each other. We believe this to be a reasonable assumption if comparative advantage is driven primarily by differences in factor productivity across countries (as in the Ricardian model) rather than factor accumulation (as in the traditional Heckscher-Ohlin model, which does not perform well empirically unless augmented by differences in factor productivity). We note in the conclusion below that creating a better MPL measurement would be a desirable area for future research.

The above null hypothesis is tested against the alternative hypothesis of a positive association between the variables.

With the average productivity ratio as the independent variable and RCA index as the dependent variable, we run separate panel regressions for each RCA index. We consider different possible specifications and compare the results generated thereby. In the first specification, we incorporate all sectors, countries and years and run random effects regressions with country fixed effects and year fixed effects. In the second specification, we use the same sample and run random effects regressions with country-year fixed effects. Thirdly, we run separate regressions for each of the 17 separate sectors and use random effects and year fixed effects.[11] The interpretation of the coefficient of average productivity ratio for the first two cases differs from the third case. For the first two bigger regressions, a positive coefficient represents, within each country how much an increase in productivity ratio causes the RCA index to increase. Due to the presence of many countries and years, the estimated coefficient effectively is an average across all such country responses. Thus for the bigger regressions positive coefficients indicate whether the Ricardian model and the relevant RCA index provide similar predictions about the sector in which countries on average have comparative advantage. For the sector specific small regressions, a positive coefficient implies that both the RCA index and the Ricardian theory provide similar predictions about the country which is most likely to export a product. To examine the validity of our results, we also consider a non-parametric test - Spearman's Rank correlation test for each of the above cases. Spearman's Rank correlation between two variables ranks the observations on the variables in a certain order and uses the differences in ranks to calculate the correlation coefficient. Relying on the order of the observations rather than the actual values makes the test results independent of the form of distribution of variables. This criterion is important given the fact that the RCA indices of Balassa, ARCA and NRCA are characterized by non-normal distributions. As a result, other parametric tests which make assumptions about the form of distribution of variables, may not be reliable in the context of those indices.

The necessary data on exports, value added, output and number of employees for 28 manufacturing sectors, classified on the basis of 3 digit ISIC (revision 2) have been collected for each country-year in our dataset from the Trade, Production and Protection Database, 1976–2004 (henceforth, TPP) by Nicita and Olarreaga (2007).[12] The same database is the source of our input–output coefficient matrices. Because the input–output matrices found in the TPP combine several ISIC sectors so that there are only 17 manufacturing sectors, we do the

[11] Country fixed effects take care of the fact that different countries may have different time invariant features which need to be estimated separately, so that we could factor out the effects that these country specific characteristics may have on the relationship between RCA index and productivity ratio. Similarly, year fixed effects factor out the variations across years for reasons such as technological changes, wars, conflicts etc., and country-year fixed effects to take care of country specific characteristics varying over time. In addition, the random effects component allows for the fact that for each sector in any particular country, the residual term would be correlated across years (country-sector is defined to be the cross sectional unit for the panel).

[12] ISIC stands for International Standard for Industrial Classification of all Economic Activities

same to the exports, value added, output and employees data.[13] In addition, it is to be noted, the input–output tables are based on GTAP version 4 for which the base year is 1995. We assume that the coefficient matrices do not vary much between years and hence apply the same coefficient matrices for all considered years. We thus use these data to construct an unbalanced panel for the years 1991 to 2001 with the number of countries varying between years due to the fact that data are not available for all 17 sectors for each country-year.

4 Comparing the domestic value-added in export-based RCA indices

In this section we analyse the different regression results, compare them with the corresponding rank correlation results, and in the process, identify the index or indices most consistent with the Ricardian theory of comparative advantage. Arguments in favour of any particular index are based upon the number of positive and significant regression coefficients and the magnitude of R-squared values in case of the regression analysis, and the number of positive and significant correlation coefficients in case of correlation analysis. In Table 1 we report the results for a large sample with all countries, years and sectors with the requisite available data. Each regression is of the form:

$$RCA_{i,j,t} = \alpha + \beta * APLR_{i,j,t} + \mu_{i,t} + \varepsilon_{i,j,t}$$

where $RCA_{i,j,t}$ is the relevant RCA index for sector j in country i in year t, $ALPR_{i,j,t}$ is the average labour productivity ratio for sector j in country i in year t (calculated in a manner similar to Eq. (10)), $\mu_{i,t}$ is a country-year fixed effect (or, in one specification, a country fixed effect and a year fixed effect) and $\varepsilon_{i,j,t}$ is a residual term. In all of the regression specifications, we allow the residual term to be correlated across time with other observations in the same country-sector. We report the estimated regression

[13] This had to be done because the input–output matrices are based on GTAP database. Due to the difficulty in matching the commodity classification of GTAP with the ISIC codes, Nicita and Olarreaga (2007) could provide a rough classification of sectors on the basis of ISIC codes in the input–output tables of Trade Production and Protection database. We use the 17 sectors from their classification and simultaneously adjust the data on exports, value added and number of employees to match those 17 sectors. While more finely disaggregated data on imports and exports are available (for example at the HS 4-digit or 8-digit levels), we cannot easily match these data to the "domestic" data on production, value added, and input–output matrices. The study is thus constrained by the non-availability of detailed disaggregated data on the relevant variables. The input–output coefficient matrices required for the adjustment of gross exports figures for the domestic value-added in export figures are provided by the TPP database only for 3 digit ISIC codes. Other sources of such data e.g., World Input Output Database, IDE-JETRO, and GTAP report data at highly aggregated sectoral classifications and matching their commodity classification into more disaggregated international sectoral classifications might not be feasible. However, at higher levels of industry aggregation, intra-industry trade might be observed while inter-industry trade is occurring. A country might be exporting a product classified under one 4 digit ISIC code while importing another product classified under a different 4 digit ISIC code, but under the same 3 digit sectoral classification. It would appear to be engaged in inter-industry trade when considering the 4 digit classifications but intra-industry trade when considering the corresponding 3 digit classification. Hence, the finer the sectoral classification, the more likely that inter-industry trade would be observed and would perhaps provide a more relevant framework for analyzing the Ricardian model of inter-industry trade. The non-availability of required input–output matrices on the basis of more disaggregated sectoral classifications is thus a major constraint in analyzing the Ricardian model.

coefficients on the average productivity ratio, which is of primary interest to us, and also the R-squared value for each index.[14] In the same table we also report the rank correlation coefficients corresponding to each index.

For the regressions with country fixed effects and year fixed effects, the estimated regression coefficients are positive and significant for all five of the RCA indices. On comparing the R-squared values we do find that the Log-of-Balassa index obtains the highest value, followed by the Balassa index. For the regression with country-year fixed effects, we observe similar results. The rank correlation coefficients are simultaneously positive and significant for all indices. As expected, they do not vary much across the different indices, as many of them are simply monotonic transformations of each other.

In Table 2, we use the following regression specification:

$$RCA_{i,t} = \alpha + \beta * APLR_{i,t} + \mu_t + \varepsilon_{i,t}$$

where the variables are defined in the same way as above, except that we run separate regressions for each of the 17 sectors in our sample.[15] Table 2 below presents the regression results for each sector, along with rank correlation coefficients.

Table 2 shows that for the sector specific regressions with year fixed effects, the ARCA index and the Log–of-Balassa index have the largest number of significant coefficients with expected sign. They are followed by the indices of RSCA, Balassa and NRCA in succession. On comparing the R-squared values averaged over 17 sectors, we find that Balassa's index outperforms the others. The Log-of-Balassa index and ARCA index have the second and third highest values for average R-squared. In terms of the positive and significant rank correlation coefficients, all except the NRCA index perform equally well.

As per the correlation results in Table 1, all the five indices are consistent with the predictions of the Ricardian theory. The rank correlation results in Table 2 show that all the indices perform better than the NRCA index. Thus across regression specifications, the four indices –Balassa, RSCA, ARCA and Log-of-Balassa – emerge as reasonably good performers.

On the basis of the regression test results in Table 1, considering both the number of significant coefficients with the expected sign and the value of the R-squared statistic, the Log-of-Balassa index is visibly the most consistent with the theory. On the basis of similar criteria, the regression results in Table 2 reveal the Log-of-Balassa index to be largely consistent with the theory. In Table 2, the Balassa and ARCA indices do perform better on some dimensions – the R-squared values are higher for Balassa than Log-of-Balassa, while ARCA has the same number of significant coefficients as Log-of-Balassa. However, these marginally better performances must be balanced against the non-normality of the distribution of these two indices, which may be problematic for some parametric tests.

[14] We do not report the estimated coefficients of fixed effects due to space constraints. However, these results are available on request.

[15] In practice, this means that we no longer subscript our variables with sectors, as a variables in a given regression are from the same sector. Also, we no longer use country-year fixed effects, but rather just year fixed effects.

Table 1 Estimated coefficients for large sample

RCA Index/ Results	Regression results with country FE and year FE		Regression results with country-year FE		Rank correlation results
	β	R^2	β	R^2	ρ
Balassa	0.6141614** (<0.001)	0.1399	0.6253481** (<0.001)	0.1453	0.3681** (<0.001)
RSCA	0.1014924** (<0.001)	0.1391	0.1021088** (<0.001)	0.1421	0.3681** (<0.001)
ARCA	0.0123098** (<0.001)	0.1110	0.0127683** (<0.001)	0.1116	0.3930** (<0.001)
NRCA	0.0001929** (<0.001)	0.0070	0.0001976** (<0.001)	0.0071	0.2656** (<0.001)
Log-of-Balassa	0.2969533** (<0.001)	0.1615	0.2996733** (<0.001)	0.1656	0.3681** (<0.001)

Note: values in parenthesis are *p* values for 1 tail when the alternative hypothesis is p>t. ** represents significant at 1 % level. Each regression and correlation results are based on a sample size of 3094

To conclude, Log-of-Balassa could be selected to be the most reliable RCA index among all because of its uniform good performance under both parametric and non-parametric tests.

5 Implausible value-added calculations and robustness checks

There is one notable feature about the above analysis. The calculated values for value-added in exports exceed the corresponding observations on domestic value-added for about 12 % of the country-sector-years in our data. In principle, it should be impossible for value-added in exports to exceed total value-added for a given sector. There are two potential explanations for this discrepancy in the data. One is that our assumptions about the input–output matrices for exports being similar to the input–output matrices for domestic production do not always hold. The other potential issue is that there may be mis-measurement in some of the variables that we use to calculate the value-added in exports numbers.

To perform a robustness check for our above reported results, we delete all country-years containing a sector with such observations and re-compute the correlation and regression results. The coefficients and R-squared values are reported in Tables 3 and 4 below in similar sequence as Tables 1 and 2.

As per the regression results of Table 3, all the indices report positive and significant regression coefficients for both types of regressions. However, the magnitude of the R-squared value is highest for Balassa's index and Log-of-Balassa index reports the second highest R-squared. So in this table, Balassa's index performs somewhat better than the Log-of-Balassa index. The rank correlation results however do not change with respect to the reported results in Table 1. In Table 4, the numbers of correlation and regression coefficients with expected sign are reduced for all but the NRCA index compared to Table 2. By considering both the magnitude of the R-squared values and

Table 2 Estimated coefficients for sector specific samples

ISIC codes	Balassa			RSCA			ARCA			NRCA			Log-of-Balassa		
	β	R^2	ρ	β	R^2	ρ	β	R^2	P	β	R^2	ρ	β	R^2	ρ
311	0.7504 (0.1710)	0.0285	−0.0549 (0.7692)	0.1663** (0.0015)	0.0121	−0.0549 (0.7692)	0.0153 (0.1170)	0.0004	−0.0832 (0.8679)	0.0006* (0.0400)	0.0011	0.0347 (0.3208)	0.3943** (0.0015)	0.0154	−0.0549 (0.7692)
313	0.2207 (0.1150)	0.0724	−0.0307 (0.6597)	0.0434** (0.0070)	0.0264	−0.0307 (0.6597)	0.0010 (0.0695)	0.0335	0.0141 (0.4250)	0.0000 (0.3270)	0.0001	−0.0905 (0.8879)	0.1348** (0.0015)	0.0391	−0.0307 (0.6597)
321	0.3566 (0.0510)	0.0083	0.0348 (0.3203)	0.0238 (0.2030)	0.0077	0.0348 (0.3203)	0.0131* (0.0480)	0.0039	0.0359 (0.3152)	0.0003 (0.3100)	0.0017	0.0887 (0.1170)	0.1202* (0.0365)	0.0125	0.0348 (0.3203)
322	0.8420* (0.0265)	0.0240	0.0388 (0.3018)	0.0374 (0.2075)	0.0028	0.0388 (0.3018)	0.0177** (0.0055)	0.0214	0.0288 (0.3496)	−0.0003 (0.9235)	0.0084	0.0757 (0.1549)	0.1725 (0.0660)	(0.0127)	0.0388 (0.3018)
323	1.0559** (0.0000)	0.2050	0.3323** (0.0000)	0.1553** (0.0000)	0.1180	0.3323** (0.0000)	0.0072** (0.0000)	0.2390	0.3523** (0.0000)	−2.66e-07 (0.5020)	0.0000	0.1939** (0.0044)	0.5003** (0.0000)	0.1462	0.3323** (0.0000)
331	1.5465** (0.0000)	0.0726	0.2583** (0.0002)	0.4124** (0.0000)	0.0568	0.2583** (0.0002)	0.0305** (0.0000)	0.0759	0.2392** (0.0006)	0.0008 (0.0570)	0.0045	0.1370* (0.0326)	1.1634** (0.0000)	0.0864	0.2583** (0.0002)
341	0.3260** (0.0000)	0.0466	0.5190** (0.0000)	0.1298** (0.0000)	0.0691	0.5190** (0.0000)	0.0193** (0.0000)	0.0439	0.4616** (0.0000)	0.0003 (0.0510)	0.0011	0.3385** (0.0000)	0.2855** (0.0000)	0.0667	0.5190** (0.0000)
351	0.4914** (0.0000)	0.2532	0.6028** (0.0000)	0.3198** (0.0000)	0.2649	0.6028** (0.0000)	0.0693** (0.0000)	0.2557	0.6035** (0.0000)	0.0010 (0.0545)	0.0727	0.4306** (0.0000)	0.6873** (0.0000)	0.2648	0.6028** (0.0000)
353	0.6835** (0.0005)	0.0809	0.5742** (0.0000)	0.0787** (0.0000)	0.2007	0.5742** (0.0000)	0.0125** (0.0000)	0.0730	0.5476** (0.0000)	0.0003** (0.0085)	0.0459	0.5393** (0.0000)	0.2240** (0.0000)	0.1946	0.5742** (0.0000)
361	0.2959** (0.0010)	0.0584	0.2808** (0.0001)	0.1742** (0.0000)	0.0722	0.2808** (0.0001)	0.0069** (0.0005)	0.0555	0.2788** (0.0001)	0.0004** (0.0085)	0.0250	0.2921** (0.0001)	0.3705** (0.0000)	0.0729	0.2808** (0.0001)
371	0.2680** (0.0000)	0.0030	0.3146** (0.0000)	0.1387** (0.0000)	0.0043	0.3146** (0.0000)	0.0149** (0.0000)	0.0028	0.3025** (0.0000)	0.0003 (0.1120)	0.0027	0.2437** (0.0005)	0.3384** (0.0000)	0.0070	0.3146** (0.0000)
372	0.7478** (0.0000)	0.4375	0.5420** (0.0000)	0.1462** (0.0000)	0.3455	0.5420** (0.0000)	0.0231** (0.0000)	0.4314	0.5407** (0.0000)	0.0002 (0.0920)	0.0781	0.5384** (0.0000)	0.4462** (0.0000)	0.3405	0.5420** (0.0000)

Table 2 (continued)

ISIC codes	Balassa			RSCA			ARCA			NRCA			Log-of-Balassa		
	β	R^2	ρ	β	R^2	ρ	β	R^2	P	β	R^2	ρ	β	R^2	ρ
381	0.4380** (0.0000)	0.0871	0.4390** (0.0000)	0.2681** (0.0000)	0.1413	0.4390** (0.0000)	0.0204** (0.0000)	0.0705	0.4078** (0.0000)	0.0006 (0.1050)	0.0159	0.2618** (0.0002)	0.5829** (0.0000)	0.1381	0.4390** (0.0000)
382	0.0889** (0.0000)	0.0798	0.6151** (0.0000)	0.0372** (0.0005)	0.0313	0.6151** (0.0000)	0.0138** (0.0000)	0.0750	0.5788** (0.0000)	0.0003 (0.2305)	0.0008	0.0795 (0.1431)	0.0894* (0.0185)	0.0236	0.6151** (0.0000)
383	0.0081 (0.4540)	0.0000	0.1468* (0.0240)	0.0286 (0.2305)	0.0003	0.1468* (0.0240)	0.0530** (0.0000)	0.0136	0.1469* (0.0239)	0.0020 (0.0955)	0.0001	−0.0507 (0.7516)	0.1497 (0.1585)	0.0071	0.1468* (0.0240)
384	0.0820* (0.0115)	0.0061	0.2702** (0.0001)	0.0944** (0.0010)	0.0157	0.2702** (0.0001)	0.0074* (0.0415)	0.0153	0.2222** (0.0013)	0.0002 (0.3500)	0.0001	−0.0212 (0.6117)	0.8304** (0.0000)	0.0337	0.2702** (0.0001)
390	0.6054** (0.0010)	0.0859	0.1989** (0.0036)	0.2052** (0.0000)	0.0573	0.1989** (0.0036)	0.0180** (0.0000)	0.0862	0.1772** (0.0084)	0.0000 (0.4125)	0.0036	−0.0133 (0.5708)	0.6480** (0.0000)	0.0761	0.1989** (0.0036)
No. of significant coefficients / average R^2	13	0.0911	13	14	0.0839	13	15	0.0881	13	3	0.0154	9	15	0.0904	13

Note: values in parenthesis are *p* values for 1 tail when the alternative hypothesis is p>t. ** represents significant at 1 %; * represents significant at 5 %. Each regression and correlation results are based on sample size of 182

Table 3 Estimated coefficients for large sample with selected observations deleted

RCA Index/ Results	Regression results with country FE and year FE		Regression results with country-year FE		Rank correlation results
	β	R^2	β	R^2	ρ
Balassa	2.035558** (<0.001)	0.2696	2.053823** (<0.001)	0.2766	0.4152** (<0.001)
RSCA	0.1362721** (<0.001)	0.1702	0.1355469** (<0.001)	0.1721	0.4152** (<0.001)
ARCA	0.0200801** (<0.001)	0.1601	0.0206628** (<0.001)	0.1604	0.4365** (<0.001)
NRCA	0.000318* (0.011)	0.0100	0.0003221* (0.011)	0.0100	0.2846** (<0.001)
Log-of-Balassa	0.385435** (<0.001)	0.1785	0.3867763** (<0.001)	0.1810	0.4152** (<0.001)

Note: values in parenthesis are *p* values for 1 tail when the alternative hypothesis is p>t. ** represents significant at 1 %; * represents significant at 5 %. Each regression and correlation results are based on sample size of 1156

the number of positive and significant regression coefficients, the ARCA index is found to outperform the Log-of-Balassa index. As in Table 2, the Balassa index generates a larger magnitude for R-squared but lower number of significant coefficients compared to the ARCA index and the Log-of-Balassa index. On the basis of the number of positive and significant rank correlation coefficients, the indices of Balassa, RSCA and Log-of-Balassa perform better than the others.

While differences between the previous and the present analyses are noticeable – particularly with respect to the regression results – we still note that many parametric tests for their validity require the assumption of normality for estimated residuals which could be violated if the underlying variables are characterized by non-normal distributions.. Hence, the Log- of-Balassa index, which is characterized by a normal distribution and consistently performs well for both parametric and non-parametric tests in this section, could still be judged to be the most reliable index among all.

6 Conclusions

This paper fills certain gaps in the existing literature on RCA indices by empirically analyzing the consistency of the RCA indices with the Ricardian theory on comparative advantage, after incorporating domestic value-added in exports of a country and providing room for a sufficiently large dataset. The panel regression results show the Log-of-Balassa index to be a strong performer and, unlike the other strong performers in our analysis,, it is characterized by normal distribution. The RSCA index also has a normal distribution, but does not perform as well as the Log-of-Balassa index in our regression analysis. The rank correlation test results show that the indices of Balassa, RSCA, and Log-of-Balassa are all equally consistent with the Ricardian model for all of the cases analysed. Thus the Log-of-Balassa index performs well in both correlation

Table 4 Estimated coefficients for sector-specific sample with selected observations deleted

ISIC	Balassa			RSCA			ARCA			NRCA			Log-of-Balassa		
	β	R^2	ρ	β	R^2	ρ	β	R^2	ρ	β	R^2	ρ	β	R^2	ρ
311	−1.8438 (0.7990)	0.1550	−0.0581 (0.6809)	0.5428** (0.0000)	0.0009	−0.0581 (0.6809)	0.0492** (0.0090)	0.0181	−0.1540 (0.8951)	0.0027 (0.0985)	0.0078	0.2151* (0.0391)	0.9653** (0.0000)	0.0108	−0.0581 (0.6809)
313	0.5232 (0.2810)	0.1425	0.0527 (0.3347)	0.0102 (0.4520)	0.0270	0.0527 (0.3347)	0.0023 (0.2650)	0.0259	0.0853 (0.2446)	0.0002 (0.1905)	0.0000	0.0644 (0.3011)	0.1208 (0.2960)	0.0452	0.0527 (0.3347)
321	0.8640 (0.2025)	0.0016	0.0558 (0.3258)	0.0654 (0.2400)	0.0000	0.0558 (0.3258)	0.0218 (0.2605)	0.0046	0.0508 (0.3404)	0.0015 (0.3065)	0.0012	0.1486 (0.1132)	0.2161 (0.1440)	0.0000	0.0558 (0.3258)
322	5.9516** (0.0050)	0.1458	0.2085* (0.0440)	−0.0137 (0.5610)	0.0164	0.2085* (0.0440)	0.0343** (0.0000)	0.0303	0.1449 (0.1193)	0.0004 (0.3145)	0.0039	0.2737* (0.0120)	0.4375* (0.0165)	0.0369	0.2085* (0.0440)
323	3.7501* (0.0265)	0.1566	0.4581** (0.0001)	0.4825** (0.0000)	0.2257	0.4581** (0.0001)	0.0151** (0.0005)	0.2215	0.5351** (0.0000)	0.0002 (0.2460)	0.0055	0.4020** (0.0004)	1.1040** (0.0000)	0.2138	0.4581** (0.0001)
331	2.2909 (0.0525)	0.0209	−0.1640 (0.9092)	0.5613* (0.0180)	0.0551	−0.1640 (0.9092)	0.0204** (0.0035)	0.0001	−0.1280 (0.8509)	0.0007 (0.0835)	0.0088	−0.3354 (0.9974)	1.3104* (0.0220)	0.0585	−0.1640 (0.9092)
341	0.0705 (0.3935)	0.0332	0.5887** (0.0000)	0.2373** (0.0020)	0.1287	0.5887** (0.0000)	0.0140** (0.0090)	0.0684	0.5523** (0.0000)	0.0004 (0.3690)	0.0000	0.1928 (0.0576)	0.6814** (0.0015)	0.1585	0.5887** (0.0000)
351	0.7680** (0.0000)	0.3886	0.6048** (0.0000)	0.4795** (0.0000)	0.4134	0.6048** (0.0000)	0.1197** (0.0000)	0.3895	0.5993** (0.0000)	0.0045* (0.0405)	0.0774	0.2553* (0.0178)	1.0068** (0.0000)	0.4149	0.6048** (0.0000)
353	1.3997** (0.0085)	0.2907	0.7222** (0.0000)	0.0718** (0.0000)	0.3553	0.7222** (0.0000)	0.0135** (0.0080)	0.3050	0.7227** (0.0000)	0.0002* (0.0220)	0.1124	0.5578** (0.0000)	0.1988** (0.0000)	0.3577	0.7222** (0.0000)
361	0.7590** (0.0005)	0.0400	0.1273 (0.1505)	0.3801** (0.0000)	0.0293	0.1273 (0.1505)	0.0158** (0.0005)	0.0333	0.1300 (0.1453)	0.0003 (0.3145)	0.0006	0.1284 (0.1484)	0.8148** (0.0000)	0.0352	0.1273 (0.1505)
371	1.0943** (0.0000)	0.4214	0.5625** (0.0000)	0.3084** (0.0000)	0.2203	0.5625** (0.0000)	0.0516** (0.0000)	0.4153	0.5624** (0.0000)	0.0025* (0.0410)	0.1006	0.6094** (0.0000)	0.8262** (0.0000)	0.2549	0.5625** (0.0000)
372	2.2461** (0.0000)	0.7299	0.5277** (0.0000)	0.0863* (0.0125)	0.2094	0.5277** (0.0000)	0.0511** (0.0000)	0.7499	0.5327** (0.0000)	0.0012** (0.0040)	0.2115	0.4655** (0.0001)	0.2682** (0.0020)	0.3026	0.5277** (0.0000)

Table 4 (continued)

ISIC	Balassa			RSCA			ARCA			NRCA			Log-of-Balassa		
	β	R^2	ρ	β	R^2	ρ	β	R^2	ρ	β	R^2	ρ	β	R^2	ρ
381	0.4489** (0.0000)	0.2855	0.6578** (0.0000)	0.2618** (0.0000)	0.2713	0.6578** (0.0000)	0.0204** (0.0000)	0.2927	0.6664** (0.0000)	0.0009 (0.0875)	0.0341	0.5248** (0.0000)	0.5375** (0.0000)	0.2579	0.6578** (0.0000)
382	0.3520** (0.0010)	0.4173	0.8121** (0.0000)	0.2919** (0.0010)	0.4511	0.8121** (0.0000)	0.0452* (0.0140)	0.3298	0.8005** (0.0000)	0.0028 (0.1820)	0.0098	0.2859** (0.0091)	1.1913* (0.0115)	0.4772	0.8121** (0.0000)
383	0.0591 (0.2380)	0.0012	0.3755** (0.0008)	0.0006 (0.4950)	0.0025	0.3755** (0.0008)	0.0915** (0.0000)	0.0427	0.3090** (0.0052)	0.0036 (0.2115)	0.0014	0.0782 (0.2632)	−0.0263 (0.5330)	0.0001	0.3755** (0.0008)
384	−0.1030 (0.8265)	0.0001	0.3929** (0.0005)	−0.0543 (0.6980)	0.0093	0.3929** (0.0005)	−0.0118 (0.7715)	0.0131	0.3653** (0.0011)	−0.0003 (0.5625)	0.0005	0.2232* (0.0337)	0.3700 (0.1490)	0.0694	0.3929** (0.0005)
390	0.1255 (0.2010)	0.0772	0.3278** (0.0032)	0.1379 (0.0695)	0.0610	0.3278** (0.0032)	0.0018 (0.2850)	0.0835	0.2989** (0.0067)	−0.0002 (0.6865)	0.0019	−0.0741 (0.7258)	0.3351 (0.0945)	0.0483	0.3278** (0.0032)
No. of significant coefficients / average R^2	9	0.1946	12	11	0.1457	12	13	0.1779	11	4	0.0340	10	12	0.1613	12

Note: values in parenthesis are *p* values for 1 tail when the alternative hypothesis is p>t. ** represents significant at 1 %; * represents significant at 5 %. Each regression and correlation results are based on sample size of 68 and regression analyses. Although the NRCA index, performs at par with the other indices in the bigger sample, it falls short with respect to the others in case of sector specific smaller samples when considering both the correlation and regression analyses. Thus, while the NRCA index is characterized by stable but non-normal distributions, its empirical performance under different circumstances does not establish it as a consistently strong performer. In this respect, the Log-of-Balassa index, while it lacks the feature of stable distribution, clearly performs well in the empirical analysis. In addition, it possesses a normal distribution.

Overall, it seems reasonable to conclude the paper by observing the fact that usage of any particular RCA index should be governed by the objective of the researcher. If one wishes to compare different sectors in a country or different countries in a sector, given the stability of its distribution the NRCA could be most reliable, although empirical results could not provide substantial evidence in support of its consistency with the Ricardian theory of comparative advantage relative to other RCA indices. But if one wishes to examine, say, different factors on which the RCA index of a country depends through regression analysis, given the performance of Log-of-Balassa index in our empirical analysis, it may be best to depend upon that index, although one should take into account the instability of its mean when making cross-country or cross-commodity comparisons. Our conclusion thus differs to some extent from that of Deb and Basu (2011). On the basis of our results, we conclude that is difficult to identify any particular index with all features of a good index and at the same time would empirically be most consistent with the theories on comparative advantage. However, the Log-of-Balassa index performs quite strongly in our test of consistency with the Ricardian model and has the desirable characteristic of a normal distribution.

Future research on this topic might focus on two areas. First, our use of the average productivity of labor as a proxy for the marginal product of labor (which is more consistent with the Ricardian model) could potentially be improved upon. Better ways of measuring the marginal product of labor at a disaggregated sectoral level would be desirable. Secondly, our data is at a relatively high level of sectoral aggregation due to the issues mentioned above with respect to matching data from different sources. More disaggregated and detailed data with appropriate concordances could improve the analysis.

References

Ando M (2006) Fragmentation and vertical intra-industry trade in East Asia. N Am J Econ Financ 17(3):257–281

Arndt SW (1997) Globalization and the open economy. N Am J Econ Financ 81(1):71–79

Balassa B (1965) Trade liberalisation and ‘revealed’ comparative advantage. Manch Sch 33(2):99–123

Balassa B, Bauwens L (1988) Changing trade patterns in manufactured goods: An econometric approach. Elsevier, New York

Bebek UG (2011) Monotonicity of additive indices revealed comparative advantage. Econ Bull 31(2):1894–1901

Bowen H, Leamer E, Sveikauskas L (1987) Multicountry, multifactor tests of the factor abundance theory. Am Econ Rev 77(5):791–809

Cantwell J (1989) Technological Innovation and Multinational Corporations. Basil Blackwell, Oxford

Dalum B, Laursen K, Villumsen G (1998) Structural change in OECD export specialisation patterns: despecialisation and ‘stickiness’. Int Rev Appl Econ 12(3):423–443

Daudin G, Rifflart C, Schweisguth D (2011) Who produces for whom in the world economy? Can J Econ 44(4):1403–1437
Davis DR (1995) Intra-industry trade: a Heckscher-Ohlin-Ricardo approach. J Int Econ 39(1995):201–226
Deardorff AV (2001) Fragmentation in simple trade models. N Am J Econ Financ 12(2):121–137
Deb K, Basu P (2011) Indices of revealed comparative advantage and their consistency with the Heckscher-Ohlin theory. Foreign Trade Rev 46(3):3–28
Eaton J, Kierzkowski H (1984) Oligopolistic Competition, Product Variety and International Trade. In: Kierzkowski H (ed) Monopolistic Competition and International Trade. Oxford University Press, Oxford, pp 69–83
Falvey RE (1981) Commercial policy and intra-industry trade. J Int Econ 11(1981):495–511
Falvey RE, Kierzkowski H (1987) Product Quality, Intra-Industry Trade and (Im)Perfect competition. In: Kierzkowski H (ed) Protection and Competition in International Trade. Basil Blackwell, New York, pp 143–161
Feenstra, RC (1998) Integration of trade and disintegration of production in the global economy. J Econ Perspectives 12(4):31–50
Feenstra RC, Hanson GH (1997) Foreign direct investment and relative wages: evidence from Mexico's maquiladoras. J Int Econ 42(3–4):371–439
Frantzen D (2008) Technology, competitiveness and specialisation in OECD manufacturing. J Econ Stud 35: 44–68
Galton F (1889) Natural Inheritance. Macmillan, London
Grubel H, Lloyd PJ (1975) Intra-Industry Trade: The Theory and Measurement of International Trade in Differentiated Product. Macmillan, London
Hart PE (1995) The convergence of labour productivity in British and German industries. Int J Econ Bus 2: 453–464
Hart PE, Prais SJ (1956) The analysis of business concentration: a statistical approach. J R Stat Soc Ser A (Gen) 119:150–191
Helpman E (1981) International trade in the presence of product differentiation, economies of scale and monopolistic competition: a Chamberlin-Heckscher-Ohlin approach. J Int Econ 11(1981):305–340
Helpman E (1985) Multinational corporations and trade structure. Rev Econ Stud 52(3):443–457
Hillman AL (1980) Observations on the relation between "revealed comparative advantage" and comparative advantage as indicated by pre-trade relative prices. Weltwirtschaftliches Arch 116:315–321
Hoen A, Oosterhaven J (2006) On the measurement of comparative advantage. Ann Reg Sci 40(3):677–691
Hummels D, Rapoport D, Yi K (1998) Vertical specialisation and the changing nature of world trade. Fed Reserv Bank N Y Econ Policy Rev 4(2):79–99
Hummels D, Ishii J, Yi K (2001) The nature and growth of vertical specialization in world trade. J Int Econ 54(2001):75–96
Johnson R, Noguera G (2012) Accounting for intermediates: production sharing and trade in value-added. J Int Econ 86(2012):224–236
Jones RW, Kierzkowski H (2001) A Framework for Fragmentation. In: Arndt SW, Kierzkowski H (eds) Fragmentation. New Production Patterns in the World Economy. Oxford University Press, Oxford, pp 17–34
Jones RW, Kierzkowski H, Leonard G (2002) Fragmentation and Intra-Industry Trade. In: Lloyd PJ, Lee HH (eds) Frontiers of Research in Intra-Industry Trade. Palgrave McMillan, New York, pp 67–86
Kinoshita Y, Campos NF (2006) Determinants and effects of foreign direct investment in transition economies. In: Nakagawa J (ed) Managing Development: Globalisation, Economic Restructuring and Social Policy. Routledge, England, pp 261–285
Koopman R, Wang Z, Wei S (2012) Estimating domestic content in exports when processing trade is pervasive. J Dev Econ 99(2012):178–189
Koopman R, Wang Z, Wei S (2014) Tracing value-added and double counting in gross exports. Am Econ Rev 104(2):459–494
Krugman P (1979) Increasing returns, monopolistic competition and international trade. J Int Econ 9(1979): 469–479
Krugman P (1980) Scale economies, product differentiation and the pattern of trade. Am Econ Rev 70(5):950–959
Lancaster K (1980) Intra-industry trade under perfect monopolistic competition. J Int Econ 10(2):151–170
Laursen K (1998) Revealed Comparative Advantage and the Alternatives as Measures of International Specialisation. DRUID Working Papers, 1998–30
Meng B, Fang Y, Yamano N (2012) Measuring Global Value Chains and Regional Economic Integration: An International Input–output Approach. IDE Discussion Paper No. 362

Miroudot S, Lanz R, Ragoussis A (2009) Trade in Intermediate Goods and Services. OECD Trade Policy Working Paper No. 93

Nicita A, Olarreaga M (2007) Trade, production and protection. World Bank Econ Rev 21(1):165–171

NRC (2006) Analyzing the U.S. content of imports and the foreign content of exports. In: National Research Council. The National Academies Press, Washington DC

Pittiglio R (2014) An essay on intra-industry trade in intermediate goods. Mod Econ 5:468–488

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

Sanidas E, Shin Y (2011) Convergence towards the revealed comparative advantage neutral point for East Asia: similarities and differences between the three countries. Seoul J Econ 24:21–50

Sharma A, Dietrich M (2007) The structure and composition of India's exports and industrial transformation (1980–2000). Int Econ J 21:207–231

Trefler D (1995) The case of missing trade and other mysteries. Am Econ Rev 85(5):1029–1046

Türkcan K (2011) Vertical intra-industry trade and product fragmentation in the auto-parts industry. J Ind Compet Trade 11(2):149–186

Türkcan K, Ates A (2011) Vertical intra-industry trade and fragmentation: an empirical examination of the U.S. auto-parts industry. World Econ 34(1):154–172

Vollrath T (1991) A theoretical evaluation of alternative trade intensity measures of revealed comparative advantage. Rev World Econ 127(2):265–280

Yi K (2003) Can vertical specialization explain the growth of world trade? J Polit Econ 111(1):52–102

Yu R, Cai J, Leung P (2009) The normalized revealed comparative advantage index. Ann Reg Sci 43(1):267–282