Compositional Analysis of Trade Flows Structure

Klára Hrůzová
Palacký University,
Czech Republic

Miroslav Rypka
Palacký University,
Czech Republic

Karel Hron
Palacký University,
Czech Republic

Abstract

Statistical analysis of trade flows structure can significantly help to reveal or to confirm important macroeconomic phenomena. Because of relative character of these multivariate observations, application of standard multivariate methods directly to raw data can lead to meaningless results, affected by trade sizes of different countries. As a way out, it is proposed to employ the logratio methodology that is able to capture interesting features through logratios between compositional parts. Particularly, the perturbation operation together with clr coefficients for coordinate representation of compositions seem to be easy to handle and to interpret for the purpose. Popular exploratory tools, principal component analysis and PARAFAC modeling of three-way data, resulting from a long-term study of the export/import structure, are applied in the compositional context for data from OECD and WIOD databases. The results show that the logratio methodology enables to reveal interesting features of world trade flows and thus provides a preferable alternative to existing exploratory tools.

Keywords: compositional data, perturbation, principal component analysis, PARAFAC, export and import.

1. Introduction

In today's globalised world, export and import play an important role in the country's economic situation. Globalisation causes growth of international trade in goods and services and two structural changes in trade patterns: the increasing importance of emerging economies and rapid growth of trade in intermediate goods as a result of vertical specialisation, meaning that each country is specialised in one or more innovation and production processes and thus it is common for the value chain of a particular final product to span several countries. Trade in intermediate goods currently represents about 56 % of total global trade in goods (Miroudot, Lanz, and Ragoussis 2009) and therefore we intend to explore trade flows broken down by end-use categories to better monitor international trade patterns.

As emphasized in Rodrik (2006) and Hausmann, Hwang, and Rodrik (2007), it is no longer important how much a country exports, but what it exports. Moreover, even manufacturing processes are fragmented, which means that tasks requiring low-skilled labour (e.g. assembling, control) are off-shored to developing countries (or countries with lower labour costs). This contributes significantly to the amount of exports while the value added to the product

in developing countries may be small. Consequently, much more interest in the part is devoted to relative structure of export rather than to its amount in absolute numbers.

The fragmentation of the manufacturing process can be analyzed using input-output tables (see Stehrer, Foster, and de Vries 2010; Timmer, Erumban, Gouma, Los, Temurshoev, de Vries, and Arto 2012; Timmer, Los, Stehrer, and de Vries 2013). The value added may be splitted up by production factors. For the purposes of this article, we distinguish capital, low-skilled, medium-skilled and high-skilled work. We will compare (relative) shares of these factors in value added exports (i.e. domestic value added embodied in final expenditures abroad). Of course, export structure is closely linked to import shares, so they cannot be analyzed separately in order to obtain concise and predicative results.

The aim of the article is to introduce appropriate statistical techniques for analysis and visualization of structure of trade flows in goods. Since we focus on the structure of trade flows, the absolute values of exports and imports are no longer relevant for the analysis. Thus we consider the data as compositional, i.e. carrying only relative information, which leads to a new perspective to the data processing. Although this perspective is recently intensively discussed in many applied fields from geochemistry and chemometrics to social sciences (Pawlowsky-Glahn and Buccianti 2011), just a few papers were published with purely an economic motivation (see Fry 2011, and references therein). On the contrary, even when the authors are aware of relative nature of the underlying economic data, this feature is mostly not (or just sloppily) taken into account for the statistical analysis (Blejer and Fernandez 1980; Devarajan, Swaroop, and Zou 1996).

In the next section, the basics of logratio methodology to compositional data analysis, essential for the purposes of this article, will be recalled together with two speficic methods, applied in the following — principal component analysis and PARAFAC. A particular focus will be devoted to the operation of perturbation, linked to the geometrical structure of compositional data, that enables easily to link the export and import structure of countries. Accordingly, in Section 3, the theoretical contributions are applied to the real-world data of exports and imports, where their structure is explored with respect to end-use categories and factors in value added. In the last section the results are briefly discussed.

2. Logratio methodology to compositional data analysis

To motivate the concept of compositional data, the basic idea will be explained with an example. Let household expenditures on housing, foodstuffs, other goods (including clothing, footwear and durable goods) and services in various countries are of interest. Obviously, their absolute values is hardly comparable due to different price level in each country. On the other hand, the relative structure of expenditures (that can be expressed, e.g., in proportions or percentages) can be quite similar. Consequently, ratios between components as a source of the relevant information, which remains unaltered with any scaling performed, can much better reflect specific situation in various countries than by processing the raw input data. Therefore, when the relative information is of main interest, the sum of components (leading to expression in the local currency, proportions, etc.) should not affect the result of statistical processing. We refer to *scale invariance* of compositional data which is completely violated when the whole analysis is based on the fixed representation of such data.

Technically, compositional data (Aitchison 1986) are strictly positive multivariate observations that carry only relative information. Accordingly, the only relevant information is contained in ratios between parts of a composition. The sample space of representations of compositional data with a prescribed constant sum constraint, the simplex, $\mathcal{S}^D$, consists of D-part compositions $\mathbf{x} = (x_1, \ldots, x_D)'$, where $\sum_{i=1}^{D} x_i = \kappa$ (which equals 100 for the case of percentages and 1 for proportions).

The specific nature of compositional data induces its own geometrical structure, called the Aitchison geometry, which has Euclidean vector space structure. Basic operations of the

Aitchison geometry (Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado 2015) are perturbation and powering, defined for compositional vectors $\mathbf{x} \in \mathcal{S}^D$ and $\mathbf{y} \in \mathcal{S}^D$ and a real number α as follows,

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}[x_1 y_1, \ldots, x_D y_D]'; \ \ \alpha \odot \mathbf{x} = \mathcal{C}[x_1^\alpha, \ldots, x_D^\alpha]',$$

where

$$\mathcal{C}(\mathbf{x}) = \left[\frac{\kappa \cdot x_1}{\sum_{i=1}^D x_i}, \frac{\kappa \cdot x_2}{\sum_{i=1}^D x_i}, \ldots, \frac{\kappa \cdot x_D}{\sum_{i=1}^D x_i} \right]'$$

stands for an arbitrarily chosen representation of the resulting composition (the closure operation). In the standard Euclidean geometry in real space, these two operations correspond to summation of vectors and multiplication of a vector by a scalar, respectively. The operation of perturbation can be also interpreted as shifting with respect to the Aitchison geometry, i.e. as a measure of difference appropriate to compositional change (Aitchison and Ng 2005). The perturbation-subtraction of $\mathbf{x}$ and $\mathbf{y}$,

$$\mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus [(-1) \odot \mathbf{y}] = \mathcal{C}[x_1/y_1, \ldots, x_D/y_D]',$$

then represents the relative difference between both compositions. In other words, how the compositions differ in terms of ratios between the corresponding components. Obviously, if all the parts in the resulting composition are the same(neutral elements), the relative contributions conveyed by both compositional vectors coincide.

The Aitchison inner product, norm and distance, defined for two compositions $\mathbf{x}$ and $\mathbf{y}$ as

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j}; \ \ \|\mathbf{x}\|_a = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} \right)^2};$$

$$d_a(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|_a = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2}, \tag{1}$$

respectively, complete the Euclidean vector space structure of the Aitchison geometry.

Although the Aitchison geometry closely follows the relative nature of compositional data, most of standard statistical methods cannot be used there as they are designed for the Euclidean geometry in real space (Pawlowsky-Glahn *et al.* 2015). Moreover, in order to apply them to compositions, any such method would need to fulfil three principles, resulting from specific character of compositional data. The first principle is the mentioned scale invariance which means that output of the processing must remain the same irrespective to the change of measurement units. The second one is called subcompositional coherence and is closely related to the previous principle. In particular, when dealing with a subcomposition, which consists only of a selected components of the original composition, results of any analysis should not be in conflict with those of processing the whole composition. The third principle is the permutation invariance, i.e. invariance with respect to change of order of parts in a composition.

Instead of developing specific methods directly in the Aitchison geometry, it is much easier to express compositions in the real space and proceed with standard statistical tools. For this purpose, the so called logratio coordinates, formed with respect to the Aitchison geometry, are utilized (Pawlowsky-Glahn and Buccianti 2011). It depends on the aim of the analysis, which coordinates are the most appropriate.

It turned out that for the purpose of dimension reduction methods, that will be further employed in this study, the clr coefficients (Aitchison 1986), defined as

$$\text{clr}(\mathbf{x}) = \left[\ln \frac{x_1}{g(\mathbf{x})}, \ldots, \ln \frac{x_D}{g(\mathbf{x})} \right]', \tag{2}$$

where $g(\mathbf{x})$ is the geometric mean of all parts of a composition $\mathbf{x}$, form the reasonable choice. The clr coefficients are symmetric in components, each of them expresses (through the corresponding logratio) dominance of a component with respect to average behaviour of the other parts, aggregated by their geometric mean; i.e., the relative contribution of each part to the other components in average is captured. On the other hand, the sum of clr coefficients is zero as they correspond to a generating system with respect to the Aitchison geometry (Pawlowsky-Glahn *et al.* 2015). The reason is that dimension of a D-part composition is just $D - 1$. This reflects the fact that it can be represented in a $(D - 1)$-dimensional subspace (the simplex of proportions, percentages) without loss of information. It also means that the corresponding covariance matrix of clr coordinates is singular. Although the clr coefficients are thus not coordinates with respect to a basis on the simplex, which would reflect the usual practice, they still possess important properties. The crucial one is an isometry between the Aitchison geometry and the Euclidean space. Concretely, for compositions $\mathbf{x} \in \mathcal{S}^D$ and $\mathbf{y} \in \mathcal{S}^D$ and real numbers α, β it holds that

$$\mathrm{clr}(\alpha \odot \mathbf{x} \oplus \beta \odot \mathbf{y}) = \alpha \cdot \mathrm{clr}(\mathbf{x}) + \beta \cdot \mathrm{clr}(\mathbf{y});$$

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \langle \mathrm{clr}(\mathbf{x}), \mathrm{clr}(\mathbf{y}) \rangle;$$

$$\|\mathbf{x}\|_a = \|\mathrm{clr}(\mathbf{x})\|; \ d_a(\mathbf{x}, \mathbf{y}) = d(\mathrm{clr}(\mathbf{x}), \mathrm{clr}(\mathbf{y})),$$

Hence, when a composition is expressed in clr coordinates, standard statistical tools (that are able to cope with singularity of the covariance matrix) can be employed.

As pointed out in the previous section, the aim of this article is to analyse the structure of export and import in the end-use categories. The question is how to compare export and import of different countries. In the standard case, one would compute simply differences between components. However, each country has different area, different size of population, different GDP and different structure of the economy. This means that if we would just subtract import from export values, the results could be completely misleading. The problem can be solved using the mentioned perturbation-subtraction, i.e. by taking the ratios of export and import for every end-use category, and further statistical processing in clr coordinates.

2.1. Principal component analysis

Principal component analysis (PCA) is one of the most popular statistical techniques when analysing the multivariate structure of a dataset. The aim of this method is to reduce the data dimension in order to preserve most of the variability which is captured by small number of new variables - principal components (PCs).

Principal components for a mean-centered data matrix $\mathbf{X}_{(n \times D)}$ are obtained through linear transformation $\mathbf{U} = \mathbf{X}\mathbf{B}$, where $\mathbf{U}_{(n \times D)}$ is the score matrix, whose columns $(\mathbf{u}_1, \ldots, \mathbf{u}_D)$ are the mentioned principal components, and $\mathbf{B}_{(D \times D)}$ stands for the loading matrix (Härdle and Simar 2012). The first PC is defined to have the largest possible variance, the second PC has to be orthogonal to the previous one and again posses the largest possible variance. Other PCs are defined in the same way.

In order to get principal components, the definiton of the matrix $\mathbf{B}$ is required. The loading matrix can be obtained via eigenvalue decomposition of the covariance matrix $\mathbf{\Sigma}$ of $\mathbf{X}$. Accordingly, $\mathbf{\Sigma} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}'$, where $\mathbf{\Lambda} = \mathrm{Diag}\{\lambda_1, \ldots, \lambda_D\}$ denotes the diagonal matrix of eigenvalues in decreasing order. In other words, the data matrix $\mathbf{X}$ can be interpreted as a product of the score and loading matrices,

$$\mathbf{X} = \mathbf{U}\mathbf{B}' \text{ with } \mathbf{U}'\mathbf{U} = \mathbf{\Lambda}^2 \text{ and } \mathbf{B}'\mathbf{B} = \mathbf{I},$$

where $\mathbf{I}_D$ is the identity matrix. Consequently, bilinear decomposition is obtained.

For representation of the results of PCA, loadings and scores, the graph called biplot (Gabriel 1971; Gower and Hand 1996) is often applied. In the biplot scores (as points) and loading

vectors (as rays) of the first two principal components are displayed. In case of standard multivariate data, the length of the rays approximates the standard deviations of the original variables and the cosine of the angle between two rays displays correlation coefficients between the corresponding variables.

The differences for the compositional biplot (Aitchison and Greenacre 2002; Kynčlová, Filzmoser, and Hron 2016) consist in applying PCA on clr coordinates of **X** defined in (2). This implies different interpretation: rays now represent variability of relative dominance of compositional parts with respect to the rest of components, conveyed by the clr variables. Instead of correlation between two clr coefficients (that might be misleading due to singularity of the corresponding covariance matrix) rather variance of the pairwise logratio, approximated by the length of a link between two vertices, is considered. In particular, when the rays (vertices) coincide, the variance $\text{var}(\ln \frac{x_i}{x_j})$ is approximately equal to zero which means that compositional parts x_i and x_j are interchangeable.

2.2. Parallel factor analysis

When in addition to the first two modes (samples, variables) also the third one, corresponding to conditions (like time or several measurement techniques, applied to the same samples), the bilinear PCA is no longer appropriate. One particular case is, when the same samples (countries) are observed for the same variables (end-use categories) in a long-term study, like for several years (occasions). Although it would be possible to analyse the data separately using PCA for each year, or even to apply PCA for the whole unfolded data set, by doing so the three-way structure could not be recognized. To analyse the complex structure of data simultaneously, the method called parallel factor analysis (PARAFAC) or canonical decomposition (CANDECOMP) needs to be applied (Harshman 1970; Carroll and Chang 1970). The data are decomposed into trilinear components where each component consists of one score vector and unlike PCA two loading vectors (though it is also usual to refer to three loading vectors). A PARAFAC model of three-way array (Carroll and Chang 1997) is thus given by three loading matrices **A**, **B** and **C** with elements a_{if}, b_{jf} and c_{kf} that minimize the sum of squares of the residuals e_{ijk} coming from expression

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \tag{3}$$

for $i = 1, \ldots, n, \ j = 1, \ldots, D$ and $k = 1, \ldots, K$.

The solution of the PARAFAC model (estimation of the loading matrices for a given number of factors F) can be found using alternating least squares (ALS) by assuming the loading vectors of two modes known and then estimating the unknown set of parameters of the last mode using the least squares regression (Carroll and Chang 1997; Kroonenberg 1983). The algorithm works in an iterative manner and under mild conditions it converges to a unique solution (Harshman and Lundy 1984; Stegeman 2006). From the compositional perspective, the rotational invariance of the ALS algorithm (Kruskal 1989) is of particular importance, because it enables to employ any logratio coordinates with the isometry property (like clr coefficients) for the estimation purposes (Di Palma, Gallo, Filzmoser, and Hron 2016). Although PARAFAC or, more generally, statistical modeling of three-way data was recently successfully employed for economic applications (Dell'Anno and Amendola 2015; Veldscholte, Kroonenberg, and Antonides 1998) and its specifics for compositional data were developed (Gallo 2013; Gardlo, Smilde, Hron, Hrdá, Karlíková, Friedecký, and Adam 2016), combination of both aspects (as far as it is known to the authors) is not available in the literature.

Similarly as of PCA, it is popular to display PARAFAC results graphically. Concretely, loading values of the first two components are displayed in terms of three scatterplots, one for each of modes. Subsequently, the obtained information can be merged together in order to get a concise view on the three-way structure. There are not specific features in case of

compositional data here, except to the fact that interpretation of clr variables needs to be taken into account.

3. Applications to trade flows structure

Theoretical considerations, introduced in the previous section, were applied on the real-world data which include the values of exported and imported goods of EU countries and 13 other largest economies of the world (regarding available data of WIOD database). These countries represented more than 85% of the world GDP in 2008. The first data set, trade flows broken down by end-use categories, is available online (`http://stats.oecd.org/index.aspx?queryid=32186`), the second database - shares of value added broken down by factors can be obtained from the WIOD database (`www.wiod.org`).

All the computations and graphical outputs were performed using the packages robCompositions (Templ, Hron, and Filzmoser 2011) and ThreeWay (Giordani, Kiers, and Del Ferraro 2014) of statistical software R (R Core Team 2016). Accordingly, the optimal number of components in the PARAFAC model was derived using the NumConvHull procedure (Ceulemans and Kiers 2006).

3.1. Trade flows in end-use categories

Breaking down trade in goods according to their end-use (OECD Directorate for Science, for Economic Analysis, and Statistics 2014) adds a new dimension to the traditional commodity-based trade statistics and provides a link to National Accounts Input-Output Tables, in which flows of goods and services are reported according to end-users. Using the basic domestic end-use categories from the System of National Accounts and the detailed classification systems of trade in goods, bilateral flows of exports and imports can be classified into intermediate goods, household consumption goods and capital goods. However, some kinds of products can be either for intermediate demand and household consumption, or for capital goods in industry and household consumption. Thus it was introduced mixed end-use category which contains personal computers, passenger cars, personal phones, packed medicines and precious goods. The last category, miscellaneous, includes commodities that don't belong to any other categories. To keep the presented study simple, we will not consider this category for further calculations. In Table 1 a small part of the data set is shown for illustration purposes.

The dataset used in this section is called The OECD STAN Bilateral Trade by Industry and End-use (Zhu, Yamano, and Cimper 2011). It firstly released in 2011 to highlight the increasing influence of export and import of intermediate goods. The values of import and export of goods are broken down by industrial sectors and, simultaneously, by end-use categories. Estimates are expressed in nominal terms, in current US dollars, and are collected from more than a hundred reporters and partners, including all 34 members of OECD and a wide range of non-members. Note that for the purpose of standard statistical analysis, without considering the relative nature of data, we would have to convert the current US dollars into constant US dollars in order to employ time. However, we are dealing with compositional data which means that just ratios between categories form the source of relevant information and thus multiplication by any constant does not affect results of the analysis. Following this idea, it is not necessary to convert the currency prior to further statistical processing using the logratio methodology.

As stated above, patterns in the relative structure of export and import of goods cannot be revealed by applying standard multivariate techniques to the raw data as the relevant information is contained exclusively in ratios between the respective components. Nevertheless, for the sake of comparison, principal component analysis was applied both to the original data and to clr coordinates for the year 2012, the most recent complete one in the database.

Obviously, when dealing with economies of different size of trade (with different population, share of trade in economy), straightforward application of PCA (see Figures 1-3) becomes

Table 1: Small part of the OECD data.

EXPORT	Intermediate	Household consumption	Capital	Mixed end-use
AUS	117409481	41592639	47616893	37627503
AUT	91332690	20919015	28060985	12265110.5
BEL	262169002	67877294	33929953	77809975.3
⋮				
TUR	85476254	45804409	13700828	6770951
USA	880162112	159919833	234236609	132854776
TWN	230018351	16278914	41361666	11368675
IMPORT				
AUS	216321786	19537887	6669449	6676966.4
AUT	103594607	30203689	19508560	15607428
BEL	269679952	58898420	33473255	70981894
⋮				
TUR	137206055	14363175	30415227	14167392
USA	1209223479	406593569	298476700	351426871
TWN	212736197	15469315	29972519	10065989
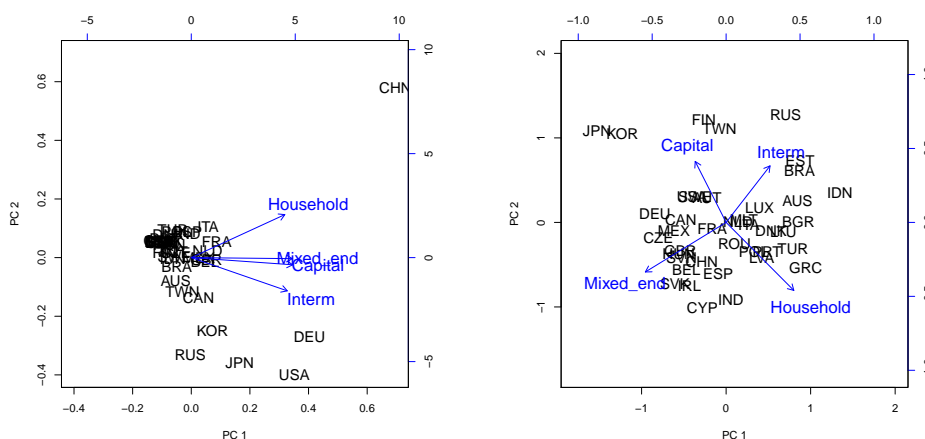
Figure 1: Biplots of export applied to the original data (on the left) and to clr coordinates (on the right).

useless. From the biplots on the left side, it is hard to recognize any structure in the dataset: it either seems that all variables are highly correlated (Figure 1 and 2), or the respective interpretation is doubtful (Figure 3).

In contrast, when relative contributions of the components, conveyed by clr coordinates (here applied to end-use categories), are considered instead, PCA and biplot diagrams are much easier to interpret (see the Figures 1 and 2 on the right). In Figure 1 (on the right), the countries exporting relatively more intermediate goods (Russia, Australia, Brazil), household (Greece, Turkey, India), mixed end-use (middle Europe countries), capital goods (Japan, Korea, Finland) can be well distinguished, no matter of their size.

Similarly, in Figure 2 on the right, the compositional biplot of import is displayed. It is evident that for Asian countries such as Korea, Taiwan, India and China dominance of intermediate and capital goods in relative structure of import can be observed. On the other hand, mixed end-use goods are imported into large countries, namely Russia, Australia, USA and Canada. Middle Europe countries are spread around the origin and Cyprus imports mostly the household consumption goods. This corresponds well to the general perspective of international trade structure of that year (UN 2012).

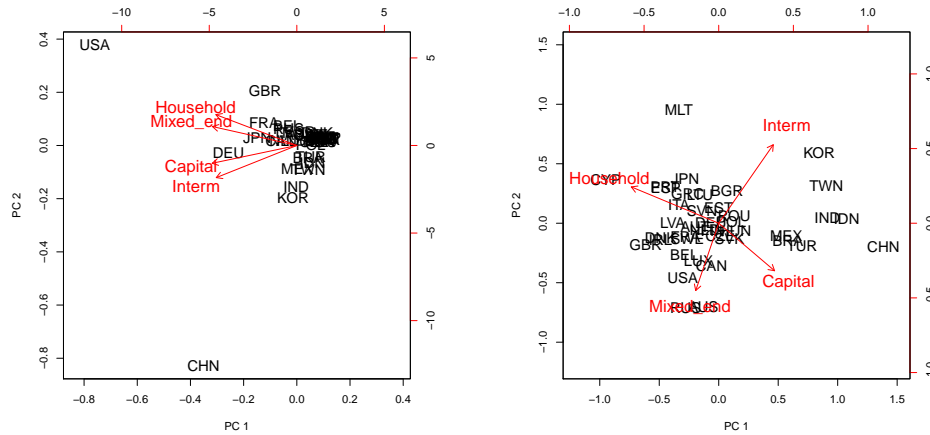The perturbation operation can be now used to capture relative differences between export

Figure 2: Biplots of import applied to the original data (on the left) and to clr coordinates (on the right).
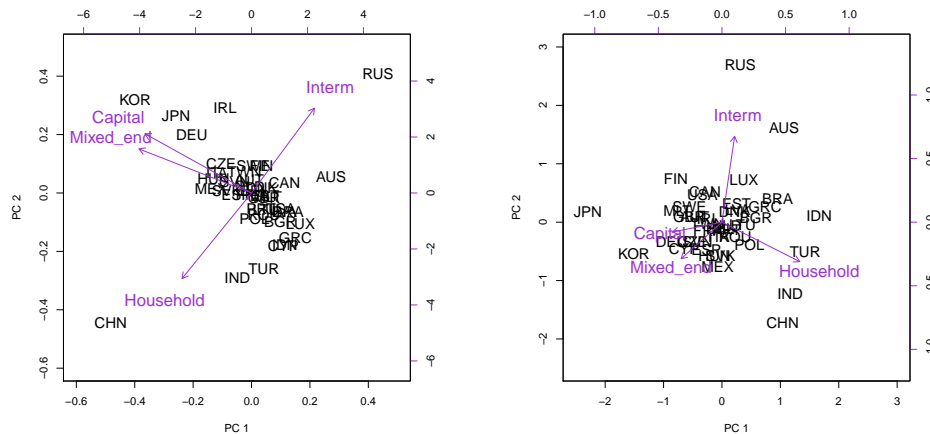

Figure 3: Biplots of differences between export and import, applied to original data (on the left) and to clr coordinates (on the right).

and import structure through ratios between the respective components. Consequently, large values of the (log-)ratios will indicate discrepancy between both international flow aspects. From the respective link in Figure 3 (right) it is visible that the variance of pairwise logratio between export/import ratios of Capital goods and Mixed end-use goods, respectively, is very small. Thus the ratios between exports and imports of these end-use categories are relatively very similar. The cluster of China, India, Indonesia and Turkey lies near the Household goods variable (in terms of its relative dominance with respect to the other categories as conveyed by the respective clr coordinate), thus these countries have the relatively largest surplus of export in this category. Russia and Australia have largest surplus in intermediate goods, while Korea and Japan in capital goods. Although these effects could be even better observed from biplot of the original data, previous results of sole export and import indicate that high variability of mixed end-use and capital goods categories is not relevant by considering relative structure of observations.

In order to include also time variable and to get a complete picture about the development in a larger time scale, also PARAFAC modeling was applied to the perturbed data, i.e. to the ratio of export and import components (after expressing them in clr coordinates) for years 2003–2012. Similar results as for the previous figures were obtained that confirm a certain stability of the export/import structure comparing to the single year 2012, considered above.
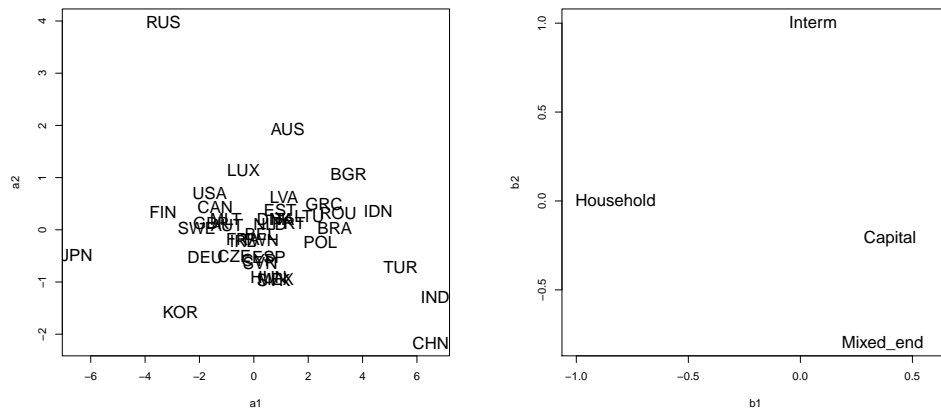
Figure 4: Results of the PARAFAC method for differences between exports and imports, mode A (on the left) and mode B (on the right), using clr coordinates.
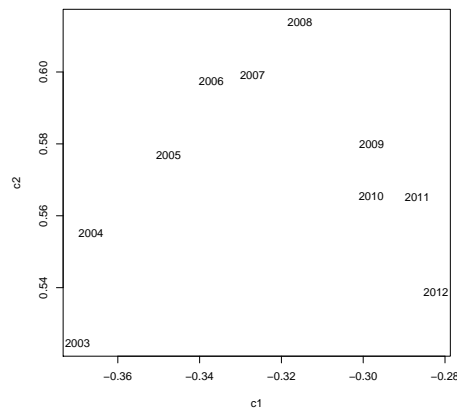

Figure 5: Results of the PARAFAC method for differences between exports and imports, mode C, using clr coordinates.

In the mode A (Figure 4 on the left), corresponding to samples, cluster of China, India, Turkey and Indonesia can be seen, as well as cluster of Japan and Korea. In the middle of the plot there is a group of middle European countries and it also seems that Russia differs significantly from the other countries. Mode B (Figure 4, right plot) confirms the result that components Capital and Mixed end-use goods are relatively very similar, when considering ratios of export and import for the years 2003–2012. And finally, mode C displayed on the Figure 5 shows the development in time, where a clear time pattern with a change point in 2008 is observed, interpretable in terms of global integration. Accordingly, this loading plot well reflects the global crisis in 2008–2009 that has temporarily brought the long-run trend of rising global integration through trade to a halt.

3.2. Trade flows of value added

Since an intensive integration process recently, the flows of value added across countries have become more relevant than the flows of goods. It is caused by the growing effect of the vertical specialization, which can be explained in a way that firms offshore activities to other countries to exploit cost advantages in particular stages of production (for more see Stehrer, Foster, and de Vries 2012; Hummels, Ishii, and Yi 2001). As discussed above, the share of intermediates in trade is significant. In order to distinguish real contribution (represented by value added)

Table 2: Small part of VA data.

VA EXPORT	Capital	High-skilled	Medium-skilled	Low-skilled
AUS	236672.49	108796.630	122716.748	109261.377
AUT	68726.10	49733.400	86969.011	12356.746
BEL	100983.53	59336.529	102739.534	27714.804
⋮				
TUR	204194.6	42477.77	31392.93	39164.41
TWN	79596.1	71722.10	34759.83	22437.47
USA	3632767.3	3073707.79	2353237.78	196386.87
VA IMPORT				
AUS	22149.071	6364.3767	9851.844	5749.7064
AUT	15143.087	6587.8375	11316.180	3869.4331
BEL	29874.281	12481.4846	19397.737	9089.2901
⋮				
TUR	8765.471	4035.702	6835.904	3442.306
TWN	29002.087	6870.004	12055.759	8311.096
USA	261998.337	46909.534	84104.518	48165.174

of each country in its exports and other countries in its imports, the composition of value added export and import was explored.
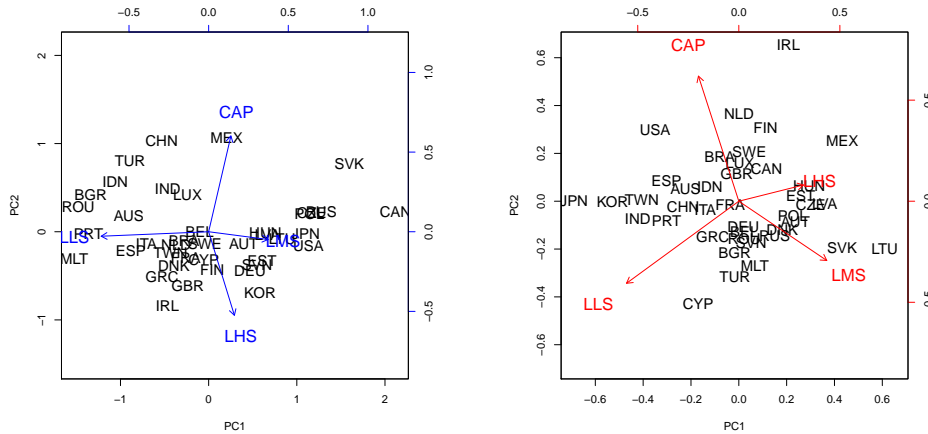

Figure 6: Compositional biplots of value added export (on the left) and import (on the right) of factors for clr coordinates.

The WIOD database (Timmer *et al.* 2012) allows to break value added of final products into factors, namely capital (CAP) and labour (low skilled (LLS), medium skilled (LMS) and high skilled (LHS)). The database comprises gross output and value added by industry for each country and the flow of products across industries and countries in a global input-output matrix. The WIOD database provides a time series of world input-output tables (WIOTs) from 1995 to 2009. The shares of factors in each industry for all considered countries may be found in the Socio Economic Accounts table (may be downloaded from `http://www.wiod.org/new_site/database/seas.htm`).

Our second data set (see Table 2) is obtained from WIOTs and Socio Economic Accounts table in the following way. From WIOTs, we can calculate value added export (VAX) (for detailed treatment see Johnson and Noguera (2012) and Timmer, Dietzenbacher, Los, Stehrer, and Vries (2015)) for each country and each industry. Employing Socio Economic Accounts table we obtain share of each factor in the calculated value added in each industry. Summing by industry we get shares of each factor in VAX for each country. Similarly we can split value added by other countries in imports to each country.

It is well known (Stehrer *et al.* 2010; Timmer *et al.* 2013) that developed countries export relatively more high skilled labour and import more capital. In contrast, developing countries are abundant with low skilled labour and import high skilled labour. This is illustrated by Figure 6 for the year 2009, for which the database provides complete data. Indeed, China, Turkey and Indonesia export relatively more low-skilled labour and capital, southern part of EU low-skilled labour (in sense of their relative contributions with respect to the other components, reflected by clr coordinates). The new countries of the EU have significant abundance in medium-skilled labour as well as United States or Japan. The opposite tendency can be seen in Figure 6 on the right, where compositional biplot of import of factors is displayed.

To see the development in time, the PARAFAC model was applied to data for years 2000–2009. In Figure 7 the results are displayed. By considering modes A and B of export (left panel) together, countries can be divided into two parts. In the left part, the countries exporting relatively more low-skilled labour are clustered (e.g., southern European countries, Turkey, India, Indonesia or China). However, in the right part clusters of countries that export relatively more capital, high- and medium skilled labour can be seen - Canada, USA, Japan, Korea and middle European countries. From the mode B it can be concluded that export of LMS and LHS is quite strongly proportional. Mode C of the left panel reflects the change in year 2004, when an intensive integration process for many European countries as new members of the European Union started.

Similarly as for the case of export, from mode B of the right panel it can be observed that import of LHS and LMS is proportional (though not so closely as for the case of export). Moreover, clusters of countries from mode A are similar to those from the biplot in Figure 6 (right). Accordingly, 1) Ireland, Finland, Sweden, Netherlands and USA, 2) Malta, Cyprus, Portugal, Turkey and Bulgaria, and 3) Japan, India, Taiwan and Korea have similar relative structure of import in terms of value added. In Mode C, the development is not so clear as for the case of export, however it still reflects the exceptional role of the year 2004.

4. Discussion

With development of detailed publicly available databases, it is possible to analyse systematically also the international trade structure. Nevertheless, it is of particular importance to consider carefully the natural properties of the observations at hand prior to their further statistical processing. The case of export and import structure shows that problems with different trade sizes can be overcome by employing the logratio methodology of compositional data. Although PCA (biplot) and PARAFAC are standard tools for analysis and visualization of multivariate data, their application in the compositional and economic contexts simultaneously form the main novelty of the paper. Results of analysing the international trade structure reflects well the general knowledge, as provided regularly by the United Nations (UN) and other institutions.

Apparently, the interpretation provided in the previous section is just illustrative capturing the main features and there is still space for its further extension. For example, differences in factors related to the export can be seen also from much broader perspective. In case of the European Union, one can distinguish "core" EU countries, its southern countries and new countries. Accordingly, the difference in technological structure of export, which is related to the level of skills, is often accounted for problems of Euro (see, e.g., Wierts, Van Kerkhoff, and De Haan 1998). We leave these issues as inspiration for those, who would employ the logratio methodology for more detailed macroeconomic analyses in the future.

Acknowledgments

References

Aitchison J (1986). *The Statistical Analysis of Compositional Data.* Chapman and Hall, London.

Aitchison J, Greenacre M (2002). "Biplots of Compositional Data." *Applied Statistics*, **51**, 375–392.

Aitchison J, Ng KW (2005). "The Role of Perturbation in Compositional Data Analysis." *Statistical Modelling*, **5**, 173–185.

Blejer MI, Fernandez RB (1980). "Effects of Unanticipated Money Growth on Prices and on Output and Its Composition in a Fixed-exchange-rate Open Economy." *Canadian Journal of Economy*, **13**, 82–95.

Carroll JD, Chang J (1970). "Analysis of Individual Differences in Multidimensional Scaling via an *n*-way Generalization of "eckartyoung" Decomposition." *Psychometrika*, **35**, 283–319.

Carroll JD, Chang J (1997). "PARAFAC. Tutorial and Applications." *Chemometrics and Intelligent Laboratory Systems*, **38**, 149–171.

Ceulemans E, Kiers HAL (2006). "Selecting among Three-mode Principal Component Models of Different Types and Complexities: A Numerical Convex Hull Based Method." *British Journal of Mathematical and Statistical Psychology*, **59**(1), 133–150.

Dell'Anno R, Amendola A (2015). "Social Exclusion and Economic Growth: An Empirical Investigation in European Economies." *Review of Income and Wealth*, **61**(2), 274–301.

Devarajan S, Swaroop V, Zou H (1996). "The Composition of Public Expenditure and Economic Growth." *Journal of Monetary Economics*, **37**, 313–344.

Di Palma MA, Gallo M, Filzmoser P, Hron K (2016). "A Robust Parafac Model for Compositional Data." Under review.

Fry T (2011). "Applications in Economics." In V Pawlowsky-Glahn, A Buccianti (eds.), *Compositional Data Analysis: Theory and Applications*, pp. 318–326. Wiley, Chichester.

Gabriel KR (1971). "The Biplot Graphic Display of Matrices with Application to Principal Component Analysis." *Biometrika*, **58**, 453–467.

Gallo M (2013). "Log-ratio and Parallel Factor Analysis: An Approach to Analyze Three-way Compositional Data." In AN Proto, M Squillante, J Kacprzyk (eds.), *Advanced Dynamic Modeling of Economic and Social Systems*, pp. 209–221. Springer, Heidelberg.

Gardlo A, Smilde AK, Hron K, Hrdá M, Karlíková R, Friedecký D, Adam T (2016). "Normalization Techniques for PARAFAC Modeling of Urine Metabolomic Data." *Metabolomics*, **12**, 117.

Giordani P, Kiers HAL, Del Ferraro MA (2014). "Three-Way Component Analysis Using the R Package ThreeWay." *Journal of Statistical Software*, **57**(7).

Gower JC, Hand DJ (1996). *Biplots.* Chapman & Hall, London.

Härdle WK, Simar L (2012). *Applied Multivariate Statistical Analysis*. Springer, Heidelberg.

Harshman RA (1970). "Foundations of the Parafac Procedure: Models and Conditions for an "Explanatory" Multi-modal Factor Analysis." *Report 10085*, University of California, Los Angeles.

Harshman RA, Lundy ME (1984). "The PARAFAC Model for Three-way Factor Analysis and Multidimensional Scaling." In HG Law (ed.), *Research Methods for Multimode Data Analysis*, pp. 122–215. Praeger, New York.

Hausmann R, Hwang J, Rodrik D (2007). "What You Export Matters." *Journal of Economic Growth*, **12**(1), 1–25.

Hummels D, Ishii J, Yi KM (2001). "The Nature and Growth of Vertical Specialization in World Trade." *Journal of International Economics*, **54**, 75–96.

Johnson RC, Noguera G (2012). "Accounting for Intermediates: Production Sharing and Trade in Value Added." *Journal of international Economics*, **86**(2), 224–236.

Kroonenberg PM (1983). *Three-mode Principal Component Analysis. Theory and Applications*. DSWO Press, Leiden.

Kruskal JB (1989). "Rank, Decomposition, and Uniqueness for 3-way and N-way Arrays." *Multiway Data Analysis*, **33**, 7–18.

Kynčlová P, Filzmoser P, Hron K (2016). "Compositional Biplots Including External Non-compositional Variables." *Statistics*, pp. 453–467. doi:10.1080/02331888.2015.1135155.

Miroudot S, Lanz R, Ragoussis A (2009). "Trade in Intermediate Goods and Services." *OECD Trade Policy Working Paper*, **93**.

OECD Directorate for Science T, for Economic Analysis ID, Statistics (2014). *OECD Bilateral Trade Database by Industry and End-use Category*. OECD Publishing.

Pawlowsky-Glahn V, Buccianti A (eds.) (2011). *Compositional Data Analysis: Theory and Applications*. Wiley, Chichester.

Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (eds.) (2015). *Modeling and Analysis of Compositional Data*. Wiley, Chichester.

R Core Team (2016). "R: A Language and Environment for Statistical Computing."

Rodrik D (2006). "What's So Special About China's Exports?" *China & World Economy*, **14**(5), 1–19.

Stegeman A (2006). "Degeneracy in Candecomp/Parafac Explained for $p \times p \times 2$ Arrays of Rank $p + 1$ or Higher." *Psychometrika*, **71**(3), 483–501.

Stehrer R, Foster N, de Vries G (2010). "Value Added and Factors in Trade: A Comprehensive Approach." *Dynamics*, **67**.

Stehrer R, Foster N, de Vries G (2012). "Value Added and Factors in Trade: A Comprehensive Approach." *wiiw Working paper*, **80**, 1–22.

Templ M, Hron K, Filzmoser P (2011). "robCompositions: an R-package for Robust Statistical Analysis of Compositional Data."

Timmer M, Erumban AA, Gouma R, Los B, Temurshoev U, de Vries GJ, Arto I (2012). "The World Input-output Database (WIOD): Contents, Sources and Methods." *WIOD Background document*, **40**. URL www.wiod.org.

Timmer MP, Dietzenbacher E, Los B, Stehrer R, Vries GJ (2015). "An Illustrated User Guide to the World Input–output Database: The Case of Global Automotive Production." *Review of International Economics*, **23**(3), 575–605.

Timmer MP, Los B, Stehrer R, de Vries GJ (2013). "Fragmentation, Incomes and Jobs: an Analysis of European Competitiveness." *Economic Policy*, **28**(76), 613–661.

UN (2012). *World Economic Situation and Prospects.* United Nations, New York.

Veldscholte CM, Kroonenberg PM, Antonides G (1998). "Three-mode Analysis of Perceptions of Economic Activities in Eastern and Western Europe." *Journal of Economic Psychology*, **19**(3), 321–351.

Wierts P, Van Kerkhoff H, De Haan J (1998). "Composition of Exports and Export Performance of Eurozone Countries." *JCMS: Journal of Common Market Studies*, **52**(4), 928–941.

Zhu S, Yamano N, Cimper A (2011). "Compilation of Bilateral Trade Database by Industry and End-use Category." In *OECD Science, Technology and Industry Working Papers.* OECD Publishing, Los Angeles.

Affiliation:

Klára Hrůzová
Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science
Palacký University
771 46 Olomouc, Czech Republic
E-mail: klara.hruzova@gmail.com

Miroslav Rypka
Department of Geoinformatics, Faculty of Science
Palacký University
771 46 Olomouc, Czech Republic
E-mail: rypkamir@seznam.cz

Karel Hron
Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science
Palacký University
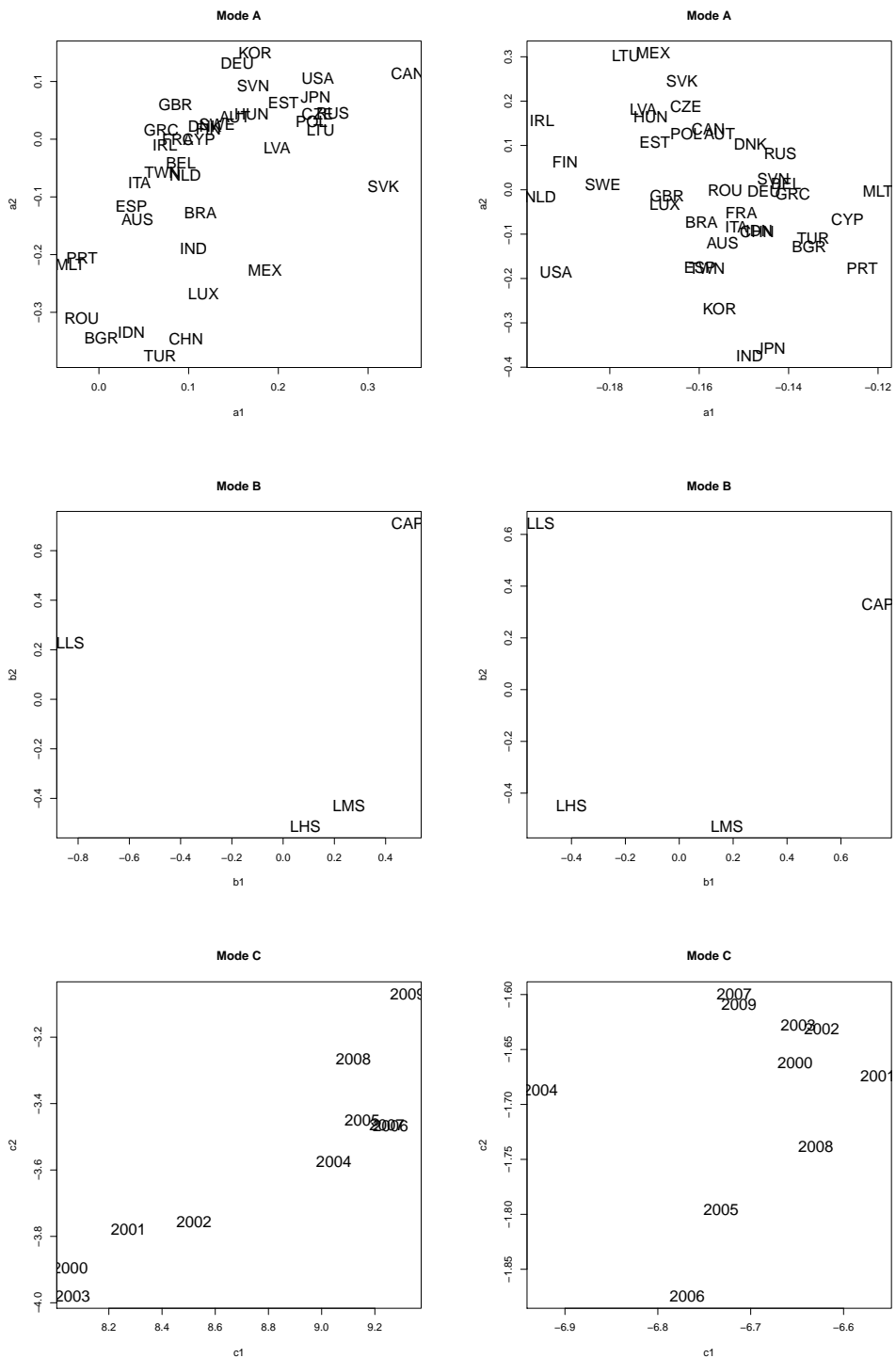771 46 Olomouc, Czech Republic
E-mail: hronk@seznam.cz

Figure 7: Results of the PARAFAC method for the export (left panel) and import (right panel) of value added using clr coordinates.