UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

BAKALÁŘSKÁ PRÁCE

Statistická analýza struktury importu a exportu

Katedra matematické analýzy a aplikací matematiky
Vedoucí bakalářské práce: **doc. Karel Hron, Ph.D.**
Vypracoval: **Tomáš Pohanka**
Studijní program: B1103 Aplikovaná matematika
Studijní obor Matematika-ekonomie se zaměřením na bankovnictví / pojišťovnictví
Forma studia: prezenční
Rok odevzdání: 2017

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Tomáš Pohanka

Název práce: Statistická analýza struktury importu a exportu

Typ práce: Bakalářská práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: doc. Karel Hron, Ph.D.

Rok obhajoby práce: 2017

Abstrakt: Struktura exportu a importu je významným ukazatelem hospodářské úrovně jednotlivých států. Relativní charakter dat přitom vyžaduje speciální přístup k jejich statistickému zpracování s využitím metodiky kompozičních dat. V úvodu práce se budeme věnovat teorii související s metodami využitými ke statistické analýze. Seznámíme se s exportem, importem, kompozičními daty a některými metodami pro zpracování trojfaktorových dat. Následuje samotná analýza, jejímž cílem bude získat malý počet nových proměnných, které zachytí co nejvíce informace o variabilitě původních dat. Následně si tyto proměnné vykreslíme do grafů a popíšeme, které státy si jsou podobné strukturou ať už exportu nebo importu, popř. která složka v této struktuře dominuje. Získané výsledky jsou v tomto případě porovnány s původními daty, abychom ověřili jejich relevanci.

Klíčová slova: Export, import, kompoziční data, knihovna ThreeWay, trojfaktorová data

Počet stran: 45

Počet příloh: 1

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Tomáš Pohanka

Title: Statistical analysis of export and import structure

Type of thesis: Bachelor's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: doc. Karel Hron, Ph.D.

The year of presentation: 2017

Abstract: The structure of export and import is an important indicator of the economic level of each country. The relative trait of such data requires a special approach to its statistical processing using compositional data. In the introduction of the bachelor's thesis, we will deal with theory related to methods used for statistical analysis. We will introduce export, import, compositional data and some of the dimension reduction methods used for analyzing three-way data where the aim is to get a small number of new variables, which preserve most of the variability of original data. Then we will plot these variables and ascertain, which countries have a similar structure of export and import, respectively which components are dominating to this structure. Obtained results will be compared with original data to check whether they are relevant.

Key words: Export, import, compositional data, package three-way, three-way data

Number of pages: 45

Number of appendices: 1

Language: Czech

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně pod vedením pana doc. Karla Hrona, Ph.D. a všechny použité zdroje jsem uvedl v seznamu literatury.

V Moravském Berouně dne ....

podpis

Obsah

Poděkování

Rád bych poděkoval panu doc. RNDr. Karlu Hronovi Ph.D. za vedení práce, trpělivost a cenné rady. Dále bych chtěl poděkovat paní Mgr. Kláře Hrůzové, Ph.D. za poskytnutí datového souboru použitého k analýze.

Úvod

Export a import mají v dnešním globalizovaném světě v rámci ekonomiky každého státu větší roli, než kdy dřív. Pro názorný příklad ani nemusíme chodit za hranice České republiky. Kolik Čechů dnes nosí české oblečení, používá českou elektroniku nebo konzumuje české potraviny? Kromě přínosu do státní pokladny můžeme na tyto toky pohlížet i jako na významné makroekonomické ukazatele. Podílí se mj. na výši HDP, ale vypovídají i o tom, jestli daná ekonomika roste.

V této práci bude hlavním cílem statistická analýza dat pro objemy těchto dvou ukazatelů. Nevyužijeme však k ní standardní statistické metody, které se nabízí. Relativní charakter těchto dat umožňuje využít kompozičních dat. Tato metodika není v ekonomických disciplínách používána tak často, jako v jiných oborech, ačkoliv se jejich použití na ekonomická data v mnoha případech vybízí. Jak se v práci dozvíme, použití kompozičních dat bude znamenat, že nebudeme porovnávat země z hlediska objemu mezinárodního obchodu, ale jeho relativní struktury. Pro tyto účely bude třeba mít obchodní toky rozděleny do kategorií.

V teoretické části práce se nejprve zaměříme na export, import a související ekonomickou interpretaci. Krátce se podíváme i do historie mezinárodního obchodu, což by čtenáři mělo pomoci uvědomit si jeho význam. Dále představíme data použitá k analýze a způsob, jakým jsou rozdělena do jednotlivých kategorií.

Svou pozornost budeme poté věnovat kompozičním datům. Nejprve je třeba seznámit se s teorií nezbytnou pro to, abychom s nimi mohli dále pracovat. Zmínka tak padne především o specifické geometrii, která tato data charakterizuje, a způsobu, jak se jí vyhnout a pracovat s kompozičními daty užitím standardních statistických metod.

Pro další práci s daty budeme používat metody redukce dimenze dat. Nejprve se podíváme na analýzu hlavních komponent, která je nejužívanější metodou a bude sloužit jako úvod do třífaktorové analýzy, která bude použita k samotné analýze. Využívat budeme její dvě konkrétní metody, a sice PARAFAC a Tucker3.

Po teoretické části bude následovat část praktická, kde využijeme nabyté znalosti k provedení samotné analýzy. K analýze bude použit statistický software R.

Kapitola 1

Export a import

Vývoz neboli export je objem statků a služeb, které daná země prodá v zahraničí. Dovoz neboli import je naopak objem statků a služeb ze zahraničí, které jsou prodány v daném státě. Obě veličiny zaujímají v ekonomice každého státu podstatnou roli. Obyvatelé různých zemí mají různé potřeby, ale geografické podmínky nebo ekonomická vyspělost státu nemusí umožňovat tyto potřeby uspokojit. Proto je potřeba některé statky, ale i služby dovážet ze zahraničí. Na druhou stranu může země disponovat zbožím a surovinami, které jsou žádané za jejími hranicemi a jsou tam tedy vyváženy.

V mezinárodním obchodě rozlišujeme u zemí, co se výroby týče, absolutní a komparativní výhody, které mohou s geografickou polohou úzce souviset. Tyto dva pojmy ovšem skrývají i další důvody, proč je pro státy lepší některé produkty dovážet. Má-li země oproti ostatním absolutní výhodu ve výrobě některého statku, znamená to, že dokáže vyrobit daného statku větší množství za stejné náklady než ostatní státy. Pro takovéto země je výhodné zaměřit svou výrobu na tyto statky a přebytek prodat za hranicemi. Ačkoliv má země absolutní výhodu v produkci více statků, může se stát, že pro něj bude výhodnější některý z těchto komodit dovážet. Komparativní výhodu má takový stát, který má pro daný statek nižší náklady obětované příležitosti. Zjednodušeně to znamená, že pokud země, ačkoliv má absolutní výhodu ve výrobě, rozloží své zdroje mezi výrobu všech statků, nedokáže vyrobit tolik, co země, která má oproti ní komparativní výhodu. Proto je pro ni výhodnější vyrábět jen některé z nich a zbytek dovážet.

Žádaným jevem je již od 17. století, kdy ve Francii začal Jean-Baptiste Colbert prosazovat politiku tzv. merkantilismu, aby export převyšoval import a tím mj. rostly příjmy do státní pokladny. Oba toky se totiž podílí na tvorbě hrubého domácího produktu. Uveďme například vzorec pro výpočet HDP výdajovou metodou,

$$\text{HDP} = \text{C+I+G+NX},$$

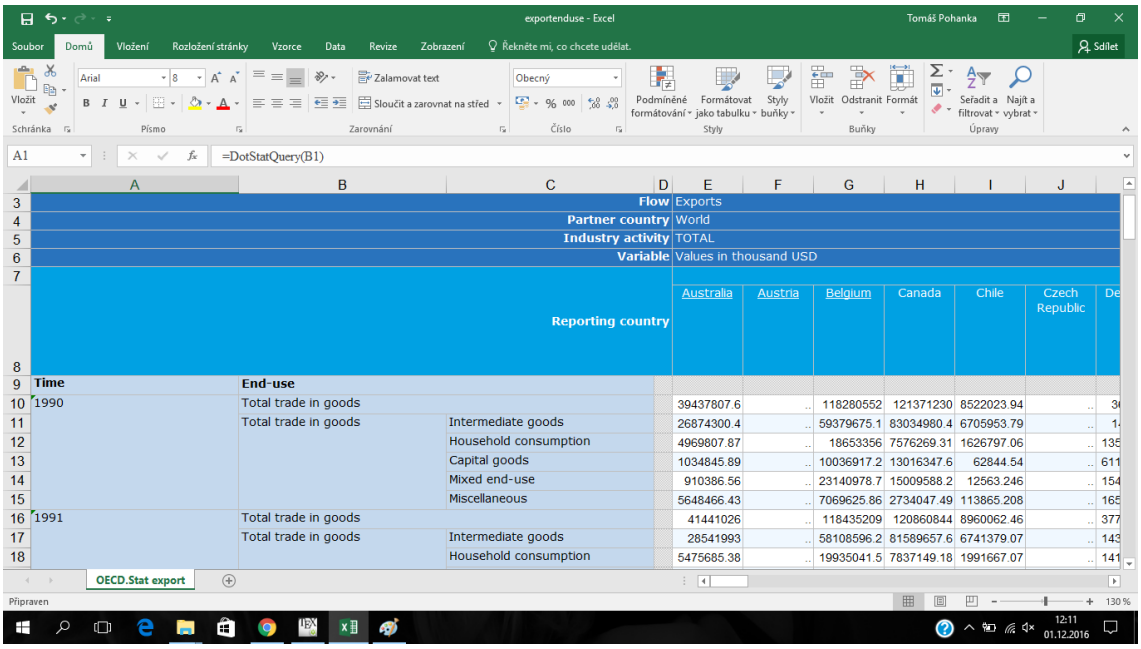
kde C značí spotřebu v rámci ekonomiky, I jsou investice podniků, G výdaje vlády a NX je čistý export, který nás zajímá.

Čistý export spočítáme jako NX = X - M, kde X je export a M import. Vidíme tedy, že převyšuje-li export import, HDP země roste. Kromě toho, že roste HDP, lze předpokládat i zvyšování zisku podniků, což má za následek růst platů, výroby a klesající nezaměstnanost. Nicméně vysoká hodnota importu nemusí být nutně negativním jevem. Vysoký objem dováženého zboží může v budoucnu znamenat potencionální růst hospodářství, zejména, jedná-li se o investiční majetek. Každopádně peníze utracené za zboží z dovozu míří do zahraničí a jako takové se na ekonomickém růstu nepodílí přímo. Objem exportu a importu přímo souvisí se sílou domácí měny. Slábnoucí měna stimuluje export, zatímco dovoz zdraží. Naopak sílící měna má za následek oslabení exportu a zlevnění importu. Je tomu tak, protože pokud domácí měna posílí proti té zahraniční, za stejné množství zboží firma dostane stejné množství peněz v zahraniční měně, ale v domácí jejich hodnota bude menší. Je ovšem zřejmé, že neméně důležitým ukazatelem je i relativní struktura exportu a importu, která může do značné míry jejich absolutní hodnoty.

Mezinárodní obchod může být i politickým nástrojem. Vážný dopad na chod země může mít vyhlášení obchodního embarga proti ní. Jiné zákony (pracovní doba, produktivní věk, ...) či cenová hladina mohou mít za následek zaplavení trhu levnými výrobky ze zahraničí (typickým příkladem jsou výrobky ze zemí jihovýchodní Asie), což by mělo negativní dopad na tuzemské výrobce. Většina států se snaží chránit vnitřní trh před negativními vlivy ze zahraničí. Tato politika se nazývá protekcionalismus [2, str.6]. Jako opatření pak zavádí vlády cla nebo například kvóty. Na jejich regulaci se snaží dohlížet Světová obchodní organizace (WTO) nebo Organizace pro hospodářskou spolupráci a rozvoj (OECD). Zboží v rámci exportu a importu můžeme různě členit, čímž dostaneme důležitou informaci o jejich struktuře. Jeden ze způsobů, který používá právě OECD, představíme v další kapitole. V rámci Evropské unie naopak funguje volný pohyb zboží bez jakýchkoliv omezení. Obdobou dohody v rámci Evropské unie pro země severní Ameriky je společnost NAFTA. Ve světě pak existují další podobné organizace.

Dříve, než vlády objevily důležitost exportu, však hrál významnější roli ve státech import. Například již ve starověku ukojil touhu Římanů po drahých tkaninách dovoz hedvábí z Asie po tzv. hedvábné stezce. Později se z Orientu dováželo např. koření, z Afriky zase exotické ovoce nebo slonovina. Velký přelom v dovozu znamenalo objevení Ameriky. Kromě nových plodin jako tabák, brambory nebo rajčata byla objevena nová ložiska drahých kovů. Jejich nadměrný dovoz do Evropy způsobil prudký pokles jejich hodnoty.

Příkladem využití zahraničního obchodu jako politického nástroje je blokáda Velké Británie během napoleonských válek. Velká Británie tak byla připravena o možnost vyvážet své zboží ze země, čímž značně klesl odbyt tamních výrobců. Pro mnohé z nich znamenala blokáda dokonce bankrot. Dalším příkladem je třeba Japonsko na počátku 20. století, které bylo na dovozu nerostných surovin závislé. Obchodní embargo vyhlášené Spojenými státy pak vedlo až ke vstupu Japonska



Obrázek 1.1: Náhled na analyzovaný datový soubor

do druhé světové války. Pro další příklad nemusíme chodit daleko do historie. V roce 2015 byly vyhlášeny ekonomické sankce vůči Rusku. Jako reakci Rusko zakázalo dovoz potravin ze zemí EU, čímž tyto země přišly o část příjmů.

Analýzou struktury exportu a importu můžeme získat mnoho informací o hospodářství jednotlivých států. Dále pak můžeme srovnat vyspělost různých ekonomik podle relativního zastoupení druhů statků, které daná země dováží nebo vyváží. Jelikož ale každá země disponuje jinými zdroji, vyrábí rozdílné objemy statků a všude ve světě je rozdílná cenová hladina, prosté aplikování standardních statistických metod by zřejmě vedlo k zcela zavádějícím výsledkům. Proto pro tyto účely budeme pracovat s tzv. kompozičními daty a logpodílovou metodikou pro jejich statistickou analýzu.

1.1. Data použitá k analýze

Použitá data je možno najít na webu OECD [8]. Na tomto odkazu je možno dohledat i další členění zkoumaných kategorií. Na obrázku 1.1 můžeme vidět náhled na použitý datový soubor s hodnotami exportu pro členské země OECD.

Jak lze z obrázku 1.1 vyčíst, hodnoty exportu, ale i importu jsou členěny dle konečného užití. Všechny částky jsou uvedeny v tisících dolarů. V případě použití standardních statistických metod by vznikl problém a to, že všechny toky by bylo třeba očistit od inflace. Mimo to se můžeme v různých zemích setkat s jinou cenovou hladinou. Jak se později můžeme dočíst v definici 2.2, použitím

kompozičních dat předejdeme nejen těmto přepočtům, ale bude dokonce irelevantní, v jaké měně či v jakých jednotkách počítáme. Data byla zaznamenávána pro 152 států po dobu 24 let. Ačkoliv nevyužijeme data v plném rozsahu, vzniká i tak specifická problematika, která povede k použití třífaktorové analýzy, se kterou čtenáře seznámíme v kapitole 3.2.

V analýze datového souboru budeme respektovat rozdělení obchodovaného zboží, jak jej provedlo OECD. Uvažovat tedy budeme pět kategorií dělených dle konečného užití výrobků a služeb. První takovou kategorií jsou meziprodukty. Do ní patří takové výrobky, které jsou použity k produkci jiných výrobků. Výrobci tohoto zboží jej mohou samy použít k další produkci nebo prodávat jiným firmám. Jako příklad lze uvést železo, motory automobilů atd. Tyto výrobky nejsou zahrnuty do HDP, pokud jsou prodávány v rámci domácího trhu, jelikož by s prodejem finálního produktu do něj byly zahrnuty dvakrát. Jako typický příklad vývozců meziproduktů, bez hlubší znalosti problematiky, nás jistě napadnou africké státy, které většinou ekonomicky závisí na těžbě nerostných surovin, ale nemají prostředky na jejich další zpracování, a tak je prodávají do zahraničí.

Druhou kategorií je spotřeba domácností, do které jsou obecně zahrnuty statky a služby, jejichž hodnota není sama o sobě ekonomicky významná, ale v součtu zaujímá významnou roli v každé ekonomice. Patří sem jídlo, oblečení, osobní služby, ale i například lístky na kulturní akce nebo do dopravních prostředků. Ačkoliv se v názvu kategorie objevují domácnosti, jsou sem zahrnuty i obchody realizované lidmi pobývajícími například v domovech důchodců, ve vězení atd. V zahraničním obchodu bude mít největší zastoupení v rámci této kategorie pochopitelně jídlo, nápoje a textil.

Třetí kategorií je investiční zboží. Sem patří všechny věci, které lze opakovaně využít k výrobě jiných statků. Jedná se zejména o dlouhodobý majetek jako jsou stroje, nářadí, některé dopravní prostředky atd. Kromě toho se však může jednat i o statky, které jsou využívány při poskytování služeb. Tím může být barva, kterou malíři vymalují pokoje, dopravní letadla aerolinek, nákladní auta dopravců nebo hudební nástroje.

Jelikož existuje zboží, které by mohlo patřit do více než jen jedné kategorie, zavedeme kategorii smíšená spotřeba. Jak už jsem uvedl, bude sem patřit zboží, které může být jak meziproduktem, tak spotřebováno v domácnostech nebo i užito ve výrobě. Jedná se o automobily, počítače nebo telefony. Poslední kategorií je ostatní zboží, kam zařadíme vše, co nepatří do žádné z předchozích.

Kapitola 2

Kompoziční data

Kompoziční data popisují části celku. Nejčastěji jsou reprezentována vektory, jejichž komponenty jsou procenta, podíly, koncentrace nebo frekvence [7, str.1]. V těchto situacích nám může metodika založená na vlastnostech kompozčních dat poskytnout lepší informace, než když na data pohlížíme klasicky, tj. s využitím absolutních hodnot jednotlivých proměnných. Obsah této kapitoly vychází z [7]. Začněme formální definicí kompozičního vektoru.

2.1. Základní pojmy

Definice 2.1. *Vektor* $\boldsymbol{x} = [x_1, x_2, ...x_D]$ *se nazývá D-rozměrná kompozice, pokud jsou všechny jeho složky kladná reálná čísla a nesou pouze relativní informaci.*

Relativní informace je obsažena v podílech složek kompozice a jejich číselná hodnota je sama o sobě bezvýznamná. Součet těchto složek je nejčastěji nějaká konstanta k. Nejčastěji se můžeme setkat s $k = 1$ pro složky vyjádřené podíly nebo $k = 100$ pro komponenty vyjádřené procenty. Jelikož pracujeme s podíly, násobení dat konstantou nemá na komponenta vliv. Tato vlastnost je popsána v následující definici.

Definice 2.2. *Vektory s D kladnými reálnými složkami* $\boldsymbol{x}, \boldsymbol{y} \in R_D^+$ $(x_i, y_i > 0,$ *pro všechna* $i = 1, 2, \ldots, D)$ *jsou kompozičně ekvivalentní, pokud existuje konstanta* $\lambda \in R_+$ *taková, že* $\boldsymbol{x} = \lambda \cdot \boldsymbol{y}$.

Tohoto lze v ekonomii využít k zjednodušení výpočtů, jelikož není potřeba data očišťovat od inflace nebo převádět částky z různých států do jednotné cenové hladiny. Pro zavedení některých operací musíme nejprve zavést operaci uzávěru, která není nic jiného, než projekcí vektorů s kladnými složkami na simplex (viz definice 2.4).

Definice 2.3. *Pro libovolný vektor s D kladnými složkami,* $\boldsymbol{z} = [z_1, z_2, \ldots, z_D] \in R_D^+$, $z_i > 0$ *pro každé i=1,2,...,D je operace uzávěru definována jako*

$$C(\boldsymbol{z}) = \left[\frac{k \cdot z_1}{\sum_{i=1}^{D} z_i}, \frac{k \cdot z_2}{\sum_{i=1}^{D} z_i}, \ldots, \frac{k \cdot z_D}{\sum_{i=1}^{D} z_i}\right].$$

Pomocí operace definované výše lze zavést jiným způsobem již zmíněnou kompoziční ekvivalenci. Vektory $\mathbf{x}$, $\mathbf{y}$ z prostoru R_+^D jsou kompozičně ekvivalentní, pokud $C(\mathbf{x}) = C(\mathbf{y})$ bez ohledu na velikost konstanty k.

Definice 2.4. *Výběrovým prostorem kompozičních dat je D-složkový simplex,*

$$S^D = \{\boldsymbol{x} = [x_1, x_2, \ldots, x_D] | x_i > 0, i = 1, 2, \ldots, D, \textstyle\sum_{i=1}^{D} x_i = k\}.$$

Každá statistická metoda aplikovaná na kompozice by měla splňovat tři podmínky. Vzhledem k tomu, že využíváme poměry, požadujeme, aby funkce, které s nimi pracují, nabývaly hodnot bez ohledu na rozšíření poměrů libovolnou konstantou. Této podmínce se říká *měřítková invariance* a můžeme ji definovat následovně:

Definice 2.5. *Mějme funkci $f(\cdot)$ definovanou na množině R_+^D. Tato funkce je měřítkově invariantní, pokud pro každou konstantu $\lambda \in R_+$ a každou kompozici $\boldsymbol{x} \in S^D$ splňuje $f(\lambda \boldsymbol{x}) = f(\boldsymbol{x})$.*

Přirozeně, jednou z těchto funkcí je podíl. Podíl $f(\mathbf{x}) = \frac{x_1}{x_2} = \frac{\lambda \cdot x_1}{\lambda \cdot x_2}$ vyhovuje definici měřítkové invarience. Nicméně záleží na pořadí, v jakém čísla uvažujeme, jelikož $\frac{x_1}{x_2} \neq \frac{x_2}{x_1}$. Vhodnou transformací těchto podílů je logpodíl $f(\mathbf{x}) = \ln(\frac{x_1}{x_2})$. Výměnou čitatele a jmenovatele dosáhneme pouze změny znaménka funkční hodnoty. Měřítkově invariantní logpodíl nazveme logkontrast.

Definice 2.6. *Mějme kompozici $\boldsymbol{x} = [x_1, x_2, \cdots, x_D] \in S^D$ a konstanty $\alpha_i \in R$ pro každé $i = 1, 2, \cdots, D$. Logkontrast je funkce*

$$f(\boldsymbol{x}) = \textstyle\sum_{i=1}^{D} \alpha_i \cdot \ln x_i, \text{ kde } \sum_{i=1}^{D} \alpha_i = 0.$$

Druhou ze zmíněných podmínek je *permutační invariance*. Funkce je permutačně invarientní, vrátí-li stejnou hodnotu pro libovolné uspořádání složek kompozice. Poslední podmínkou je *podkompoziční soudržnost*, která v rámci statistické analýzy odpovídá projekci vektoru v eukleidovském prostoru.

Specifická struktura kompozic vyžaduje zavést speciální geometrii, která k ní bude citlivá. Už od základní školy jsme všichni zvyklí pracovat v reálném prostoru s eukleidovskou geometrií. Uvažujme kompozice [20,25,40], [22,25,40] a [50,55,70], [52,55,70]. Chceme-li zjistit rozdíl prvních komponent jednotlivých dvojic vektorů, pomocí eukleidovské geometrie bychom dospěli ke stejnému číslu a to 2. Ovšem ve chvíli, kdy uvažujeme kompozice nás zajímá relativní přírůstek, který je v případě první složky 10% pro první dvojici kompozic, zatímco pro druhou 4%. Další důvody, proč nelze nadále používat eukleidovskou geometrii pro práci

s kompozicemi, jsou uvedeny v [7, str.24]. Ukázalo se, že v vhodným nástrojem pro práci s těmito daty je tzv. Aitchisonova geometrie, která respektuje uvedené tři podmínky pro relevantní analýzu kompozic. V praxi ovšem preferujeme práci v eukleidovské geometrii, ve které je zavedena většina populárních statistických metod.

2.2. Clr koeficienty

Existuje několik způsobů, jak se vyhnout práci na Aitchisonově geometrii a kompoziční data převést na nové proměnné, se kterými lze pracovat v reálném prostoru s eukleidovskou geometrií. V praxi nejpoužívanější jsou aditivní logpodílové souřadnice (alr), centrované logpodílové koeficienty (clr), izometrické logpodílové souřadnice (ilr). První souřadnicový systém nezachovává vzdálenosti mezi souřadnicemi. Jinými slovy je to souřadnicový systém izomorfní mezi S^D a R^{D-1}, ale nikoli izometrický. Clr koeficienty jsou izomorfní i izometrický systém prostoru S^D na R^D, ovšem obsahuje o jednu proměnnou více, než je dimenze Aitchisonovy geometrie (D-1). Poslední souřadnicový systém je izomorfní i izometrický prostoru S^D na R^{D-1} a navíc odstraňuje nevýhody obou předchozích. Ilr souřadnice se jeví geometricky nejvýhodněji, pro účely práci však bude stačit dále pracovat s clr koeficienty. Jakmile jsou kompozice převedeny na některé z těchto nových proměnných, lze na ně aplikovat standardní statistické metody.

Clr koeficienty získáme pomocí následujícího vztahu

$$clr(\mathbf{x}) = \left[\ln \frac{x_1}{g_m(x)}, \ln \frac{x_2}{g_m(x)}, \ldots, \ln \frac{x_D}{g_m(x)}\right],$$

kde $g_m(\mathbf{x}) = (\prod_{i=1}^{D} x_i)^{1/D}$ je geometrický průměr složek vektoru $\mathbf{x}$.

Specifikum tohoto zobrazení je nulový součet výsledných proměnných. Z toho plyne, že varianční a korelační matice jsou singulární. Můžeme podotknout, že clr souřadnice nejsou subkompozičně soudržné, pokud bychom je chtěli vnímat ve smyslu původních složek. Geometrický průměr subkompozice se zřejmě nemusí shodovat s tím celé kompozice, takže clr koeficienty komponentů podkompozice nejsou shodné s těmi, které mají stejné komponenty celé kompozice. Z toho důvodu je třeba si uvědomit, že clr koeficienty vyjadřují dominance jednotlivých složek vůči ostatním složkám v kompozici, agregovaných pomocí geometrického průměru.

Kapitola 3

Metody redukce dimenze dat

Nyní se přesuneme k metodám redukce dimenze mnohorozměrných dat, pomocí kterých budou v této práci data analyzována. Jejich užitím získáme malý počet nových proměnných, které obsahují co nejvíce množství informace, obsažené v celkovém rozptylu datového souboru. Následně je možno tyto proměnné např. zobrazit, a tak získat představu o mnohorozměrné struktuře původních dat.

3.1. Analýza hlavních komponent

Nejprve se seznámíme s analýzou hlavních komponent (angl. Principal Component Analysis, PCA), která k samotné statistické analýze v této práci využitá nebude, nicméně představuje jisté zjednodušení následně představených metod.

Analýza hlavních komponent, jako ostatní podobné metody, má za úkol zmenšit dimenzi daných dat a zároveň zachovat co nejvíce variability, která je zachycena novými proměnnými, tzv. hlavními komponentami. Ty jsou seřazeny podle množství informace (velikosti rozptylu) od nejdůležitější po nejméně důležitou, většina informace je tedy obsažena jen v několika prvních komponentách.

Vstupní datový soubor uspořádáme do matice $\mathbf{X}_{(I\times J)}$, kde I je počet pozorování a J počet znaků (proměnných). Alternativně se pozorování nazývají módem A a znaky módem B. Prvky této matice poté redukujeme do malého počtu nových proměnných, zvaných hlavní komponenty, které jsou tvořeny lineárními kombinacemi původních proměnných. Redukce dimenze dat spočívá v tom, že využijeme jen několik málo prvních komponent, které nesou největší informaci o celkovém rozptylu.

Před samotnou transformací je ještě třeba matici $\mathbf{X}$ centrovat dle průměru jednotlivých proměnných.

Následně vyjádříme

$$\mathbf{U} = \mathbf{XB},$$

kde $\mathbf{U}_{(I\times J)}$ je tzv. matice skórů. Skóry jsou pak pozice původních pozorování v novém souřadnicovém systému. Matice $\mathbf{B}_{(J\times J)}$ je matice zátěží. Zátěže můžeme interpretovat jako váhy, kterými musíme vynásobit původní proměnné, abychom dostali skóry. Sloupcové vektory $(\mathbf{u}_1, \ldots, \mathbf{u}_D)$ tedy vyjadřují skóry odpovídající zmíněným hlavním komponentám [4].

Matici $\mathbf{B}$ získáme pomocí vlastních čísel varianční matice $\mathbf{\Sigma}$ matice $\mathbf{X}$, $\mathbf{\Sigma} = \mathbf{B\Lambda B}'$, kde $\mathbf{\Lambda} = \{\alpha_1, \ldots, \alpha_D\}$ značí diagonální matici vlastních čísel v sestupném pořadí. V praxi je matice $\mathbf{\Sigma}$ reprezentována výběrovou varianční maticí. Matice $\mathbf{X}$ je tedy součin matic skórů a zátěží

$$\mathbf{X} = \mathbf{UB}' \text{ a platí } \mathbf{U}'\mathbf{U} = \Lambda^2, \mathbf{B}'\mathbf{B} = \mathbf{I}.$$

$\mathbf{I}_J$ značí jednotkovou matici řádu J.

3.2. Třífaktorová analýza kompozičních dat

Nyní uvažujme situaci, kdy takových datových matic $\mathbf{X}$, jako jsme uvažovali v předchozí kapitole, máme K. Tato situace může nastat v případě, kdy máme k dispozici I pozorování J znaků za K let. V tomto případě můžeme tyto matice uspořádat do trojrozměrného pole $\underline{\mathbf{X}}_{(I\times J\times K)}$. Jednotlivé dimenze pak odpovídají horizontální, vertikální a hloubkové ose soustavy souřadnic.

Vícefaktorová analýza komponent zahrnuje statistické metody pro případ, kdy je datová množina více než dvourozměrná. Třífaktorová analýza dat je konkrétní případ pro počet dimenzí (faktorů) roven třem. S další dimenzí je potřeba zavést ke stávajícím dvěma módům i mód C odpovídající vývoji pozorování v čase. Před samotnou analýzou je třeba všechny matice $\mathbf{X}_{..k}$ pro $k = 1, \ldots, K$ uspořádat do jedné dvourozměrné matice, kterou označíme $\mathbf{X}_A = [\mathbf{X}_{..1}, \ldots, \mathbf{X}_{..k}, \ldots, \mathbf{X}_{..K}]$, kde index k značí k-tou entitu módu C. Dva rozměry nové matice umožňují použít např. PCA, nicméně ta nebere v potaz mód C jako specifický faktor, což znamená neúplné využití informace obsažené v datech. Kvůli tomu je třeba uvažovat metody, které budou ke zpracování této matice vhodnější.

Existuje mnoho způsobů, jak analyzovat trojrozměrné matice (pole) pozorování, nejčastěji užívané techniky dle [3] pak jsou Tucker3 a PARAFAC (zkratka anglického Parallel Factor Analysis).

3.2.1. Tucker3

První technikou, kterou představíme, se nazývá Tucker3 dle Ledyarda R. Tuckera, který ji navrhl v roce 1966. Kromě verze pro redukci dimenze ve třech faktorech, existují i modely Tucker2 a Tucker1, kdy redukujeme dimenzi pole dvou,

resp. jednoho faktoru. Model Tucker3 je charakteristický tím, že umožňuje využít k analýze rozdílný počet komponent pro každý mód. Mějme matici pozorování $\mathbf{X}_A$ tak, jak je definovaná na začátku kapitoly 3.2.

Metodu Tucker3 s P komponenty pro mód A, Q komponenty pro mód B a R komponenty pro mód C můžeme vyjádřit jako

$$\mathbf{X}_A = \mathbf{A}\mathbf{G}_A(\mathbf{C}^T \otimes \mathbf{B}^T) + \mathbf{E}_A,$$

kde $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ jsou matice komponent pro odpovídající módy, čili matice skórů. Jejich prvky a_{ip} odpovídají i-té entitě p-té komponenty módu A, b_{jq} j-té entitě q-té komponenty módu B a c_{kr} k-té entitě r-té komponenty módu C. $\mathbf{G}_A$ je matice o rozměrech $I \times JK$ získaná z pole $\underline{\mathbf{G}}$ o rozměrech $(P \times Q \times R)$. Jeho prvek g_{pqr} udává interakci mezi komponenty všech tří módů. $\mathbf{E}_A$ udává matici chyb. Operace $\otimes$ se nazývá Kroneckerův součin a pro matice $\mathbf{U}$ a $\mathbf{V}$ je definován jako

$$\mathbf{U} \otimes \mathbf{V} = \begin{bmatrix} u_{11}\mathbf{V} & \dots & u_{1J}\mathbf{V} \\ \vdots & \ddots & \vdots \\ u_{I1}\mathbf{V} & \dots & u_{IJ}\mathbf{V} \end{bmatrix}.$$

Model Tucker3 lze alternativně vyjádřit skalárně následujícím způsobem:

$$x_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} + e_{ijk}.$$

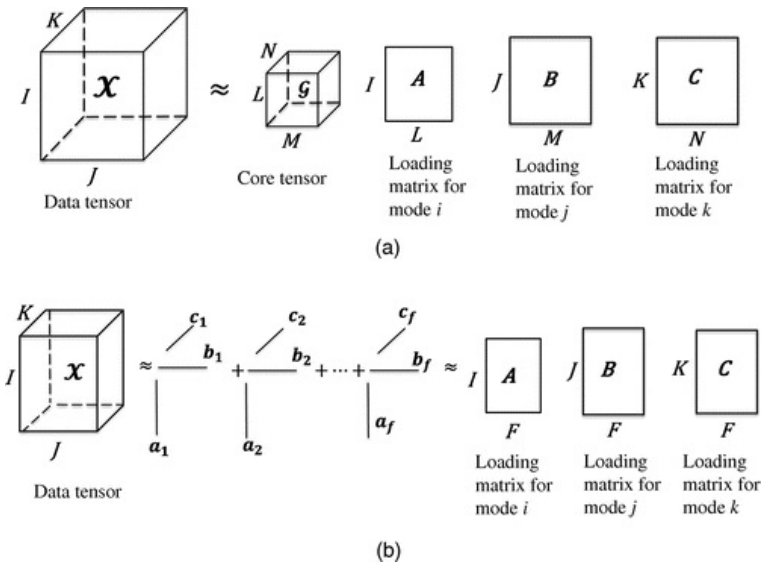
V případě, kdy bychom chtěli redukovat dva nebo pouze jeden mód, definovali bychom obdobně i Tucker 2 vypuštěním např. módu C, resp. Tucker1 vypuštěním i módu B z výše psaného vzorce. Konkrétně lze tyto vztahy nalézt např. v [3].

Optimální řešení modelu dosáhneme minimalizováním

$$||\mathbf{E}_A||^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} e_{ijk}^2,$$

čehož lze dosáhnout metodou nejmenších čtverců. Minimalizovat součet čtverců pro tři neznámé matice je obtížný úkol, který by se řešil jen těžce. Proto se pro výpočet matic skórů používá speciální algorimus, označován jako ALS (Alternating Least Squares). Principem ALS je zafixovat hodnoty pro dvě matice skórů a odhadnout prvky té třetí. Postup se následně opakuje pro zbylé dvě matice s již odhadnutými hodnotami. Algoritmus ukončíme, pokud je rozdíl hodnot ztrátové funkce ve dvou po sobě jdoucích iteracích menší než předem daný práh. Je dokázáno, že po konečném počtu iterací algoritmus konverguje k lokálnímu minimu matice chyb. Aby byla šance nalezení pouze lokálního minima minimální, je doporučeno více průchodů algoritmem při různých počátečních (většinou náhodných) volbách matic A, B, C.

Získané řešení není jediným. Jiné optimální řešení můžeme získat, pokud budeme uvažovat $\mathbf{A}$=$\mathbf{AS}$, $\mathbf{B}$=$\mathbf{BT}$, $\mathbf{C}$=$\mathbf{CU}$ a $\tilde{\mathbf{G}}_A = \mathbf{S}^{-1}\mathbf{G}_A((\mathbf{U}^T)^{-1} \otimes (\mathbf{T}^T)^{-1})$, kde matice $\mathbf{S}$, $\mathbf{T}$ a $\mathbf{U}$ jsou ortogonální matice. Změna hodnot matic $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ se vykompenzuje změnou matice $\underline{\mathbf{G}}$.



Obrázek 3.1: Grafická reprezentace modelu Tucker3 (a) a modelu PARAFAC (b) [9]

3.2.2. PARAFAC

Model PARAFAC na rozdíl od předchozího modelu umožňuje pouze redukci na stejný počet komponent pro každý mód. Skalárně lze tento model vyjádřit takto:

$$x_{ijk} = \sum_{s=1}^{S} a_{is} b_{js} c_{ks} + e_{ijk}.$$

Jak jsme již zmínili, PARAFAC a Tucker3 jsou obecně vzato rozdílné modely, nicméně z výše uvedeného vztahu lze vidět, že PARAFAC můžeme považovat za speciální případ modelu Tucker3 pro $P = Q = R = S$ a platí $g_{pqr} = 1$, pokud $p = q = r$ a $g_{pqr} = 0$ jinak.

Vztah mezi těmito dvěma modely lépe uvidíme, pokud PARAFAC vyjádříme maticově:

$$\mathbf{X}_A = \mathbf{A}\mathbf{I}_A(\mathbf{C}^T \otimes \mathbf{B}^T) + \mathbf{E}_A,$$

kde $\mathbf{I}_A$ je maticová verze trojrozměrného pole $\mathbf{I}$ ($i_{pqr} = 1$, pokud $p = q = r$ a $i_{pqr} = 0$ jinak). Vidíme tedy, že PARAFAC lze chápat jako speciální případ modelu Tucker3 pro $\mathbf{I}_A = \mathbf{G}_A$. Řešení lze opět získat metodou ALS a na rozdíl od Tucker3 tato konverguje k jedinému řešení. Jelikož je pole $\underline{\mathbf{G}}$ v tomto případě pevně dané, řešení není možné dále rotovat.

Názorně můžeme rozklad vstupního trojrozměrného pole $\underline{\mathbf{X}}$ pomocí obou metod vidět na obrázku 3.1. Datové soubory jsou rozloženy na jednotlivé matice skórů. U rozkladu pomocí metody Tucker3 můžeme vidět, že počty komponent jednotlivých módů mohou být různé (na obrázku označené jako L, M a N),

zatímco v případě PARAFAC musí být stejné (F). Rozklad pomocí Tucker3 obsahuje kromě těchto matic ještě jádrové pole $\underline{\mathbf{G}}$. V případě metody PARAFAC je, jak víme, toto pole rovno jednotkové matici. Obě rovnosti jsou pouze přibližné, jelikož v rozkladech není obsažena matice chyb.

Pro odhady parametrů a zobrazení příslušných skórů obou zmíněných modelů budeme využívat software R a odpovídající funkce `T3` pro Tucker3, resp. `CP` pro PARAFAC knihovny ThreeWay. Pokud jsou v jednotlivých vrstvách $\mathbf{X}_{..1}, \ldots, \mathbf{X}_{..K}$ trojrozměrného pole $\mathbf{X}$ obsažena kompoziční data, stačí je jednoduše vyjádřit v cl souřadnicích a dále pak pokračovat ve výše zpracování. Je přitom ovšem třeba mít na paměti interpretaci clr koeficientů ve smyslu dominance jednotlivých původních složek kompozice vůči průměrnému chování ostatních složek.

Kapitola 4

Analýza reálných dat

Datový soubor, který jsme si představili v kapitole 1.1, obsahuje velké množství chybějících pozorování zejména v prvních a posledních sledovaných letech. Z toho důvodu pozorování z těchto let v analýze vynecháme. V případě, že chybí data některých zemí i v analyzovaných letech, nezbývá než vynechat z analýzy pozorování i pro takovéto země. V souboru s daty se pro každý rok nachází i řádek s celkovým množstvím vyvezeného, respektive dovezeného zboží, který do analýzy nesmí být pochopitelně taktéž zahrnut. I po provedení těchto úprav datového souboru se může stát, že chybí pozorování v jedné nebo dvou kategoriích u některých zemí. Tyto chybějící hodnoty nahradíme průměrnými hodnotami daných proměnných.

Software R umožňuje prostřednictvím knihovny `robCompositions` používat funkce pro práci s kompozičními daty. Pracovat budeme s clr koeficienty. Soubor kompozic s kladnými složkami můžeme převést na clr koeficienty použitím funkce `cenLR`. Výstup této funkce však mimo koeficienty zahrnuje i vektor geometrických průměrů složek kompozic. Pokud chceme zobrazit pouze koeficienty, musíme přidat k funkci argument `$x.clr`. Knihovna `ThreeWay` pak nabízí funkce, které interaktivně provází prostředím analýzy pomocí metod PARAFAC i Tucker3. Pro analýzu je možno využít objekty třídy *matice*, *data.frame* nebo *array*, v prvních dvou případech je třeba zadat do příslušné funkce počet komponent pro jednotlivé módy.

Použijeme-li funkci `T3`, zobrazí se následující prostředí:

```
WELCOME to the interactive TUCKER3 analysis program
Warning: If you insert an object of mode CHARACTER when not requested,
an error occurs and the program stops!

Specify the number of A-mode entities
1:
Read 1 item
Specify the number of B-mode entities
```

```
1:
Read 1 item
Specify the number of C-mode entities
1:
Read 1 item
```

Do předchozích řádků musíme manuálně zadat rozměry jednotlivých módů. Následně se můžeme rozhodnout, zda chceme chceme data nejprve zpracovat pomocí analýzy rozptylu neboli ANOVA. Názvy proměnných pro jednotlivé módy můžeme v R zadat do funkcí jako jedny z parametrů, v případě, že tak neučiníme, prostředí nás vyzve, abychom popis zadali manuálně, popřípadě použije implicitní značení.

```
To see ANOVA results, specify 1:
1:

How do you want to center your array?
0 = none (default)
1 = across A-mode
2 = across B-mode
3 = across C-mode
12 = across A-mode and across B-mode
13 = across A-mode and across C-mode
23 = across B-mode and across C-mode
1:

How do you want to normalize your array?
0 = none (default)
1 = within A-mode
2 = within B-mode
3 = within C-mode
1:
```

Předchozí dvě možnosti umožní centrovat, popřípadě normovat analyzovaný soubor. Tyto možnosti můžeme použít, pokud chceme předejít velkému rozdílu v měřítku vstupních proměnných. Centrování výsledná data upraví jako poměr ku módu, respektive módům, který do prostředí zadáme. Normování slouží pro případy, kdy je značný rozdíl v měřítku proměnných jednoho či více módů.

```
To see PCA of MEAN, specify '1':
1:
```

Poslední volbou před samotnou analýzou rozhodneme, zda chceme provést analýzu hlavních komponent matice průměrů přes třetí mód. Následuje samotná trojfaktorová analýza pomocí modelu Tucker3.

```
You can now do a TUCKER3 ANALYSIS

How many A-mode components do you want to use?
1:
Read 1 item
How many B-mode components do you want to use?
1:
Read 1 item
How many C-mode components do you want to use?
1:
Read 1 item

Specify convergence criterion (default=1e-6)
1:
```

Nejprve musíme určit, kolik komponent jednotlivých módů chceme vypočítat. Tato čísla můžeme zadat intuitivně, nebo můžeme použít funkce `T3runsApproxFit` a `DimSelectorx`, které nám napoví nejvhodnější počet komponent. Následně můžeme určit práh konvergence.

```
If you want additional runs, specify how many (e.g., 4):
1:
```

Touto volbou můžeme zadat dodatečné průchody algoritmem, abychom zmenšili šanci, že mineme optimální řešení, resp. dosáhneme pouze lokálního optima. Nutno podotknout, že těmito dodatečnými průchody se k optimálnímu řešení Tucker3 přibližuje velmi pomalu a lepším způsobem je změnit počty uvažovaných komponent.

V dalším kroku můžeme zjistit mimo jiné hodnotu minimalizačního kritéria pro model Tucker3 po každém průchodu algoritmem a to, jak moc jsme se přiblížili optimálnímu řešení. Následně se objeví několik možností, které nejsou pro výše popsaný algoritmus důležité, až na následující, která nám umožní zadat váhy pro komponenty jednotlivých módů, a tedy rotovat obdržené řešení:

```
Specify (range of) relative weight(s) for A (default=0):
1:
Read 0 items

Specify (range of) relative weight(s) for A (default=0):
1:
Read 0 items

Specify (range of) relative weight(s) for C (default=0):
1:
```

```
Read 0 items
```

Použijeme-li funkci CP, objeví se velmi podobné prostředí, které nás bude provádět výpočtem skórů u jednotlivých módů pro daný počet komponent pomocí modelu PARAFAC.

```
WELCOME to the interactive CANDECOMP/PARAFAC analysis program
Warning: If you insert an object of mode CHARACTER when not requested,
an error occurs and the program stops!

Specify the number of A-mode entities
1:
Read 1 item
Specify the number of B-mode entities
1:
Read 1 item
Specify the number of C-mode entities
1:
Read 1 item
```

Budeme opět vyzváni k zadání rozměrů jednotlivých módů. Následně nám prostředí obdobně jako v případě Tucker3 nabídne zobrazit výsledky ANOVA a centrovat, popř. normovat vstupní data a provést PCA pro matici průměrů.

Následně se přesuneme k samotné analýze. Budeme vyzváni k zadání pouze jednoho počtu komponent, jelikož, jak jsme se dozvěděli dříve, PARAFAC neumožňuje použít rozdílný počet komponent pro jednotlivé módy. Další volbou je možnost zadat omezení. Ta jsou dobrá v případě, kdy je úloha degenerovaná, a v průběhu konečného počtu iterací bychom se vzdalovali od optimálního řešení.

```
You can now do a CANDECOMP/PARAFAC ANALYSIS

How many components do you want to use?
1:
Read 1 item

Do you want to use constraints? If so, enter '1':
1:
Read 0 items
No constraints imposed
```

Dále můžeme opět zadat práh konvergence a počet dodatečných průchodů algoritmem. Poslední možností je zadat ručně maximální přípustný počet iterací.

```
Specify convergence criterion (default=1e-6)
1:
Read 0 items

By default, only a rationally started analysis run will be carried out.
To decrease the chance of missing the optimal solution, you may use
additional, randomly started runs.
If you want additional runs, specify how many (e.g., 4):
1:
Read 0 items

Specify the maximum number of iterations you allow (default=10000).
1:
Read 0 items
```

Další text uvádí průběh algoritmu již pro konkrétní data. Popíšeme, jak přesného řešení jsme dosáhli, a samozřejmě též vypíšeme a zobrazíme vypočtené komponenty.

4.1. Analýza exportu

4.1.1. Pomocí metody Tucker3

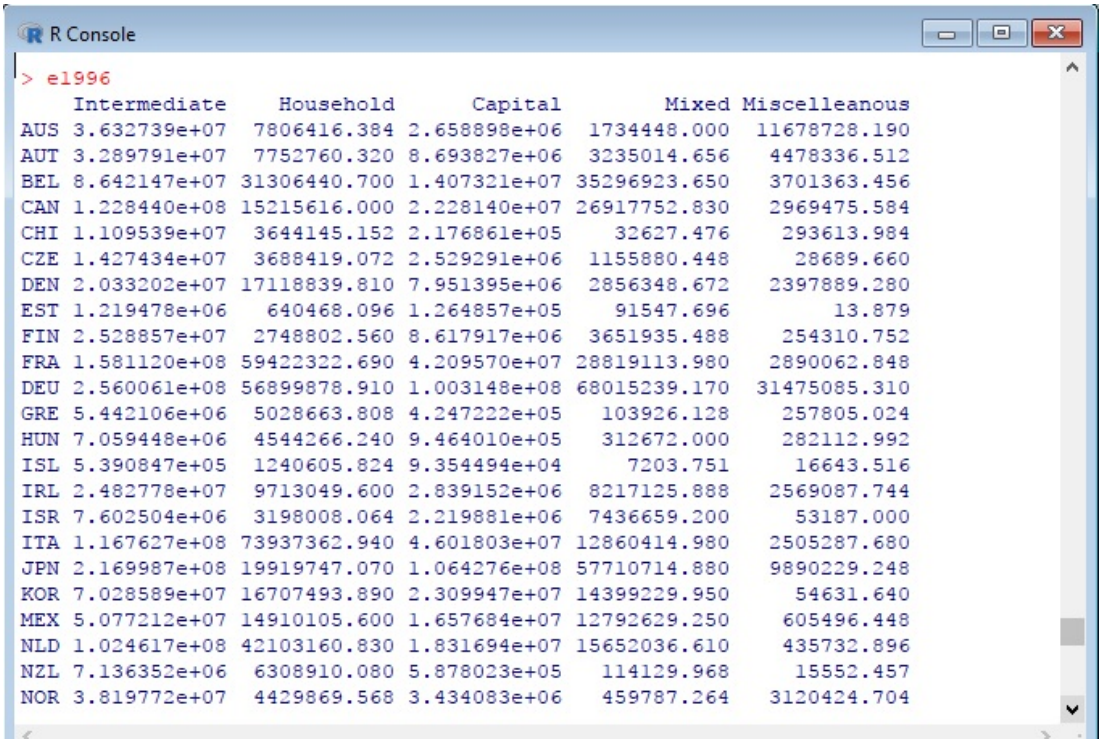
Nejprve se budeme věnovat analýze exportu metodou Tucker3. Pro analýzu exportu využijeme pozorování za roky 1996 - 2011, abychom tak z analýzy nevynechali pozorování pro příliš mnoho států nebo let.

Než přistoupíme k samotné analýze, je potřeba provést několik operací. V R musíme v první řadě načíst obě potřebné knihovny.

```
>library('ThreeWay')
>library('robCompositions')
```

Data za jednotlivé roky převedeme do formátu .txt a postupně načteme do R. Abychom respektovali pozici jednotlivých módů v datové matici, je třeba matici transponovat. Následně převedeme vstupní hodnoty na clr koeficienty. Samotná funkce vrátí kromě souřadnic i geometrický průměr, který je pro účel analýzy irelevantní, což vyřešíme přidáním argumentu `$x.clr` k funkci, abychom se tak dostali k požadované položce v seznamu. Tento postup demonstrujeme na prvním analyzovaném roce.

```
>e1996=read.table('export1996.txt', header=FALSE)
>e1996=t(e1996)
>e1996=cenLR(e1996)$x.clr
```



Obrázek 4.1: Náhled na matici původních dat v softwaru R

Na obrázcích 4.1 a 4.2 můžeme vidět části tabulek s původními daty resp. clr koeficienty v softwaru R. Řádky jsou pozorování pro jednotlivé státy, sloupce pak proměnné (kompoziční složky) odpovídající jednotlivým kategoriím.

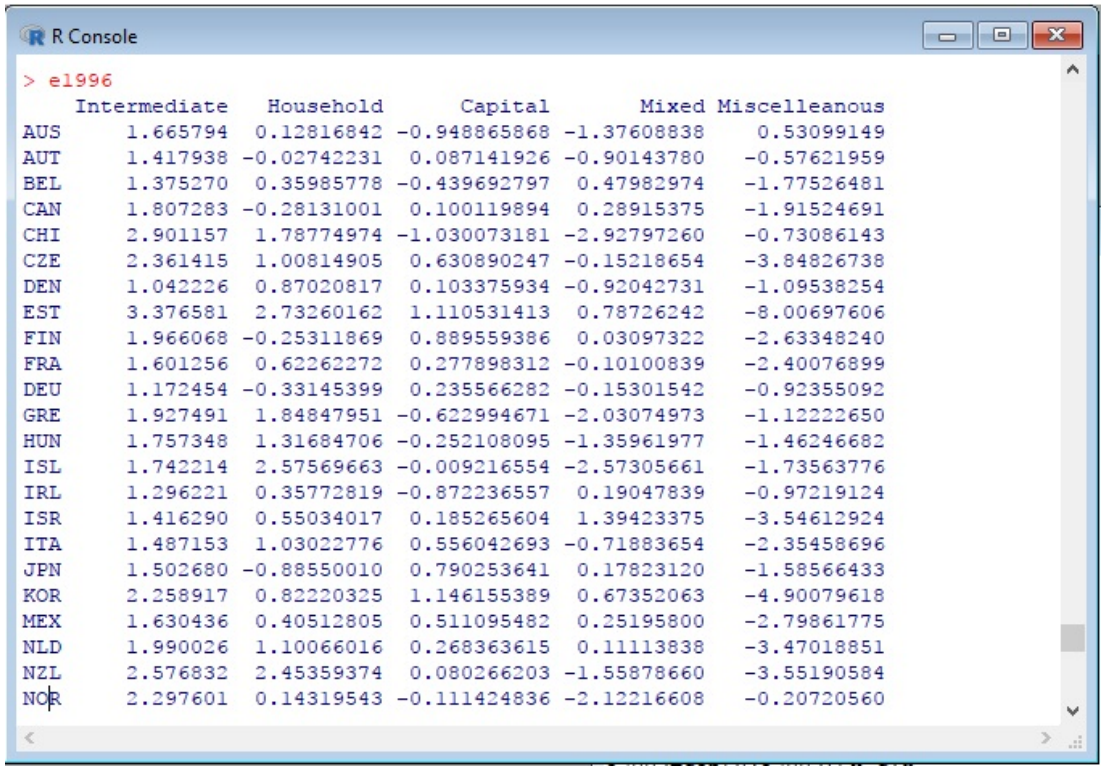
Jakmile výše popsaný postup aplikujeme na všechny analyzované roky, spojíme matice do jedné.

```
>e=cbind(e1996,e1997,e1998,e1999,e2000,e2001,e2002,e2003,e2004,e2005,
e2006,e2007,e2008,e2009,e2010,e2011)
```

Do programu zadáme vektory s popisky pro všechny tři módy, které pojmenujeme $laba, labb, labc$ a na novou matici aplikujeme funkci Tucker3. Popisky módu A jsou kódy států dle ISO 3166-1 dostupných např. na [6]. Popisky módu B jsou anglické názvy kategorií zboží, jak je uvádí OECD. Popisky módu C pak jednotlivé roky, ze kterých používáme data k analýze.

```
>et3=T3(e, laba, labb, labc)
```

V tomto momentě se spustí průvodce funkcí, který zpřístupní různé funkce a možnosti spojené s metodou Tucker3.



Obrázek 4.2: Náhled na matici clr koeficientů v softwaru R

```
WELCOME to the interactive TUCKER3 analysis program
Warning: If you insert an object of mode CHARACTER when not requested,
an error occurs and the program stops!

Specify the number of A-mode entities
1: 84
Read 1 item
Specify the number of B-mode entities
1: 5
Read 1 item
Specify the number of C-mode entities
1: 16
Read 1 item
```

Průvodce nás vyzve, abychom zadali počet entit pro jednotlivé módy. Mód A odpovídá státům, kterých máme 84. Mód B pak zastupuje jednotlivé kategorie, kterých je 5. V neposlední řadě pak mód C zastupuje počet všech sledovaných let, tedy 16. Soubor vycentrujeme přes mód A, ale nebudeme ho normovat. Po zamítnutí několika dalších nabídek nás bude zajímat až volba počtu komponent pro každý mód. Nejvíce informací bývá obvykle zachyceno v prvních dvou

komponentách, které navíc budeme moci snadno vykreslit do grafu, proto zvolíme tento počet.

```
You can now do a TUCKER3 ANALYSIS


How many A-mode components do you want to use?
1: 2
Read 1 item
How many B-mode components do you want to use?
1: 2
Read 1 item
How many C-mode components do you want to use?
1: 2
Read 1 item
```

Přeskočíme další nabídky a zastavíme se až u možnosti zadat počet dodatečných startů algoritmu, abychom potenciálně snížili možnost, že nezískáme optimální řešení. Zkusíme zadat čtyři dodatečné průchody algoritmem, abychom ukázali, jak se bude měnit procento, s jakým řešení vyhovuje

```
By default, only a rationally started analysis run will be carried out.
To decrease the chance of missing the optimal solution, you may use
additional, randomly started runs.
If you want additional runs, specify how many (e.g., 4):
1: 4

Run no. 5
Random ORTHONORMALIZED starts
Tucker3 function value at start is 19731.4668221652
Tucker3 function value after iteration 10 is 8303.26879759983
Tucker3 function value is 8303.26057824844 after 11 iterations
Fit percentage is 57.9273866910791 %
Procedure used 0.03 seconds


Start n.1 Start n.2 Start n.3 Start n.4 Start n.5
57.93 57.93 57.93 57.93 57.93
```

Řešení po pěti startech algoritmu vyhovuje stále na 57,93%, proto se spokojíme s tímto výsledkem (uvedené číslo představuje trojfaktorovou obdobu koeficientu determinace u regresních modelů). Řešení by se přiblížilo tomu optimálnímu více, zvolili-li bychom větší počet komponent. Jelikož by tím byla znemožněna interpretace výsledků z grafu, ponecháme 2 komponenty pro každý mód.

Všechny další nabízené možnosti přeskočíme a získáme výsledky analýzy. Po srovnání původního datového souboru se získanými výsledky, jsme došli k závěru, že je potřeba pro správnou interpretaci obě komponenty módu B vynásobit číslem *-1*. Následně si komponenty vykreslíme v grafu. Výstup funkce T3 nabízí velké množství různých hodnot (viz [1]). Nás zajímají nejprve skóry módu A, které chceme vykreslit do grafu, čehož docílíme přidáním argumentu `$Aplot` k výsledku analýzy. Tento argument zobrazí souřadnice komponent s jednotným měřítkem, což značně usnadní jejich interpretaci v interakci s dalšími módy. Jakmile vykreslíme graf, je třeba pojmenovat množinu bodů, aby byly blíže identifikovatelné. Použijeme výstup funkce T3 udávající popis složek módu A, který získáme přidáním argumentu `$laba`. Obdobně docílíme i znázornění módu B, který potřebujeme pro bližší porozumění hodnotám pro mód A, a nakonec zobrazíme i mód C. Poznamenejme ještě, že funkce `identify` použitá pro popis grafu s komponentami módu B, se oproti funkci `text` liší v tom, že pozice popisků zadáváme do grafu individuálně kliknutím.

```
>plot(et3$Aplot, main="Analýza exportu")
>text(et3$Aplot, labels=et3$laba, cex=0.9, pos=3)
```

```
>plot(et3$Bplot)
>textidentify(et3$Bplot, labels=et3$labb, cex=0.9)
```

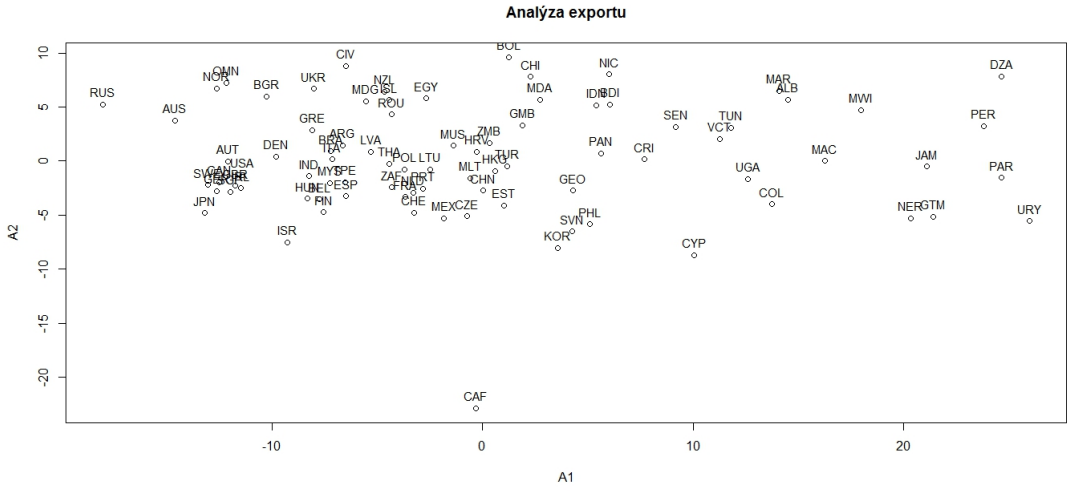
```
>plot(et3$Cplot)
>text(et3$Cplot, labels=et3$labc)
```

Na obrázku 4.3 můžeme vidět výsledky analýzy. V dolní části grafu se nachází pouze Středoafrická republika. Při pohledu na graf na obrázku 4.4 můžeme konstatovat, že tato země vyváží významný podíl zboží smíšené spotřeby.

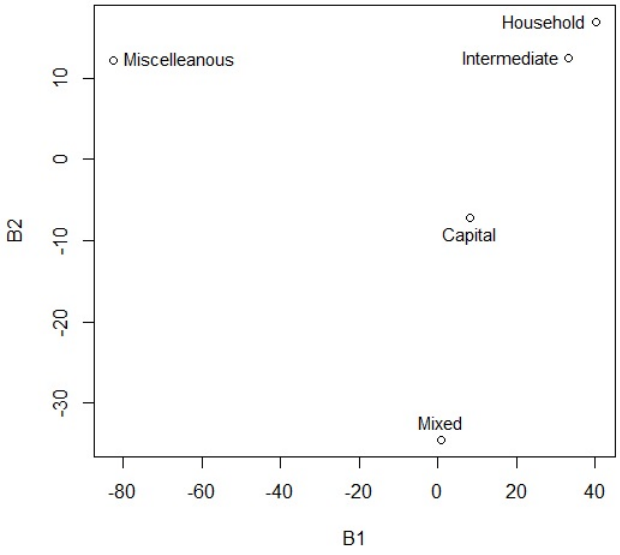
Na samotném pravém kraji obrázku 4.3 můžeme vidět Uruguay, Paraguay, Peru a Alžírsko, o kterých tak můžeme říct, že vyváží velký podíl meziproduktů a relativně významný podíl zboží domácí spotřeby. První tři zmíněné státy mají podobnou geografickou polohu, oproti tomu s Alžírskem na první pohled žádnou spojitost nemají. Podobnou strukturu exportu lze očekávat i u států nalevo od nich, tj. Jamajka, Guatemala a Niger.

S tím, jak budeme postupovat grafem dále doleva dále ke středu, mělo by u zemí dle grafu 4.4 zastávat větší podíl v exportu kapitálové zboží. Nicméně pro státy ležící nad osou $A2 = 0$ bude i nadále platit, že vyváží největší podíl meziproduktů a zboží domácí spotřeby.

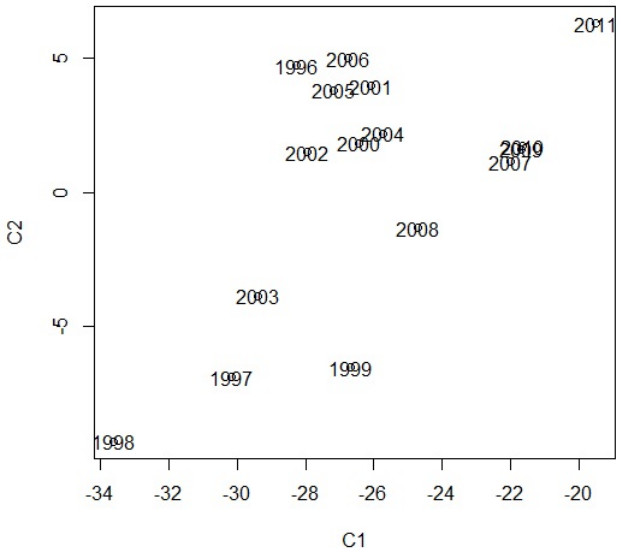
V grafu můžeme vidět jistou podobnost struktury exportu Ugandy, Kolum-



Obrázek 4.3: Výsledek analýzy exportu metodou Tucker3 v clr koeficientech, mód A



Obrázek 4.4: Výsledek analýzy exportu metodou Tucker3 v clr koeficientech, mód B



Obrázek 4.5: Výsledek analýzy exportu metodou Tucker3 v clr koeficientech, mód C

bie, Tuniska a Svatého Vincence a Grenadiny, čili země bez zjevné spojitosti. Dále vlevo můžeme vidět Kostariku a Panamu, nacházející se shodně ve střední Americe, doprovázené Senegalem Gruzií. Větší podobnost struktury vývozu pak můžeme očekávat u Indonésie, Burundi, Nikaragui nebo u Korey, Slovinska a Filipín, u kterých můžeme právě očekávat vyšší podíl kapitálové zboží, což bychom u Korey nejspíše tipovali i bez výsledků analýzy.

Podobnou strukturu s vysokým zastoupení meziproduktů, zbožím domácí spotřeby a relativně vyšším zastoupení kapitálového zboží můžeme očekávat u států vlevo v horní části grafu – Chile, Bolívie a Moldávii.

Dostáváme se do levé části obrázku, ve které můžeme pozorovat větší shluky zemí tvořené zeměmi s velmi podobnou strukturou exportu. Tu můžeme očekávat u Číny a Hong Kongu, což nás nejspíše nepřekvapí, doprovázené Chorvatskem, Estonskem, Tureckem a např. Maltou. Hned vedle se nachází Česká republika společně s Mexikem. U těchto dvou zemí bychom podle polohy v grafu mohli očekávat relativně vyrovnanou strukturu exportu samozřejmě s nízkým podílem ostatního zboží, které se ve svém grafu nachází dál, než kterýkoliv analyzovaný stát.

Podíváme-li se nalevo od Mexika, uvidíme shluk tvořený Polskem, Litvou a Lotyšskem, tedy státy nacházející se u Baltského moře, doprovázené vyspělými evropskými ekonomikami jako je Portugalsko, Francie, Nizozemí nebo Švýcarsko a dále Jihoafrickou republikou, tedy ekonomicky jednou z nejvyspělejších af-

rických zemí, a Thajskem.

Nad tímto shlukem můžeme pozorovat poměrně rozmanitou skupinu států - Nový Zéland, Island, Madagaskar, Rumunsko a Egypt, u kterých lze očekávat výraznější podíl zboží domácí spotřeby než u předchozí zmíněné skupiny.

Dále vlevo můžeme vidět další velký shluk, který je tvořen Brazílií, Argentinou, Itálií, Indií, Španělskem, Maďarskem, Finskem, Belgií nebo Tchaj-wanem. V jejich okolí se nachází ještě Řecko, Pobřeží slonoviny, Ukrajina a Izrael.

V horní části dále můžeme pozorovat státy, u kterých bychom podle pozice v grafu mohli očekávat, že budou vyvážet větší podíl ostatního zboží, než všechny ostatní státy. Jedná se o Bulharsko, Rusko, Austrálii, Norsko a Omán.

Pod nimi se nachází velký shluk států, u které lze očekávat vyrovnanou strukturu exportu s malým podílem ostatního zboží, nicméně vyšším než u zemí nacházejících se napravo od nich. Jedná se o Japonsko, Švédsko, Rakousko, USA, Kanada, Německo, Velkou Británii a Irsko, tedy o ekonomicky velmi vyspělé země.

Ještě se podívejme na graf 4.5. Najdeme-li si rok 1996, zjistíme, že se, co se týče exportu, liší od let 97-99 a je si více podobný s lety 2001 - 2006. Na obrázku je patrný bod zvratu na přelomu tisíciletí.

4.1.2. Pomocí metody PARAFAC

Analýzu exportu provedeme ještě jednou s využitím metody PARAFAC. Výsledky obou metod následně porovnáme.

Postup analýzy v R je zpočátku obdobný. Nejprve si nachystáme vstupní data do podoby, ve které můžeme použít funkci `CP`, což obnáší načtení souborů s daty, transpozice matice, převedení hodnot na clr koeficienty a spojení datových matic pro jednotlivé roky do jedné. Následně spustíme průvodce metodou `PARAFAC`:

```
>ecp=CP(e, laba, labb, labc)


   WELCOME to the interactive CANDECOMP/PARAFAC analysis program
Warning: If you insert an object of mode CHARACTER when not requested,
an error occurs and the program stops!

Specify the number of A-mode entities
1: 84
Read 1 item
Specify the number of B-mode entities
1: 5
Read 1 item
Specify the number of C-mode entities
1: 16
Read 1 item
```

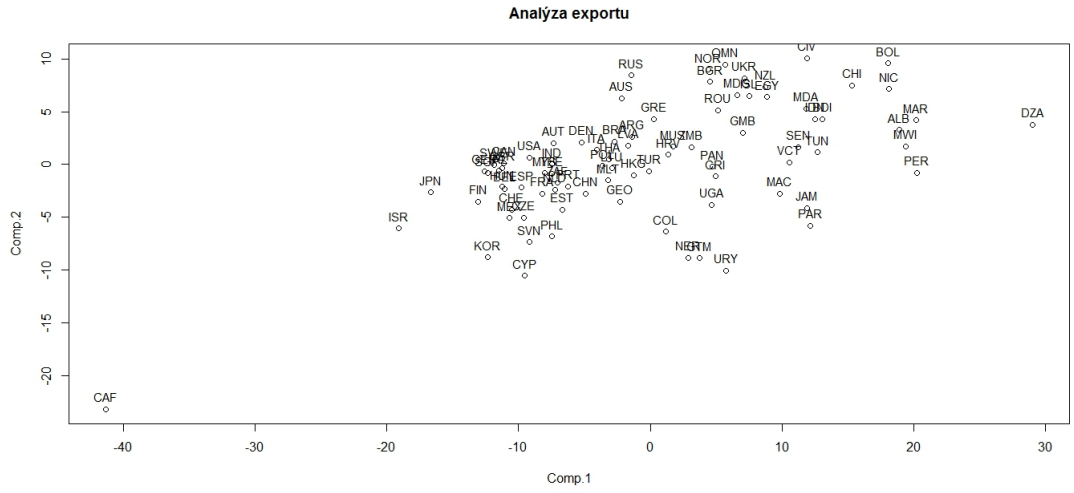
Stejně jako u Tucker3 zadáme počty entit pro jednotlivé módy a výsledek budeme opět centrovat přes mód A. Přeskočíme další nabízené možnosti a zadáme až počet komponent, které použít. Na rozdíl od `T3`, který umožňuje použít rozdílný počet komponent pro každý mód, musíme zadat pouze jedno číslo.

```
How many components do you want to use?
1: 2
Read 1 item
```

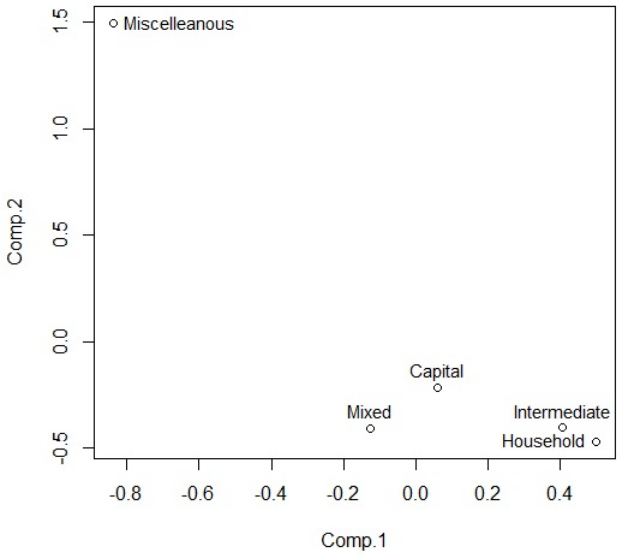
Opět přeskočíme další nabídky a získáme tak řešení statistické analýzy exportu pomocí metody PARAFAC.

```
Run no. 1
Candecomp/Parafac function value at Start is 19549.4744700224
f= 8305.55727801487 after 50 iters; diff.= 0.113778807532071
f= 8303.81410052823 after 100 iters; diff.= 0.0118452382048417
Candecomp/Parafac function value is 8303.67790401752 after 114 iterations
Fit percentage is 57.9252721018114
Procedure used 0.41 seconds
```

Vektor prvních komponent módu B opět vynásobíme -1, aby interpretace grafu odpovídala skutečnosti. Všechny tři grafy si poté vykreslíme do grafu obdobně jako v případě Tucker3.



Obrázek 4.6: Výsledek analýzy exportu metodou PARAFAC v clr koeficientech, mód A



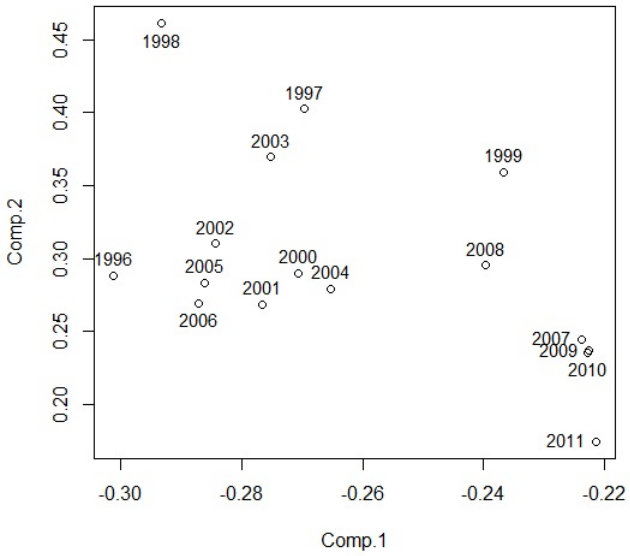
Obrázek 4.7: Výsledek analýzy exportu metodou PARAFAC v clr koeficientech, mód B

Stejně jako při použití Tucker3 můžeme vidět naprostou odlišnost Středoafrické republiky, nacházející se v levém dolním rohu, od vzorku ostatních zemí. To odpovídá dominanci složky zboží smíšené spotřeby a významného podílu meziproduktů. Blíže ke středu se budou nacházet státy s větším podílem zboží smíšené spotřeby a kapitálového zboží. Konkrétně můžeme pozorovat Japonsko, Izrael, Kypr, Korea a Finsko.

Vpravo vidíme obdobně jako v předchozí kapitole shluk států tvořený Švédskem, USA, Belgií, Maďarskem, Německem nebo Kanadou. Ještě o něco dále vidíme další státy s podobnou strukturou jako třeba Španělsko, Francii, Estonsko, Nizozemí, Portugalsko, Jihoafrickou republiku, Indii nebo Čínu. Pod nimi se nachází Česká republika opět společně s Mexikem, tentokrát doprovázené Švýcarskem a ještě o něco níže Slovinsko a Filipíny.

Dále vpravo by se měly nacházet země s relativně vysokým zastoupením kapitálového zboží v jejich exportu. Jedná se o velkou skupinu států, ve které můžeme najít např. Malta, Litva, Thajsko, Polsko, Itálie, Brazílie, Argentina, Lotyšsko, Hong Kong a Turecko. Nad nimi se nachází Rusko a Austrálie, což potvrzuje domněnku, že v jejich struktuře má vyšší podíl ostatní zboží, než u ostatních států.

Dole se oproti tomu nachází Uruguay, Guatemala a Niger, které by měly mít vysoké zastoupení meziproduktů a zboží domácí spotřeby. Budeme-li stoupat grafem na stejné úrovni, bude o států opět nepatrně růst význam kategorie zboží ostatní spotřeby. Namátkou můžeme vybrat Ugandu, Panamu a Kostariku a dále



Obrázek 4.8: Výsledek analýzy exportu metodou PARAFAC v clr koeficientech, mód C

Rumunsko, Bulharsko, Norsko, Omán, Ukrajina, Island, Egypt a Nový Zéland.

Dále vpravo se budou nacházet státy s dominantní rolí meziproduktů. Jako příklad uveďme Chile, Nikaraguu, Bolívii, Maroko, Albánii a nejvíce vpravo Alžírsko.

Na obrázku 4.8 můžeme vidět velké skoky mezi lety 1996, 1997, 1998 a 1999. V letech 2000-2006 se struktura zboží ustálila, ale ve zbylých letech opět můžeme rozdíl. V průběhu let tak nelze vidět jednoznačný trend ve vývozu zboží.

Řešení získaná pomocí obou metod jsou v tomto případě celkem odlišná. Z řešení získaného metodou PARAFAC plyne podoba struktury exportu většiny zemí. Komponenty módu B získané oběma metodami naznačují silnější vztah mezi dominancemi meziproduktů a zbožím domácí spotřeby. Po pohledu na původní datový soubor se výsledky analýzy metodou PARAFAC mohou zdát více podobné skutečnosti. Z toho, co víme o obou metodách z kapitoly 3.2, pak můžeme očekávat, že tyto výsledky budou skutečně přesnější, než v případě metody Tucker3. Z polohy komponent získaných metodou PARAFAC jasně plyne, že ostatní zboží nedominuje exportu žádné země. Kromě toho i poloha odlehlých pozorování v grafu více odpovídá struktuře zboží odpovídajících států.

4.2. Analýza importu

4.2.1. Pomocí metody Tucker3

Oproti analýze exportu využijeme z datové množiny s hodnotami importu jen roky 1997-2011, abychom z analýzy vynechali co nejméně států. Do analýzy zahrneme, jelikož to data umožňují, i některé nové státy zejména z Afriky, Jižní a Střední Ameriky, a kromě nich pak i našeho východního souseda Slovensko. Jako v předešlém případě nahradíme chybějící data průměrnými hodnotami.

V analýze budeme postupovat obdobně jako v případě exportu. Proces načtení a přípravy dat opět ukážu na jednom příkladu.

```
library('robCompositions')
library('ThreeWay')
i1997=read.table("import1997.txt",header=FALSE)
i1997=t(i1997)
i1997=cenLR(i1997)$x.clr
```

Následně přichystaná data za jednotlivé roky opět spojíme do jedné matice:

```
i=cbind(i1997,i1998,i1999,i2000,i2001,i2002,i2003,i2004,i2005,i2006,i2007,
i2008,i2009,i2010,i2011)
```

A po zadání vektorů se jmény entit jednotlivých módů, které pojmenujeme opět *laba*, *labb*, *labc*, můžeme přistoupit k použití funkce `T3`:

```
it3=T3(i, laba, labb, labc)
```

Následně budeme pokračovat v průvodci funkcí.

```
WELCOME to the interactive TUCKER3 analysis program
Warning: If you insert an object of mode CHARACTER when not requested,
an error occurs and the program stops!

Specify the number of A-mode entities
1: 98
Read 1 item
Specify the number of B-mode entities
1: 5
Read 1 item
Specify the number of C-mode entities
```

```
1: 15
Read 1 item
```

Počet entit je rozdílný. Mód A má nyní 98 entit, což je počet států zahrnutých do analýzy, mód B má opět 5 entit, jelikož dělení do kategorií zůstává stejné, a mód C má nyní 15 entit, což je počet let využitých k analýze.

Výsledek budeme opět centrovat přes mód A a dále se zastavíme až u výběru počtu komponent. Volit opět budeme po dvou komponentách v každém módu.

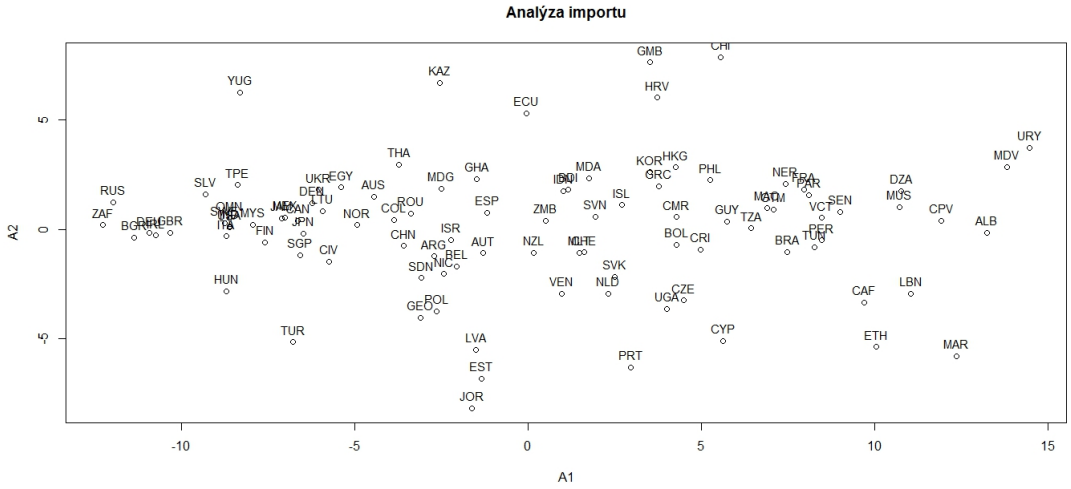
```
You can now do a TUCKER3 ANALYSIS

How many A-mode components do you want to use?
1: 2
Read 1 item
How many B-mode components do you want to use?
1: 2
Read 1 item
How many C-mode components do you want to use?
1: 2
Read 1 item
```

Všechny další nabídky přeskočíme, načež získáme hledané výsledky. Oba vektory komponent módu B vynásobíme číslem (-1), tak jako předtím je vykreslíme do grafu a body pojmenujeme.

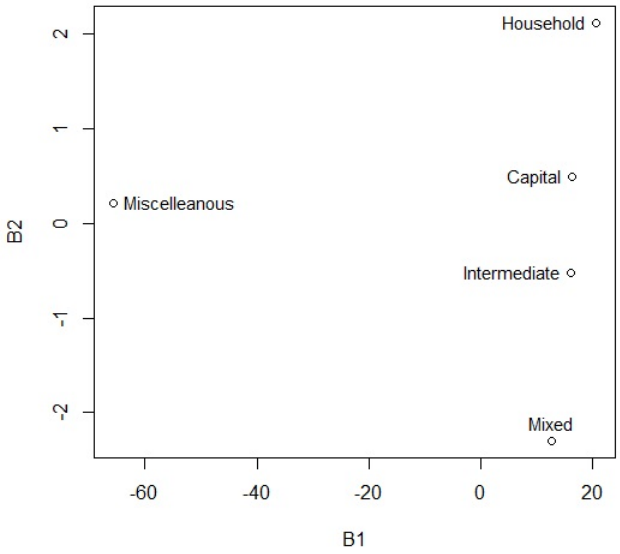
```
>plot(it3$Aplot, main="Analýza importu")
>text(it3$Aplot, labels=it3$laba, cex=0.9, pos=3)

>plot(it3$Bplot)
>textidentify(it3$Bplot, labels=it3$labb, cex=0.9)

>plot(it3$Cplot)
>text(it3$Cplot, labels=it3$labc, cex=0.9, pos=3)
```

Na obrázku 4.9, na kterém je graf s výsledky analýzy importu pomoci Tucker3, můžeme vidět, že státy jsou v grafu rozmístěny relativně symetricky kolem osy $A2 = 0$.

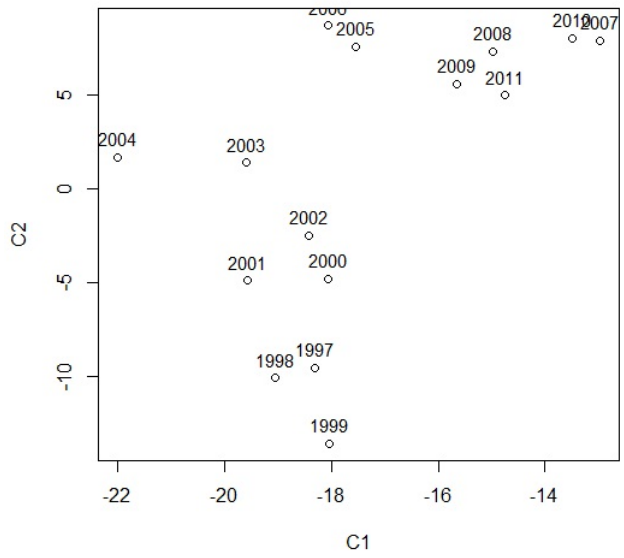
Začněme s popisem odlehlým pozorování. Vlevo nahoře můžeme vidět Jugoslávii, o které tak můžeme říct, že dováží relativně vysoký podíl ostatního



Obrázek 4.9: Výsledek analýzy importu metodou Tucker3 v clr koeficientech, mód A



Obrázek 4.10: Výsledek analýzy importu metodou Tucker3 v clr koeficientech, mód B



Obrázek 4.11: Výsledek analýzy importu metodou Tucker3 v clr koeficientech, mód C

zboží, kapitálového zboží a zboží domácí spotřeby. Ještě vyšší podíl zboží domácí spotřeby dle grafu dováží Kazachstán a Ekvádor. Největší podíl těchto produktů pak z analyzovaných států dováží Chile, Chorvatsko a Gambie.

Dále se budeme zabývat státy nacházejících se v grafu níže. Dle 4.10 můžeme u těchto zemí očekávat relativně rovnoměrné zastoupení všech kategorií kromě ostatního zboží. Dále můžeme říct, že státy nahoře od osy A2=0 budou dovážet více kapitálového zboží a zboží domácí spotřeby, než státy nacházející se dole od této osy, jejichž importu pravděpodobně dominují meziprodukty a zboží smíšené spotřeby. Vlevo se nachází Rusko a Jihoafrická republika, které z celého vzorku dováží největší podíl ostatního zboží. O něco menší podíl tohoto zboží dováží Velká Británie, Irsko, Bulharsko a Německo.

Vpravo vidíme první velký shluk států, ve kterém lze rozeznat Salvador, Tchaj-wan, Omán, Malajsii, Švédsko, Itálii a Spojené státy americké. Vedle nich pak můžeme pozorovat Finsko, Mexiko, Kanadu, Japonsko, Dánsko, Litvu a Ukrajinu. Podíváme-li se dále, najdeme Singapur, Pobřeží slonoviny, Austrálii, Egypt a Norsko, které tedy doplňuje předešlé dvě skandinávské země, se kterými tak tvoří prozatím jedinou pozorovanou skupinou geograficky a kulturně příbuzných zemí s podobnou strukturou dovozu.

S tím, jak v popisu grafu budeme pokračovat dále vpravo, bude více klesat podíl dováženého ostatního zboží. Můžeme tedy vidět Kolumbii, Rumunsko, Čína, dále pak Argentina, Izrael, Belgie, Nikaragua, Rakousko a Španělsko.

Další menší shluk tvoří opět země bez zjevné spojitosti - Indonésie, Burundi, Moldávie, Island, Slovinsko, Švýcarsko, ale i Slovensko a Nizozemí.

Z dalších zemí tentokrát už geograficky si blízkých můžeme pozorovat podobu v struktuře importu Korey s Hong Kongem a Bolívií s Kostarikou. Pozorovat můžeme ale i podobnost struktury vývozu Ugandy a České republiky.

V další části grafu se již nachází státy, které dováží zanedbatelné množství ostatního zboží a dominantní složkou struktury jejich importu jsou meziprodukty. Jedná se o Brazílii, Tunisko, Niger, Francii, Paraguay, Senegal, Alžírsko, Kapverdy a Albánie. O něco výše se nachází Uruguay a Maledivy, které dováží vyšší podíl kapitálového zboží.

V dolní části grafu můžeme pozorovat státy, jejichž import závisí na meziproduktech a zboží smíšené spotřeby. Vlevo jsou pak státy které dováží větší podíl ostatního zboží, než státy, které leží napravo. Postupně tak můžeme zleva vidět Maďarsko a Turecko. Dále můžeme pozorovat podobnost struktury importu pobaltských států Lotyšska a Estonska. Dále vpravo se pak nachází Portugalsko a Kypr. Podobnou strukturu importu mají dle grafu i Libanon, Středoafrická republika, Etiopie a Maroko, které dováží největší podíl meziproduktů a zboží smíšené spotřeby.

Pohledem na obrázek 4.11 zjistíme, že pozorování v letech vykazují jistý trend. To dle komponent módu B naznačuje, že v čase rostl podíl dovozu meziproduktů a klesal podíl importu zboží smíšené spotřeby, i když tyto změny nebyly nikterak zásadní.

4.2.2. Pomocí metody PARAFAC

Jako poslední část analýzy ještě zpracujeme soubor dat pomocí metody PARAFAC. Nejprve si opět data připravíme v R tak, jak jsme uvedli v předchozí části. Na matici aplikujeme funkci CP.

```
>icp=CP(i, laba, labb, labc)


   WELCOME to the interactive CANDECOMP/PARAFAC analysis program
Warning: If you insert an object of mode CHARACTER when not requested,
an error occurs and the program stops!

Specify the number of A-mode entities
1: 98
Read 1 item
Specify the number of B-mode entities
1: 5
Read 1 item
Specify the number of C-mode entities
```

```
1: 15
Read 1 item
```

Do průvodce metodou PARAFAC zadáme počet jednotlivých komponent, který se shoduje s těmi v analýze pomocí Tucker3. Opět využijeme centrování přes mód A, normování používat nebudeme.

```
You can now do a CANDECOMP/PARAFAC ANALYSIS


How many components do you want to use?
1: 2
Read 1 item
```

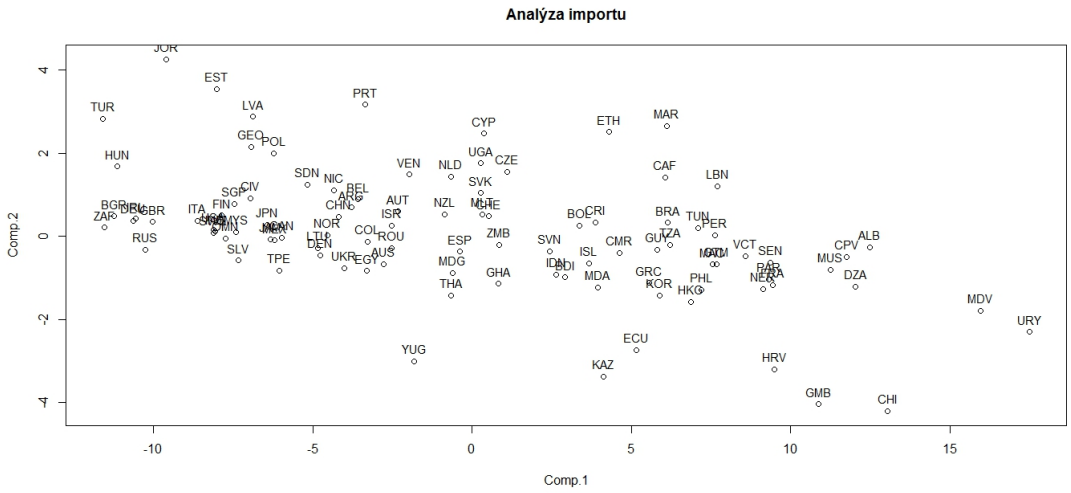
Pro účely analýzy nám opět bude stačit vypočítat dvě komponenty, které nesou největší část informace. Žádnou další nabízenou možnost nevyužijeme a získáme výsledky. Oba vektory módu B vynásobíme -1 a všechny tři módy si vykreslíme do grafu.

```
>plot(icp$A, main="Analýza importu")
>text(icp$A, labels=it3$laba, cex=0.9, pos=3)


>plot(icp$B)
>text(icp$B, labels=it3$labb, cex=0.9, pos=3)


>plot(icp$C)
>text(icp$C, labels=it3$labc, cex=0.9, pos=3)
```

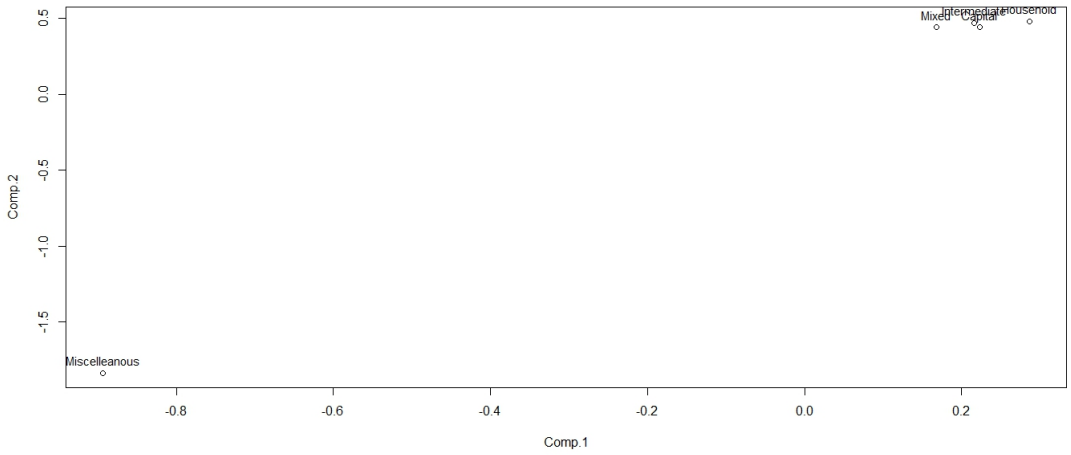
Rozmístění složek komponent módu A je podobné tomu v případě metody Tucker3. Podíváme-li se ale na obrázek 4.13 vidíme, že dle metody PARAFAC existuje silný vztah mezi dominancemi všech složek importu kromě ostatního zboží. To značně ztíží interpretaci výsledků porovnáním komponent jednotlivých módů.

Pohledem na 4.13 zjistíme, že v levém spodním rohu by se měly nacházet státy s dominancí ostatního zboží. Jak vidíme v 4.12 žádný takový stát se v analyzovaném souboru nenachází. Výraznější podíl ostatního zboží můžeme pravděpodobně pozorovat u Ruska, Jihoafrické republiky, Velké Británie, Bulharsko a Německo, které se nachází v levé části grafu.

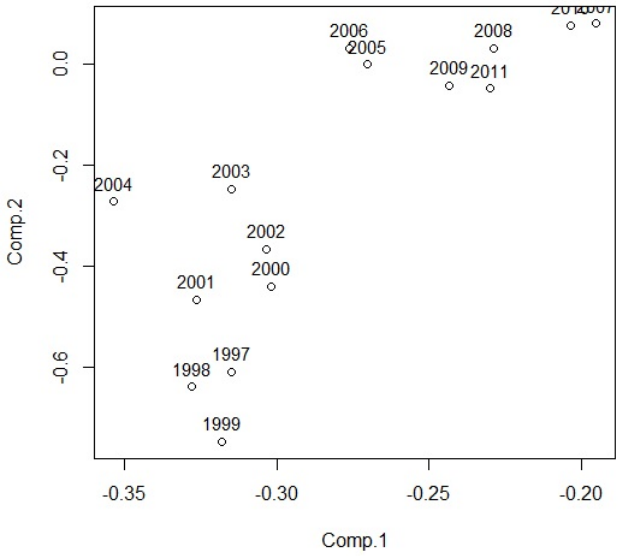
Podobnou strukturu lze očekávat i u států ve shluku po pravé straně. Jedná se o Itálii, Finsko, Malajsii a třeba USA. V okolí se dále nachází Salvador, Pobřeží



Obrázek 4.12: Výsledek analýzy importu za využití clr souřadnic metodou PARAFAC, mód A



Obrázek 4.13: Výsledek analýzy importu za využití clr souřadnic metodou PARAFAC, mód B



Obrázek 4.14: Výsledek analýzy importu za využití clr souřadnic metodou PARAFAC, mód C

slonoviny, Singapur, Japonsko, Mexiko nebo Tchaj-wan.

Dále dle grafu 4.12 můžeme očekávat podobnou strukturu importu Argentiny, Belgie, Číny, Nikaraguy, Norska, Litvy a Dánska. Dále vpravo můžeme vidět, že kromě Ugandy má Česká republika podobnou strukturu importu i se Slovenskem, což bychom asi očekávali.

V pravé části grafu se budou nacházet státy se zanedbatelným podílem dovezeného ostatního zboží. Převažovat u nich budou zbylé čtyři složky, nicméně vzhledem k jejich pozici v grafu s komponentami módu B, nelze přesně určit, která složka bude u jednotlivých států dominovat. Pozorovat i tak můžeme podobu struktury importu Filipín, Korey a Hong Kongu, tedy východoasijských zemí. Dále třeba Paraguay, Francie, Niger a Senegal, které jsou naopak zcela odlišné. U pravého okraje grafu se nachází Mauricius, Kapverdy, Albánie, Alžírsko a ještě dále Uruguay a Maledivy.

Opět se ještě krátce podíváme na graf s komponentami módu C, ve kterém můžeme vidět podobný trend jako v případě metody Tucker3.

V případě analýzy importu považuji za relevantnější výsledky získané metodou Tucker3. Jak jsem již zmínil složky komponent módu B jsou umístěny tak blízko u sebe, že je těžké určit dominující složky pro složky módu A. Metoda PARAFAC tak v tomto případě postačí pro určení, které země mají podobnou strukturu vývozu. Kromě toho výsledky metody PARAFAC v případě exportu i importu naznačují silný vztah mezi dominancemi kategorie meziproduktů, zbožím domácí

spotřeby, kapitálového zboží a zboží smíšené spotřeby, zatímco kategorie ostatní zboží se v tomto směru vyjímá. Do této kategorie je zahrnuto opravdu jen malé množství zboží, které nezapadá do žádné jiné kategorie, a tak je jeho podíl v obchodních tocích většiny států minimální.

Srovnáme-li ještě výsledky pro export a import, můžeme říci, že u vývozu se více odrážela geografická a kulturní podoba jednotlivých zemí v jeho struktuře. Důvodem může být např. to, že zdroje jednotlivých zemí nacházející se ve stejné části světa se projeví v podobné produkci zboží, které je pak mimo jiné vyváženo, zatímco potřeba zboží ze zahraničí je již čistě individuální záležitost, která závisí i na ostatních faktorech.

Závěr

Cílem práce bylo analyzovat datové soubory týkající se struktury exportu a importu států z databáze OECD. Odtud plyne, že na tato data bylo možno pohlížet jako na kompozice. Nebyly pro nás tedy relevantní objemy exportu a importu jako takové, nýbrž podíly jednotlivých složek na celku. Použití kompozičních dat tedy navíc eliminovalo vliv inflace a rozdíl v cenových hladinách různých zemí. Čtenář byl seznámen s ekonomickým pozadím dovozu a vývozu. Historie mezinárodního obchodu mu pak přiblížila důležitost těchto veličin v politice a pro všechny občany. Pro analýzu dat bylo nezbytné seznámit se s log-podílovou metodikou kompozičních dat a metodami redukce dimenze dat. Princip těchto metod byl představen na nejpoužívanější z nich, analýze hlavních komponent, která byla dále rozšířena pro případ třífaktorových datových souborů. K jejich analýze pak byly použity dva modely, PARAFAC a Tucker3.

K získání výsledků byl využit software R, konkrétně funkce CP a T3 v knihovně ThreeWay. Všechny podstatné informace o zkoumaných státech byly obsaženy v komponentách zachycujících informaci o rozptylu dat, kterou jsme dále interpretovali. Pro lepší názornost byly výsledky vykresleny do grafů, ze kterých jsme se dozvěděli, které země jsou si strukturou exportu popř. importu podobné a jak je která země, co se týče mezinárodního obchodu, zaměřená.

Cíle práce bylo dosaženo, výsledky analýzy navíc potvrzují vhodnost použitých nástrojů a metod. Je ovšem nutné si přitom uvědomit, že invariance na změnu měřítka u kompozičních dat nezohledňuje skutečné objemy peněžních toků u exportu a importu, výsledky metod tedy mohly vést k poněkud jiným závěrům, než by člověk očekával při pohledu na původní datový soubor. V případě exportu jsme oběma modely docílili podobných výsledků, což nabízí možnost volby takového, jehož teoretické vlastnosti nám více vyhovují, bez dramatického vlivu na kvalitu zpracování vstupní informace. V případě importu byly získané výsledky poněkud rozdílné, zejména pak v případě komponent módu B. Je tedy třeba srovnání výsledků s původními daty, abychom vhodně zvolili model, jehož výsledky budou více odpovídat skutečnosti. Text může sloužit jako inspirace pro zpracování jiných ekonomických dat způsobem představeným v práci. Navíc skýtá prostor k další statistické analýze, například rozdílů v obchodních tocích.

Literatura

[1] Del Ferraro, M. A., Kiers, H. A. L., Giordani, P.: *Package 'ThreeWay'* [online]. [cit. 2017-02-28]. Dostupné z: https://cran.r-project.org/web/packages/ThreeWay/ThreeWay.pdf.

[2] Fojtíková, L.: *Zahraničně obchodní politika ČR. Historie a současnost (1945-2008).* C. H. Beck, Praha, 2009.

[3] Giordani, P., Kiers, H. A. L., Del Ferraro, M. A.: *Three-way component Analysis Using the R.* Journal of Statistical Software **7/57** (2014), s. 1–23.

[4] Hebák, P. a kol.: *Vícerozměrné statistické metody (3).* Informatorium, Praha, 2010.

[5] Hrůzová, K., Rypka, M., Hron, K.: *Compositional Analysis Of Trade Flows Structure.* Austrian Journal of Statistics **6/46** (2017), s. 49–63.

[6] Kódy států [online]. [cit. 2017-03-07]. Dostupné z: http://martik.cz/kody-statu/?i=1.

[7] Pawlowsky-Glahn, V., Eqozcue, J. J., Tolosana-Delgado, R.: *Modeling and Analysis of Compositional Data.* Wiley, Chichester, 2015.

[8] STAN Bilateral Trade Database by Industry and End-use category [online]. [cit. 2017-03-07]. Dostupné z: http://stats.oecd.org/index.aspx?queryid=32186.

[9] Using Tensor Factorization to Predict Network-Level Performance of Bridges [online]. [cit. 2017-03-07]. Dostupné z: http://ascelibrary.org/cms/attachment/415052/3351550/figure3.gif.