

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

DIPLOMOVÁ PRÁCE

Regresní modely s prediktory relativní i absolutní
povahy



Katedra matematické analýzy a aplikací matematiky
Vedoucí diplomové práce: **Mgr. Kamila Fačevicová, Ph.D.**
Vypracovala: **Bc. Hana Janečková**
Studijní program: B1103 Aplikovaná matematika
Studijní obor Aplikace matematiky v ekonomii
Forma studia: prezenční
Rok odevzdání: 2020

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Bc. Hana Janečková

Název práce: Regresní modely s prediktory relativní i absolutní povahy

Typ práce: Diplomová práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: Mgr. Kamila Fačevicová, Ph.D.

Rok obhajoby práce: 2020

Abstrakt: Práce se zabývá regresními modely, ve kterých jsou na straně vysvětlujících proměnných zahrnuty kompoziční proměnné společně s jejich součtem. Cílem práce je představit možné způsoby zahrnutí složky součtu do regresních modelů a tento přístup demonstrovat na vlastním příkladu. V teoretické části práce jsou definovány základní pojmy týkající se kompozičních dat a práce s nimi společně s převodem do souřadnicové reprezentace. Na to navazuje teorie regresních modelů, která postupně zahrnuje klasický přístup užívající k modelování závisle proměnné logaritmus složek kompozice, a poté tzv. T-prostory. Ukázány jsou různé možnosti zahrnutí součtu složek kompozice do modelu a dále samotné modely s jejich vlastnostmi a interpretací parametrů společně s jednotlivými podmodely. V následující kapitole je vysvětlena problematika ekologické stopy, která představuje v praktické části vysvětlující proměnnou, a také ukazatel HDP na 1 obyvatele v amerických dolarech, který je závisle proměnnou. Praktická část práce zahrnuje výsledky aplikace regresních modelů z teoretické části při zkoumání závislosti ukazatele HDP na struktuře ekologické stopy. Jsou porovnány výsledky jednotlivých modelů a vybrán model nejlepší pro danou problematiku.

Klíčová slova: kompoziční data, součet složek, bilance, regrese

Počet stran: 85

Počet příloh: 1 CD

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Bc. Hana Janečková

Title: Regression Models with Compositional Predictors and Total

Type of thesis: Master's

Department: Department of Mathematical Analysis and Applications of Mathematics

Supervisor: Mgr. Kamila Fačevicová, Ph.D.

The year of presentation: 2020

Abstract: The thesis deals with regression models that comprise compositional variables with their total as explanatory variables. The aim is to present methods of implementing the total into regression models and demonstrate them on designated examples. As part of the theoretical framework, basic terms regarding compositional data and their analysis are defined together with coordinate representation conversion. Subsequently, the author discusses the regression model theory including the classical approach that uses component logarithms for dependent variable modelling, and T-space. Various possibilities of including the component total into the model are presented as well as the model itself with its characteristics and the interpretation of parameters, together with the respective submodels. The following chapter focuses on the problem of ecological footprint that represents the explanatory variable and the GDP per capita that represents the dependent variable in the practical part of the thesis. This part incorporates results from regression models that examine the dependence of the GDP on the structure of ecological footprint. Results from the individual models are compared and the author then selects the most suitable one.

Key words: compositional data, total, balances, regression

Number of pages: 85

Number of appendices: 1 CD

Language: Czech

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracovala samostatně pod vedením paní Mgr. Kamily Fačevicové, Ph.D. a všechny použité zdroje jsem uvedla v seznamu literatury.

V Olomouci dne

.....

podpis

Obsah

| | |
|--|-----------|
| Úvod | 9 |
| 1 Kompoziční data | 11 |
| 1.1 Základní pojmy | 11 |
| 1.2 Základní principy analýzy kompozičních dat | 13 |
| 1.3 Aitchisonova geometrie | 15 |
| 2 Souřadnicová reprezentace | 19 |
| 2.1 Clr a alr souřadnice | 19 |
| 2.2 Ortonormální souřadnice | 22 |
| 2.2.1 Bilance | 23 |
| 3 Lineární modely s kompozičními daty a součtem | 29 |
| 3.1 Klasický přístup | 29 |
| 3.2 T-prostory | 31 |
| 3.2.1 Model se součtem τ | 33 |
| 3.2.2 Model se součtem η | 37 |
| 3.3 Podmodely | 38 |
| 3.3.1 Model s ilr souřadnicemi bez součtu | 39 |
| 3.3.2 Model pouze se součtem | 41 |
| 4 Data o ekologické stopě | 42 |
| 4.1 Proč právě ekologická stopa | 42 |
| 4.2 Základní informace o ekologické stopě | 44 |
| 4.3 Části kompozičního vektoru | 45 |
| 4.3.1 Cropland | 45 |
| 4.3.2 Grazing Land | 46 |
| 4.3.3 Forest Products | 46 |
| 4.3.4 Fishing Grounds | 46 |
| 4.3.5 Built-up Land | 47 |
| 4.3.6 Carbon | 47 |
| 4.4 Metoda výpočtu ekologické stopy | 48 |
| 4.5 Použitá závisle proměnná | 50 |

| | | |
|----------|--|-----------|
| 4.5.1 | Vztah k ekologické stopě | 51 |
| 4.6 | Úprava dat | 52 |
| 5 | Analýza dat | 54 |
| 5.1 | Nezávisle proměnné | 54 |
| 5.2 | Klasický přístup | 55 |
| 5.3 | Modely s ilr souřadnicemi | 56 |
| 5.3.1 | Grafické srovnání závisle a nezávisle proměnných | 57 |
| 5.3.2 | Model bez součtu | 65 |
| 5.3.3 | Model se součtem τ | 68 |
| 5.3.4 | Modely se součtem η | 70 |
| 5.3.5 | Využití pivotových souřadnic | 72 |
| 5.3.6 | Model pouze se součtem | 75 |
| 5.4 | Srovnání modelů | 77 |
| | Závěr | 79 |
| | Literatura | 82 |

Seznam obrázků

| | | |
|-----|--|----|
| 5.1 | Mapa s HDP na 1 obyvatele | 58 |
| 5.2 | Mapa s první ilr souřadnicí | 59 |
| 5.3 | Mapa s druhou ilr souřadnicí | 60 |
| 5.4 | Mapa se třetí ilr souřadnicí | 61 |
| 5.5 | Mapa se čtvrtou ilr souřadnicí | 62 |
| 5.6 | Mapa s pátou ilr souřadnicí | 63 |
| 5.7 | Mapa s prvním typem součtu | 64 |
| 5.8 | Mapa s druhým typem součtu | 65 |

Poděkování

Ráda bych poděkovala paní Mgr. Kamile Fačevicové, Ph.D. za vedení mojí diplomové práce a za ochotu a cenné připomínky, které mi při jejím zpracování poskytla.

Úvod

V analýze kompozičních dat se běžně předpokládá, že pro kompoziční vektor je důležitá pouze relativní informace obsažená ve složkách tohoto vektoru. Mohou však nastat případy, kdy nejen relativní, ale také absolutní velikost je relevantní pro analýzu daného problému. V této práci jsou zkoumána data o ekologické stopě na 1 obyvatele pro rok 2014. Naším cílem je využít regresní modely, kde na straně vysvětlujících proměnných je zastoupena nejen relativní informace vyjádřená pomocí bilancí, ale také absolutní informace, vyjádřená členem, který se v anglické literatuře označuje jako tzv. total (v této práci bude tento člen označován jako součet). Vysvětlovanou proměnnou poté bude ukazatel HDP na 1 obyvatele v amerických dolarech.

První kapitola práce shrnuje základní pojmy týkající se kompozičních dat, principy jejich analýzy společně s Aitchisonovou geometrií a operacemi na simplexu. Kapitola je doplněna příklady na smyšlených datech i na datech o ekologické stopě. Druhá kapitola popisuje jednotlivé typy souřadnicové reprezentace užívané v analýze kompozičních dat. Nejdříve jsou představeny alr a clr souřadnice společně s příklady jejich výpočtu pro kompoziční vektor ekologické stopy vybrané země. Druhá část kapitoly se věnuje ortonormálním souřadnicím. Blíže je pojednáno o bilancích a jejich speciálním typu, kterým jsou pivotové souřadnice. Konstrukce těchto souřadnic je také ilustrována na příkladech pro ekologickou stopu. Ve třetí kapitole práce jsou představeny regresní modely, které lze využít pro model s prediktory relativní i absolutní povahy. Jsou zavedeny dva typy součtu a jejich vlastnosti. U modelů jsou rozebrány jejich vzájemné vztahy, vlastnosti, interpretace parametrů a jednotlivé podmodely.

Čtvrtá kapitola představuje datový soubor využitý v regresních modelech. Popisuje základní informace o ekologické stopě a její jednotlivé složky. Dále jsou stručně ukázány způsoby výpočtu ekologické stopy a její vztah s vybranou závisle proměnnou, kterou je HDP na 1 obyvatele. Kapitulu uzavírá shrnutí o nutných úpravách dat před samotnou analýzou.

V páté kapitole, tvořící praktickou část práce, jsou data o ekologické stopě a ukazatel HDP na 1 obyvatele použity v regresních modelech představených v kapitole 3. Nejdříve jsou pomocí map vytvořených v softwaru R zkoumány hodnoty \ln souřadnic, HDP, součtů a jejich vzájemné vztahy. V dalších částech jsou poté postupně představeny a vyhodnoceny výsledky jednotlivých modelů aplikovaných v softwaru R. Na závěr kapitoly jsou shrnuty výsledky modelů a vyhodnoceny s cílem najít ten nejlepší.

Kapitola 1

Kompoziční data

V této kapitole bude popsána základní teorie týkající se kompozičních dat a principy jejich analýzy. V druhé části budou představeny operace s kompozičními daty, které umožňují zavést na simplexu vektorový prostor a posléze také euklidovský vektorový prostor pomocí tzv. Aitchisonovy geometrie.

1.1. Základní pojmy

Kompozičním vektorem s D složkami se nazývá řádkový vektor \mathbf{x} , pro který platí

$$\mathbf{x} = (x_1, x_2, \dots, x_D) \in \mathbb{R}_+^D, \quad x_j > 0 \quad \forall j = 1, \dots, D. \quad (1.1)$$

Kompoziční data se využívají v případech, kdy není důležitá informace o absolutní velikosti složek vektoru, ale zajímá nás pouze relativní informace obsažená v datech. Tato relativní informace může být vyjádřena v různých jednotkách, často jsou to procenta, nebo proporce. U dat nicméně nezáleží na tom, v jakých jednotkách jsou vyjádřena, protože při vynásobení kompozice kladnou konstantou se nemění poměry mezi složkami, a nemění se tedy ani relativní informace v ní obsažená. [7, 24]

Můžeme tedy říct, že dva vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^D$ jsou kompozičně ekvivalentní, když existuje pozitivní konstanta $\lambda \in \mathbb{R}_+$ taková, že

$$\mathbf{x} = \lambda \mathbf{y}. \quad (1.2)$$

Při práci s kompozičními daty se jako reprezentant dané kompozice běžně používá vektor s konstantním součtem. Nový vektor má poté součet složek $\kappa > 0$ a je získán pomocí operace uzávěr kompozičního vektoru \mathbf{x} definované jako

$$C(\mathbf{x}) = \left(\frac{\kappa x_1}{\sum_{i=1}^D x_i}, \frac{\kappa x_2}{\sum_{i=1}^D x_i}, \dots, \frac{\kappa x_D}{\sum_{i=1}^D x_i} \right) \quad (1.3)$$

$$\text{pro } \forall \mathbf{x} = (x_1, \dots, x_D) \in \mathbb{R}_+^D, x_i > 0 \forall i = 1, \dots, D.$$

Odtud lze odvodit jiný zápis kompoziční ekvivalence. Dva kompoziční vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^D$ jsou kompozičně ekvivalentní, pokud $C(\mathbf{x}) = C(\mathbf{y})$ pro každou konstantu κ .[\[24\]](#)

Příklad 1.1 *Mějme data o ekologické stopě jednotlivých zemí za rok 2014 na jednoho obyvatele. Ekologická stopa popisuje množství biologicky produktivní půdy a vody v globálních hektarech potřebných pro vyprodukování zdrojů spotřebovaných lidmi v daném roce a pro následnou obnovu zdrojů a vypořádání se se vzniklými odpady. Stopa se dále dělí na šest jednotlivých částí, a to cropland, grazing land, forest land, fishing grounds, built-up land a carbon. Jednotlivé části kompozice a další informace o datech budou blíže popsány v kapitole 4. Uvažujme nyní jeden konkrétní kompoziční vektor z datového souboru vyjadřující ekologickou stopu České republiky ve zkoumaném roce, zaokrouhlenou na 4 desetinná místa*

$$\mathbf{x} = (0, 8727; 0, 3028; 0, 7856; 0, 0348; 0, 1524; 3, 4516).$$

Chceme ukázat převedení našeho kompozičního vektoru na různé součty a dokázat, že skutečně nezáleží na jednotkách, v jakých jsou data vyjádřena, a že tedy platí kompoziční ekvivalence. Všechny výsledné kompoziční vektory budou opět zaokrouhleny na 4 desetinná místa. Nejdříve provedeme převod na součet $\kappa = 100$ (procenta)

$$\begin{aligned} C(\mathbf{x}) &= \left(\frac{100 \cdot 0, 8727}{5, 5999}, \frac{100 \cdot 0, 3028}{5, 5999}, \dots, \frac{100 \cdot 3, 4516}{5, 5999} \right) = \\ &= (15, 5842; 5, 4072; 14, 0288; 0, 6214; 2, 7215; 61, 6368). \end{aligned}$$

Při převodu na proporce ($\kappa = 1$) dostaneme

$$C(\mathbf{x}) = (0, 1558; 0, 05407; 0, 1403; 0, 0062; 0, 0272; 0, 6164).$$

Nyní máme různé reprezentace stejného kompozičního vektoru. Pokud bychom vzali u každého z vektorů např. podíl druhé a první složky, dostali bychom pokaždé hodnotu 0,3470. To by platilo také u podílů ostatních složek, kdy by tyto podíly byly u všech tří vektorů totožné. Je tedy vidět, že podíly složek se nezmění, i když dojde k vyjádření kompozičního vektoru v odlišných jednotkách. Proto lze říct, že tyto tři vektory jsou kompozičně ekvivalentní.

Dále můžeme definovat výběrový prostor kompozičních dat jako množinu S^D , pro kterou platí

$$S^D = \left\{ \mathbf{x} = (x_1, x_2, \dots, x_D) \mid x_i > 0, i = 1, 2, \dots, D; \sum_{i=1}^D x_i = \kappa \right\}. \quad (1.4)$$

Tato množina má tvar simplexu a nachází se v podprostoru \mathbb{R}_+^{D-1} . [24]

Často se však nepracuje přímo s původní kompozicí \mathbf{x} , ale s tzv. podkompozicí. Důvodem může být příliš mnoho složek v původní kompozici, nebo některé z nich nemusí být vůbec k dispozici. Máme-li kompozici \mathbf{x} a množinu indexů $S = \{i_1, \dots, i_s\}$ (ne nutně prvních s indexů), potom podkompozici \mathbf{x}_s o s složkách získáme jako $C(x_{i_1}, \dots, x_{i_s})$. [24]

Další operací, kterou můžeme provést, pokud nechceme pracovat s původní kompozicí, je použití amalgamace, neboli sečtení několika složek do jedné nové složky. Pokud máme kompoziční vektor $\mathbf{x} \in S^D$, množinu indexů $A = \{i_1, \dots, i_a\}$ (ne nutně prvních a indexů), $D - a \geq 1$ a množinu zbývajících indexů \bar{A} , potom amalgovaná část je hodnota

$$x_A = \sum_{j \in A} x_j. \quad (1.5)$$

Amalgovanou kompozicí se nazývá vektor $\mathbf{x} = (\mathbf{x}_{\bar{A}}, \mathbf{x}_A)$, kde v $\mathbf{x}_{\bar{A}}$ jsou obsaženy složky s indexy z množiny \bar{A} , zatímco \mathbf{x}_A je amalgovaná část. Amalgovaná kompozice leží v prostoru S^{D-a+1} . [24]

1.2. Základní principy analýzy kompozičních dat

Metody, které jsou využívány při analýze kompozičních dat, by měly splňovat tři základní principy - škálovou invarianci, permutační invarianci a podkom-

poziční koherenci.[24]

První z uvedených vlastností je škálová invariance. Pro danou funkci $f(\cdot)$, definovanou na \mathbb{R}_+^D , je tato vlastnost splněna, pokud pro libovolnou kompozici $\mathbf{x} \in S^D$ a pro libovolnou hodnotu $\lambda \in \mathbb{R}_+$ platí

$$f(\lambda\mathbf{x}) = f(\mathbf{x}), \quad (1.6)$$

tedy pokud pro všechny vektory, které jsou kompozičně ekvivalentní, dává funkce stejné výsledky. V praxi se pro zachování této vlastnosti používají logratio transformace složek kompozice \mathbf{x} . Pracujeme tedy s logaritmy podílů jednotlivých složek kompozice.[7, 24]

Máme-li kompozici $\mathbf{x} \in S^D$ (danou v procentech), poté funkce podílu jejích složek je škálově invariantní, tedy $f(\mathbf{x}) = x_1/x_2 = (\lambda x_1)/(\lambda x_2)$. Tyto podíly zachovávají pozitivitu, ale výsledek záleží na tom, v jakém pořadí podíl vytvoříme, tzn. $x_1/x_2 \neq x_2/x_1$; zatímco při použití logratio transformace $f(\mathbf{x}) = \ln(x_1/x_2)$ nezáleží na pořadí složek, při obráceném podílu se mění pouze znaménko. Funkce f je tedy symetrická vzhledem k řazení prvků. Často se používají spíše složitější logratio transformace. Logratio transformace však mají tu nevýhodu, že složky nesmí být nulové, protože bychom mohli dostat nulu ve jmenovateli. V tom případě je nutné použít některé ze speciálních metod, které tento problém řeší. Složky kompozice by tedy měly být vždy striktně kladné, jak je vyžadováno v definici kompozičního vektoru.[7, 24]

Druhou vlastností je permutační invariance. Její splnění znamená, že funkce dává stejné výsledky při různém uspořádání částí v kompozici. Tato vlastnost je rovněž vidět na příkladu s podíly z předchozího odstavce.[24]

Posledním principem je podkompoziční koherence. Ta spočívá ve dvou vlastnostech. První z nich je podkompoziční dominance, která platí, pokud vzdálenost mezi dvěma kompozicemi je větší než nebo stejná jako vzdálenost mezi dvěma jejich libovolnými podkompozicemi. Dále poměry mezi libovolnými složkami v podkompozici musí být stejné jako jejich poměry v původní kompozici.[7, 24]

Z podkompoziční dominance vyplývá potřeba nějakým způsobem měřit vzdálenost mezi složkami kompozice. V případě kompozičních dat však není možné

použít euklidovskou vzdálenost, poněvadž nesplňuje škálovou invarianci, ani podkompoziční dominanci. Z toho plyne potřeba speciální geometrie na D -prvkovém simplexu S^D , která se nazývá Aitchisonova geometrie.[22]

1.3. Aitchisonova geometrie

Pro práci s kompozičními daty je potřeba speciální geometrie, která umožní provádět matematické operace obdobně jako v euklidovském prostoru. Dvě základní operace, díky kterým lze na simplexu definovat vektorový prostor, jsou perturbace a mocninná transformace.[24]

Perturbace je analogií ke sčítání v reálném prostoru. Pro daný kompoziční vektor $\mathbf{x} \in S^D$ je jeho perturbace kompozičním vektorem $\mathbf{y} \in S^D$ definována jako

$$\mathbf{x} \oplus \mathbf{y} = C(x_1y_1, x_2y_2, \dots, x_Dy_D) \in S^D. \quad (1.7)$$

Pro perturbační rozdíl se pak užívá následující značení

$$\mathbf{x} \oplus \mathbf{y}^{-1} = \mathbf{x} \ominus \mathbf{y}. \quad (1.8)$$

Mocninná transformace kompozičního vektoru $\mathbf{x} \in S^D$ s konstantou $\alpha \in \mathbb{R}$ je definována jako

$$\alpha \odot \mathbf{x} = C(x_1^\alpha, x_2^\alpha, \dots, x_D^\alpha) \in S^D. \quad (1.9)$$

Jedná se o analogii k násobení vektoru skalárem v reálném prostoru. Tyto dvě operace na simplexu S^D tvoří vektorový prostor (S^D, \oplus, \odot) . [24]

(S^D, \oplus) dohromady tvoří komutativní grupu, je tedy splněna komutativita a asociativita. Neutrální prvek splňující $\mathbf{x} \oplus \mathbf{n} = \mathbf{x}$ se značí \mathbf{n} a platí pro něj

$$\mathbf{n} = \mathbf{x} \ominus \mathbf{x} = C(1, \dots, 1). \quad (1.10)$$

Prvek \mathbf{n} je tedy tvořen kompozicí s identickými částmi a je barycentrem simplexu. Tento prvek existuje pouze jeden. Inverze je dána jako

$$\mathbf{x}^{-1} = C(x_1^{-1}, x_2^{-1}, \dots, x_D^{-1}) \quad (1.11)$$

a splňuje $\mathbf{x} \oplus \mathbf{x}^{-1} = \mathbf{n}$. [7, 22, 24]

Mocninná transformace má stejné vlastnosti jako smíšený součin a splňuje tedy asociativitu a distributivitu. Neutrální prvek splňující $\mathbf{x} \odot 1 = \mathbf{x}$ existuje opět pouze jediný. [24]

Pro kompoziční data je možné zavést euklidovský vektorový prostor. K tomu je však nutné mimo perturbace a mocninné transformace definovat také skalární součin, normu a vzdálenost. [24]

Nejdříve definujme Aitchisonův skalární součin, který je pro dva vektory \mathbf{x} , $\mathbf{y} \in S^D$ zadán následovně

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j}. \quad (1.12)$$

Aitchisonova norma kompozičního vektoru $\mathbf{x} \in S^D$ je dána vztahem

$$\|\mathbf{x}\|_a = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} \right)^2}. \quad (1.13)$$

Aitchisonova vzdálenost mezi kompozičními vektory \mathbf{x} a $\mathbf{y} \in S^D$ je

$$d_a(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|_a = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2}. \quad (1.14)$$

Lze využít odlišné verze zápisu skalárního součinu, normy i vzdálenosti, které jsou s původně definovanými ekvivalentní. Například u skalárního součinu je možné použít tři verze zápisu

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{D} \sum_{i=1}^{D-1} \sum_{j=i+1}^D \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j} = \quad (1.15)$$

$$= \sum_{i=1}^D \ln x_i \ln y_i - \frac{1}{D} \left(\sum_{j=1}^D \ln x_j \right) \left(\sum_{k=1}^D \ln y_k \right) = \quad (1.16)$$

$$= \sum_{i=1}^D \ln \frac{x_i}{g_m(\mathbf{x})} \ln \frac{y_i}{g_m(\mathbf{y})}, \quad (1.17)$$

kde $g_m(\cdot)$ značí geometrický průměr.[24]

Zápis 1.17 je ekvivalentní součinu dvou reálných vektorů, vyjadřujících kompozice \mathbf{x} a \mathbf{y} v tzv. centrovaných logratio (clr) souřadnicích (viz kapitola 2).[24]

Operace perturbace a mocninná transformace společně s Aitchisonovým skalárním součinem, normou a vzdáleností tvoří euklidovskou vektorovou strukturu na simplexu, která se označuje jako Aitchisonova geometrie na simplexu.[22, 24]

Příklad 1.2 *Mějme dva kompoziční vektory $\mathbf{x} = (2, 4, 6, 8)$ a $\mathbf{z} = (9, 5, 2, 4)$ se součtem složek rovným 20. Pro tyto vektory si nyní ukážeme výpočet perturbace, mocninné transformace, Aitchisonovy normy, skalárního součinu a vzdálenosti, přičemž všechny výsledné kompoziční vektory budou opět převedeny na součet složek 20.*

Výpočet perturbace kompozičních vektorů \mathbf{x} a \mathbf{z} :

$$\begin{aligned}\mathbf{x} \oplus \mathbf{z} &= C(2 \cdot 9, 4 \cdot 5, 6 \cdot 2, 8 \cdot 4) = \left(\frac{20 \cdot 18}{82}, \frac{20 \cdot 20}{82}, \frac{20 \cdot 12}{82}, \frac{20 \cdot 32}{82} \right) = \\ &= (4, 3902; 4, 8780; 2, 9268; 7, 8049).\end{aligned}$$

Mocninná transformace vektoru \mathbf{x} s konstantou $\alpha = 5$:

$$\begin{aligned}\alpha \odot \mathbf{x} &= C(2^5, 4^5, 6^5, 8^5) = \left(\frac{20 \cdot 32}{41600}, \frac{20 \cdot 1024}{41600}, \frac{20 \cdot 7776}{41600}, \frac{20 \cdot 32768}{41600} \right) = \\ &= (0, 0154; 0, 4923; 3, 7385; 15, 7538).\end{aligned}$$

Následující tři operace budou vypočteny pomocí upravených vztahů, kdy pro skalární součin se jedná o vzorec 1.15, u normy a vzdálenosti by se vztahy odvodily analogicky.

Norma kompozičního vektoru \mathbf{x} :

$$\begin{aligned}\|\mathbf{x}\|_A &= \sqrt{\frac{1}{4} \cdot \left(\left(\ln \frac{x_1}{x_2} \right)^2 + \left(\ln \frac{x_1}{x_3} \right)^2 + \dots + \left(\ln \frac{x_3}{x_4} \right)^2 \right)} = \\ &= \sqrt{\frac{1}{4} \cdot \left(\left(\ln \frac{2}{4} \right)^2 + \left(\ln \frac{2}{6} \right)^2 + \dots + \left(\ln \frac{6}{8} \right)^2 \right)} = 1,0413.\end{aligned}$$

Skalární součin vektorů \mathbf{x} a \mathbf{z} :

$$\begin{aligned}\langle \mathbf{x}, \mathbf{z} \rangle_A &= \frac{1}{4} \cdot \left(\ln \frac{x_1}{x_2} \ln \frac{z_1}{z_2} + \ln \frac{x_1}{x_3} \ln \frac{z_1}{z_3} + \ln \frac{x_1}{x_4} \ln \frac{z_1}{z_4} + \dots + \ln \frac{x_3}{x_4} \ln \frac{z_3}{z_4} \right) = \\ &= \frac{1}{4} \cdot \left(\ln \frac{2}{4} \ln \frac{9}{5} + \ln \frac{2}{6} \ln \frac{9}{2} + \ln \frac{2}{8} \ln \frac{9}{4} + \dots + \ln \frac{6}{8} \ln \frac{2}{4} \right) = -0,8777.\end{aligned}$$

Vzdálenost kompozičních vektorů \mathbf{x} a \mathbf{z} :

$$\begin{aligned}d_a(\mathbf{x}, \mathbf{z}) &= \sqrt{\frac{1}{4} \left[\left(\ln \frac{x_1}{x_2} - \ln \frac{z_1}{z_2} \right)^2 + \left(\ln \frac{x_1}{x_3} - \ln \frac{z_1}{z_3} \right)^2 + \dots + \left(\ln \frac{x_3}{x_4} - \ln \frac{z_3}{z_4} \right)^2 \right]} = \\ &= \sqrt{\frac{1}{4} \left[\left(\ln \frac{2}{4} - \ln \frac{9}{5} \right)^2 + \left(\ln \frac{2}{6} - \ln \frac{9}{2} \right)^2 + \dots + \left(\ln \frac{6}{8} - \ln \frac{2}{4} \right)^2 \right]} = \\ &= 1,9996.\end{aligned}$$

(Výpočty je možné spočítat také v R , jednou z možností je využít funkce z balíčku `compositions` [3]. Nejdříve je nutné převést vektor pomocí funkce `acomp()` na objekt třídy `acomp`, na který lze poté použít funkce `perturbe()`, `power.acomp()`, `norm()`, `scalar()` a `dist()`.)

Kapitola 2

Souřadnicová reprezentace

V této kapitole budou postupně představeny tři typy souřadnicové reprezentace, užívané ve statistických metodách při práci s kompozičními daty, společně s jejich vlastnostmi, výhodami a nevýhodami. Transformace původního kompozičního vektoru do souřadnicové reprezentace je výhodná hned z několika důvodů. Kompozice jsou tímto převedeny do standardní euklidovské geometrie, se kterou pracují běžné statistické metody, a při výběru vhodných souřadnic lze získat výhodnou interpretaci výsledků.[\[2\]](#)

2.1. Clr a alr souřadnice

Prvním typem transformace je centrovaná logratio transformace (clr) daná jako

$$\text{clr}(\mathbf{x}) = \left(\ln \frac{x_1}{g_m(\mathbf{x})}, \ln \frac{x_2}{g_m(\mathbf{x})}, \dots, \ln \frac{x_D}{g_m(\mathbf{x})} \right) = \xi. \quad (2.1)$$

Pomocí inverzní transformace lze poté získat koeficienty v kanonické bázi reálného prostoru

$$\text{clr}^{-1}(\xi) = C(\exp(\xi_1), \exp(\xi_2), \dots, \exp(\xi_D)) = \mathbf{x}. \quad (2.2)$$

Clr souřadnice mají symetrické složky. Pro dva kompoziční vektory $\mathbf{x}_1, \mathbf{x}_2 \in S^D$

a konstanty $\alpha, \beta \in \mathbb{R}$ dále platí

$$clr(\alpha \odot \mathbf{x}_1 \oplus \beta \odot \mathbf{x}_2) = \alpha \cdot clr(\mathbf{x}_1) + \beta \cdot clr(\mathbf{x}_2); \quad (2.3)$$

$$\langle \mathbf{x}_1, \mathbf{x}_2 \rangle_a = \langle clr(\mathbf{x}_1), clr(\mathbf{x}_2) \rangle; \quad (2.4)$$

$$\|\mathbf{x}_1\|_a = \|clr(\mathbf{x}_1)\|; \quad (2.5)$$

$$d_a(\mathbf{x}_1, \mathbf{x}_2) = d(clr(\mathbf{x}_1), clr(\mathbf{x}_2)). \quad (2.6)$$

Clr transformace tedy převádí operace ze simplexu do celého reálného prostoru. Dané vlastnosti však neplatí pro dvojice složek, pouze pro celé clr vektory. Problémem je, že clr nejsou ortonormální souřadnice s ohledem na bázi simplexu. Protože je součet složek clr transformace roven 0, je varianční matice $clr(\mathbf{x}) = \xi$ pro náhodnou kompozici singulární, což znemožňuje využití některých statistických metod. Clr dále nespĺňují princip podkompoziční koherence.[24]

Příklad 2.1 *Mějme stejnou kompozici ekologické stopy České republiky jako v předchozích příkladech, tj. $\mathbf{x} = (0, 8727; 0, 3028; 0, 7856; 0, 0348; 0, 1524; 3, 4516)$. Clr souřadnice pro tento vektor jsou*

$$clr(\mathbf{x}) = (0, 7926; -0, 2659; 0, 6875; -2, 4294; -0, 9525; 2, 1676) .$$

Vzhledem k tomu, že jednotlivé složky ekologické stopy dělíme geometrickým průměrem všech složek, můžeme vlastně srovnávat jednotlivé části stopy s průměrnou hodnotou přes všechny složky ekologické stopy. Části grazing land, fishing grounds a built-up land mají záporné clr souřadnice, a proto se jedná o podprůměrné hodnoty stopy. Nejnižší hodnotu má fishing grounds a jedná se tedy o nejméně významnou část ekologické stopy. Nejvyšší hodnoty naopak dosahuje stopa carbon, která má výrazně nadprůměrné hodnoty.

(U výpočtu souřadnic v softwaru R byla použita funkce `geoMean()` z balíčku `EnvStats` [20]. Pro výpočet clr souřadnic lze využít funkci `cenLR()` z balíčku `robCompositions` [28].)

Dalším typem souřadnic jsou aditivní logratio (alr) souřadnice. Jedná se o původní souřadnice zavedené Johnem Aitchisonem a pro kompoziční vektor $\mathbf{x} \in$

S^D jdou dány jako

$$alr(\mathbf{x}) = \left(\ln \frac{x_1}{x_D}, \ln \frac{x_2}{x_D}, \dots, \ln \frac{x_{D-1}}{x_D} \right) = \zeta, \quad (2.7)$$

kde x_D je tzv. referenční část. Původní kompoziční vektor \mathbf{x} lze získat pomocí inverzní transformace

$$\mathbf{x} = alr^{-1}(\zeta) = C(\exp(\zeta_1), \exp(\zeta_2), \dots, \exp(\zeta_{D-1}), 1). \quad (2.8)$$

Alr transformace splňuje pro každé dva vektory $\mathbf{x}_1, \mathbf{x}_2 \in S^D$ a koeficienty $\alpha, \beta \in \mathbb{R}$

$$alr(\alpha \odot \mathbf{x}_1 \oplus \beta \odot \mathbf{x}_2) = \alpha \cdot alr(\mathbf{x}_1) + \beta \cdot alr(\mathbf{x}_2). \quad (2.9)$$

Je vidět, že tyto souřadnice nemají zdaleka tolik výhodných vlastností jako předchozí clr transformace. Důvodem je to, že alr transformace nejsou izometrické. Nejsou tedy vhodné pro použití v metodách, kde je potřeba zachování úhlů a vzdáleností, protože by došlo k jejich deformaci. Dále nesplňují permutační invarianci, což může být problém při použití některých statistických metod. Alr transformace nemá oproti clr transformaci symetrické složky, a to kvůli výběru referenční části. Při výběru odlišné složky jako referenční dostaneme odlišné alr souřadnice. Výhodou je však snadná interpretace vyjadřující dominanci jednotlivých složek kompozice na složce referenční.[2, 22, 24]

Příklad 2.2 *Nyní spočítáme alr souřadnice pro ekologickou stopu České republiky. Použijeme vzorec 2.9, kde je jako referenční složka použit poslední člen kompozičního vektoru, což je v našem případě carbon*

$$alr(\mathbf{x}) = (-1, 3750; -2, 4335; -1, 4801; -4, 5970; -3, 1201).$$

Vzhledem k výběru referenční složky získáme pomocí těchto souřadnic určité srovnání ostatních složek kompozičního vektoru s touto vybranou složkou. Je vidět, že ve všech případech vychází pro souřadnice záporná znaménka, a proto lze říct, že složka carbon dominuje všechny ostatní části ekologické stopy. Nejvýraznější je tato dominance u složek fishing grounds a built-up land. Po nich následuje grazing land a forest products. Nejvyšší podíl ku složce carbon má část cropland.

(Pro výpočet *alr* souřadnic lze využít funkci `addLR()` z balíčku `robCompositions` [28].)

Kvůli nedostatkům předchozích typů transformací je potřeba zavést další typ souřadnic, a to isometrické logratio souřadnice (*ilr*) konstruované vzhledem k určité ortonormální bázi na simplexu.[24]

2.2. Ortonormální souřadnice

Množina kompozičních vektorů $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{D-1}\}$, které tvoří ortonormální bázi, splňuje

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle_a = 0 \text{ pro } i \neq j \text{ a zároveň } \langle \mathbf{e}_i, \mathbf{e}_i \rangle_a = 1 \text{ pro } i = j. \quad (2.10)$$

Ortonormální bázi lze získat pro $D - 1$ nezávislých kompozičních vektorů například pomocí Gram-Schmidtova procesu, kterým dostaneme jednu z nekonečně mnoha ortonormálních bází na euklidovském vektorovém prostoru.[6, 24]

Danou kompozici $\mathbf{x} \in S^D$ je poté možné vyjádřit ve tvaru

$$\mathbf{x} = \bigoplus_{i=1}^{D-1} z_i \odot \mathbf{e}_i, \quad z_i = \langle \mathbf{x}, \mathbf{e}_i \rangle_a, \quad (2.11)$$

kde $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{D-1}\}$ je zvolená ortonormální báze a $\mathbf{z} = (z_1, z_2, \dots, z_{D-1})$ je vektor souřadnic s ohledem na tuto bázi.[24]

Isometrická logratio transformace (*ilr*) je funkce $\text{ilr}: S^D \rightarrow \mathbb{R}^{D-1}$, která kompozici $\mathbf{x} \in S^D$ přiřazuje souřadnice \mathbf{z} s ohledem na danou ortonormální bázi $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{D-1}\}$. Funkce $S^D \rightarrow \mathbb{R}^{D-1}$ je izometrie vektorových prostorů.[24]

Pro dva kompoziční vektory $\mathbf{x}_1, \mathbf{x}_2 \in S^D$ a konstanty $\alpha, \beta \in \mathbb{R}$ platí

$$\text{ilr}(\alpha \odot \mathbf{x}_1 \oplus \beta \odot \mathbf{x}_2) = \alpha \cdot \text{ilr}(\mathbf{x}_1) + \beta \cdot \text{ilr}(\mathbf{x}_2) = \alpha \cdot \mathbf{z}_1 + \beta \cdot \mathbf{z}_2; \quad (2.12)$$

$$\langle \mathbf{x}_1, \mathbf{x}_2 \rangle_a = \langle \text{ilr}(\mathbf{x}_1), \text{ilr}(\mathbf{x}_2) \rangle = \langle \mathbf{z}_1, \mathbf{z}_2 \rangle; \quad (2.13)$$

$$\|\mathbf{x}_1\|_a = \|\text{ilr}(\mathbf{x}_1)\| = \|\mathbf{z}_1\|; \quad (2.14)$$

$$d_a(\mathbf{x}_1, \mathbf{x}_2) = d(\text{ilr}(\mathbf{x}_1), \text{ilr}(\mathbf{x}_2)) = d(\mathbf{z}_1, \mathbf{z}_2). \quad (2.15)$$

Rozdíl mezi vlastnostmi clr a ilr transformací je v tom, že clr vektory koeficientů se nachází v \mathbb{R}^D prostoru, zatímco ilr v \mathbb{R}^{D-1} . Při využití ilr transformace nedostaneme singulární varianční matici jako u clr souřadnic. [2, 24]

Pro definování ortonormální báze v simplexu existuje více možností. Hlavním kritériem pro výběr vhodné metody je snaha o lepší interpretovatelnost souřadnicové reprezentace kompozičního vektoru. Jednou z možností je využití bilancí.[22]

2.2.1. Bilance

Bilance jsou ilr souřadnice získané prostřednictvím postupného binárního dělení kompozičního vektoru (SBP).[4]

Proces SBP začíná rozdělením složek kompozice na dvě části, z nichž jedna bude umístěna v čitateli, druhá ve jmenovateli. V dalším kroku se to stejné provede u každé z těchto dvou skupin a v dělení se pokračuje tak dlouho, dokud ve skupinách nezůstane pouze jedna složka.[24]

V k -tém kroku je možné určit bilanci mezi dvěma podskupinami, které byly vytvořeny na dané úrovni. První podskupina je tvořena r_k částmi i_1, i_2, \dots, i_{r_k} umístěnými v čitateli. Druhá podskupina má s_k částí j_1, j_2, \dots, j_{s_k} umístěných ve jmenovateli. Bilance mezi těmito podskupinami je dána jako

$$\begin{aligned} z_k &= \sqrt{\frac{r_k s_k}{r_k + s_k}} \ln \frac{(x_{i_1} x_{i_2} \cdots x_{i_{r_k}})^{1/r_k}}{(x_{j_1} x_{j_2} \cdots x_{j_{s_k}})^{1/s_k}} = \\ &= \sqrt{\frac{r_k s_k}{r_k + s_k}} \left(\frac{\ln(x_{i_1}) + \cdots + \ln(x_{i_{r_k}})}{r_k} - \frac{\ln(x_{j_1}) + \cdots + \ln(x_{j_{s_k}})}{s_k} \right), \end{aligned} \quad (2.16)$$

kde $\sqrt{\frac{r_k s_k}{r_k + s_k}}$ je faktor normující souřadnice.[4, 7]

Nyní se vrátíme zpět k logratio transformacím, a to přesně k logkontrastu, který je škálově invariantní logratio transformace. Pro danou kompozici $\mathbf{x} = (x_1, \dots, x_D) \in S^D$ a koeficienty $\psi_j \in \mathbb{R} \forall j = 1, \dots, D$ je logkontrast funkce f , pro kterou platí

$$f(\mathbf{x}) = \sum_{j=1}^D \psi_j \ln x_j, \text{ kde } \sum_{j=1}^D \psi_j = 0. \quad (2.17)$$

U logkontrastu jsou koeficienty k -té souřadnice pro část v čitateli $\psi_{jk} = \sqrt{\frac{s_k}{r_k(r_k+s_k)}}$, pro část ve jmenovateli $\psi_{jk} = -\sqrt{\frac{r_k}{s_k(r_k+s_k)}}$. Pro část, která se nevyskytuje ani v čitateli, ani ve jmenovateli je $\psi_{jk} = 0$. Jakoukoli ze souřadnicových reprezentací z této kapitoly je možné vyjádřit v podobě logkontrastu.[4, 7, 24]

Lze definovat transformační matici Ψ jako

$$\Psi = \begin{pmatrix} \psi_{11} & \cdots & \psi_{1D-1} \\ \vdots & & \vdots \\ \psi_{D1} & \cdots & \psi_{DD-1} \end{pmatrix} \quad (2.18)$$

s D řádky a $(D - 1)$ ortonormálními sloupci. Pomocí této matice je možné získat bilance $\mathbf{z}_i = (z_{i1}, \dots, z_{iD-1})$ k log-absolutním složkám $\ln(\mathbf{x}_i) = (\ln(x_{i1}), \dots, \ln(x_{iD}))$ následujícím způsobem[4]

$$\mathbf{z}_i = \ln(\mathbf{x}_i) \cdot \Psi. \quad (2.19)$$

Dle [4] je možné matici interpretovat jako matici ortonormální projekce z prostoru \mathbb{R}^D do prostoru \mathbb{R}^{D-1} , který je ortogonální k vektoru $\mathbf{1} = (1, 1, \dots, 1)$ izometrickému k simplexu. Z různého postupného binárního dělení je získána odlišná Ψ matice.

Příklad 2.3 *Na datech o ekologické stopě provedeme postupné binární dělení. V prvním kroku oddělíme carbon od ostatních složek, protože se jedná o emise CO_2 vyjádřené jako lesní plocha potřebná pro jejich absorbování, zatímco ostatní složky jsou přímo vázány na využití daného typu půdy. V dalším kroku oddělíme built-up land, který se týká lidské infrastruktury. Ostatní složky popisují přímé využití přírodních zdrojů. Ve třetím kroku se oddělí fishing grounds, protože se na rozdíl od zbylých částí zabývá využitím vodních zdrojů pro rybolov, nikoli pevninou. Poté oddělíme forest land, využívanou pro těžbu dřeva, nikoli pro pěstování potravy pro člověka či zvířata jako zbylé dvě složky. V následujícím kroku se oddělí poslední dvě části, a to cropland a grazing land. Postup dělení je znázorněn v následující tabulce, kde + značí umístění složky v čitateli, - ve jmenovateli.*

| | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 |
|---|-------|-------|-------|-------|-------|-------|
| 1 | + | + | + | + | + | - |
| 2 | + | + | + | + | - | 0 |
| 3 | + | + | + | - | 0 | 0 |
| 4 | + | + | - | 0 | 0 | 0 |
| 5 | + | - | 0 | 0 | 0 | 0 |

Odovídajících pět ilr souřadnic (bilancí) vypadá následovně

$$z_1 = \sqrt{\frac{5}{6}} \ln \frac{(x_1 x_2 x_3 x_4 x_5)^{1/5}}{x_6}$$

$$z_2 = \sqrt{\frac{4}{5}} \ln \frac{(x_1 x_2 x_3 x_4)^{1/4}}{x_5}$$

$$z_3 = \sqrt{\frac{3}{4}} \ln \frac{(x_1 x_2 x_3)^{1/3}}{x_4}$$

$$z_4 = \sqrt{\frac{2}{3}} \ln \frac{(x_1 x_2)^{1/2}}{x_3}$$

$$z_5 = \sqrt{\frac{1}{2}} \ln \frac{x_1}{x_2}.$$

Nyní si vezměme vektor ekologické stopy pro některou konkrétní zemi, např. Českou republiku: $\mathbf{x} = (0,8727; 0,3028; 0,7856; 0,0348; 0,1524; 3,4516)$. Pro ni spočítáme odpovídajících pět bilancí

$$z_1 = \sqrt{\frac{5}{6}} \ln \frac{0,2560472}{3,451645} = -2,3746,$$

$$z_2 = \sqrt{\frac{4}{5}} \ln \frac{0,2915034}{0,1524143} = 0,5800,$$

$$z_3 = \sqrt{\frac{3}{4}} \ln \frac{0,5921448}{0,0347769} = 2,4550,$$

$$z_4 = \sqrt{\frac{2}{3}} \ln \frac{0,5140829}{0,7856287} = -0,3463,$$

$$z_5 = \sqrt{\frac{1}{2}} \ln \frac{0,8727034}{0,3028305} = 0,7484.$$

Z vypočtených bilancí jsou tři s kladným znaménkem, zbylé dvě se záporným. U záporných bilancí je tedy vidět, že část ve jmenovateli je vyšší než skupina složek v čitateli. U z_1 to značí, že carbon dominuje nad všemi zbylými složkami ekologické stopy. U z_4 poté forest products dominují cropland a grazing land. Dominance složky ve jmenovateli však není tak výrazná jako v předchozím případě. Z kladných bilancí má nejvyšší vliv čítec u z_3 , kde se jedná o převahu přírodních suchozemských částí ekologické stopy na části vodní. U z_5 se jedná o dominanci části cropland na grazing land. Nejnižší kladná souřadnice je z_2 , značící převahu všech přírodních složek ekologické stopy k built-up land.

(Bilance je možné spočítat v R například pomocí funkce `balances()` z balíčku `robCompositions` [28].)

Pivotové souřadnice

Dalším typem bilancí jsou pivotové souřadnice. Název je odvozen z postupu tvorby souřadnic, kde se vždy vezme jedna z částí kompozice jako tzv. pivot a je použita pouze v první souřadnici. Jednotlivé souřadnice při použití x_1 v roli pivota vypadají

$$z_j = \sqrt{\frac{D-j}{D-j+1}} \ln \frac{x_j}{\sqrt[D-j]{\prod_{k=j+1}^D x_k}}, \text{ kde } j = 1, \dots, D-1. \quad (2.20)$$

Souřadnice z_1 obsahuje veškerou relativní informaci o části x_1 s ohledem k ostatním složkám v kompozici.[7]

Příklad 2.4 *Mějme opět stejná data o ekologické stopě jako v předchozím případě. Tentokrát však pro tvorbu souřadnic použijeme pivotové souřadnice se vzorcem 2.20, kde je jako pivot použit první člen kompozice. Nejdřív tedy oddělíme část cropland, poté grazing land a postupně se pokračuje až po poslední část carbon. Kódovací tabulka vypadá následovně*

| | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 |
|---|-------|-------|-------|-------|-------|-------|
| 1 | + | - | - | - | - | - |
| 2 | 0 | + | - | - | - | - |
| 3 | 0 | 0 | + | - | - | - |
| 4 | 0 | 0 | 0 | + | - | - |
| 5 | 0 | 0 | 0 | 0 | + | - |

Samotné souřadnice jsou uvedeny níže

$$z_1 = \sqrt{\frac{5}{6}} \ln \frac{x_1}{(x_2 x_3 x_4 x_5 x_6)^{1/5}}$$

$$z_2 = \sqrt{\frac{4}{5}} \ln \frac{x_2}{(x_3 x_4 x_5 x_6)^{1/4}}$$

$$z_3 = \sqrt{\frac{3}{4}} \ln \frac{x_3}{(x_4 x_5 x_6)^{1/3}}$$

$$z_4 = \sqrt{\frac{2}{3}} \ln \frac{x_4}{(x_5 x_6)^{1/2}}$$

$$z_5 = \sqrt{\frac{1}{2}} \ln \frac{x_5}{x_6}.$$

Souřadnice z_1 označuje celkovou relativní informaci o části cropland ke všem ostatním složkám ekologické stopy. z_2 značí opět celkovou relativní informaci grazing land ku všem dalším složkám kromě x_1 apod. S použitím výše uvedeného je vektor pivotových souřadnic pro Českou republiku $\mathbf{z} = (0, 8684; -0, 1198; 0, 9461; -2, 4802; -2, 2062)$. Tentokrát jsou tři souřadnice záporné, zbývající dvě kladné. U první skupiny má nejvyšší zápornou hodnotu z_4 , kde je vidět, že části carbon a built-up land dominují fishing grounds. Ze souřadnice z_5 plyne, že pro Českou republiku je vyšší hodnota carbon proti stopě built-up land. U souřadnice z_2 je vidět, že ekologické stopy lesních produktů (forest products), fishing grounds, built-up land a carbon mají mírnou převahu nad grazing land. Nejvyšší kladné hodnoty dosahuje souřadnice z_3 , která ukazuje dominanci forest products nad fishing grounds, built-up land a uhlíkovou stopou (carbon). U první souřadnice z_1 je dominance první složky ekologické stopy, kterou je cropland, na všech ostatních složkách.

(Pivotové souřadnice lze spočítat pomocí funkce `pivotCoord()` z balíčku `robCompositions` [28].)

Kapitola 3

Lineární modely s kompozičními daty a součtem

V této kapitole budou postupně představeny jednotlivé lineární modely, které lze použít v situacích, kdy nás nezajímá pouze závislost zvolené proměnné na daném kompozičním vektoru, ale předpokládáme důležitost také absolutní informace obsažené v kompozici. Tato informace může být vyjádřena více způsoby, které budou blíže představeny.

3.1. Klasický přístup

Pokud chceme zkoumat relativní i absolutní informaci obsaženou v datech, nejjednodušší možností je použít klasický přístup pro modelování závisle proměnné. Ten spočívá v použití původních hodnot kompozičního vektoru, a to po provedení jejich log transformace. Užítí log transformace je výhodné hned z několika důvodů. Z proměnných s pozitivními hodnotami získáme hodnoty na celém reálném prostoru a zbavíme se pozitivní šikmosti a velkých odlehlých pozorování (outlierů), které se po transformaci přiblíží středu rozdělení. Model pro modelování závisle metrické proměnné y v závislosti na proměnných (x_1, x_2, \dots, x_D) pak vypadá pro i -té pozorování následovně

$$y_i = \gamma_0 + \gamma_1 \ln(x_{i1}) + \dots + \gamma_D \ln(x_{iD}) + u_i = \gamma_0 + \ln(\mathbf{x}_i) \cdot \boldsymbol{\gamma} + u_i, \quad (3.1)$$

kde u_i je chyba modelu. Regresní parametry γ_1 až γ_D lze interpretovat jako efekt zvýšení logaritmu jedné ze složek kompozice na hodnotu závisle proměnné

při konstantních hodnotách zbylých složek. Protože všechny koeficienty γ_1 až γ_D nesou absolutní i relativní informaci, jedná se vlastně o růst relativní významnosti dané složky při současném růstu absolutní velikosti kompozice.[4]

Pokud bychom měli kompoziční vektor o $D = 6$ složkách (což je případ, který bude využit v praktické části pro data o ekologické stopě), dostali bychom následující model

$$y = \gamma_0 + \gamma_1 \ln(x_1) + \cdots + \gamma_6 \ln(x_6). \quad (3.2)$$

Parametr γ_0 nám říká, jaká je hodnota závisle proměnné, pokud všechny složky kompozičního vektoru mají hodnotu 1, tj. $x_1 = x_2 = \cdots = x_6 = 1$. Při interpretaci ostatních parametrů se např. pro γ_1 vychází ze dvou modelů

$$\begin{aligned} y &= \gamma_0 + \gamma_1 \ln(x_1) + \cdots + \gamma_6 \ln(x_6), \\ y' &= \gamma_0 + \gamma_1 \ln(kx_1) + \cdots + \gamma_6 \ln(x_6), \end{aligned}$$

kde y a y' jsou stejné modely, ale hodnota složky x_1 se u y' zvýšila k -krát. Výsledný efekt změny se poté odvozuje z následujících úprav

$$\begin{aligned} y' - y &= \gamma_1 \ln(kx_1) - \gamma_1 \ln(x_1), \\ y' - y &= \gamma_1 \ln \frac{kx_1}{x_1} = \gamma_1 \ln k. \end{aligned}$$

Pro parametr γ_1 nám poslední řádek odvození říká, o kolik se změní závisle proměnná y , pokud se složka x_1 zvýší k -krát a všechny ostatní složky kompozičního vektoru zůstanou nezměněny. Pro $k = e$, potom γ_i odpovídá tomu, o kolik se zvýší y , pokud se přílušná složka kompozičního vektoru x_i zvýší e -krát a všechny ostatní se nezmění.

V kapitole 5, obsahující praktickou část práce, však nebudeme používat přímo tento model, ale jeho upravenou verzi, kde jsou nejen na straně vysvětlujících proměnných logaritmy složek kompozičního vektoru, ale také u zkoumané závisle proměnné jsou její hodnoty po log transformaci. Tento model vypadá pro i -té pozorování

$$\ln y_i = \gamma_0 + \gamma_1 \ln(x_{i1}) + \cdots + \gamma_D \ln(x_{iD}) + u_i, \quad (3.3)$$

a opět pro $D = 6$ složek kompozičního vektoru dostaneme

$$\ln y = \gamma_0 + \gamma_1 \ln(x_1) + \cdots + \gamma_6 \ln(x_6). \quad (3.4)$$

Pokud by složky kompozice nabývaly všechny totožné hodnoty, a to $x_1 = x_2 = \cdots = x_6 = 1$, hodnota závisle proměnné y by byla e^{γ_0} . U ostatních parametrů by se vztah pro výpočet efektů změn odvodil podobně jako v předchozím případě, výsledkem by byl vztah

$$\begin{aligned} \ln \frac{y'}{y} &= \gamma_1 \ln(kx_1) - \gamma_1 \ln(x_1), \\ \ln \frac{y'}{y} &= \gamma_1 \ln \frac{kx_1}{x_1} = \ln k^{\gamma_1}, \\ \frac{y'}{y} &= k^{\gamma_1}. \end{aligned}$$

Pro parametr γ_1 vyjadřuje výraz k^{γ_1} kolikrát se změní y , pokud se zvýší složka x_1 k -krát a všechny ostatní složky kompozice se nezmění. Pokud bychom opět za k vzali hodnotu e , dostali bychom v posledním kroku e^{γ_1} , což by vyjadřovalo kolikrát se změní y , pokud se hodnota x_1 zvýší e -krát a všechny ostatní složky budou mít stále stejné hodnoty. To odpovídá vzrůstu $\ln x_1$ o 1, protože platí $\ln x_1 + 1 = \ln x_1 + \ln e = \ln(ex_1)$. Stejná interpretace by byla u změn všech ostatních složek kompozičního vektoru, kde např. u γ_2 by se při k -násobné změně složky x_2 změnilo y k^{γ_2} krát.

3.2. T-prostory

Dalším přístupem je tzv. T-prostor, o kterém se lze dočíst podrobněji v [23]. T-prostor je tvořen kartézským součinem prostoru \mathbb{R}_+ a simplexu S^D . Principem je, že místo původního kompozičního vektoru se použije $(D - 1)$ ilr souřadnic a součet složek (total) vyjádřený v různých formách. Jedním z možných vyjádření je ve tvaru $\tau_\delta(\mathbf{x}) = \prod_{i=1}^D x_i^\delta$. Pokud za δ vezmeme $1/\sqrt{D}$, dostaneme při použití log transformace součet ve formě logaritmu geometrického průměru

$$\tau_i = \sqrt{D} \ln(\sqrt[D]{x_{i1}x_{i2} \cdots x_{iD}}) = \frac{1}{\sqrt{D}} (\ln(x_{i1}) + \ln(x_{i2}) + \cdots + \ln(x_{iD})). \quad (3.5)$$

V této formě jsou vzdálenosti mezi pozorováními stejné jako u logaritmu absolutních hodnot. Je to vlastně projekce $\ln(\mathbf{x}_i)$ na vektor $(1/\sqrt{D}) \cdot \mathbf{1}$. [4, 23]

Druhou formou je vyjádření pomocí logaritmu součtu složek kompozice jako

$$\eta_i = \ln(x_{i1} + x_{i2} + \dots + x_{iD}). \quad (3.6)$$

Jedná se o stejný logaritmus součtu užívaný v operaci uzávěr. [4]

Příklad 3.1 *Nyní si ukážeme výpočet součtu pro kompoziční vektor $\mathbf{x} = (0, 8727; 0, 3028; 0, 7856; 0, 0348; 0, 1524; 3, 4516)$ vyjadřující ekologickou stopu České republiky. Nejdřív spočítáme součet τ_i*

$$\tau_{CZ} = \frac{1}{\sqrt{6}}(\ln(0, 8727) + \dots + \ln(3, 4516)) = -2, 2750.$$

Při použití součtu η_i dostaneme následující výsledek

$$\eta_{CZ} = \ln(0, 8727 + \dots + 3, 4516) = 1, 7227.$$

Samotný člen s absolutní informací nelze příliš dobře interpretovat. Při pohledu na součty více zemí je však možné je mezi sebou porovnat. To stejné proto nyní spočítáme pro Slovensko, které má vektor ekologické stopy $\mathbf{y} = (0, 6502; 0, 1328; 0, 6749; 0, 0421; 0, 1252; 2, 5703)$ a součty jsou

$$\tau_{SK} = \frac{1}{\sqrt{6}}(\ln(0, 6502) + \dots + \ln(2, 5703)) = -2, 9166,$$

$$\eta_{SK} = \ln(0, 6502 + \dots + 2, 5703) = 1, 4340.$$

Z výpočtů je vidět, že u obou typů součtu je výsledná hodnota vyšší u České republiky. Lze tedy říct, že Česká republika má vyšší celkovou úroveň ekologické stopy než Slovensko.

Výše uvedené formy součtu nejsou jedinou možností, použít se může také např. aritmetický průměr či jiné formy součtu vhodné pro daný problém. Nyní se však budeme zabývat pouze modelem se součtem ve formě geometrického průměru, který je z hlediska kompoziční teorie optimální variantou, jak si ukážeme v následující podkapitole. [4, 23]

3.2.1. Model se součtem τ

Při použití součtu τ (viz 3.5) vypadá model pro i -té pozorování následovně

$$y_i = \beta_0 + \beta_1 z_{i1} + \cdots + \beta_{D-1} z_{iD-1} + \beta_D \tau_i + u_i = \beta_0 + (\mathbf{z}_i \tau_i) \cdot \boldsymbol{\beta} + u_i, \quad (3.7)$$

kde \mathbf{z}_i jsou jednotlivé bilance.[4]

Pokud nyní opět budeme uvažovat kompozici s šesti složkami a k nim příslušných pět bilancí z příkladu 2.3, dostaneme model

$$\begin{aligned} y = & \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 \sqrt{\frac{4}{5}} \ln \frac{\sqrt[4]{x_1 \cdots x_4}}{x_5} + \beta_3 \sqrt{\frac{3}{4}} \ln \frac{\sqrt[3]{x_1 \cdots x_3}}{x_4} + \\ & + \beta_4 \sqrt{\frac{2}{3}} \ln \frac{\sqrt{x_1 x_2}}{x_3} + \beta_5 \sqrt{\frac{1}{2}} \ln \frac{x_1}{x_2} + \beta_6 \sqrt{6} \ln \sqrt[6]{x_1 \cdots x_6}. \end{aligned} \quad (3.8)$$

Zde parametr β_0 udává hodnotu závisle proměnné y , když všechny složky kompozičního vektoru mají hodnoty 1, tj. $x_1 = x_2 = \cdots = x_6 = 1$. Efekt změny y při změně v hodnotě \ln souřadnic, lze opět odvodit podobným způsobem jako u klasického přístupu ze dvou modelů. Budeme uvažovat změnu u první \ln souřadnice

$$\begin{aligned} y &= \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 \ln r_2 + \cdots + \beta_5 \ln r_5 + \beta_6 \tau, \\ y' &= \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln k \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 \ln r_2 + \cdots + \beta_5 \ln r_5 + \beta_6 \tau, \\ y' - y &= \beta_1 \sqrt{\frac{5}{6}} \ln \frac{k \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6}}{\frac{\sqrt[5]{x_1 \cdots x_5}}{x_6}} = \beta_1 \sqrt{\frac{5}{6}} \ln k, \end{aligned}$$

kde $\beta_1 \sqrt{\frac{5}{6}} \ln k$ vyjadřuje, o kolik se změní y , vzroste-li hodnota podílu první \ln souřadnice k -krát a ostatní bilance a součet zůstanou nezměněny. Podobné odvození poté platí také pro ostatní bilance. U součtu se poté při k -násobném vzrůstu geometrického průměru uvnitř logaritmu zvýší závisle proměnná při konstantních bilancích o $\beta_6 \sqrt{6} \ln k$.

V kapitole 5 se budeme zabývat modelem s logaritmovanými hodnotami závisle proměnné. Obecně by tento model pro i -té pozorování vypadal

$$\ln y_i = \beta_0 + \beta_1 z_{i1} + \cdots + \beta_{D-1} z_{iD-1} + \beta_D \tau_i + u_i = \beta_0 + (\mathbf{z}_i \tau_i) \cdot \boldsymbol{\beta} + u_i. \quad (3.9)$$

Pro tento případ si nyní ukážeme interpretaci pro $D = 6$ složek kompozičního vektoru. Budeme vycházet z modelu

$$\begin{aligned} \ln y = & \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 \sqrt{\frac{4}{5}} \ln \frac{\sqrt[4]{x_1 \cdots x_4}}{x_5} + \beta_3 \sqrt{\frac{3}{4}} \ln \frac{\sqrt[3]{x_1 \cdots x_3}}{x_4} + \\ & + \beta_4 \sqrt{\frac{2}{3}} \ln \frac{\sqrt{x_1 x_2}}{x_3} + \beta_5 \sqrt{\frac{1}{2}} \ln \frac{x_1}{x_2} + \beta_6 \sqrt{6} \ln \sqrt[6]{x_1 \cdots x_6}. \end{aligned} \quad (3.10)$$

Pro $x_1 = \cdots x_6 = 1$ je hodnota závisle proměnné rovna e^{β_0} . Efekty zvýšení bilancí a součtu na hodnotu závisle proměnné lze např. pro první ilr souřadnici odvodit následujícím způsobem

$$\begin{aligned} \ln y = & \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 ilr_2 + \cdots + \beta_5 ilr_5 + \beta_6 \tau, \\ \ln y' = & \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln k \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 ilr_2 + \cdots + \beta_5 ilr_5 + \beta_6 \tau. \end{aligned}$$

Po odečtení modelů dostaneme

$$\begin{aligned} \ln y' - \ln y = & \beta_1 \sqrt{\frac{5}{6}} \ln k \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} - \beta_1 \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6}, \\ \ln \frac{y'}{y} = & \beta_1 \sqrt{\frac{5}{6}} \ln \frac{k \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6}}{\ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6}}, \\ \ln \frac{y'}{y} = & \ln [k]^{\beta_1 \sqrt{\frac{5}{6}}}, \\ \frac{y'}{y} = & k^{\beta_1 \sqrt{\frac{5}{6}}}, \end{aligned}$$

kde $k^{\beta_1 \sqrt{\frac{5}{6}}}$ odpovídá efektu, kolikrát se změní y , zvýší-li se podíl v první ilr souřadnici k -krát a přitom ostatní bilance a součet zůstanou na stejných hodnotách. Obdobným způsobem by se poté odvodily efekty pro zbylé souřadnice a součet. Vzhledem k tomu, že jednotlivé složky kompozičního vektoru se objevují ve všech bilancích i součtu, musí u příslušného podílu vždy dojít k takové změně, aby ostatní bilance a součet zůstaly nezměněny. Toho docílíme proporční změnou jak v čitateli, tak ve jmenovateli podílu.[4]

Například u první bilance, musí při k -násobném nárůstu podílu uvnitř logaritmu dojít jak k růstu všech pěti složek v čitateli bilance, tak k poklesu složky ve jmenovateli. Složky v čitateli se musí všechny zvýšit $\sqrt[6]{k}$ krát, zatímco x_6 se změní $1/\sqrt[6]{k^5}$ krát.

U druhé bilance se při zvýšení podílu složek na k -násobek změní závisle proměnná y $k^{\beta_2}\sqrt[4]{5}$ krát. Aby se tato bilance k -krát zvýšila a současně ostatní bilance a součet zůstaly nezměněny, musí se složky v jejím čitateli změnit $\sqrt[5]{k}$ násobně a x_5 vynásobit $1/\sqrt[5]{k^4}$. Složka x_6 , která není součástí druhé bilance, zůstane nezměněna.

Výše efektu změny závisle proměnné na k -násobném růstu zlomku uvnitř třetí bilance, je $k^{\beta_3}\sqrt[3]{4}$. Pro tak velkou změnu této bilance je třeba vynásobit složky v čitateli $\sqrt[4]{k}$ krát, x_4 se změní $1/\sqrt[4]{k^3}$ krát. Složky x_5 a x_6 se nezmění.

Při k -násobném zvýšení podílu složek čtvrté bilance dojde ke $k^{\beta_4}\sqrt[2]{3}$ násobné změně proměnné y . K tomu se musí změnit složky v čitateli $\sqrt[3]{k}$ krát a x_3 vynásobit $1/\sqrt[3]{k^2}$. Složky x_4 , x_5 a x_6 zůstanou nezměněny.

U poslední bilance se při vzrůstu podílu složek x_1 a x_2 na k -násobek změní závisle proměnná $k^{\beta_5}\sqrt[1]{2}$ krát. K tomuto efektu je nutná vynásobit složku x_1 číslem \sqrt{k} a x_2 číslem $1/\sqrt{k}$. Složky x_3 , x_4 , x_5 a x_6 se nezmění.

U všech předchozích vyčíslení změn složek kompozice se jedná o jediné možné změny, které zaručí, že ostatní bilance a součet zůstanou nezměněny. Jiné možnosti v modelu se součtem τ nejsou.

Při k -násobném zvýšení součtu, který ve formě geometrického průměru představuje vlastně celkovou úroveň, dojde ke změně hodnoty y na $k^{\beta_6}\sqrt[6]{6}$ násobek. Všechny složky kompozičního vektoru se musí změnit stejně, a to k -krát. Tak dojde nejen k zachování hodnot všech bilancí, ale znamená to také zvýšení celkové úrovně při zachování všech relativních podílů mezi složkami. Opět se jedná o jedinou možnost, jak zvýšit součet při zachování stejné úrovně bilancí.

Jednou z velmi výhodných vlastností modelů 3.7 a 3.9 je SBP invariance. Znamená to, že dané výsledky nezávisí na vytvořených bilancích, protože při výběru jakéhokoli SBP dělení bude dosaženo totožných výsledků. Tato vlastnost

platí jak pro výsledky F statistiky, tak také pro celkový efekt parametrů β_0 a β_D a charakteristiky goodness of fit, jako jsou např. R^2 a Akaikeho informační kritérium (AIC). Nezávisí přitom na zvoleném typu součtu, pro modely s oběma v této práci užívanými formami součtu je vlastnost SBP invariance splněna.[4]

Další vlastností je totožnost výsledků F statistiky a charakteristik goodness of fit modelu 3.7, resp. 3.9 s výsledky u klasického přístupu 3.1, resp. 3.3. Model s ilr souřadnicemi má však výhodnější interpretaci. Zde však závisí také na zvoleném typu součtu, poněvadž daná vlastnost platí pouze pro součet τ . Pro druhý typ součtu η , tvořený logaritmem součtu složek kompozičního vektoru, model nedává stejné výsledky jako klasický přístup.[4]

Totožnost výsledků F statistiky u klasického přístupu a modelu se součtem τ lze ukázat následujícím způsobem. Můžeme definovat novou transformační matici \mathbf{U} , která bude obsahovat sloupce pro součet

$$\mathbf{U} = \begin{pmatrix} \psi_{11} & \cdots & \psi_{1D-1} & 1/\sqrt{D} \\ \vdots & & \vdots & 1/\sqrt{D} \\ \psi_{D1} & \cdots & \psi_{DD-1} & 1/\sqrt{D} \end{pmatrix}. \quad (3.11)$$

Pro matici platí $\mathbf{U}^T \mathbf{U} = \mathbf{I}$, $\mathbf{U}^{-1} = \mathbf{U}^T$ a také

$$(\mathbf{z}_i \ \tau_i) = \ln(\mathbf{x}_i) \cdot \mathbf{U} \text{ a } \ln(\mathbf{x}_i) = (\mathbf{z}_i \ \tau_i) \cdot \mathbf{U}^T. \quad (3.12)$$

Z uvedených vlastností plyne, že platí

$$\begin{aligned} y_i &= \gamma_0 + \ln(\mathbf{x}_i) \cdot \boldsymbol{\gamma} + u_i = \gamma_0 + (\mathbf{z}_i \ \tau_i) \cdot \mathbf{U}^T \cdot \boldsymbol{\gamma} + u_i = \\ &= \beta_0 + (\mathbf{z}_i \ \tau_i) \cdot \boldsymbol{\beta} + u_i, \end{aligned} \quad (3.13)$$

$$y_i = \beta_0 + (\mathbf{z}_i \ \tau_i) \cdot \boldsymbol{\beta} + u_i = \beta_0 + \ln(\mathbf{x}_i) \cdot \mathbf{U} \cdot \boldsymbol{\beta} + u_i = \gamma_0 + \ln(\mathbf{x}_i) \cdot \boldsymbol{\gamma} + u_i, \quad (3.14)$$

a že tedy $\gamma_0 = \beta_0$, $\boldsymbol{\gamma} = \mathbf{U} \cdot \boldsymbol{\beta}$, $\boldsymbol{\beta} = \mathbf{U}^T \cdot \boldsymbol{\gamma}$. Z předchozího plyne, že klasický přístup a model s geometrickým průměrem jsou skutečně ekvivalentní se stejnými chybami a dávají stejné hodnoty předpovědi.[4]

V klasickém modelu a v modelu se součtem ve formě logaritmu geometrického průměru složek kompozice (τ) mají každý globální test a míry goodness of fit totožné výsledky. Každá metoda odhadu optimalizující funkci predikovaných hodnot, pozorovaných hodnot a chyb má pro oba modely ekvivalentní odhady se

stejným optimem. Celkový β_D efekt je podle zobecněné transformační matice roven $\beta_D = \frac{1}{\sqrt{D}} \cdot \gamma_1 + \dots + \frac{1}{\sqrt{D}} \cdot \gamma_D$ a nezávisí na matici Ψ ani na použitém SBP. Předpovědní síla součtu roste se vzdáleností součtu koeficientů γ od nuly. Pokud je součet všech parametrů v γ roven 0, nepodává složka součtu žádnou další informaci a není třeba ji do modelu zahrnout.[4]

Model s ilr souřadnicemi a součtem τ má oproti klasickému přístupu několik výhod. Při testování lze lépe provádět hypotézy o absolutní a relativní významnosti odděleně, přitom hypotézy o relativní významnosti lze přímo formulovat pomocí výběru SBP. Můžeme také využít různé lineární modely, a to i zobecněné, např. s ordinální nebo binární závisle proměnnou. Dle [4] součet τ nejlépe odděluje absolutní a relativní informaci.[4]

3.2.2. Model se součtem η

Pokud v modelech využijeme druhý typ součtu 3.6, označovaný jako η , vypadá model pro i -té pozorování následovně

$$y_i = \beta_0 + \beta_1 z_{i1} + \dots + \beta_{D-1} z_{iD-1} + \beta_D \eta_i + u_i = \beta_0 + (\mathbf{z}_i \eta_i) \cdot \boldsymbol{\beta} + u_i, \quad (3.15)$$

kde \mathbf{z}_i jsou bilance. Budeme-li uvažovat náš model s kompozicí s šesti složkami a bilance z příkladu 2.3, dostaneme model

$$\begin{aligned} y = & \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 \sqrt{\frac{4}{5}} \ln \frac{\sqrt[4]{x_1 \cdots x_4}}{x_5} + \beta_3 \sqrt{\frac{3}{4}} \ln \frac{\sqrt[3]{x_1 \cdots x_3}}{x_4} + \\ & + \beta_4 \sqrt{\frac{2}{3}} \ln \frac{\sqrt{x_1 x_2}}{x_3} + \beta_5 \sqrt{\frac{1}{2}} \ln \frac{x_1}{x_2} + \beta_6 \ln(x_1 + \dots + x_6). \end{aligned} \quad (3.16)$$

Parametr β_0 udává hodnotu závisle proměnné y , pokud všechny složky kompozičního vektoru mají stejné hodnoty, a to $x_1 = x_2 = \dots = x_6 = \frac{1}{6}$. Efekty změny závisle proměnné při změně v hodnotách ilr souřadnic jsou stejné jako v modelu s prvním typem součtu. Například pro parametr β_5 dojde při změně podílu v páté bilanci na k -násobek ke změně proměnné y o $\beta_5 \sqrt{\frac{1}{2}} \ln k$. Liší se však efekt změny součtu η , kde při jeho k -násobném růstu dojde ke změně hodnoty y o $\beta_6 \ln k$.

Opět se budeme zabývat modelem s logaritmovanou proměnnou y

$$\ln y_i = \beta_0 + \beta_1 z_{i1} + \cdots + \beta_{D-1} z_{iD-1} + \beta_D \eta_i + u_i = \beta_0 + (\mathbf{z}_i \eta_i) \cdot \boldsymbol{\beta} + u_i, \quad (3.17)$$

který pro případ kompozičního vektoru o šesti složkách vypadá následovně

$$\begin{aligned} \ln y = & \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 \sqrt{\frac{4}{5}} \ln \frac{\sqrt[4]{x_1 \cdots x_4}}{x_5} + \beta_3 \sqrt{\frac{3}{4}} \ln \frac{\sqrt[3]{x_1 \cdots x_3}}{x_4} + \\ & + \beta_4 \sqrt{\frac{2}{3}} \ln \frac{\sqrt{x_1 x_2}}{x_3} + \beta_5 \sqrt{\frac{1}{2}} \ln \frac{x_1}{x_2} + \beta_6 \ln(x_1 + \cdots + x_6). \end{aligned} \quad (3.18)$$

Hodnota proměnné y v případě, že $x_1 = \cdots = x_6 = \frac{1}{6}$, by byla e^{β_0} . U β_1 by při k -násobném růstu podílu v první bilanci a nezměněných hodnotách ostatních bilancí a součtu došlo ke změně y $k^{\beta_1 \sqrt{5/6}}$ krát. Podobně by to bylo také při změnách ostatních bilancí a u součtu, u kterého by při jeho k -násobném vzrůstu došlo ke změně závisle proměnné k^{β_6} krát.

Pokud bychom měli model, kde je součet η na straně závisle proměnné, dostali bychom stejné výsledky jako při užití součtu τ . V roli vysvětlující proměnné však dává odlišné výsledky, než model se součtem tvořeným geometrickým průměrem složek kompozice. Dochází ke změně v interpretaci, která je obtížnější než v prvním případě a absolutní a relativní informace nejsou dostatečně odděleny.[4]

U této formy součtu ovšem není jasné, jak přesně se musí změnit při k -násobné změně bilancí či součtu všechny složky kompozice tak, aby ostatní bilance, popř. součet zůstaly beze změny. Důvodem je odlišná forma součtu, která již neobsahuje součin složek kompozice, ale jejich součet, který je mnohem obtížnější pro interpretaci. Zároveň model s tímto součtem není ekvivalentní s klasickým přístupem. V případě snahy o co nejpřesnější predikce závisle proměnné autoři článku [4] navrhuji použít ten typ součtu, který zajistí co nejpřesnější odhady.[4]

3.3. Podmodely

Jednotlivé modely můžeme použít pro testování, zda je relativní a absolutní informace relevantní pro predikci závisle proměnné y . Pro testování, zda

je vhodnější použít plný model, nebo pouze čistě kompoziční model, slouží hypotéza o $\beta_D = 0$, a tedy že součet není důležitý. Pro testování plného modelu proti modelu se součtem se testuje $\beta_1 = \beta_2 = \dots = \beta_{D-1} = 0$, tedy že nezáleží na kompozičním vektoru.[4]

Dle [4] je však doporučeno vždy ponechat celý model, který má jasnou interpretaci. Odstranění nevýznamných bilancí může vést ke změně v interpretaci bilancí ponechaných v modelu. Jednotlivé podmodely by se tedy měly použít pro testování hypotéz, ne pro výběr modelu. Pomoci můžou také charakteristiky goodness of fit, jako např. AIC, BIC nebo upravený koeficient determinace, které by nám měly říct do jaké míry velikost a kompozice přispívají k předpovědní síle.[4]

3.3.1. Model s ilr souřadnicemi bez součtu

Prvním podmodelem je model, u kterého jsou na straně vysvětlujících proměnných pouze bilance tvořené složkami zkoumaného kompozičního vektoru. Jedná se tedy o model se součtem, kde $\beta_D = 0$, nebo také klasický model, ve kterém $\sum \gamma_i = 0$. U modelu s čistě kompozičním prediktorem je součet nekorelovaný s bilancemi i se závisle proměnnou. Model pro i -té pozorování vypadá následovně

$$y_i = \beta_0 + \beta_1 z_{i1} + \dots + \beta_{D-1} z_{iD-1} + u_i, \quad (3.19)$$

kde \mathbf{z}_i jsou jednotlivé bilance. Pokud nyní opět budeme uvažovat náš model s kompozicí s šesti složkami a k nim příslušných pět bilancí z příkladu 2.3, dostaneme model

$$\begin{aligned} y = & \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 \sqrt{\frac{4}{5}} \ln \frac{\sqrt[4]{x_1 \cdots x_4}}{x_5} + \beta_3 \sqrt{\frac{3}{4}} \ln \frac{\sqrt[3]{x_1 \cdots x_3}}{x_4} + \\ & + \beta_4 \sqrt{\frac{2}{3}} \ln \frac{\sqrt{x_1 x_2}}{x_3} + \beta_5 \sqrt{\frac{1}{2}} \ln \frac{x_1}{x_2}. \end{aligned} \quad (3.20)$$

Zde parametr β_0 udává hodnotu závisle proměnné y , když všechny složky kompozičního vektoru mají totožné hodnoty, a tedy $x_1 = x_2 = \dots = x_6$. Efekt změny y při změně v hodnotě první ilr souřadnice lze opět odvodit podobným způsobem jako u modelů se součtem. $\beta_1 \sqrt{\frac{5}{6}} \ln k$ vyjadřuje, o kolik se změní y ,

vzroste-li hodnota podílu první ilr souřadnice k -krát. Podobně to platí také pro ostatní bilance.[4]

V praktické části práce však opět budeme pracovat s modelem s logaritmovanou hodnotou závisle proměnné. Model tedy bude pro i -té pozorování vypadat následovně

$$\ln y_i = \beta_0 + \beta_1 z_{i1} + \dots + \beta_{D-1} z_{iD-1} + u_i, \quad (3.21)$$

kde \mathbf{z}_i jsou opět bilance a i zde nás zajímá konkrétní model s bilancemi z příkladu 2.3 pro kompoziční vektor o šesti složkách

$$\begin{aligned} \ln y = & \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 \sqrt{\frac{4}{5}} \ln \frac{\sqrt[4]{x_1 \cdots x_4}}{x_5} + \beta_3 \sqrt{\frac{3}{4}} \ln \frac{\sqrt[3]{x_1 \cdots x_3}}{x_4} + \\ & + \beta_4 \sqrt{\frac{2}{3}} \ln \frac{\sqrt{x_1 x_2}}{x_3} + \beta_5 \sqrt{\frac{1}{2}} \ln \frac{x_1}{x_2}. \end{aligned} \quad (3.22)$$

U tohoto modelu odpovídá e^{β_0} hodnotě y v situaci, kdy mají všechny složky kompozičního vektoru stejnou hodnotu. $k^{\beta_1 \sqrt{5/6}}$ vyjadřuje efekt, kolikrát se změnil y , zvýší-li se podíl v první ilr souřadnici k -krát a přitom ostatní bilance zůstanou na stejných hodnotách. Obdobným způsobem by se poté počítaly efekty také pro zbylé souřadnice (viz odvození pro model 3.10). Vzhledem k tomu, že jednotlivé složky se objevují ve všech bilancích, dojde u nich také ke změně. Tato změna však musí být taková, aby se změna složek v čitateli bilance přesně vyrovnala změně složky ve jmenovateli.

Pokud se podíl v první bilanci změnil k -krát, znamená to, že $k \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} = \frac{\sqrt[5]{k x_1 \cdots k x_5}}{x_6}$. Odtud je vidět, že všechny složky v čitateli se změnil k -krát, zatímco složka x_6 zůstane nezměněna. Další možností je například vynásobení prvních pěti složek kompozice číslem $\sqrt[6]{k}$ a složky x_6 číslem $\frac{1}{\sqrt[6]{k^5}}$, jak bylo uvedeno u modelu se součtem τ . Protože v tomto modelu není obsažen součet, existuje zde více možností, jakým způsobem lze dosáhnout totožné změny bilance při zachování hodnot všech ostatních.

Pokud při testování vychází parametr β_D jako významný, je člen vyjadřující absolutní informaci důležitý a měl by být v modelu ponechán. Při vynechání τ

nebo η , pokud jsou nenulové, můžeme dostat vychýlené odhady efektu bilancí, a to ještě ve větší míře, pokud jsou bilance a součet korelované (přitom součet τ je vždy ortonormální na všechny bilance). Navíc bude ovlivněna interpretace parametrů.[4]

3.3.2. Model pouze se součtem

Druhým podmodelem je model, kde na straně vysvětlující proměnné je pouze člen s absolutní informací, neboli součet. Není tedy důležitá relativní informace a model pro i -té pozorování vypadá v závislosti na užití formě součtu následovně

$$y_i = \beta_0 + \beta_D \tau_i + u_i, \text{ popř. } y_i = \beta_0 + \beta_D \eta_i + u_i, \quad (3.23)$$

přičemž pro kompoziční vektor o $D = 6$ složkách jako v předchozích uvažovaných modelech dostaneme

$$y = \beta_0 + \beta_6 \sqrt[6]{\ln \sqrt[6]{x_1 \cdots x_6}}, \text{ popř. } y = \beta_0 + \beta_6 \ln(x_1 + \cdots + x_6). \quad (3.24)$$

Parametr β_0 odpovídá hodnotě y , když v modelu se součtem τ jsou všechny složky kompozičního vektoru rovny 1, u modelu se součtem η v situaci, kdy jsou všechny složky kompozice rovny $\frac{1}{6}$. Při k -násobném vzrůstu u součtu se závisle proměnná změní u prvního z modelů o $\beta_6 \sqrt[6]{\ln k}$, u druhého modelu o $\beta_6 \ln k$.

V praktické části však opět budeme používat modely, ve kterých je závisle proměnná logaritmována. Tyto modely vypadají pro i -té pozorování

$$\ln y_i = \beta_0 + \beta_D \tau_i + u_i, \text{ popř. } \ln y_i = \beta_0 + \beta_D \eta_i + u_i, \quad (3.25)$$

a konkrétní model dále využitý v kapitole 5 poté vypadá následovně

$$\ln y = \beta_0 + \beta_6 \sqrt[6]{\ln \sqrt[6]{x_1 \cdots x_6}}, \text{ popř. } \ln y = \beta_0 + \beta_6 \ln(x_1 + \cdots + x_6). \quad (3.26)$$

Pokud by všechny složky kompozice měly hodnotu 1 u prvního z modelů a $\frac{1}{6}$ u druhého, byla by hodnota závisle proměnné e^{β_0} . Při vzrůstu součtu na k -násobek by se y změnilo $k^{\beta_6 \sqrt[6]{\ln k}}$ krát u modelu s prvním typem součtu a k^{β_6} krát u druhého.

U modelu pouze se součtem však opět není jasné, jak se při jeho změně přesně musí změnit jednotlivé složky kompozice, zda se všechny změní stejně či nikoli, a to ani u modelu se součtem τ . [4]

Kapitola 4

Data o ekologické stopě

V této kapitole budou postupně představena data, která byla použita pro následnou analýzu s využitím regresních modelů zavedených v kapitole 3. Budou vysvětleny základní pojmy důležité pro pochopení významu ekologické stopy a metody jejího výpočtu. Dále budou představeny její jednotlivé části tvořící kompoziční vektor a zvolená závisle proměnná společně s úpravami, které byly na datech provedeny před samotnou analýzou.

4.1. Proč právě ekologická stopa

Globální oteplování a s ním související změny klimatu se v 21. století stávají stále důležitějšími a naléhavějšími tématy. I když oteplování planety je částečně přirozený proces provázející planetu odedávna, bylo zjištěno, že současné oteplování je rychlejší než kdykoli v minulosti. O větší zájem o danou problematiku se zasadila v loňském roce také švédská aktivistka Greta Thunberg, která inspirovala studenty po celém světě ke stávkám snažícím se upozornit na potřebu změny přístupu politiků, ale také veřejnosti k otázce globálního oteplování a změn klimatu. Hlavními důvody tohoto procesu jsou lidské aktivity jako např. spalování fosilních paliv, uhlí, ropy a plynu, a také odlesňování. Kvůli těmto činnostem je vypouštěno do ovzduší obrovské množství CO_2 a následně dochází k oteplování planety. S tím souvisí také pojem uhlíková stopa. Ačkoli se téma uhlíkové stopy často v článcích objevuje samostatně, jedná se pouze o jednu z

částí každoročně měřené celkové ekologické stopy, která zkoumá lidské požadavky na přírodní zdroje a je v podstatě ukazatelem udržitelnosti. V této práci jsou proto použity i zbylé složky ekologické stopy týkající se také odlišných způsobů využívání půdy lidskou činností, aby bylo možné prozkoumat i vlivy těchto částí na zvolené externí proměnné, a to nejen samostatně, ale rovněž v různých kombinacích, získaných vhodným SBP dělením.[11, 31]

Data o ekologické stopě použitá v této práci pochází z internetové stránky Kaggle [15]. Jejich původním autorem je však mezinárodní nezisková organizace Global Footprint Network [9], která každoročně publikuje data o ekologické stopě, *National Footprint Accounts*. Ta jsou volně dostupná na internetu a mohou být kýmkoli použita ke zkoumání lidského dopadu na planetu. Některé státy, např. Spojené Arabské Emiráty, Německo nebo Francie, si společností Global Footprint Network nechávají zpracovat shrnutí svojí ekologické stopy a biokapacity a tyto údaje poté využívají k určení cílů k dosažení udržitelného rozvoje.[5, 18]

Použitý datový soubor obsahuje hodnoty ekologické stopy a biokapacity jednotlivých zemí od roku 1961 do roku 2014 a byl uveřejněn v roce 2018. K datu zveřejnění bývá posledním rokem v datech ten, který je čtyři roky před publikováním souboru, a to kvůli dostupnosti použitých vstupních dat. Dataset obsahuje 274 zemí, a také pozorování popisující souhrnné hodnoty pro celý svět. V práci se dále zkoumají pouze naměřené údaje pro poslední dostupný rok, a to 2014.[18]

Vzhledem k tomu, že společnost Global Footprint Network neustále pracuje na zlepšování výpočtů a zahrnutí dalších produktů a vlivů majících dopad na výši ekologické stopy, dochází s každým zveřejněním nové edice také s přepočítáním všech předchozích zkoumaných let, aby bylo možné výsledky srovnat. V této kapitole bude dále u popisu jednotlivých kategorií ekologické stopy použito označení NFA 2018 odkazující čtenáře přímo na dataset, který je zkoumán v této práci, a který byl publikován v roce 2018. V odlišných vydáních, pocházejících z jiného roku, se mohou zde popsané metody výpočtu ekologické stopy, uvedené v publikaci [17], a zahrnuté produkty více či méně lišit.[18]

4.2. Základní informace o ekologické stopě

Ekologická stopa měří velikost plochy v globálních hektarech, která je potřeba pro vyprodukování daných zdrojů využívaných lidmi či aktivitami v dané zemi v daném roce a k absorpci generovaného odpadu pomocí běžných technologií. Plochou se zde rozumí biologicky produktivní půda, ale také vodní plochy.[17]

V *National Footprint Accounts* se ve výpočtech zohledňuje rozdělení ekologické stopy podle odlišných druhů využití půdy na 6 různých kategorií, a to *cropland*, *grazing land*, *fishing grounds*, *built-up land*, *forest products* a *carbon*. Jednotlivé kategorie budou blíže popsány v podkapitole 4.3.[17]

Dalším důležitým pojmem je biokapacita. Pod tímto pojmem se rozumí množství biologicky produktivní půdy a vodních zdrojů dostupných na území dané země. Biokapacita se počítá pro 5 druhů využití půdy, které jsou totožné jako u ekologické stopy s výjimkou kategorie *carbon* a to z toho důvodu, že biokapacita uhlíku je začleněna společně s *forest products* do *forest land*, protože absorpce uhlíku vytváří požadavky na lesní plochu a pokud by byla uvedena zvlášť, došlo by k jejímu dvojitému započítání. (V datech má tedy proměnná *carbon* pro biokapacitu hodnoty 0 a pro ekologickou stopu tedy proměnná s názvem *forest land* značí ekologickou stopu lesních produktů.)[17]

Při samotném výpočtu ekologické stopy se využívají dva principy - aditivita a ekvivalence, ze které vyplývá použití jednotek globální hektary. Princip aditivity znamená, že celková ekologická stopa je získána ze součtu všech činností vykonávaných lidmi, které vyžadují biologicky produktivní půdu. To poté umožňuje ekologickou stopu srovnat s biokapacitou. Druhý princip vychází z odlišné výnosnosti jednotlivých druhů půdy, tedy schopnosti produkovat zdroje a služby využívané lidmi. Proto je nutné převádět jednotlivé typy půdy na společnou jednotku, kterou jsou globální hektary (gha). Ty pro každý zkoumaný rok reprezentují hektar biologicky produktivní půdy, která má průměrnou světovou biologickou produktivitu. Například globální hektar půdy vhodné pro pěstování plodin je ve skutečnosti menší plocha, než u méně produktivní půdy vhodné pro pastvu, a to z důvodu nižší bioproduktivity u druhého typu půdy.[17]

U jednotlivých zemí lze srovnat hodnoty ekologické stopy a biokapacity. Pokud má země vyšší hodnotu ekologické stopy než biokapacity, mluvíme o ekologickém deficitu. V takovém případě má daná země nedostatek vlastní biokapacity a musí ji importovat odjinud, nebo např. vypouští odpad do atmosféry a jiných společných přírodních zdrojů na planetě. V roce 2014, který bude naším zkoumaným rokem, měla většina zemí ekologický deficit. Výjimku tvořily pouze země v zalesněných oblastech, jako např. země Jižní Ameriky, které mají na svém území deštné pralesy, nebo např. Kanada a jiné severské oblasti. U takových zemí se říká, že mají ekologickou rezervu.[10, 18]

4.3. Části kompozičního vektoru

Jak již bylo uvedeno výše, celková ekologická stopa se dělí na 6 kategorií, a tedy 6 částí kompozičního vektoru. Tyto části jsou *cropland*, *grazing land*, *fishing grounds*, *built-up land*, *forest products* a *carbon*, které budou dále blíže představeny.

4.3.1. Cropland

Proměnná *cropland* popisuje množství globálních hektarů půdy potřebných pro pěstování plodin pro lidskou obživu a jako krmivo pro dobytek. V NFA 2018 se tato stopa počítá z celkových 177 kategorií zemědělských produktů. Pro každý z nich je stopa spočítána jako podíl množství vyprodukované plodiny k celkové ploše sklizené půdy. Jednotlivé produkty lze zařadit do 4 kategorií - plodiny pro přímou lidskou obživu, pro krmení dobytka (např. vojtěška), vlákna (např. bavlna) a plodiny pro ostatní využití (např. tabák). Platí následující vztah mezi proměnnou *cropland* a *grazing land*: při růstu množství plodin užívaných jako krmivo pro dobytek může docházet k poklesu trávy pro pastvu.[17]

4.3.2. Grazing Land

Grazing land popisuje množství globálních hektarů půdy potřebných jako pastva pro nakrmení dobytka, a také pro požadavky na pastvu zahrnutou v obchodovaném zboží. Při výpočtu se začíná určením počtu dobytka v dané zemi a množstvím krmiva potřebným pro jeho nakrmení. Potřebné krmivo je podle předpokladu nejdříve nalezeno v podobě plodin pěstovaných přímo za účelem krmení zvířat, pomocí zbytků plodin, které nelze použít jako potravu pro člověka a jako tráva rostoucí na půdě užívané k pěstování plodin a užívaná po sklizení pouze jako krmivo pro dobytek. Zbývající potřeba krmiva, která nebyla uspokojena předchozími způsoby, je brána jako poskytnutá prostřednictvím pastvin, a tedy *grazing land*. Jedná se tedy o krmivo pro dobytek, které nebylo zahrnuto v předchozím typu půdy, v *cropland*.[\[17\]](#)

4.3.3. Forest Products

Tato část kompozičního vektoru popisuje množství globálních hektarů půdy potřebných pro uspokojení lidské poptávky po lesních produktech. Při určení ekologické stopy produkce (EF_P , viz podkapitola 4.4) se vyčleňují dva typy primárních produktů, a to palivové dřevo a dřevo a celulóza užívané při výrobě dřevěných produktů. V NFA 2018 je tedy sledována produkce těchto dvou typů primárních lesních produktů a produktů z nich odvozených. Ve výpočtech však nejsou zohledněny některé další lesní produkty, jako např. lesní ořechy, byliny a kůra. Neexistují totiž data, která by bylo možné ve výpočtech použít k určení množství sběru těchto přírodnin.[\[17\]](#)

4.3.4. Fishing Grounds

Ekologická stopa *fishing grounds* vyjadřuje globální hektary půdy tvořící ekvivalent k nároku rybolovu na vodní ekosystém, a to takový, aby byl podpořen udržitelný odlov ryb daného státu. Do výpočtů stopy v NFA 2018 je zahrnuto celkem 1941 sladkovodních a mořských druhů ryb, savců, bezobratlých živočichů a vodních rostlin. Počítá se podíl množství primární produkce zkonsumované

vodními druhy za jejich život a odhadu sklízitelné primární produkce na hektar vodní plochy. Sklízitelná primární produkce je založena na globálním odhadu udržitelného odlovu několika vodních druhů. Údaje o udržitelném odlovu jsou převedeny na ekvivalenty primární produkce a poděleny celkovou plochou kontinentálního mořského dna. Stejná kalkulace se používá pro vnitrozemské ryby.[17]

4.3.5. Built-up Land

Tato složka kompozičního vektoru popisuje globální hektary bioproduktivní půdy zastavěné lidskou infrastrukturou. V NFA 2018 se jedná o oblasti vyžadované pro lidská obydlí, průmysl a přepravu. Pojmem bioproduktivní půda je myšleno, že pro lidskou infrastrukturu se používá nejvíce produktivní půda, kterou je *cropland*. Vychází se tedy z předpokladu, že veškerá lidská infrastruktura pokrývá původní půdu vhodnou pro pěstování plodin. Při výpočtech se proto využívá stejný ekvivalenční faktor (viz podkapitola 4.4) jako u první kategorie. Z výše uvedených důvodů je hodnota ekologické stopy (i biokapacity) nadhodnocena při infrastruktuře nacházející se na méně úrodné půdě. Správně by také mělo docházet k započítání této stopy u importovaných a exportovaných produktů vyráběných v továrnách započtených do této kategorie. Protože tomu tak zatím není, je stopa u exportních zemí vyšší než odpovídá skutečnému stavu, zatímco u zemí, kde převažuje import, je stopa nižší. Při výpočtech dochází také k celkovému podhodnocení vzhledem k nemožnosti započítat více rozptýlenou lidskou infrastrukturu, jako např. domy nebo silnice.[17]

4.3.6. Carbon

Uhlíková stopa zahrnuje globální hektary lesa potřebné pro vstřebání emisí CO_2 produkovaných lidskou činností. Do výpočtu se zahrnuje několik kategorií - domácí spalování fosilních paliv, užití elektřiny, uhlík zahrnutý v prodávaném zboží a elektřině, podíl dané země na emisích pocházejících z mezinárodní dopravy a další zdroje nepocházející z fosilních paliv. Uhlíková ekologická stopa tvoří největší podíl z celkové ekologické stopy. Jedná se také o složku, které nejrychleji

roste.[17, 18]

4.4. Metoda výpočtu ekologické stopy

V této podkapitole budou uvedeny základní vzorce užívané pro výpočet ekologické stopy. Pro jednoduchost se budeme zabývat pouze vzorci platnými pro všechny, nebo alespoň většinu typů využití půdy, podrobnější vzorce a vzorce týkající se např. pouze jednoho z typů využití půdy zde uvedeny nebudou. Lze je však najít v [17].

Pod pojmem ekologická stopa je většinou (i v našem případě) myšlena ekologická stopa spotřeby (ecological footprint of consumption). Ta se skládá ze tří částí: ekologická stopa produkce, ekologická stopa exportu a ekologická stopa importu. Vztah mezi těmito složkami a celkovou ekologickou stopou je

$$EF_C = EF_P + EF_I - EF_E, \quad (4.1)$$

kde EF_C je ekologická stopa spotřeby, EF_P produkce a EF_I , resp. EF_E jsou ekologické stopy importu, resp. exportu. Při výpočtu se nejdříve spočítají ekologické stopy spotřeby pro jednotlivé produkty nebo odpady podle výše uvedeného vzorce, které se potom agregují dohromady pro jednotlivé typy využití půdy. Ty lze poté sečíst a získat ekologické stopy jednotlivých zemí, při jejich sečtení poté celého světa.[17]

Pro úplnost uvedeme také výpočet biokapacity pro jednotlivý typ využití půdy (v gha), která se počítá na základě vzorce

$$BC = A \cdot YF \cdot IYF \cdot EQF, \quad (4.2)$$

kde A je oblast pro dané využití půdy ve zkoumané zemi (v národních hektarech, nha), YF je faktor výnosu pro daný typ využití půdy v dané zemi ($wha \cdot nha^{-1}$, kde wha je světový průměrný hektar), IYF je srovnávací faktor výnosu (inter-temporal yield factor, bez jednotek) zahrnující změny v průměrném světovém výnosu stejného typu využití půdy v čase a EQF je ekvivalenční faktor pro daný typ využití půdy ($gha \cdot wha^{-1}$), který slouží pro samotný převod na globální hektary. Ekvivalenční faktory jsou v daném roce stejné pro všechny státy.[17]

Ekologická stopa roční produkce EF_P , stejně jako ekologické stopy importu a export se získávají následujícím způsobem

$$EF_P = \frac{P}{Y_N} \cdot YF \cdot EQF \cdot IYF = \frac{P}{Y_W} \cdot EQF \cdot IYF, \quad (4.3)$$

kde P je množství vyrobeného produktu nebo generovaného odpadu (tun/rok), Y_N průměrný výnos daného státu pro výrobu produktu nebo absorpci odpadu ($t \cdot nha^{-1}$ za rok), YF faktor výnosu pro daný typ využití půdy v dané zemi ($wha \cdot nha^{-1}$), EQF ekvivalenční faktor pro daný typ využití půdy ($gha \cdot wha^{-1}$) a IYF je srovnávací faktor výnosu pro daný typ využití půdy. Druhý z uvedených vztahů plyne ze skutečnosti, že $YF = \frac{Y_N}{Y_W}$, tedy faktor výnosu je podíl domácího a světového výnosu (viz vztah 4.5).[17]

U sekundárních a terciárních produktů, neboli tzv. odvozených produktů, je výpočet jejich výnosu založen na původním primárním produktu a míře extrakce. Ta slouží jako faktor vyjadřující množství primárního produktu zahrnutého v odvozeném produktu. Výnos odvozeného produktu je tedy dán jako

$$Y^D = Y^P \cdot EXTR, \quad (4.4)$$

kde Y^D je výnos odvozeného produktu ($t \cdot ha^{-1}$ za rok), Y^P výnos primárního produktu ($t \cdot ha^{-1}$ za rok) a $EXTR$ míra extrakce (odvozené tuny/tuny primárního produktu).[17]

Dále je potřeba spočítat výnosový faktor pro daný typ půdy. Ten se počítá odlišně při existenci pouze jednoho primárního produktu a při více primárních produktech. Zde si uvedeme pouze první příklad, kterým se míra výnosu počítá pro *grazing land, fishing grounds a forest products*

$$YF_N^L = \frac{Y_N^L}{Y_W^L}, \quad (4.5)$$

kde YF_N^L je výnosový faktor pro danou zemi a typ půdy ($wha \cdot nha^{-1}$), Y_N^L , resp. Y_W^L výnos pro danou zemi a typ půdy ($t \cdot nha^{-1}$), resp. průměrný světový výnos pro daný typ půdy ($t \cdot wha^{-1}$).[17]

Srovnávací faktor výnosu IYF, zachycující změny v průměrných ročních výnosech jednotlivých typů využití půdy v čase, se počítá z následujícího vzorce

$$IYF_{w,j} = \frac{\sum_i \frac{P_{w,i,j}}{Y_{w,i,b}}}{\sum_i \frac{P_{w,i,j}}{Y_{w,i,j}}}, \quad (4.6)$$

kde $IYF_{w,j}$ je průměrný světový srovnávací faktor výnosu v roce j , $P_{w,i,j}$ světové množství sklizeného produktu nebo vypuštěného CO_2 u produktu i v j -tém roce, $Y_{w,i,b}$ světový výnos produktu i v základním roce b a $Y_{w,i,j}$ světový výnos produktu i v j -tém roce.[17]

Poslední veličinou ze vzorce pro výpočet ekologické stopy produkce je ekvivalenční faktor. Ten pro různé typy půdy vyjadřuje relativní produktivitu průměrných světových hektarů. Při výpočtech se vychází z indexů vhodnosti a informací o aktuálních plochách typů využití půdy jako *cropland*, *forest land* a *grazing land*. Indexy vhodnosti dělí půdu do celkem pěti kategorií od nejvíce vhodných (Very Suitable) s indexem 0,9 po nevhodné (Not Suitable) s indexem 0,1. Podrobnější rozpis indexů a organizace údaje vydávající lze najít v [17]. Při výpočtech se poté vychází z předpokladu, že nejvíce produktivní půda je použita k nejproduktivnějšímu účelu, a tedy nejvíce produktivní půda bude využita pro růst plodin (*cropland*), poté jako lesní plocha a nejméně produktivní bude sloužit pro růst pastvy pro dobytek (*grazing land*). *Built-up land* je brána jako stejně produktivní jako *cropland* vzhledem k již zmíněnému předpokladu, že veškerá lidská infrastruktura se nachází na bývalé půdě pro pěstování plodin. Ekvivalenční faktor se určí poměrem, kde v čitateli se nachází průměrný index vhodnosti daného typu půdy a ve jmenovateli průměrný index vhodnosti pro všechny typy půdy.[17]

4.5. Použitá závisle proměnná

V následující kapitole bude při aplikaci regresních modelů použit na straně závisle proměnné ukazatel HDP na 1 obyvatele za rok 2014 (v amerických dolarech). HDP popisuje celkovou hodnotu zboží a služeb nově vyrobených v dané zemi ve sledovaném období. Popisuje v podstatě výkonnost dané ekonomiky.[12,

14]

Existují tři metody výpočtu HDP, a to výrobní metoda, výdajová metoda a důchodová metoda. Správně by všechny tyto metody měly dávat totožné výsledky, což ale v praxi platit nemusí, protože ne vždy jsou dostupná všechna data potřebná pro získání kompletního obrazu celé ekonomiky.[32]

Uvedeme si zde pouze výdajovou metodu, která je nejvíce známým způsobem výpočtu HDP a počítá se následovně[12]

$$\begin{aligned} HDP = & \text{Výdaje na konečnou spotřebu} + \text{Tvorba hrubého kapitálu} + \\ & + \text{Vývoz výrobků a služeb} - \text{Dovoz výrobků a služeb.} \end{aligned} \quad (4.7)$$

HDP spočítané v cenách platných v daném roce se nazývá nominální HDP. Pro srovnání vývoje HDP v průběhu let je však potřeba ukazatel přepočítat k vybranému základnímu roku, aby došlo k odstranění vlivu inflace. Takové HDP se poté nazývá reálné HDP. V našem případě, kdy nepotřebujeme údaje z více let za účelem srovnání, byl zvolen ukazatel HDP v běžných cenách v amerických dolarech.[13]

K následnému modelování bylo zvoleno HDP na jednoho obyvatele, které je získáno vydělením nominálního HDP počtem obyvatel dané země. Tento počet obyvatel je většinou průměrný počet obyvatel dané země v daném roce. Získáme tak vlastně průměrný ekonomický výstup na jednoho obyvatele země, nebo také průměrný příjem. Nejvyšší HDP na osobu mají malé bohaté země a více rozvinuté industriální země. Data k HDP byla získána ze stránek World Bank Open Data [8].[13, 14]

4.5.1. Vztah k ekologické stopě

V mnoha odborných článcích (např. [16, 19]) byl již zkoumán vliv ekonomických ukazatelů, jako je např. HDP na 1 obyvatele, na výši ekologické stopy. Bylo dokázáno, že ekologická stopa daného státu je funkcí úrovně jeho ekonomického rozvoje, nerovnosti příjmů, přesunu obyvatel z venkova do měst a závislosti na exportu. Z výsledků studií také vyšlo, že ukazatel HDP má významný

pozitivní vliv na výši ekologické stopy. Naším cílem je však modelovat závislost HDP na struktuře ekologické stopy. Dá se předpokládat, že výše HDP bude na ekologické stopě záviset, jak vyplývá z výše zmíněných výsledků. Zůstává však otázkou, zda má smysl zkoumat pouze celkovou výši ekologické stopy, nebo také její části. Zajímá nás také, které části mají největší vliv, popřípadě zda mají významný vliv všechny složky, nebo zda je pro HDP určující pouze nejnámější část ekologické stopy, kterou je uhlíková stopa.[16, 19]

4.6. Úprava dat

Z datového souboru obsahujícího hodnoty ekologické stopy i biokapacity jednotlivých zemí od roku 1961 až po rok 2014 byla vyfiltrována pouze data týkající se ekologické stopy pro poslední dostupný rok 2014. Pro následné použití v regresních modelech byly vybrány hodnoty ekologické stopy jednotlivých států na jednoho obyvatele, aby byl odstraněn vliv velikosti jednotlivých zemí. To je důležité kvůli použití modelu se součtem, u čistě kompozičního modelu by tento krok nebyl potřeba.

V datech byly dále provedeny následující úpravy. Po získání dat k závislé proměnné, kterou je HDP na jednoho obyvatele, bylo zjištěno, že ne pro všechny státy, pro které jsou dostupné hodnoty ekologické stopy, jsou k dispozici také odpovídající hodnoty vysvětlované proměnné. Proto byly tyto státy z dat odstraněny. Těchto států bylo celkově pět. Vzhledem k tomu, že rozdělení HDP je značně sešikmené z důvodu velkého množství velmi nízkých hodnot a pouze málo vysokých, nejsou v modelech použity původní hodnoty HDP, ale hodnoty logaritmované.

Dalším problémem, který se v datech vyskytoval, byly nulové hodnoty v kompozičním vektoru. Tyto hodnoty se vyskytovaly pouze u dvou zemí, a to v části *built-up land*. Jednalo se o Spojené Arabské Emiráty a Francouzskou Polynésii. U první z těchto dvou zemí je nulová hodnota dle [30] způsobena umístěním většiny lidské infrastruktury v poušti místo na úrodně půdě. Přitom ve výpočtech ekologické stopy v části *built-up land* se předpokládá, že lidská infrastruktura se

nachází právě na bývalé *cropland*. Není přitom jasné, zda se v hodnotách jedná skutečně o nulovou plochu, nebo zda byly např. hodnoty pod určitou prahovou hodnotou zaokrouhleny na nulu. Z tohoto důvodu, a také proto, že náhrada nulových hodnot v kompozičních datech není předmětem této práce, byla tato pozorování taktéž odstraněna. Metody týkající se problematiky nahrazování nulových hodnot v kompozičních datech lze najít např. v [7]. (K úpravám a nahrání dat byly použity funkce z balíčků `readxl` [34], `dplyr` [35] a `tidyr` [36].)

Po provedených úpravách a odstraněných pozorováních v datovém souboru zbylo celkově 158 zemí, které byly použity v regresních modelech uvedených v následující kapitole.

Kapitola 5

Analýza dat

V této kapitole budou postupně použity regresní modely představené v kapitole 3. V roli závisle proměnné bude využit ukazatel HDP za rok 2014 na jednoho obyvatele v amerických dolarech, který budeme uvažovat ve formě logaritmu. Kvalita jednotlivých modelů bude hodnocena na základě koeficientu determinace, upraveného koeficientu determinace a Akaikeho informačního kritéria, o nichž se lze více dočíst např. v [1].

5.1. Nezávisle proměnné

Na straně vysvětlujících proměnných je ve všech následujících modelech kompoziční vektor \mathbf{x} tvořený jednotlivými složkami ekologické stopy, blíže popsány v kapitole 4. Tyto části jsou následující:

- $x_1 \dots$ cropland (ekologická stopa půdy pro pěstování plodin),
- $x_2 \dots$ grazing land (ekol. stopa pastvin),
- $x_3 \dots$ forest products (ekol. stopa lesních produktů),
- $x_4 \dots$ fishing grounds (ekol. stopa oblastí pro rybolov),
- $x_5 \dots$ built-up land (ekol. stopa lidské infrastruktury),
- $x_6 \dots$ carbon (uhlíková stopa).

Tyto složky jsou v závislosti na použitém modelu buď logaritmovány, popřípadě převedeny na ortonormální souřadnice. Celková úroveň ekologické stopy je zachována prostřednictvím součtu v podobě logaritmu geometrického průměru slo-

žek kompozice, nebo logaritmu jejich součtu, značené v kapitole 3 jako τ , resp. η .

5.2. Klasický přístup

Prvním použitým modelem je klasický přístup 3.4, kde jsou na straně vysvětlujících proměnných logaritmy jednotlivých částí ekologické stopy

$$\ln y = \gamma_0 + \gamma_1 \ln(x_1) + \gamma_2 \ln(x_2) + \gamma_3 \ln(x_3) + \gamma_4 \ln(x_4) + \gamma_5 \ln(x_5) + \gamma_6 \ln(x_6).$$

Tento model nemá žádný člen samostatně vyjadřující absolutní informaci, neboť všechny koeficienty již nesou současně absolutní i relativní informaci.

Call:

```
lm(formula = log(hdp) ~ log(crop_land) + log(grazing_land) +
    log(forest_land) + log(fishing_ground) + log(built_up_land) +
    log(carbon), data = ef)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|----------|----------|---------|---------|---------|
| -1.54984 | -0.42541 | 0.02513 | 0.34414 | 1.63143 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|---------------------|----------|------------|---------|--------------|
| (Intercept) | 10.15559 | 0.22140 | 45.870 | < 2e-16 *** |
| log(crop_land) | 0.16769 | 0.12320 | 1.361 | 0.175505 |
| log(grazing_land) | 0.14774 | 0.03917 | 3.772 | 0.000232 *** |
| log(forest_land) | 0.07959 | 0.06168 | 1.290 | 0.198870 |
| log(fishing_ground) | 0.16415 | 0.03847 | 4.267 | 3.48e-05 *** |
| log(built_up_land) | 0.16385 | 0.06084 | 2.693 | 0.007883 ** |
| log(carbon) | 0.82110 | 0.04475 | 18.348 | < 2e-16 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5576 on 151 degrees of freedom

Multiple R-squared: 0.8618, Adjusted R-squared: 0.8563

F-statistic: 157 on 6 and 151 DF, p-value: < 2.2e-16

U tohoto modelu jsou významné koeficienty všech částí ekologické stopy kromě *cropland* a *forest land*, přičemž všechny jsou kladné. Znamená to tedy, že at

vzroste jakákoli část ekologické stopy při konstantní úrovni ostatních, dojde ke zvýšení HDP. Tento efekt si můžeme ukázat například na složce *carbon*. Pokud by uhlíková stopa vzrostla na dvojnásobek, zvýšilo by se HDP asi o 77 %

$$\frac{HDP'}{HDP} = 2^{0,82110} = 1,76675,$$

kde HDP značí původní úroveň ukazatele a HDP' značí novou úroveň při změně uhlíkové stopy. Stejným způsobem, představeným v kapitole 3, lze vypočítat také efekty u ostatních složek. U těch však při růstu těchto částí nedochází k tak výrazné změně HDP. Nejvyšší změna u dalších významných složek by při stejném růstu nastala u ekologické stopy *fishing grounds* a *built-up land*, a to nárůst asi o 12 %. U *grazing land* by HDP vzrostlo pouze o 11 %.

Model má koeficient determinace na hodnotě 0,8618, upravený koeficient 0,8563. Akaikeho informační kritérium je 272,6188. Vypadá to tedy, že model je poměrně kvalitní. Nezohledňuje ale strukturu ekologické stopy a vztahy mezi jejími jednotlivými složkami.

5.3. Modely s ilr souřadnicemi

Abychom zachovali informaci o struktuře ekologické stopy, vyjádříme v následujících modelech její jednotlivé složky v ilr souřadnicích zavedených v příkladu 2.3, ze kterých bude následně vycházet interpretace výsledných modelů

$$ilr_1 = \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 x_2 x_3 x_4 x_5}}{x_6}$$

$$ilr_2 = \sqrt{\frac{4}{5}} \ln \frac{\sqrt[4]{x_1 x_2 x_3 x_4}}{x_5}$$

$$ilr_3 = \sqrt{\frac{3}{4}} \ln \frac{\sqrt[3]{x_1 x_2 x_3}}{x_4}$$

$$ilr_4 = \sqrt{\frac{2}{3}} \ln \frac{\sqrt{x_1 x_2}}{x_3}$$

$$ilr_5 = \sqrt{\frac{1}{2}} \ln \frac{x_1}{x_2}.$$

Pro připomenutí uvedme, že první souřadnice poskytuje srovnání uhlíkové stopy se všemi ostatními částmi ekologické stopy. Další ze souřadnic je tvořena podílem všech „přírodních“ složek ekologické stopy ke stopě lidské infrastruktury. U třetí souřadnice jsou poté odděleny všechny suchozemské složky „přírodních“ složek stopy od vodní složky. Čtvrtá souřadnice srovnává půdu využívanou pro obživu člověka a dobytka s lesními oblastmi. Poslední souřadnice vyjadřuje, zda převládá ekologická stopa půdy pro pěstování plodin, nebo pastvin. (K výpočtu souřadnic v R byla využita funkce `balances()` z balíčku `robCompositions` [28].)

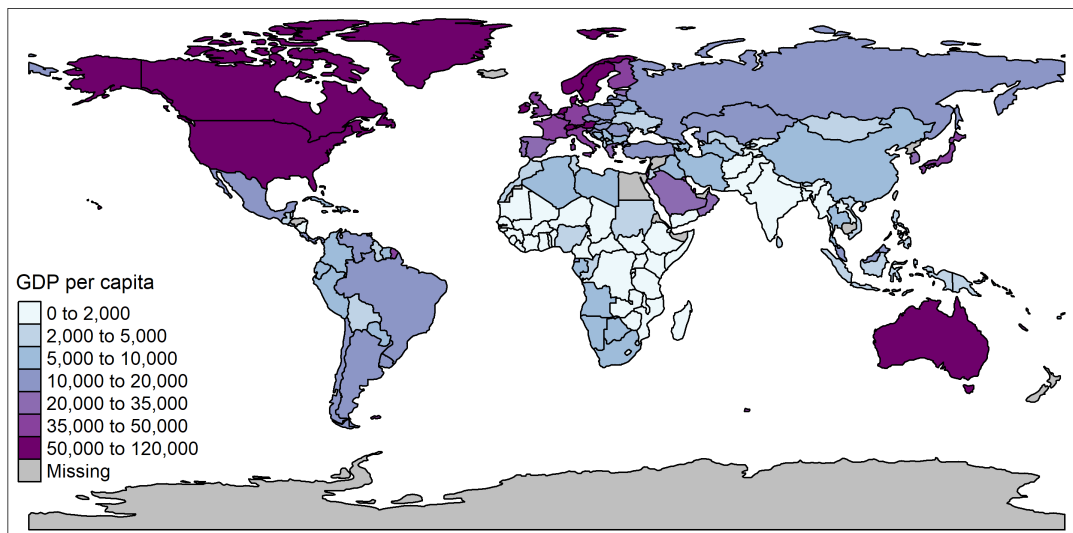
Z důvodu přehlednosti a snadnější orientace v jednotlivých bilancích jsou v R v modelech využity pro jejich značení vždy zkratky odkazující na složky kompozice, ze kterých jsou složeny. Například první souřadnice je značena jako `cr.gr.fo.fi.bu.ca`, protože v čitateli obsahuje stopu půdy pro pěstování plodin (cropland), pastvin (grazing land), lesních produktů (forest products), rybolovu (fishing grounds) a stopu lidské infrastruktury (built-up land). Ve jmenovateli bilance je uhlíková stopa (carbon). V modelech se součtem je poté součet τ značen jako `souc_gp`, zatímco součet η jako `souc_sum`.

5.3.1. Grafické srovnání závisle a nezávisle proměnných

Protože se data o ekologické stopě, a proto také výsledné souřadnice, vztahují k jednotlivým státům, je možné pro znázornění využít zakreslení souřadnic do mapy světa. Do mapy lze zakreslit také hodnoty HDP na 1 obyvatele. To nám umožní jeho srovnání s hodnotami η souřadnic a se součtem. (K vytvoření map byly použity balíčky `sp` [25], `tmap` [29], `rnaturalearth` [27], `stringr` [33] a `RColorBrewer` [21].)

Z mapy 5.1 vidíme, že nejvyšší hodnoty HDP na 1 obyvatele mají USA, Kanada, Austrálie a evropské státy, přitom v Evropě dosahují nejvyšších hodnot skandinávské země kromě Finska, dále poté Rakousko, Švýcarsko, Nizozemsko a Irsko. Nejnižší hodnoty mají země jižní Asie a většina afrických zemí.

Na následujících mapách jsou vykresleny všechny státy z našeho datasetu,

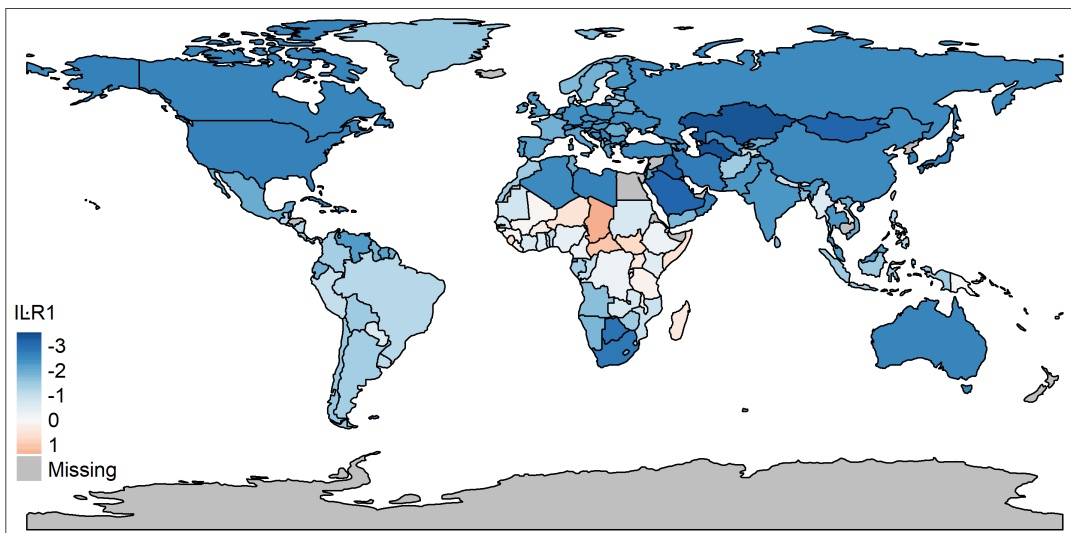


Obrázek 5.1: Mapa s HDP na 1 obyvatele v amerických dolarech v roce 2014

kteří zbyly po úpravách uvedených v kapitole 4, a to postupně pro všechny bilance a posléze také pro členy vyjadřující absolutní velikost ekologické stopy, kterými jsou postupně oba typy součtů.

U mapy 5.2 je zobrazena první ilr souřadnice, která představuje podíl ekologické stopy půdy pro pěstování plodin, pastvy, lesních produktů, oblastí pro rybolov a lidské infrastruktury proti uhlíkové stopě. Červená barva značí převahu prvních pěti složek ekologické stopy nad uhlíkovou, a naopak modrá barva značí převahu uhlíkové stopy. Je vidět, že u většiny zemí převažuje uhlíková stopa nad ostatními složkami. Největší převahy je dosaženo u Kazachstánu a Turkmenistánu. Další státy s výraznější převahou uhlíku jsou např. Saudská Arábie, Irák a Mongolsko. Přibližně vyrovnané hodnoty jsou např. u států Papua-Nová Guinea, Mali, Guinea, Tanzánie a Malawi. Převaha zbylých složek ekologické stopy k uhlíku je pouze u několika afrických zemí, přičemž není příliš výrazná. Jedná se především o státy jako Čad a Středoafriická republika, u kterých je převaha čitatele nejvyšší. Je vidět, že nejvyšší podíly uhlíkové stopy ku ostatním složkám mají evropské a asijské státy společně se severní Amerikou a Austrálií. Nižší hodnoty mají naopak státy Jižní Ameriky a Afriky.

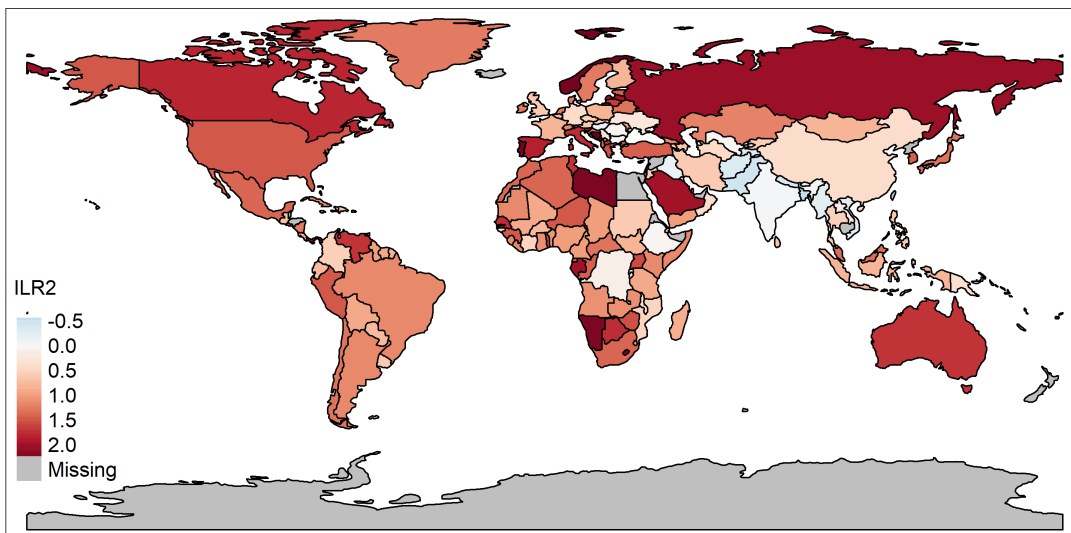
Ze srovnání mapy 5.2 s 5.1 vyplývá, že u velkého množství zemí, u kterých



Obrázek 5.2: Mapa s první ilr souřadnicí vyjadřující podíl prvních pěti složek ek. stopy na uhlíkové stopě

převládá uhlíková stopa, je také vysoká hodnota HDP. Jedná se mimo jiné o Kanadu, USA, Austrálii, Saudskou Arábii, Omán a u evropských zemí například o Švýcarsko nebo Německo. U Kazachstánu, Turkmenistánu, Mongolska a Pákistánu je naopak nízké HDP, i když uhlíková stopa těchto zemí je v porovnání s ostatními složkami vysoká. U zemí s převahou složek v čitateli bilance jsou vždy nízké hodnoty HDP.

U druhé souřadnice je z mapy 5.3 na první pohled vidět, že zde převládají kladné hodnoty, což znamená, že u většiny států je převaha prvních čtyř složek ekologické stopy (stopy půdy pro pěstování plodin, pastvin, lesních produktů a oblastí pro rybolov) ke stopě lidské infrastruktury. Nejvyšších hodnot je přitom dosaženo pro státy Belize (6,0798), Trinidad a Tobago (4,7322) a Svatá Lucie (4,0578), které byly z mapy odstraněny, aby byly vidět rozdíly mezi ostatními státy, které mají mnohem nižší hodnoty. Další státy s vyššími hodnotami složek v čitateli jsou např. Rusko, Kanada, Libye, Namibie, Portugalsko a Norsko. Přibližně nulových hodnot, a tedy vyrovnanosti složek v čitateli a jmenovateli souřadnice, je dosaženo u států Indie, Etiopie a Demokratická republika Kongo. Mezi státy, u kterých převažuje stopa lidské infrastruktury, patří třeba Pákistán,

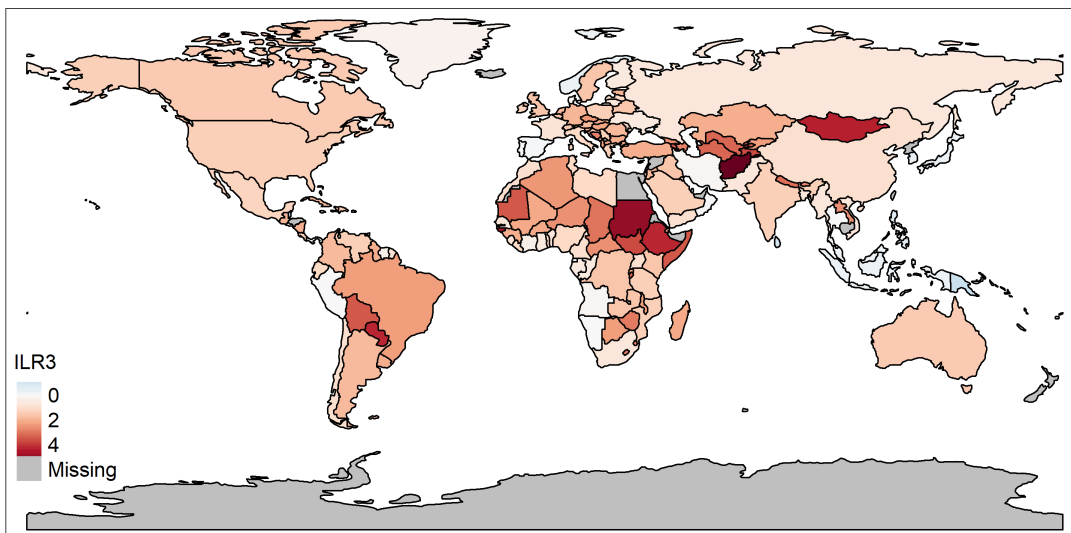


Obrázek 5.3: Mapa s druhou ilr souřadnicí tvořenou podílem ek. stop *cropland*, *grazing land*, *forest products* a *fishing grounds* na *built-up land* po odstranění států Belize, Trinidad a Tobago a Svatá Lucie

Afghánistán, Bangladéš nebo Myanmar. Jedná se tedy pouze o státy jižní Asie.

Opět lze srovnat mapu 5.3 s mapou 5.1. Země, které mají převahu stopy lidské infrastruktury, mají všechny nízkou hodnotu HDP. Mezi státy, u kterých převažují složky v čitateli, nemají všechny stejný vztah k HDP. Kanada a Norsko mají například vysoké HDP, zatímco Libye a Bosna a Hercegovina nízké, přičemž všechny tři výše jmenované státy s nejvyššími hodnotami složek v čitateli mají nízké hodnoty HDP.

U mapy 5.4 je opět převaha kladných hodnot. U většiny zemí je tedy vyšší podíl prvních tří složek ekologické stopy (stopy půdy pro pěstování plodin, pastvin a lesních produktů) než ekologické stopy oblastí pro rybolov. Záporné hodnoty se nacházejí pouze u pár států, přičemž nejvyšší dominance *fishing grounds* je u Papuy-Nové Guiney. Další státy s méně výraznými zápornými hodnotami souřadnice jsou např. Norsko, Španělsko a Filipíny. Vyrovnané hodnoty jsou u zemí Namíbie, Peru, Angola, Omán, Írán a Thajsko. Země s vysokou převahou složek v čitateli souřadnice jsou např. Afghánistán, u kterého je dominance těchto složek největší, dále potom Mongolsko, Súdán, Etiopie a Paraguay.

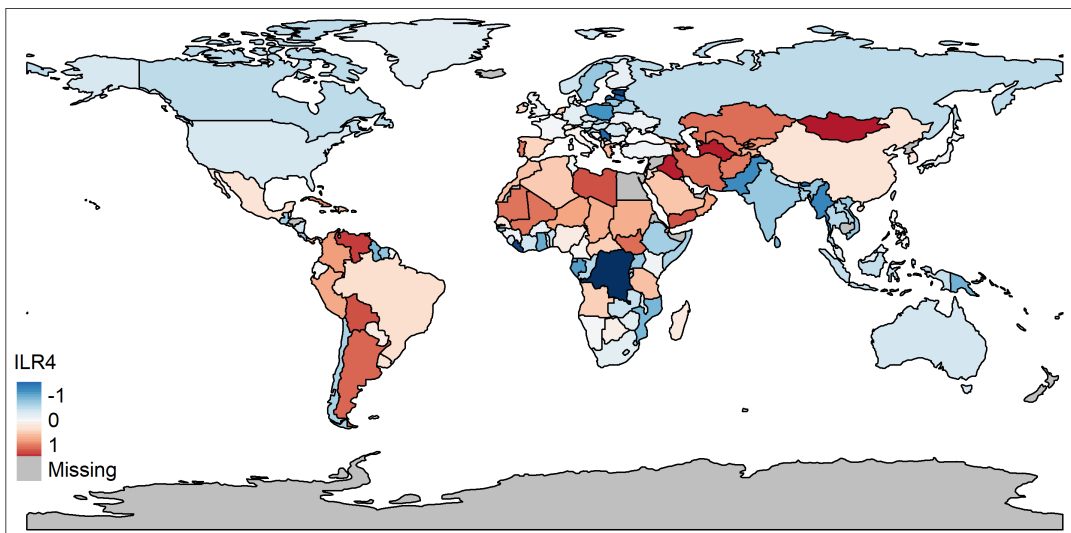


Obrázek 5.4: Mapa se třetí ilr souřadnicí tvořenou podílem složek *cropland*, *grazing land* a *forest products* na *fishing grounds*

Mezi státy, které mají převahu ekologické stopy rybolovu, mají nízké HDP Papua-Nová Guinea, nebo Šrí Lanka. Vysoké HDP mají naopak Japonsko, Jižní Korea a Norsko. Státy, které mají nejvyšší hodnotu složek v čitateli, mají nízké hodnoty HDP. Jsou to např. Bolívie, Mongolsko, Afghánistán a Etiopie. Státy s převahou čitatele a vysokým HDP jsou např. USA, Kanada, Austrálie a evropské státy.

Čtvrtá souřadnice u mapy 5.5 má na první pohled větší rozmanitost hodnot pro jednotlivé státy než u dvou předcházejících. Záporné hodnoty značí převahu ekologické stopy lesních produktů na stopách půdy pro pěstování plodin a pastvin. Tato převaha je nejvíce výrazná u Demokratické republiky Kongo, Estonska a Libérie. Přibližně vyrovnané hodnoty obou skupin stop jsou např. u Turecka, Francie, Velké Británie, Ukrajiny, Japonska, Namibie, Keni a Ekvádoru. Dominance složek v čitateli je poté nejvýraznější např. u Mongolska, Turkmenistánu a Iráku. Je vidět, že kladné hodnoty nejvíce převažují u států Jižní Ameriky, Afriky a u velké části Asie.

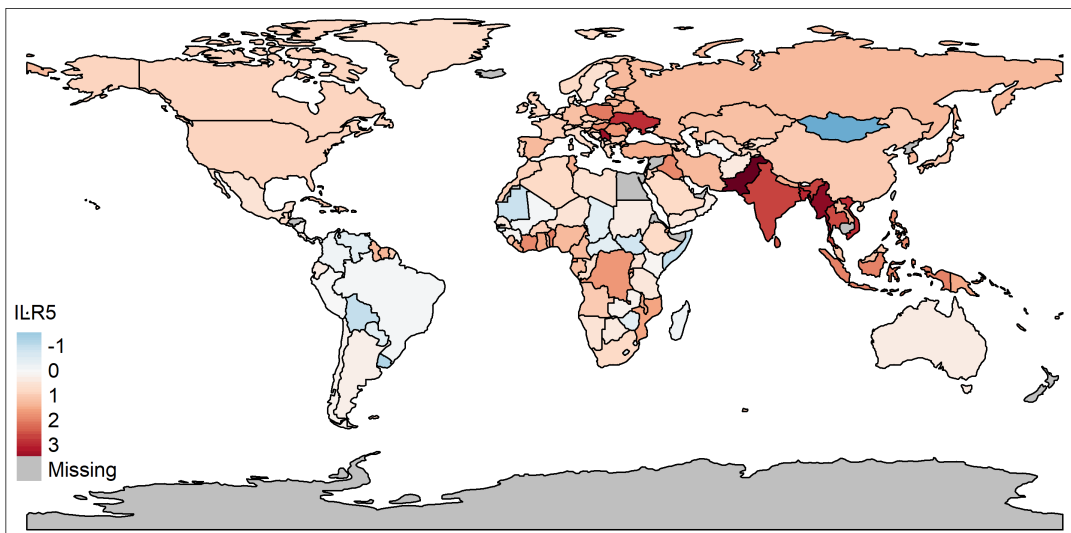
Při srovnání mapy čtvrté souřadnice 5.5 s mapou HDP 5.1 opět platí, že u států s nejvyššími převahami čitatele jsou nízké hodnoty HDP. Některé státy



Obrázek 5.5: Mapa se čtvrtou ilr souřadnicí tvořenou podílem *cropland* a *grazing land* na *forest products*

s převahou *cropland* a *grazing land*, které mají zároveň vysoké HDP, jsou Portugalsko, Španělsko, Nizozemsko a Omán. Státy s vysokými hodnotami těchto částí ekologické stopy a nízkým HDP jsou Mongolsko, Libye, Jemen, Uzbekistán a Bolívie. Státy Estonsko, USA, Kanada a země Skandinávie mají převahu ekologické stopy lesních produktů a současně vysoké HDP. Nízké HDP a převahu složky ve jmenovateli bilance mají např. Pákistán, Myanmar a Demokratická Republika Kongo.

Pátá souřadnice na mapě 5.6 srovnává dvě první složky kompozičního vektoru, a to ekologickou stopu půdy pro pěstování plodin a stopu pastvin. Záporné hodnoty tedy značí dominanci stopy pastvin a nastávají pouze u malého množství států. Nejvýraznější je tento jev u Mongolska. Mezi další takové země poté patří například Bolívie, Mauritanie a Somálsko. Přibližně vyrovnaný podíl těchto složek je u Peru, Brazílie, Turkmenistánu a Ománu. Část *cropland* poté dominuje nejvýrazněji u Pákistánu, Myanmaru a Srbska. U evropských a asijských zemí je povětšinou kladná hodnota souřadnice, zatímco u Jižní Ameriky je hodně států s mírně zápornými hodnotami či s rovnováhou složek. U afrických zemí poté převládají kladné hodnoty, ale nachází se zde také několik zemí se zápornými.



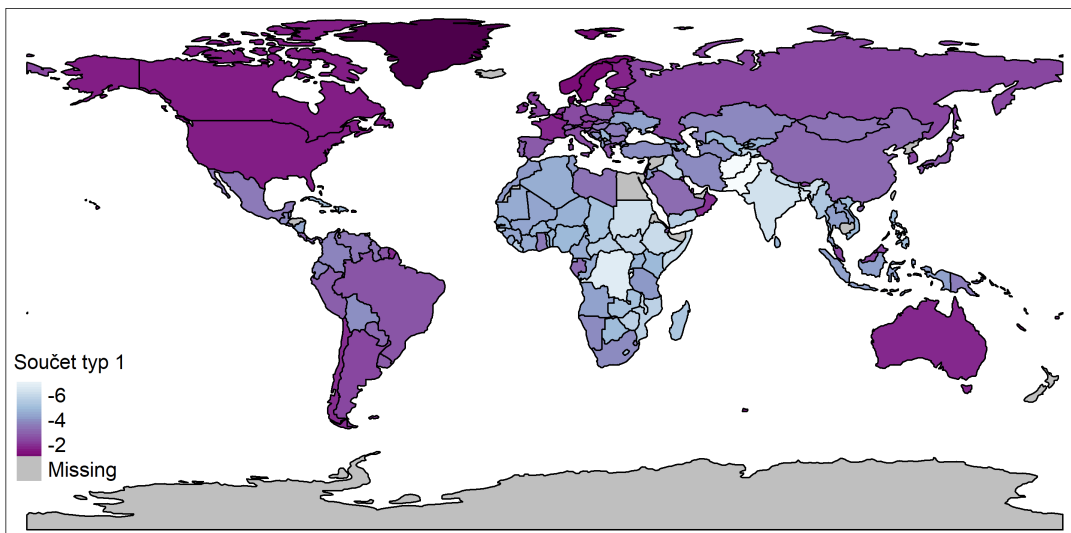
Obrázek 5.6: Mapa s pátou ilr souřadnicí tvořenou podílem *cropland* a *grazing land*

Ze srovnání map 5.6 a 5.1 lze vyvodit, že země s převahou ekologické stopy pastvin mají vždy nízké HDP. Jedná se například o Mongolsko a Bolívii. Země s větším podílem stopy půdy pro pěstování plodin a vysokým HDP jsou např. Kanada, USA a evropské státy. U hodně vysokých hodnot poměru *cropland* ku *grazing land* však mají státy také nízké HDP. Je to vidět například u Indie, Pákistánu nebo Myanmaru.

Nyní budou uvedeny mapy, ve kterých budou vykresleny hodnoty součtu, a to postupně prvního typu součtu tvořeného logaritmem geometrického průměru složek kompozice, τ , následně poté ve tvaru logaritmu součtu složek, η .

Jak je vidět z mapy 5.7, jsou u součtu τ všechny hodnoty ekologické stopy záporné, přičemž platí, že čím vyšší je hodnota součtu (a tedy čím více se součet blíží nule), tím je stopa dané země vyšší. Nejlépe jsou na tom tedy jihoasijské a africké země. Z nich mají nejnižší ekologickou stopu Pákistán a Afghánistán. Nejvyšší ekologickou stopu lze najít u Dánska, dále potom u amerických zemí, dalších evropských zemích a v Austrálii.

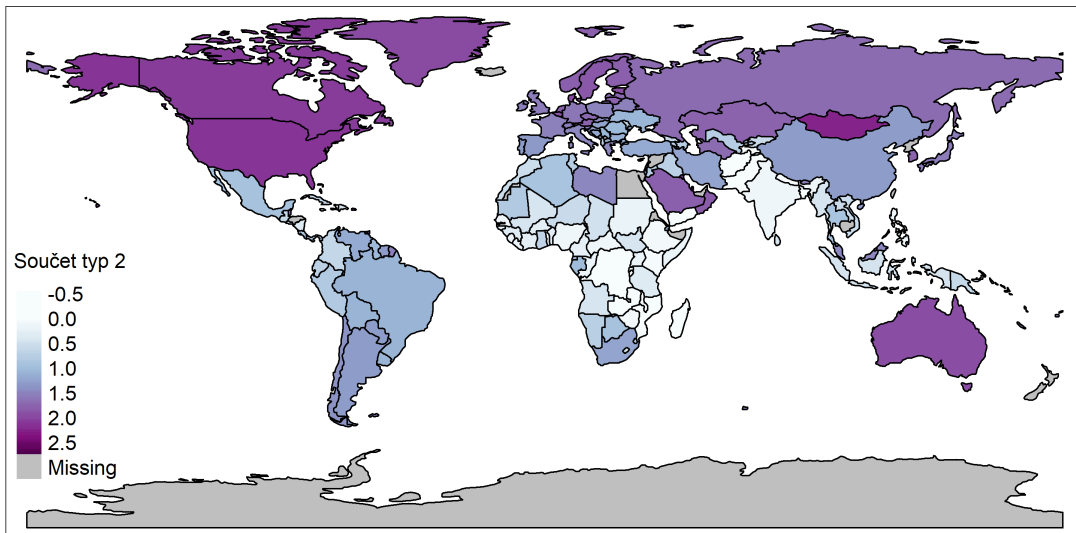
U druhé formy součtu z obr. 5.8, který dosahuje téměř samých kladných hodnot, je ekologická stopa tím větší, čím vyšší je hodnota součtu. Je vidět, že



Obrázek 5.7: Mapa s prvním typem součtu τ , ve formě logaritmu geometrického průměru složek kompozičního vektoru

výsledky jsou v podstatě stejné jako v předchozím případě. U jihoamerických zemí však hodnoty ekologické stopy patřily u předchozího typu součtu k nejvyšším, zatímco zde se jedná spíše o hodnoty ve středu škály. Nejvyšší ekologickou stopu zde mají Spojené státy americké a Kanada, dále Evropa, Austrálie a Mongolsko. Nejnižší hodnoty má Afrika a jižní Asie.

Při srovnání mapy HDP s předchozími dvěma mapami součtů, je vidět, že lze předpokládat kladný vztah mezi výší HDP a celkovou velikostí ekologické stopy vyjádřenou prostřednictvím součtů. Státy, které mají vysokou ekonomickou úroveň by tedy měly dosahovat také vysokých požadavků na přírodní zdroje pro produkci výrobků a služeb a pro eliminaci následně vzniklých odpadů, a měly by tedy mít vyšší ekologickou stopu.



Obrázek 5.8: Mapa s druhým typem součtu η , ve formě logaritmu součtu složek kompozičního vektoru

5.3.2. Model bez součtu

Následující model je model 3.22 s ilr souřadnicemi bez součtového členu

$$\ln y = \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 \sqrt{\frac{4}{5}} \ln \frac{\sqrt[4]{x_1 \cdots x_4}}{x_5} + \beta_3 \sqrt{\frac{3}{4}} \ln \frac{\sqrt[3]{x_1 \cdots x_3}}{x_4} + \beta_4 \sqrt{\frac{2}{3}} \ln \frac{\sqrt{x_1 x_2}}{x_3} + \beta_5 \sqrt{\frac{1}{2}} \ln \frac{x_1}{x_2}.$$

Na straně nezávisle proměnných jsou tedy použity jednotlivé části ekologické stopy, transformované do ilr souřadnic. Rozdíl oproti minulému modelu je tedy v tom, že modelujeme HDP pomocí relativní struktury ekologické stopy. Díky vytvořeným ilr souřadnicím se také nezaměřujeme na samostatné části ekologické stopy, ale na jejich kombinace dané zvoleným SBP dělením.

Call:

```
lm(formula = log(hdp) ~ cr.gr.fo.fi.bu_ca + cr.gr.fo.fi_bu +
    cr.gr.fo_fi + cr.gr_fo + cr_gr, data = df_ilr)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|--------|--------|
| -2.2740 | -0.5681 | -0.1613 | 0.5698 | 2.1378 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|-------------------|----------|------------|---------|----------|-----|
| (Intercept) | 8.29820 | 0.28744 | 28.869 | < 2e-16 | *** |
| cr.gr.fo.fi.bu_ca | -1.08780 | 0.07433 | -14.634 | < 2e-16 | *** |
| cr.gr.fo.fi_bu | -0.22687 | 0.10812 | -2.098 | 0.03753 | * |
| cr.gr.fo_fi | -0.41252 | 0.07635 | -5.403 | 2.48e-07 | *** |
| cr.gr_fo | -0.33936 | 0.11702 | -2.900 | 0.00429 | ** |
| cr_gr | -0.72047 | 0.12165 | -5.922 | 2.04e-08 | *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8748 on 152 degrees of freedom

Multiple R-squared: 0.6576, Adjusted R-squared: 0.6463

F-statistic: 58.38 on 5 and 152 DF, p-value: < 2.2e-16

Zde jsou významné parametry u všech bilancí. Zároveň mají všechny stejná znaménka, a to záporná. U první ilr souřadnice bude při růstu složek v čitateli bilance, což je skupina prvních pěti složek ekologické stopy, klesat hodnota HDP. To samé by se dělo také při poklesu výše uhlíkové stopy. Obojí samozřejmě při konstantní výši ostatních bilancí. Naopak při zvyšování uhlíkové stopy při nezměněné úrovni ostatních bilancí, by docházelo k růstu HDP. Stejným způsobem lze na základě vytvořeného SBP dělení popsat také změny u ostatních koeficientů, přičemž nejvýraznější změna by nastala u výše uvedené první bilance. Vzhledem k tomu, že u celého β jsou totožná znaménka, je vidět, že vždy při růstu složek v čitateli či poklesu složky ve jmenovateli dojde k poklesu HDP, zatímco při růstu složky ve jmenovateli či poklesu čitatele by došlo k růstu HDP.

Můžeme nyní porovnat výsledky modelu se srovnáním map ilr souřadnic a HDP z předchozí podkapitoly. U první ilr souřadnice byly v mapách u všech zemí, kde byla převaha složek v čitateli, což je prvních pět složek ekologické stopy, nízké hodnoty HDP. Z výsledků modelu lze usoudit, že skutečně s rostoucím podílem složek v čitateli bude docházet k poklesu HDP. U jmenovatele byly v mapách různé vztahy s HDP. Z modelu však plyne, že pro danou zemi bude tím vyšší HDP, čím vyšší bude hodnota uhlíkové stopy. U druhé souřadnice nesedí předpoklad, že s rostoucím podílem části ve jmenovateli bude klesat HDP. Z modelu vyplývá

trend přesně opačný. U čitatele byly v mapě různé vztahy mezi HDP a složkami kompozice, v modelu vychází záporná závislost HDP na těchto složkách. Ve třetí souřadnici se ve výsledcích modelu potvrdilo, že pro rostoucí převahu složek v čitateli klesá HDP. U jmenovatele žádný trend potvrdit nelze, v mapách byly různé vztahy s HDP, zde vychází růst HDP při růstu jmenovatele. Čtvrtá souřadnice podle srovnání map naznačovala, že u států s nejvyšší převahou složek v čitateli bilance bude také nízké HDP. Podle odhadnuté hodnoty parametru opravdu s rostoucím čitatelem klesá HDP. Vztah však v mapě nebyl naprosto jednoznačný, nízké HDP měly také státy s nejvyšší hodnotou jmenovatele. U páté souřadnice některé předpokládané vztahy týkající se složek v čitateli sedí a některé ne. U států, které měly nejvyšší hodnotu čitatele, bylo nízké HDP, což by sedělo s tím, že při růstu čitatele vychází v modelu pokles HDP. U jmenovatele měla jeho vyšší hodnota přispívat k nízkému HDP, což se nepotvrdilo.

Také v tomto případě lze porovnat velikost efektů změn hodnot souřadnic jako v předchozím modelu, který je v tomto případě počítaný na základě vztahů z podkapitoly 3.3.1. U první ilr souřadnice by při dvojnásobném nárůstu podílu složek uvnitř logaritmu, a tedy vzrůstu geometrického průměru prvních pěti složek ekologické stopy ku uhlíkové stopě, došlo k poklesu HDP o asi 50 %. Při stejném růstu by HDP pokleslo u druhé souřadnice o asi 13 %, u třetí asi o 22 %, u čtvrté o přibližně 17 % a u páté o asi 30 %. Největší vliv na HDP by tedy měla změna u první ze souřadnic. Přitom pokud bychom zde uvažovali naopak vzrůst uhlíkové stopy na dvojnásobnou velikost, vzrostlo by HDP o přibližně 99 %.

Koeficient R^2 dosahuje hodnoty 0,6576. Upravený koeficient má hodnotu 0,6463 a AIC 414,0078. Je vidět, že předchozí model dosahoval lepších hodnot těchto měř kvality modelu. Důvodem může být skutečnost, že klasický přístup pracoval s relativní i absolutní informací, zatímco v tomto modelu, který obsahuje na straně nezávisle proměnných pouze ilr souřadnice, je využita pouze relativní informace. To by mohlo znamenat, že také absolutní informace je v našem případě s ekologickou stopou důležitá a měla by být v modelu zahrnuta.

5.3.3. Model se součtem τ

Tentokrát se již jedná o model 3.10, který nás zajímá ze všech nejvíc, a to model, který neobsahuje na straně vysvětlujících proměnných pouze ilr souřadnice, ale také určitou formu součtu, přičemž nyní se budeme zabývat modelem obsahujícím součet ve formě logaritmu geometrického průměru složek ekologické stopy, značený jako τ

$$\begin{aligned} \ln y = & \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 \sqrt{\frac{4}{5}} \ln \frac{\sqrt[4]{x_1 \cdots x_4}}{x_5} + \beta_3 \sqrt{\frac{3}{4}} \ln \frac{\sqrt[3]{x_1 \cdots x_3}}{x_4} + \\ & + \beta_4 \sqrt{\frac{2}{3}} \ln \frac{\sqrt{x_1 x_2}}{x_3} + \beta_5 \sqrt{\frac{1}{2}} \ln \frac{x_1}{x_2} + \beta_6 \sqrt{6} \ln \sqrt[6]{x_1 \cdots x_6}. \end{aligned}$$

Při modelování HDP tedy zohledňujeme nejen strukturu ekologické stopy, ale také její absolutní výši.

Call:

```
lm(formula = log(hdp) ~ cr.gr.fo.fi.bu_ca + cr.gr.fo.fi_bu +
    cr.gr.fo.fi + cr.gr_fo + cr_gr + souc_gp, data = df_gp)
```

Residuals:

| | Min | 1Q | Median | 3Q | Max |
|--|----------|----------|---------|---------|---------|
| | -1.54984 | -0.42541 | 0.02513 | 0.34414 | 1.63143 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------------|----------|------------|---------|------------|
| (Intercept) | 10.15559 | 0.22140 | 45.870 | <2e-16 *** |
| cr.gr.fo.fi.bu_ca | -0.61755 | 0.05688 | -10.858 | <2e-16 *** |
| cr.gr.fo.fi_bu | -0.02152 | 0.07026 | -0.306 | 0.760 |
| cr.gr.fo.fi | -0.02813 | 0.05504 | -0.511 | 0.610 |
| cr.gr_fo | 0.06379 | 0.07931 | 0.804 | 0.423 |
| cr_gr | 0.01411 | 0.09181 | 0.154 | 0.878 |
| souc_gp | 0.63039 | 0.04219 | 14.940 | <2e-16 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5576 on 151 degrees of freedom

Multiple R-squared: 0.8618, Adjusted R-squared: 0.8563

F-statistic: 157 on 6 and 151 DF, p-value: < 2.2e-16

Zde je významný pouze koeficient β_1 první ilr souřadnice a součtu vyjadřujícího v podstatě celkovou úroveň ekologické stopy. Na rozdíl od předchozího modelu zde nemají všechny koeficienty totožná znaménka, první tři mají záporná a další dva kladná. U první bilance je vidět, že při růstu všech složek v čitateli, nebo poklesu uhlíkové stopy by při konstantních ostatních bilancích a součtu došlo k poklesu HDP. Naopak při vzrůstu uhlíkové stopy by došlo k růstu HDP. Tento výsledek je tedy totožný s předchozím modelem. U součtu by poté při růstu celkové velikosti stopy, avšak při zachování stejných bilancí, a tedy zachování poměrů mezi složkami ekologické stopy, došlo k růstu HDP. To odpovídá našim předpokladům o vztahu ukazatele HDP a ekologické stopy.

Vzhledem k tomu, že první tři souřadnice mají stejná znaménka parametrů jako u předchozího modelu, platí pro ně také stejné srovnání s mapami souřadnic a HDP. Tyto parametry však mají nižší hodnoty než u předchozího modelu, proto při změně v souřadnicích, nebude tak vysoká změna závisle proměnné. U čtvrté ilr souřadnice měly v mapách hodnoty s vysokou převahou složek v čitateli nízké HDP, zatímco z výsledků modelu plyne opačná tendence. Parametr páté souřadnice odpovídá předpokladu z map, že rostoucí převaha pastvin nad půdou pro pěstování plodin zvyšuje hodnotu HDP, zatímco s rostoucí převahou půdy pro pěstování plodin dochází k poklesu HDP. Sedí dále předpoklad vztahu mezi součtem a HDP, kdy s rostoucím součtem se zvyšuje také hodnota HDP, jak bylo uvedeno v předchozím odstavci.

Nyní si můžeme porovnat rozdílné efekty změny u čtvrté bilance pro tento model a pro model předchozí, kde byla odlišná znaménka koeficientů souřadnic. U předchozího modelu by při růstu zvolené bilance na dvojnásobek, a tedy vzrůstu části *cropland* a *grazing land* proti *forest products* došlo k poklesu HDP skoro o 17 %. Naopak u druhého modelu, který obsahuje také součet, by při růstu stejné bilance došlo k růstu HDP, a to zhruba o 4 %, přičemž v tomto modelu nevyšel koeficient jako významný, takže celková změna závisle proměnné je poměrně nízká.

Pokud bychom si opět vyčíslili efekty u nezávisle proměnných, jejichž koefici-

enty vyšly v modelu jako významné, došli bychom k tomu, že při změně podílu částí ekologické stopy u první ilr souřadnice na dvojnásobek by opět došlo k poklesu HDP, a to tentokrát o asi 32 %. Tento pokles je tedy při stejné změně bilance nižší, než při modelu bez součtového členu. Naopak při vzrůstu uhlíkové stopy na dvojnásobnou výši by se HDP zvýšilo o asi 48 %. U součtu by poté při vzrůstu geometrického průměru složek ekologické stopy na dvojnásobek vzrostlo HDP o asi 192 %.

Koeficient determinace a upravený koeficient determinace jsou zde 0,8618 a 0,8563. AIC dosahuje hodnoty 272,6188. Tyto hodnoty jsou totožné jako u klasického přístupu s logaritmy složek kompozice. Totožný je také absolutní člen u obou modelů a hodnoty F-statistiky. Platí tedy tvrzení, které bylo uvedeno v podkapitole 3.2.1, že klasický přístup s logaritmy složek kompozice a přístup s ilr souřadnicemi a součtem ve formě logaritmu geometrického průměru složek mají totožné výsledky F statistiky a charakteristik goodness of fit.

V tomto modelu je však na rozdíl od klasického přístupu oddělena relativní a absolutní informace. Můžeme navíc díky ilr souřadnicím sledovat nejen vliv jednotlivých částí ekologické stopy na závisle proměnnou, jak tomu bylo u klasického přístupu, ale také vlivy současných změn kombinací těchto dílčích ekologických stop. Vše záleží na použitém SBP dělení. Dále je vidět, že oproti předchozímu modelu, který zkoumal pouze relativní informaci, došlo ke zvýšení R^2 , upraveného R^2 a k poklesu AIC. Vypadá to tedy, že absolutní informace je opravdu důležitá a celková velikost ekologické stopy má také vliv na HDP, nejen relativní informace obsažená v kompozičním vektoru. Jedná se tedy o lepší model než model, který součet členů ekologické stopy neuvažoval.

5.3.4. Modely se součtem η

Nyní je použit opět model s totožnými ilr souřadnicemi jako v předchozích příkladech. Oproti předchozí podkapitole je však použit součet ve tvaru logaritmu

součtu složek kompozičního vektoru. Jedná se tedy o případ 3.18

$$\ln y = \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{\sqrt[5]{x_1 \cdots x_5}}{x_6} + \beta_2 \sqrt{\frac{4}{5}} \ln \frac{\sqrt[4]{x_1 \cdots x_4}}{x_5} + \beta_3 \sqrt{\frac{3}{4}} \ln \frac{\sqrt[3]{x_1 \cdots x_3}}{x_4} + \beta_4 \sqrt{\frac{2}{3}} \ln \frac{\sqrt{x_1 x_2}}{x_3} + \beta_5 \sqrt{\frac{1}{2}} \ln \frac{x_1}{x_2} + \beta_6 \ln(x_1 + \cdots + x_6).$$

Call:

```
lm(formula = log(hdp) ~ cr.gr.fo.fi.bu_ca + cr.gr.fo.fi_bu +
    cr.gr.fo.fi + cr.gr_fo + cr_gr + souc_sum, data = df_sum)
```

Residuals:

| | Min | 1Q | Median | 3Q | Max |
|--|----------|----------|---------|---------|---------|
| | -2.45032 | -0.35208 | 0.00977 | 0.39014 | 1.74344 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------------|----------|------------|---------|--------------|
| (Intercept) | 7.47567 | 0.21691 | 34.464 | < 2e-16 *** |
| cr.gr.fo.fi.bu_ca | -0.37513 | 0.07957 | -4.714 | 5.47e-06 *** |
| cr.gr.fo.fi_bu | -0.19545 | 0.07749 | -2.522 | 0.0127 * |
| cr.gr.fo.fi | -0.27354 | 0.05589 | -4.894 | 2.51e-06 *** |
| cr.gr_fo | -0.02061 | 0.08790 | -0.234 | 0.8149 |
| cr_gr | -0.18842 | 0.09769 | -1.929 | 0.0556 . |
| souc_sum | 1.39844 | 0.11604 | 12.052 | < 2e-16 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6266 on 151 degrees of freedom

Multiple R-squared: 0.8255, Adjusted R-squared: 0.8185

F-statistic: 119 on 6 and 151 DF, p-value: < 2.2e-16

U tohoto modelu jsou významné parametry součtu a všech ilr souřadnic kromě čtvrté a páté. U všech souřadnic jsou navíc záporné koeficienty, což se liší oproti modelu s předchozím typem součtu, kde byly některé koeficienty bilancí kladné a některé záporné. Zde však vždy při růstu složek v čitateli nebo poklesu složky ve jmenovateli klesá při konstantních ostatních bilancích HDP, při růstu složky ve jmenovateli roste HDP. U součtu je opět kladný koeficient, a tedy při růstu celkové velikosti ekologické stopy při zachování konstantních bilancí dojde k růstu HDP.

Vzhledem ke stejným znaménkům u tohoto modelu jako u modelu bez součtu zde platí stejné závěry při srovnání se vztahy odvozenými z map.

Pokud u ilr souřadnic s významnými parametry opět zkusíme jejich zvýšení na dvojnásobnou velikost, dostaneme u první souřadnice pokles HDP o 21 %, u třetí asi 15 % a nejnižší efekt u druhé, kde by došlo k poklesu HDP o 11 %. U součtu by při totožném růstu došlo k narůstu HDP o přibližně 164 %. To je o 30 p.b. menší efekt než u součtu s geometrických průměrem, použitým v předchozím modelu.

Koeficient R^2 má hodnotu 0,8255, upravený R^2 0,8185 a AIC je 309,5316. Tento model tedy dává odlišné a horší výsledky než předchozí model s jiným typem součtu. Zároveň je vidět, že pouze model se součtem ve formě geometrického průměru složek kompozice dává totožné výsledky F statistiky a goodness of fit charakteristik jako klasický přístup. U modelu se zde použitým typem součtu tato vlastnost neplatí.

5.3.5. Využití pivotových souřadnic

U následujících dvou modelů jsou použity odlišné ilr souřadnice než v modelech předcházejících, a to pivotové souřadnice použité v příkladu 2.4

$$p_1 = \sqrt{\frac{5}{6}} \ln \frac{x_1}{\sqrt[5]{x_2 x_3 x_4 x_5 x_6}}$$

$$p_2 = \sqrt{\frac{4}{5}} \ln \frac{x_2}{\sqrt[4]{x_3 x_4 x_5 x_6}}$$

$$p_3 = \sqrt{\frac{3}{4}} \ln \frac{x_3}{\sqrt[3]{x_4 x_5 x_6}}$$

$$p_4 = \sqrt{\frac{2}{3}} \ln \frac{x_4}{\sqrt{x_5 x_6}}$$

$$p_5 = \sqrt{\frac{1}{2}} \ln \frac{x_5}{x_6}.$$

První souřadnice vyjadřuje dominanci půdy pro pěstování plodin na všech ostatních složkách ekologické stopy. Druhá souřadnice popisuje převahu stopy past-

vin na všech zbývajících složkách kromě první. U souřadnice p_3 se jedná o dominanci lesních produktů na následujících třech složkách. U p_4 lze sledovat převahu stopy oblastí pro rybolov na stopě lidské infrastruktury a uhlíkové stopě. Poslední souřadnice vyjadřuje dominanci ekologické stopy lidské infrastruktury na uhlíkové stopě. (K výpočtu souřadnic byla využita funkce `pivotCoord()` z balíčku `robCompositions` [28].)

Níže jsou postupně uvedeny modely s pivotovými souřadnicemi a oběma typy součtu. Je použito stejné značení jednotlivých složek kompozice v bilancích, jako u předchozích souřadnic. První model vypadá následovně

$$\ln y = \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{x_1}{\sqrt[5]{x_2 \cdots x_6}} + \beta_2 \sqrt{\frac{4}{5}} \ln \frac{x_2}{\sqrt[4]{x_3 \cdots x_6}} + \beta_3 \sqrt{\frac{3}{4}} \ln \frac{x_3}{\sqrt[3]{x_4 \cdots x_6}} + \beta_4 \sqrt{\frac{2}{3}} \ln \frac{x_4}{\sqrt{x_5 x_6}} + \beta_5 \sqrt{\frac{1}{2}} \ln \frac{x_5}{x_6} + \beta_6 \sqrt{6} \ln \sqrt[6]{x_1 \cdots x_6}.$$

Call:

```
lm(formula = log(hdp) ~ cr_gr.fo.fi.bu.ca + gr_fo.fi.bu.ca + fo-fi.bu.ca + fi_bu.ca + bu_ca + souc_gp, data = df_piv)
```

Residuals:

| | Min | 1Q | Median | 3Q | Max |
|--|----------|----------|---------|---------|---------|
| | -1.54984 | -0.42541 | 0.02513 | 0.34414 | 1.63143 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------------|----------|------------|---------|--------------|
| (Intercept) | 10.15559 | 0.22140 | 45.870 | < 2e-16 *** |
| cr_gr.fo.fi.bu.ca | -0.09822 | 0.12367 | -0.794 | 0.42831 |
| gr_fo.fi.bu.ca | -0.14260 | 0.04304 | -3.313 | 0.00116 ** |
| fo-fi.bu.ca | -0.26279 | 0.06347 | -4.140 | 5.75e-05 *** |
| fi_bu.ca | -0.26808 | 0.04593 | -5.837 | 3.14e-08 *** |
| bu_ca | -0.46474 | 0.05134 | -9.052 | 6.61e-16 *** |
| souc_gp | 0.63039 | 0.04219 | 14.940 | < 2e-16 *** |

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5576 on 151 degrees of freedom

Multiple R-squared: 0.8618, Adjusted R-squared: 0.8563

F-statistic: 157 on 6 and 151 DF, p-value: < 2.2e-16

Model má významné parametry všech ilr souřadnic kromě první. Všechny parametry jsou záporné, což je opět rozdíl oproti modelu s odlišnými ilr souřadnicemi. Koeficient součtu je kladný, a tedy opět platí, že při růstu celkové úrovně ekologické stopy při zachování poměrů mezi částmi dochází k růstu HDP. Koeficient determinace modelu je 0,8618, upravený koeficient determinace 0,8563 a AIC dosahuje hodnoty 272,6188.

U tohoto modelu a u předchozího modelu s totožným vyjádřením součtu jsou stejné výsledky F statistiky, stejné R^2 , ale také koeficienty β_0 u absolutního členu a β_D u součtu jsou totožné. Je tedy vidět, že model s ilr souřadnicemi je skutečně SBP invariantní a dává stejné výsledky F statistiky při různém výběru bilancí, a to včetně součtu a charakteristik goodness of fit. Proto má tento model stejně jako model s odlišnými souřadnicemi také stejné výsledky F statistiky jako klasický přístup. Při výpočtu změny HDP při dvojnásobném nárůstu geometrického průměru uvnitř součtu bychom dostali opět nárůst o 192 %, což je stejné jako u modelu s odlišnými souřadnicemi.

Pro srovnání uvedeme také model s totožnými souřadnicemi, ale se součtem η , který vypadá

$$\begin{aligned} \ln y = & \beta_0 + \beta_1 \sqrt{\frac{5}{6}} \ln \frac{x_1}{\sqrt[5]{x_2 \cdots x_6}} + \beta_2 \sqrt{\frac{4}{5}} \ln \frac{x_2}{\sqrt[4]{x_3 \cdots x_6}} + \beta_3 \sqrt{\frac{3}{4}} \ln \frac{x_3}{\sqrt[3]{x_4 \cdots x_6}} + \\ & + \beta_4 \sqrt{\frac{2}{3}} \ln \frac{x_4}{\sqrt{x_5 x_6}} + \beta_5 \sqrt{\frac{1}{2}} \ln \frac{x_5}{x_6} + \beta_6 \ln(x_1 + \cdots + x_6). \end{aligned}$$

Call:

```
lm(formula = log(hdp) ~ cr_gr.fo.fi.bu.ca + gr_fo.fi.bu.ca +
    fo.fi.bu.ca + fi_bu.ca + bu_ca + souc_sum, data = df_piv_s)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|----------|----------|---------|---------|---------|
| -2.45032 | -0.35208 | 0.00977 | 0.39014 | 1.74344 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------------|----------|------------|---------|-------------|
| (Intercept) | 7.47567 | 0.21691 | 34.464 | < 2e-16 *** |
| cr_gr.fo.fi.bu.ca | -0.36457 | 0.13171 | -2.768 | 0.00635 ** |

```

gr_fo.fi.bu.ca    -0.14858    0.04840   -3.070    0.00254 **
fo_fi.bu.ca      -0.31652    0.07093   -4.462 1.58e-05 ***
fi_bu.ca         -0.08139    0.05513   -1.476    0.14192
bu_ca            -0.16696    0.07049   -2.369    0.01912 *
souc_sum         1.39844    0.11604   12.052 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 0.6266 on 151 degrees of freedom
Multiple R-squared:  0.8255, Adjusted R-squared:  0.8185
F-statistic:  119 on 6 and 151 DF,  p-value: < 2.2e-16

```

Model má významné všechny koeficienty kromě β_4 . R^2 má hodnotu 0,8255, upravený R^2 0,8185. Akaikeho informační kritérium je 309,5316. U charakteristik goodness of fit je dosaženo horších hodnot než u předchozího modelu. Pokud však srovnáme tento model s modelem s odlišnými souřadnicemi, ale stejným typem součtu, je vidět, že je opět dosaženo totožných hodnot F statistiky, měř goodness of fit i parametrů u absolutního členu a u součtu. Také model s tímto typem součtu je tedy SBP invariantní.

5.3.6. Model pouze se součtem

Posledním modelem je model obsahující na straně vysvětlujících proměnných pouze součet, vyjadřující absolutní velikost ekologické stopy. Jedná se tedy o modely [3.26](#)

$$\ln y = \beta_0 + \beta_6 \sqrt{6} \ln \sqrt[6]{x_1 \cdots x_6} \text{ a } \ln y = \beta_0 + \beta_6 \ln(x_1 + \cdots + x_6).$$

Nejdříve bude ukázán model pro součet v podobě logaritmu geometrického průměru složek kompozičního vektoru.

Call:

```
lm(formula = log(hdp) ~ souc_gp, data = df_gp)
```

Residuals:

```

      Min       1Q   Median       3Q      Max
-2.01865 -0.54169 -0.00077  0.57925  2.06999

```

Coefficients:

```

                Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.91576    0.18401   64.75  <2e-16 ***
souc_gp      0.83357    0.04379   19.04  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8095 on 156 degrees of freedom
Multiple R-squared:  0.6991, Adjusted R-squared:  0.6972
F-statistic: 362.4 on 1 and 156 DF,  p-value: < 2.2e-16

```

Ve výsledcích modelu je vidět, že jako významný vyšel koeficient absolutního členu i součtu, přičemž koeficient u součtu je opět kladný. Při růstu ekologické stopy by tedy došlo k růstu HDP, což odpovídá vztahu HDP a součtu, který jsme mohli vidět při porovnání map. Koeficient R^2 modelu je 0,6991, upravený 0,6972. AIC má hodnotu 385,5802.

Nyní zkusíme opět model obsahující pouze součet, ale tentokrát ve formě logaritmu součtu složek kompozice.

```

Call:
lm(formula = log(hdp) ~ souc_sum, data = df_sum)

Residuals:
    Min       1Q   Median       3Q      Max
-2.71924 -0.51732  0.08704  0.53844  2.28486

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.90592    0.09614   71.83  <2e-16 ***
souc_sum      1.84172    0.08233   22.37  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7194 on 156 degrees of freedom
Multiple R-squared:  0.7623, Adjusted R-squared:  0.7608
F-statistic: 500.4 on 1 and 156 DF,  p-value: < 2.2e-16

```

U tohoto modelu je opět koeficient součtu významný a kladný. Koeficient determinace má hodnotu 0,7623, upravený koeficient determinace 0,7608. AIC je

348,2996. Pokud by se zvýšil výraz uvnitř logaritmu součtů na dvojnásobou hodnotu jako v předcházejících modelech, došlo by v modelu se součtem τ k růstu HDP o asi 312 %, zatímco zde o přibližně 258 %. Opět se tedy jedná o větší efekt u prvního typu součtu.

Model se součtem η má lepší hodnoty charakteristik goodness of fit. Zároveň jsou tyto hodnoty horší než u modelu, kde byly na straně prediktorů bilance i součet ve formě logaritmu geometrického průměru složek kompozice. Proto lze předpokládat, že u ekologické stopy závisí hodnota HDP nejen na absolutní velikosti ekologické stopy, ale také na relativních velikostech jednotlivých složek.

5.4. Srovnání modelů

Hodnocení jednotlivých modelů bylo uvedeno již u popisu jejich výsledků, zde však tyto výsledky budou ještě shrnuty. V tabulce níže jsou pro srovnání zapsány výsledky charakteristik goodness of fit všech modelů. Modely o odlišných typech souřadnic jsou uvedeny jako jeden model vzhledem k SBP invarianci.

| Model | R^2 | uprav. R^2 | AIC |
|----------------------------------|--------|--------------|----------|
| Klasický přístup | 0,8618 | 0,8563 | 272,6188 |
| Ilr souřadnice bez součtu | 0,6576 | 0,6463 | 414,0078 |
| Ilr souřadnice se součtem τ | 0,8618 | 0,8563 | 272,6188 |
| Ilr souřadnice se součtem η | 0,8255 | 0,8185 | 309,5316 |
| Model pouze se součtem τ | 0,6991 | 0,6972 | 385,5802 |
| Model pouze se součtem η | 0,7623 | 0,7608 | 348,2996 |

Nejlépe hodnot všech ukazatelů dosahuje klasický přístup, a také model s ilr souřadnicemi a součtem tvořeným logaritmem geometrického průměru složek ekologické stopy. Vzhledem k tomu, že model s klasickým přístupem neumožňuje odděleně zkoumat absolutní a relativní informaci, upřednostňujeme druhý z modelů, který využívá ilr souřadnice společně s dalším členem vyjadřujícím absolutní informaci o ekologické stopě. Tento model vychází lépe než jeho podmodely, u kterých jsou na straně vysvětlujících proměnných buď pouze ilr souřadnice, nebo pouze součet, a tedy zkoumají buď jen relativní nebo absolutní informaci o eko-

logické stopě. Proto lze usoudit, že při zkoumání ekologické stopy je vhodné zaměřit se nejen na absolutní velikost této stopy, ale také relativní vyjádřenou pomocí kompozičního vektoru jednotlivých složek a pro využití v lineárních modelech vyjádřenou pomocí bilancí vytvořených vhodným SBP dělením. I když jsou výsledky z tabulky totožné pro modely s odlišnými bilancemi, jako lepší se jeví modely s SBP dělením z příkladu 2.3, kde bylo při tvorbě souřadnic bráno do úvahy dělení složek kompozičního vektoru do skupin podle jejich významu a společných vlastností.

Závěr

Práce byla zaměřena na teorii regresních modelů, kdy na straně prediktorů vystupuje mimo kompozičních proměnných také jejich součet. Cílem bylo nastudovat možnosti zahrnutí součtu složek do regresních modelů s kompozičními regresory a výhody, nevýhody a specifika tohoto přístupu demonstrovat na vlastním příkladu.

První kapitola se věnovala základním pojmům kompozičních dat společně se základními principy jejich analýzy - škálovou invariancí, permutační invariancí a podkompoziční koherencí. Závěrečná část kapitoly se zabývala speciální geometrií na simplexu. Byly představeny operace umožňující definovat vektorový prostor a posléze euklidovský vektorový prostor, nazývaný Aitchisonova geometrie na simplexu. Tyto operace byly následně aplikovány v ukázkovém příkladu.

Druhá kapitola popisovala možnosti souřadnicové reprezentace kompozičních dat. Nejdříve byly představeny clr a alr transformace společně s jejich výhodami a nevýhodami. Oba typy reprezentace byly použity v názorných příkladech pro vybraný kompoziční vektor. Druhá část kapitoly se věnovala ortonormálním souřadnicím a blíže poté bilancím, získaným postupným binárním dělením kompozičního vektoru. Krátce byly představeny také pivotové souřadnice. Oba tyto přístupy k tvorbě ortonormálních souřadnic byly demonstrovány na dvou příkladech, ve kterých byly vytvořeny souřadnice pro kompoziční vektor ekologické stopy na jednoho obyvatele pro rok 2014, a které byly využity v regresních modelech v poslední kapitole této práce.

Ve třetí kapitole byly popsány regresní modely, které je možné využít pro analýzu kompozičních dat v případě, kdy chceme modelovat závisle proměnnou

na základě vysvětlujících proměnných nesoucích jak absolutní, tak relativní informaci. Jako první byl zkoumán klasický přístup spočívající v použití původních hodnot kompozičního vektoru po log transformaci. Dále byly představeny tzv. T-prostory, u kterých dochází k vyjádření absolutní velikosti ve formě tzv. součtu. Dvě z forem součtu byly blíže rozebrány a popsány rozdílné vlastnosti modelů tyto členy zahrnující. Jako hlavní byl zkoumán regresní model, u kterého jsou na straně vysvětlujících proměnných bilance společně se součtem tvořeným logaritmem geometrického průměru složek kompozice. Byly zkoumány jak vlastnosti tohoto modelu, tak interpretace parametrů a jeho srovnání s klasickým přístupem a jednotlivé podmodely.

Čtvrtá kapitola se věnovala datovému souboru o ekologické stopě za rok 2014. Byl popsán význam ekologické stopy a jejích jednotlivých částí a způsob jejího výpočtu. Dále se kapitola zabývala zvolenou závisle proměnnou, kterou bylo HDP na 1 obyvatele v amerických dolarech, a jejím vztahem k ekologické stopě. Kapitulu uzavíralo shrnutí provedených úprav na datech před samotnou aplikací modelů.

V závěrečné kapitole byly postupně použity jednotlivé regresní modely z kapitoly 3 na ekologické stopě na 1 obyvatele v roli vysvětlujících proměnných a HDP na 1 obyvatele v roli závisle proměnné. Modely byly konstruovány v softwaru R a u každého z nich bylo v práci zobrazeno *summary* s výsledky, které byly následně interpretovány. Prvním modelem byl klasický přístup, který dosahoval dobrých výsledků AIC i R^2 . Před samotnými modely s ilr souřadnicemi byly hodnoty souřadnic i oba typy součtu představeny pomocí jejich vykreslení do map světa. Ty byly poté srovnány s mapou HDP na 1 obyvatele za účelem nalezení vzájemných vztahů. Z výsledků modelů využívajících ilr souřadnice se ukázalo, že u modelů s oběma typy součtu skutečně platí vlastnost SBP invariance, a také že model se součtem ve formě logaritmu geometrického průměru složek kompozice má totožné výsledky jako klasický přístup. Dle hodnot koeficientu determinace a Akaikého informačního kritéria bylo zjištěno, že nejlepších výsledků dosahuje v našem případě model zahrnující absolutní i relativní informaci a užívající pro vyjádření absolutní velikosti ekologické stopy součet tvořený logaritmem geo-

metrického průměru složek kompozičního vektoru. Tento model měl díky svým vlastnostem sice totožné výsledky jako klasický přístup, ale protože umožňuje zkoumat odděleně relativní a absolutní informaci, byl vyhodnocen jako nejvíce vyhovující.

Literatura

- [1] AGRESTI, Alan. Foundations of linear and generalized linear models. Hoboken, New Jersey: John Wiley & Sons, 2015 [cit. 2020-05-03]. ISBN 978-1-118-73003-4.
- [2] BOOGAART, K. Gerald van den a Raimon TOLOSANA-DELGADO. Analyzing Compositional Data with R. Berlin: Springer, 2013 [cit. 2020-05-03]. ISBN 9783642368080.
- [3] BOOGART, K. Gerald van den, Raimon TOLOSANA-DELGADO a Matevz BREN (2019). compositions: Compositional Data Analysis [online]. R package version 1.40-3 [cit. 2020-05-03]. Dostupné z: <https://CRAN.R-project.org/package=compositions>.
- [4] COENDERS, Germa, Josep A MARTÍN-FERNÁNDEZ a Berta FERRER-ROSELL. When relative and absolute information matter: Compositional predictor with a total in generalized linear models. Statistical Modelling: An International Journal [online]. 2017, 17(6), 494-512 [cit. 2020-04-11]. DOI: 10.1177/1471082X17710398. ISSN 1471-082X. Dostupné z: <http://journals.sagepub.com/doi/10.1177/1471082X17710398>.
- [5] Country Work. In: Global Footprint Network [online]. Oakland, c2003-2020 [cit. 2020-01-01]. Dostupné z: <https://www.footprintnetwork.org/our-work/countries/>.
- [6] EGOZCUE, J. J. Isometric logratio transformations for compositional data analysis. Mathematical Geology [online]. 2003, 35(3), 279-300 [cit. 2019-10-28]. DOI: 10.1023/A:1023818214614. ISSN 08828121. Dostupné z: <http://link.springer.com/10.1023/A:1023818214614>.
- [7] FILZMOSER, Peter, Karel HRON a Matthias TEMPL. Applied Compositional Data Analysis: With Worked Examples in R. Springer, 2018 [cit. 2020-05-03]. ISBN 978-3-319-96420-1.
- [8] GDP per capita (current US\$). In: World Bank Open Data [online]. The World Bank Group, c2019 [cit. 2020-03-08]. Dostupné z: <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?view=chart>.

- [9] Global Footprint Network [online]. Oakland: Global Footprint Network, c2003-2020 [cit. 2020-05-03]. Dostupné z: <https://www.footprintnetwork.org/>.
- [10] Glossary - Global Footprint Network. Global Footprint Network [online]. Oakland, c2003-2019 [cit. 2019-12-27]. Dostupné z: <https://www.footprintnetwork.org/resources/glossary/>.
- [11] HOUGHTON, J. T. Global warming: the complete briefing. Fifth edition. Cambridge: Cambridge University Press, 2015 [cit. 2020-05-03]. ISBN 978-1-107-46379-0.
- [12] Hrubý domácí produkt (HDP) - Metodika. Český statistický úřad [online]. Praha, 2015 [cit. 2020-02-13]. Dostupné z: https://www.czso.cz/csu/czso/hruby_domaci_produk_t_-hdp-.
- [13] CHAPPELOW, Jim. Gross Domestic Product—GDP. Investopedia: Sharper insight, better investing [online]. New York, Jun 27, 2019 [cit. 2020-02-15]. Dostupné z: <https://www.investopedia.com/terms/g/gdp.asp>.
- [14] CHAPPELOW, Jim. Per Capita GDP. Investopedia: Sharper insight, better investing [online]. New York, Aug 30, 2019 [cit. 2019-12-30]. Dostupné z: <https://www.investopedia.com/terms/p/per-capita-gdp.asp>.
- [15] National Footprint Accounts 2018: The Ecological Footprint of 196 Countries. Kaggle: Your Machine Learning and Data Science Community [online]. San Francisco: Kaggle, c2019 [cit. 2020-04-30]. Dostupné z: <https://www.kaggle.com/footprintnetwork/national-footprint-accounts-2018>.
- [16] JORGENSON, Andrew K. a Thomas J. BURNS. The political-economic causes of change in the ecological footprints of nations, 1991–2001: A quantitative investigation. Social Science Research [online]. 2007, 36(2), 834-853 [cit. 2020-02-04]. DOI: 10.1016/j.ssresearch.2006.06.003. ISSN 0049089X. Dostupné z: <https://linkinghub.elsevier.com/retrieve/pii/S0049089X06000482>.
- [17] LIN, David, Laurel HANSCOM, Jon MARTINDILL, et al. Working Guidebook to the National Footprint Accounts [online]. Oakland: Global Footprint Network, 2018 [cit. 2019-12-26]. Dostupné z: <https://www.footprintnetwork.org/content/uploads/2018/05/2018-National-Footprint-Accounts-Guidebook.pdf>.
- [18] LIN, David, Laurel HANSCOM, Adeline MURTHY, et al. Ecological Footprint Accounting for Countries: Updates and Results of the National Footprint Accounts, 2012–2018. Resources [online]. 2018, 1856-1868, 7(3) [cit.

- 2019-12-26]. DOI: 10.3390/resources7030058. ISSN 2079-9276. Dostupné z: <http://www.mdpi.com/2079-9276/7/3/58>.
- [19] MARQUART-PYATT, Sandra T. Environmental Sustainability. *International Journal of Sociology* [online]. 2014, 40(2), 65-84 [cit. 2020-02-09]. DOI: 10.2753/IJS0020-7659400204. ISSN 0020-7659. Dostupné z: <https://www.tandfonline.com/doi/full/10.2753/IJS0020-7659400204>.
- [20] MILLARD, SP (2013). EnvStats: Package for Environmental Statistics, Including US EPA Guidance [online]. R package version 2.3.1 [cit. 2020-05-03]. Dostupné z: <https://cran.r-project.org/package=EnvStats>.
- [21] NEUWIRTH, Erich (2014). RColorBrewer: ColorBrewer Palettes [online]. R package version 1.1-2 [cit. 2020-05-03]. Dostupné z: <https://CRAN.R-project.org/package=RColorBrewer>.
- [22] PAWLOWSKY-GLAHN, Vera a Antonella BUCCIANTI. *Compositional data analysis: theory and applications*. Chichester, West Sussex, U.K.: Wiley, 2011 [cit. 2020-05-03]. ISBN 978-0470711354.
- [23] PAWLOWSKY-GLAHN, Vera, Juan José EGOZCUE a David LOVELL. Tools for compositional data with a total. *Statistical Modelling: An International Journal* [online]. 2015, 15(2), 175-190 [cit. 2019-11-24]. DOI: 10.1177/1471082X14535526. ISSN 1471-082X. Dostupné z: <http://journals.sagepub.com/doi/10.1177/1471082X14535526>.
- [24] PAWLOWSKY-GLAHN, Vera, Juan J. EGOZCUE a Raimon TOLOSANA-DELGADO. *Modelling and analysis of compositional data*. Chichester, West Sussex, UK: John Wiley & Sons, 2015 [cit. 2020-05-03]. ISBN 978-1-118-44306-4.
- [25] PEBESMA, E.J., R.S. BIVAND (2005). sp: Classes and Methods for Spatial Data [online]. R package version 1.3-2 [cit. 2020-05-03]. Dostupné z: <https://cran.r-project.org/package=sp>.
- [26] PICARDO, Elvis. The Importance of the GDP. Investopedia: Sharper insight, better investing [online]. New York, May 18, 2019 [cit. 2019-12-30]. Dostupné z: <https://www.investopedia.com/articles/investing/121213/gdp-and-its-importance.asp>.
- [27] SOUTH, Andy (2017). rnaturalearth: World Map Data from Natural Earth [online]. R package version 0.1.0 [cit. 2020-05-03]. Dostupné z: <https://CRAN.R-project.org/package=rnaturalearth>.
- [28] TEMPL, Matthias, Karel HRON, Peter FILZMOSER (2011). robCompositions: Compositional Data Analysis [online]. R package version 2.1.0

- [cit. 2020-05-03]. Dostupné z: <https://cran.r-project.org/package=robCompositions>.
- [29] TENNEKES, M (2018). tmap: Thematic Maps [online]. R package version 2.3-2 [cit. 2020-05-03]. Dostupné z: <https://cran.r-project.org/package=tmap>.
- [30] THE UAE ECOLOGICAL FOOTPRINT INITIATIVE: SUMMARY REPORT 2007-2010 [online]. Abu Dhabi: EWS-WWF Head Office, 2011 [cit. 2019-12-28]. Dostupné z: http://awsassets.panda.org/downloads/en_final_report_ecological_footprint.pdf.
- [31] Vše, co teď bereme jako samozřejmost, bude minulostí. Týdeník Respekt [online]. 2019, (16) [cit. 2020-02-04]. Dostupné z: <https://www.respekt.cz/tydenik/2019/16/klima-vstupujeme-do-temneho-veku>
- [32] What is GDP? Bank of England [online]. London, c2020 [cit. 2020-02-13]. Dostupné z: <https://www.bankofengland.co.uk/knowledgebank/what-is-gdp>.
- [33] WICKHAM, Hadley (2019). stringr: Simple, Consistent Wrappers for Common String Operations [online]. R package version 1.4.0 [cit. 2020-05-03]. Dostupné z: <https://CRAN.R-project.org/package=stringr>.
- [34] WICKHAM, Hadley a Jennifer BRYAN (2019). readxl: Read Excel Files [online]. R package version 1.3.1 [cit. 2020-05-03]. Dostupné z: <https://CRAN.R-project.org/package=readxl>.
- [35] WICKHAM, Hadley, Romain FRANCOIS, Lionel HENRY a Kirill MÜLLER (2019). dplyr: A Grammar of Data Manipulation [online]. R package version 0.8.3 [cit. 2020-05-03]. Dostupné z: <https://CRAN.R-project.org/package=dplyr>.
- [36] WICKHAM, Hadley a Lionel HENRY (2019). tidyr: Tidy Messy Data [online]. R package version 1.0.0 [cit. 2020-05-03]. Dostupné z: <https://CRAN.R-project.org/package=tidyr>.