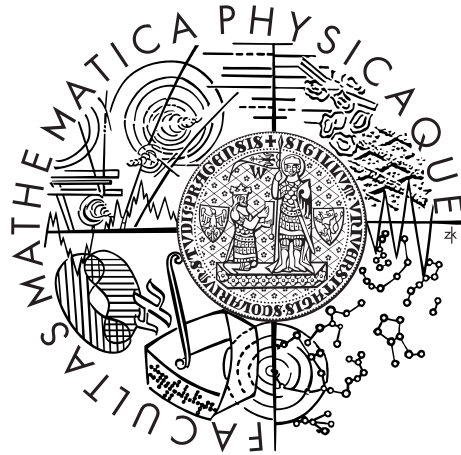


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Bc. Martin Ondruš

Zdrojové faktory indexů ekonomické svobody

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Mgr. Michal Vrabec, Csc.

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

Praha 2015

Na tomto mieste by som sa rád poďakoval Mgr. Michalovi Vrabcovi Csc. za odborné rady a pripomienky pri písaní tejto diplomovej práce.

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V dne

Podpis autora

Název práce: Zdrojové faktory indexů ekonomické svobody

Autor: Bc. Martin Ondruš

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Mgr. Michal Vrabec, Csc., Vysoká škola ekonomická v Praze

Abstrakt: V rámci této práce se budeme zabývat detekcí latentních proměnných, které se podílí na tvorbě indexů ekonomické svobody. Nejprve si představíme v současnosti nejznámější indexy ekonomické svobody (IEF, EFW). Dále se podíváme na mnohorozměrnou statistickou metodu - faktorovou analýzu, kterou použijeme k detekci latentních proměnných. Ukážeme si, jaké různé metody odhadu existují ve faktorové analýze a při jejich charakteristice se zaměříme především na metodu hlavních faktorů a metodu maximální věrohodnosti. Popsané metody následně mezi sebou srovnáme pomocí analýzy EFW indexu. Na základě odhadnutých modelů budeme interpretovat zjištěné latentní proměnné. Faktorovou analýzu provedeme v prostředí statistických softwarových programů SPSS a R.

Klíčová slova: index ekonomické svobody, faktorová analýza, latentná proměnná

Title: Factors of economical freedom indices

Author: Bc. Martin Ondruš

Department: Department of Probability and Mathematical Statistics

Supervisor: Mgr. Michal Vrabec, Csc., University of Economics, Prague

Abstract: This work discusses the detection of latent variables, which create indices of economic freedom. Firstly, we present the most well-known indices of economic freedom (IEF, EFW). Secondly, this work discusses multivariate statistical method – factor analysis, which we use to detect latent variables. We show different methods of estimates in factor analysis and we focus on principal factor method. Furthermore, we compare already defined methods by analysing the structure of EFW index. According to estimated models, we interpret detected latent variables. We use statistical software SPSS and R for factor analysis of EFW index.

Keywords: economical freedom index, factor analysis, latent variable

Názov práce: Zdrojové faktory indexov ekonomickej slobody

Autor: Bc. Martin Ondruš

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedúci diplomovej práce: Mgr. Michal Vrabec, Csc., Vysoká škola ekonomická v Praze

Abstrakt: V rámci tejto práce sa budeme zaoberať detekciou latentných premenných, ktoré sa podieľajú na tvorbe indexu ekonomickej slobody. Najprv si predstavíme v súčasnosti najznámejšie indexy ekonomickej slobody (IEF, EFW). Ďalej sa pozrieme na mnohorozmernú štatistickú metódu - faktorovú analýzu, ktorú použijeme na detekciu latentných premenných. Ukážeme si, aké rôzne metódy odhadu existujú vo faktorovej analýze a pri ich charakteristike sa zameriame predovšetkým na metódu hlavných faktorov a metódu maximálnej vierohodnosti. Popísané metódy následne medzi sebou porovnáme na analýze EFW indexu. Na základe odhadnutých modelov budeme interpretovať zistené latentné premenné. Na faktorovú analýzu EFW indexu použijeme štatistické softvéry SPSS a R.

Kľúčové slová: index ekonomickej slobody, faktorová analýza, latentná premenná

Obsah

1	Index ekonomickej slobody	3
1.1	Index of Economic Freedom	3
1.1.1	IEF index 2015	4
1.2	Economic Freedom of the World Index	5
1.2.1	EFW index 2014	7
2	Faktorová analýza	9
2.1	Motivácia	9
2.2	Matematický model	9
2.2.1	Interpretácia faktorov	12
2.2.2	Nejednoznačnosť faktorových váh	12
2.3	Odhad faktorového modelu	15
2.3.1	Metóda hlavných faktorov	16
2.3.2	Metóda hlavných komponentov	19
2.3.3	Metóda maximálnej vierohodnosti	19
2.3.4	Metóda najmenších štvorcov	22
2.3.5	Zobecnená metóda najmenších štvorcov	22
2.4	Faktorové skóre	23
2.4.1	Regresná metóda	23
3	Faktorová analýza EFW indexu	25
3.1	Predpoklady pre faktorovú analýzu	25
3.1.1	Overenie predpokladov faktorovej analýzy	27
3.2	Hľadanie počtu faktorov	31
3.3	Odhad faktorového modelu	33
3.3.1	Metóda hlavných faktorov	34
3.3.2	Metóda najmenších štvorcov	42
3.3.3	Zobecnená metóda najmenších štvorcov	45
3.4	Výber faktorového modelu	47
3.4.1	Analýza komunalít	49
3.4.2	Odhad faktorového skóre	50
	Literatúra	55
	Zoznam obrázkov	57
	Zoznam tabuliek	58

Úvod

Ekonomická sloboda sa javí ako pomerne široký a komplexný pojem. Môžeme povedať, že je akosi nevyhnutnou podmienkou pre demokratický vývoj v krajine. Oslobodzuje ľudí od závislosti na vláde a dovoľuje robiť ľuďom ich vlastné ekonomické a politické rozhodnutia.

V súčasnosti niekoľko svetových inštitúcií každoročne publikuje index ekonomickej slobody jednotlivých krajín sveta. Medzi najznámejšie patria Index of Economic Freedom, ktorý publikuje Heritage Foundation v spolupráci s Wall Street Journal a Economic Freedom of the World Index, ktorý vydáva Fraser Institute. Oba indexy su založené na tom, že sledujú právne, ekonomické a iné oblasti jednotlivých krajín a tie sa následne snažia kvantifikovať, ako vplývajú na ekonomickú slobodu. Týchto sledovaných oblastí (indikátorov ekonomickej slobody) je pomerne veľa a často je obtiažne im porozumieť. Tvorcovia indexu preto sledované indikátory ďalej zoskupujú do zrozumiteľnejších širších oblastí.

Keďže medzi viacerými indikátormi je často úzky súvis, tak môžeme predpokladať, že ich vieme vysvetliť pomocou menšieho počtu nepozorovaných latentných premenných. K ich detekcii nám vie poslúžiť faktorová analýza, mnohorozmerná štatistická metóda, ktorá vznikla v oblasti psychológie. Za jej zakladateľa je považovaný Spearman (1904), ktorý v článku o povahovej inteligencii navrhol hypotézu o existencii spoločného faktoru - obecnej intelektuálnej schopnosti, ktorý spôsobuje koreláciu medzi výsledkami rôznych intelligenčných testov. Neskôr Thurstone (1947) rozšíril Spearmanov jednofaktorový model na viacfaktorový. O ďalší rozvoj sa takisto zaslúžil štatistik D. N. Lawley (Lawley a Maxwell (1962)), ktorý previedol pôvodný problém získavania faktorov na problém maximálnej vierohodnosti.

Táto práca tak spočíva v predstavení v súčasnosti najznámejších indexov ekonomickej slobody. V druhej kapitole predstavíme model faktorovej analýzy, jednotlivé metódy odhadov faktorov, kde sa upriamime predovšetkým na metódu hlavných komponentov a metódu maximálnej vierohodnosti. V tretej kapitole sa už budeme venovať aplikácii faktorovej analýzy, kde budeme detekovať spoločné faktory indikátorov vstupujúcich do výpočtu indexu ekonomickej slobody. Pozrieme sa taktiež na to, ako sa náš štatistický prístup k problematike líši od idealistického a intuitívneho zoskupovania indikátorov tvorcami indexu.

Kapitola 1

Index ekonomickej slobody

Cieľom tejto kapitoly je predstaviť v súčasnosti publikované dva celosvetové indexy, ktoré slúžia na meranie ekonomickej slobody jednotlivých krajín sveta. Týmito indexami sú:

1. Index of Economic Freedom (ďalej iba IEF)
2. Economic Freedom of the World Index (ďalej iba EFW)

Ešte pred ich predstavením si zadefinujeme pojem ekonomickej slobody (prevzatý z Miller (2009)).

Definícia 1. *Najvyššia forma ekonomickej slobody poskytuje úplné právo na vlastníctvo majetku, na voľný pohyb pracovnej sily, kapitálu a tovarov. Ďalej takisto zabráňuje nátlaku a obmedzovanie ekonomickej slobody nad rámec nevyhnutný pre ochranu občanov.*

1.1 Index of Economic Freedom

IEF index každoročne publikuje Heritage Foundation v spolupráci s Wall Street Journal od roku 1995. Veľkým pozitívom tohto indexu je jeho aktuálnosť. Je to dané jeho konštrukciou, ktorá umožňuje tvorbu indexu na najaktuálnejších dátach a hodnotení v čo najreálnejšom čase. Výsledný index jednotlivých krajín je určený priemerom desiatich ekonomických indikátorov ekonomickej slobody (viď Miller (2009)). Týmito ekonomickými indikátormi sú:

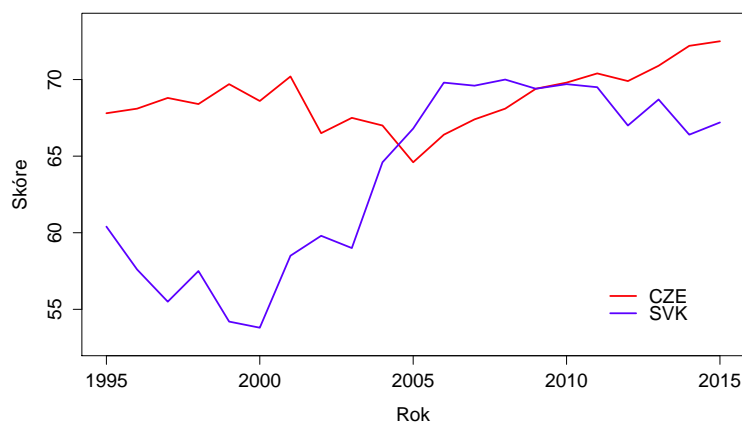
- sloboda podnikat' - právo jednotlivca otvoriť si podnikanie, spravovať ho prípadne zatvoriť bez zásahu štátu
- sloboda obchodovať - odráža otvorenosť hospodárstva importovať tovary a služby z celého sveta a taktiež schopnosť obyvateľov pôsobiť ako odberateľ a aj ako dodávateľ
- fiškálna sloboda - sloboda jednotlivcov a podnikov mať príjem a iné bohatstvo na vlastný úžitok
- veľkosť štátnej správy - nadbytočná veľkosť štátnej správy má za následok privysoké náklady

- monetárna sloboda - nutnosť mať stabilnú a spoľahlivú menu
- sloboda investovať - štátne reštrikcie na zahraničné investície limitujú prítok aj odliv kapitálu
- finančná sloboda - zabezpečenie bezpečnosti a solídnosti finančného systému
- vlastnícke práva - hlavným motivačným nástrojom v tržnej ekonomii je sloboda akumulovať súkromný majetok
- sloboda od korupcie - nezávislosť od sledovania vlastných záujmov nad záujmami spoločnosti
- sloboda práce - schopnosť jednotlivcov pracovať koľko chcú a kde chcú

Každý jeden z týchto desiatich indikátorov sa hodnotí od 0 do 100, kde 100 predstavuje maximálnu slobodu, tj. ekonomické prostredie, ktoré je najviac nápomocné k ekonomickej slobode.

1.1.1 IEF index 2015

Posledná hodnota IEF indexu je známa z roku 2015, kedy bol publikovaný index pre 186 krajín sveta. Pre 8 krajín však aspoň jedna hodnota z desiatich ekonomických indikátorov nie je známa a tak pre nich nie je publikované výsledné skóre IEF indexu. Každopádne, od prvej publikácie indexu v roku 1995, kedy boli známe hodnoty všetkých indikátorov pre 101 krajín, je to nárast o 78 krajín.



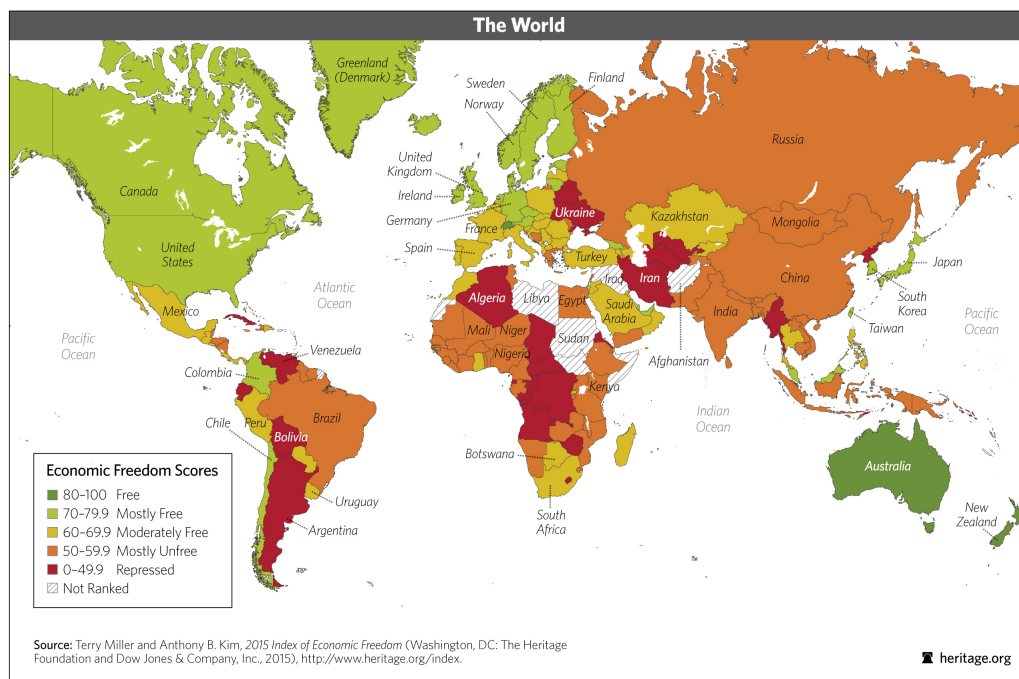
Obr. 1.1: Vývoj IEF indexu Českej republiky a Slovenskej republiky

Veľkou nevýhodou tohto indexu je, že dochádza k používaniu dát z predošlých rokov, ak pre súčasný rok potrebné dáta nie sú k dispozícii. Týmto sa index stáva nekonzistentným a výsledné skóre je miestami skreslené. Konštrukcia tohto indexu je akosi "Pandorinou skrinkou", pretože nie je verejne k dispozícii úplný súbor dát, na základe ktorého by sme mohli zistiť, ako sa jednotlivé indikátory merajú, respektíve nie je známy proces ich subjektívneho určovania.

Pre ilustráciu IEF indexu, obrázok 1.1 zobrazuje vývoj IEF indexu Českej republiky a Slovenskej republiky od roku 1995. Môžeme vidieť, že IEF index Českej

republiky od roku 2005 rastie, čo znamená, že prostredie v krajine je ekonomicky slobodnejšie. Naopak, index Slovenskej republiky od roku 2006 mierne klesá, čo znamená, že prostredie sa stáva menej slobodným. Tento pokles je zrejme spôsobený zhoršujúcim sa prostredím pre investorov kvôli reštrikciám štátu voči investorom.

Najvyššie skóre IEF indexu každoročne dosahujú Hong Kong, Singapúr, Nový Zéland, Švajčiarsko a Austrália. Naopak najhoršie skóre dosahujú predovšetkým africké krajiny a takisto napríklad aj Ukrajina, Bolívia alebo Irán (viď obrázok 1.2 prevzatý z Heritage Foundation (2015)).



Obr. 1.2: Mapa IEF indexu 2015

1.2 Economic Freedom of the World Index

V porovnaní s IEF indexom, začal EFW index byť publikovaný od roku 1996. Od tohto roku ho každoročne publikuje Fraser Institute. Ďalej sa EFW index považuje za konzistentnejší ako IEF index, pretože sa určuje na základe dlhšieho časového rámca v porovnaní s jednoročným časovým rámcom IEF indexu. Taktiež hodnoty vstupných indikátorov nie sú určované subjektívnym pozorovaním, ale jednotlivé hodnoty sa preberajú zo svetových inštitúcií ako napríklad Svetová banka, Medzinárodný menový fond alebo z reportov od PricewaterhouseCoopers, PRS Group a podobne.

Do výpočtu EFW indexu vstupuje 42 rôznych ekonomických indikátorov, ktorými sa meria stupeň ekonomickej slobody v piatich sledovaných oblastiach:

1. vládny sektor
2. právny systém a vlastnícke práva
3. mena

4. sloboda medzinárodného obchodu
5. regulácia

Konštrukcia prebieha tak, že týchto 42 vstupných premenných sa zoskupuje do 24 komponentov a tie následne do vyššie uvedených sledovaných piatich oblastí. Skóre vstupných premenných sa hodnotí na škále od 0 do 10 a zoskupovanie do jednotlivých komponentov respektíve sledovaných oblastí prebieha ich priemerovaním.

Celková štruktúra vstupných premenných, komponentov a sledovaných oblastí EFW indexu vyzerá nasledovne:

1. vládny sektor
 - (a) vládna spotreba
 - (b) transféry a dotácie
 - (c) vládne investície
 - (d) najvyššia hraničná sadzba zdanenia
 - i. daň z príjmu
 - ii. daň zafažujúca vlastníctvo majetku
2. právny systém a vlastnícke práva
 - (a) nezávislosť súdnictva
 - (b) nestrannosť súdov
 - (c) ochrana vlastníckych práv
 - (d) vojenský zásah do súdnictva a politiky
 - (e) integrita právneho systému
 - (f) právna vymáhateľnosť záväzkov plynúcich zo zmlúv
 - (g) regulačné reštrikcie na nakladanie s nehnuteľnosťami
 - (h) spoľahlivosť polície
 - (i) kriminálna činnosť v danej zemi
3. peňažné slobody
 - (a) rast peňažnej zásoby
 - (b) smerodajná odchýlka inflácie
 - (c) inflácia za posledný kalendárny rok
 - (d) sloboda vlastníť bankový účet v cudzej mene
4. sloboda medzinárodného obchodu
 - (a) tarifné bariéry medzinárodného obchodu
 - i. príjem z daní zafažujúcich medzinárodný obchod
 - ii. priemerná sadzba dane uvalenej na medzinárodný obchod
 - iii. smerodajná odchýlka daní uvalených na medzinárodný obchod

- (b) regulačné prekážky medzinárodného obchodu
 - i. netarifované obchodné prekážky
 - ii. náklady na dodržanie obchodných predpisov
- (c) ceny na čiernom trhu
- (d) kontrola pohybu osôb a kapitálu
 - i. reštrikcie zahraničného vlastníctva a investovania
 - ii. kontrola kapitálu
 - iii. možnosť zahraničných osôb navštíviť danú krajinu

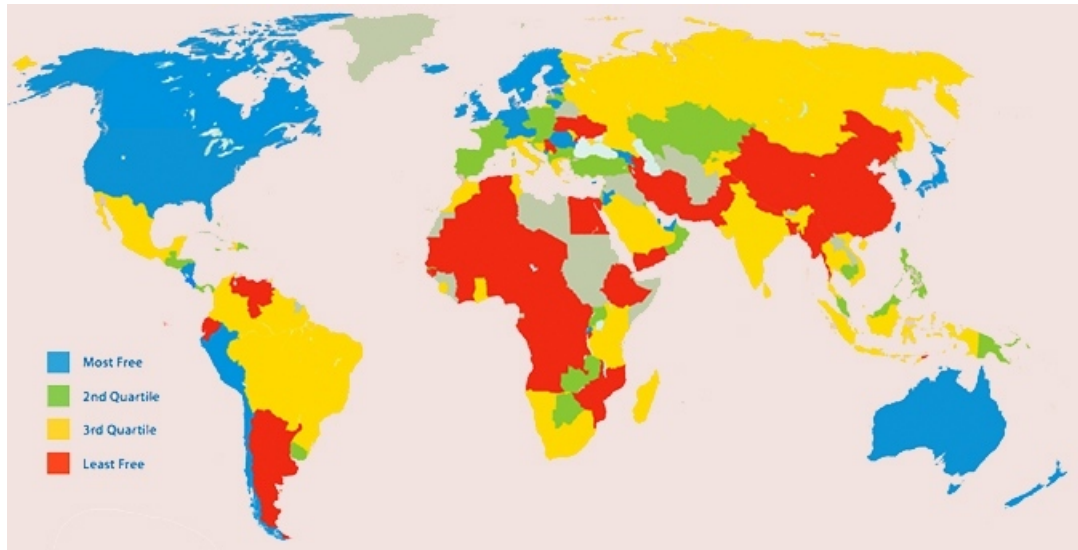
5. regulácia

- (a) regulácia finančného trhu
 - i. vlastníctvo bánk
 - ii. sektor súkromných úverov
 - iii. kontrola a regulácia úrokových sadzieb
- (b) regulácia trhu práce
 - i. regulácia minimálnej mzdy
 - ii. regulácia najímania a prepúšťania pracovnej sily
 - iii. sila centrálného vyjednávania a odborov
 - iv. predpisy stanovujúce pracovnú dobu
 - v. náklady na prepustenie pracovníka
 - vi. odvody štátu za najímanie pracovnej sily
- (c) regulácia podnikateľskej sféry
 - i. administratívne obmedzenia
 - ii. náklady na byrokráciu
 - iii. založenie podnikania
 - iv. miera korupcie a extra platby za zvýhodňovanie oproti ostatným
 - v. licenčné zákony
 - vi. náklady na dodržiavanie daňových predpisov

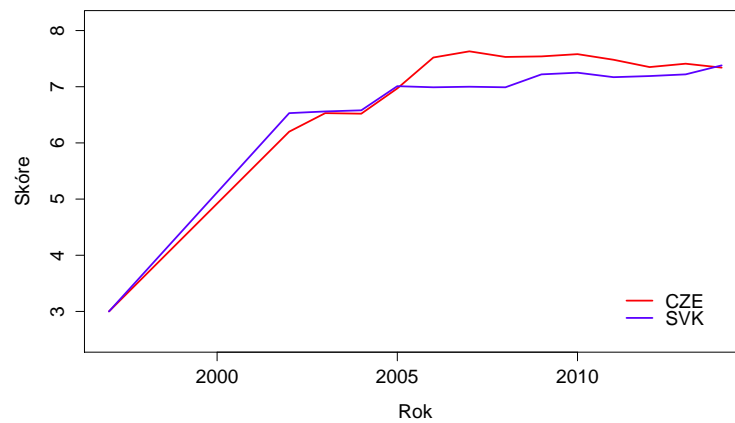
1.2.1 EFW index 2014

Posledná známa hodnota EFW indexu je z roku 2014, kedy Fraser Institute publikoval skóre EFW indexu pre 153 krajín na základe dát z roku 2012. Podobne ako pri IEF indexe, nie sú známe všetky hodnoty indikátorov pre všetky krajiny. Kompletné hodnoty indikátorov má 101 krajín. Popredné miesta opäť obsadili ázijské krajiny Hong Kong a Singapúr, ďalej Nový Zéland, Švajčiarsko a taktiež Maurícius. Najhoršie skóre EFW indexu dosahujú znovu africké krajiny a takisto napríklad Argentína, Venezuela či Ekvádor (viď obrázok 1.3 prevzatý z Fraser Institute (2015)).

Pre ilustráciu EFW indexu je na obrázku 1.4 opäť zobrazený vývoj hodnoty indexu pre Slovenskú republiku a Českú republiku. V porovnaní s vývojom IEF indexu v 1.1 vidíme, že hodnota indexu pre Slovenskú republiku neklesá, ale



Obr. 1.3: Mapa EFW indexu 2014



Obr. 1.4: Vývoj EFW indexu Českej republiky a Slovenskej republiky

naopak mierne stúpa, zatiaľ čo index Českej republiky zaznamenal za posledné roky mierny pokles.

Tento súbor dát bude ďalej pre účely tejto diplomovej práce kľúčový, pretože si na ňom ukážeme použitie mnohorožmernej štatistickej metódy na určenie zdrojových faktorov týchto indikátorov.

Kapitola 2

Faktorová analýza

Táto kapitola je venovaná faktorovej analýze, ktorá sa radí medzi mnohorozmerné štatistické metódy. V prvej podkapitole ukážeme, aká motivácia stojí za použitím faktorovej analýzy a mnohorozmerných štatistických metód vo všeobecnosti. Druhá podkapitola už bude pojednávať o samotnom štatistickom a matematickom pozadí jej použitia.

2.1 Motivácia

Motiváciu využitia faktorovej analýzy si môžeme ilustrovať na jednoduchom príklade z Hurt (2014). Majme súbor dát s ôsmimi morfológickými charakteristikami človeka, ako sú výška, rozpätie rúk, dĺžka predlaktí, dĺžka nôh, váha, obvod stehna, obvod hrudi a šírka hrudi. Z biologickej stavby človeka je zjavné, že medzi pozorovanými charakteristikami bude silná pozitívna kolerácia, ako napríklad, že vyšší človek bude mať väčšie rozpätie rúk resp. dlhšie nohy a tak ďalej. Môžeme tak predpokladať, že týchto osem charakteristík dostatočne vysvetľuje menší počet faktorov ako napríklad obecný faktor vzrastu a obecný faktor šírky človeka.

Dostávame sa tak k cieľu faktorovej analýzy, ktorým je popísať pozorované premenné pomocou menšieho počtu latentných premenných, tzv. faktorov, ktoré už často nie sú priamo merateľné. Inak povedané, cieľom je zredukovať dimenzionalitu dát na základe kovariančnej štruktúry dát a vysvetliť tým vzťahy medzi pozorovanými premennými.

Faktorová analýza má tak bohaté uplatnenie či už v sociológii (meranie politickej orientácie na základe zodpovedaných otázok), psychológii (úsudok o verbálnej a matematickej inteligencii na základe zodpovedaných otázok) a v iných odvetviach.

2.2 Matematický model

Na základe motivačného príkladu z predošlej podkapitoly tak vidíme hlavný rozdiel medzi faktorovou analýzou a štandardnou metódou hlavných komponentov pre zredukovanie dimenzionality dát. Zatiaľ, čo u metódy hlavných komponentov sa pozorované premenné agregujú do komponentov, tak u faktorovej analýzy sú faktory latentnými premennými, ktoré zapríčiňujú kovariančnú štruktúru medzi pozorovanými premennými. Ďalej sa v tejto práci už metóde hlavných

komponentov venovať nebudeme a preto tento model nechávame bez matematického zadefinovania.

V tejto podkapitole si predstavíme model faktorovej analýzy, ktorého základné princípy sú zhrnuté v Hebák a kol. (2004), prípadne v Härdle (2007). Nech $\mathbf{X} = (X_1, X_2, \dots, X_p)^T$ je p -rozmerný vektor pozorovaných premenných. Nech ďalej Σ je kovariančná matica vektoru \mathbf{X} definovaná ako

$$\Sigma = \begin{pmatrix} \text{Var}X_1 & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_p) \\ \text{Cov}(X_1, X_2) & \text{Var}X_2 & \cdots & \text{Cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_1, X_p) & \text{Cov}(X_2, X_p) & \cdots & \text{Var}X_p \end{pmatrix}.$$

Bez ujmy na obecnosti predpokladajme, že $\mathbb{E}X_j = 0$ pre $j = 1, 2, \dots, p$. Inak by sme ďalej miesto \mathbf{X} uvádzali $\mathbf{X} - \boldsymbol{\mu}$, kde $\boldsymbol{\mu}$ by bol vektor stredných hodnôt pozorovaný premenných X_j , $\mathbb{E}X_j = \mu_j$.

Faktorová analýza slúži na vysvetlenie vzájomnej lineárnej závislosti pozorovaných premenných X_j pomocou obecných faktorov f_1, \dots, f_q a špecifických faktorov $\epsilon_1, \dots, \epsilon_p$, ktoré zas zachytávajú rozptyl jednotlivých pozorovaných premenných. Tieto špecifické faktory môžeme tiež nazvať ako reziduálne zložky. Cieľom faktorovej analýzy sú však predovšetkým obecné faktory, ktorých počet q chceme, aby bol omnoho menší ako počet p pozorovaných premenných ($q \ll p$), ale na druhej strane aby počet obecných faktorov stále dostatočne dobre vysvetľoval závislosť medzi pozorovanými premennými. Matematický model tak môžeme zapísať v tvare

$$X_j = \sum_{r=1}^q \lambda_{jr} f_r + \epsilon_j, \quad j = 1, \dots, p$$

alebo podobne sústavou rovníc

$$\begin{aligned} X_1 &= \lambda_{11}f_1 + \lambda_{12}f_2 + \cdots + \lambda_{1q}f_q + \epsilon_1 \\ X_2 &= \lambda_{21}f_1 + \lambda_{22}f_2 + \cdots + \lambda_{2q}f_q + \epsilon_2 \\ &\vdots \\ X_p &= \lambda_{p1}f_1 + \lambda_{p2}f_2 + \cdots + \lambda_{pq}f_q + \epsilon_p \end{aligned}$$

alebo v maticovom tvare

$$\mathbf{X} = \mathbf{\Lambda} \mathbf{f} + \boldsymbol{\epsilon} \tag{2.1}$$

Koeficienty λ_{jk} sa nazývajú faktorové váhy j -tej pozorovanej premennej u k -teho faktoru, kde $k = 1, 2, \dots, q$ a $\mathbf{\Lambda}_{(p \times q)}$ je matica týchto váh a \mathbf{f} je q -rozmerný vektor obecných faktorov f_1, \dots, f_q , respektíve $\boldsymbol{\epsilon}$ je p -rozmerný vektor špecifických faktorov $\epsilon_1, \dots, \epsilon_p$. Predpokladáme, že matica $\mathbf{\Lambda}$ má hodnotu q .

Ďalšími predpokladmi pre nepozorovateľné náhodné vektory \mathbf{f} a $\boldsymbol{\epsilon}$ sú:

$$\begin{aligned}\mathbb{E}\mathbf{f} &= \mathbf{0} \\ \text{Var}\mathbf{f} &= \mathbb{E}\mathbf{f}\mathbf{f}^T = \mathbb{I}(q) \\ \mathbb{E}\boldsymbol{\epsilon} &= \mathbf{0} \\ \text{Var}\boldsymbol{\epsilon} &= \mathbb{E}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T = \boldsymbol{\Psi} = \text{diag}(\Psi_i), \text{ kde } \text{Var}\epsilon_i = \Psi_i \\ \text{Cov}(\mathbf{f}, \boldsymbol{\epsilon}) &= \mathbb{E}\mathbf{f}\boldsymbol{\epsilon}^T = \mathbf{O}\end{aligned}\tag{2.2}$$

O obecných faktoroch teda predpokladáme, že sú centrované, nekorelované a štandardizované. O špecifických faktoroch zas predpokladáme, že majú nulovú strednú hodnotu, sú vzájomne nekorelované a že sú taktiež nekorelované s obecnými faktormi. Všeobecný faktorový model spoločne s týmito predpokladmi nazývame *ortogonálny faktorový model*.

Na základe týchto predpokladov vyplýva, že kovariančnú maticu Σ vieme vyjadriť v tvare

$$\Sigma = \mathbb{E}(XX^T) = \mathbb{E}[(\boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\epsilon})(\boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\epsilon})^T] = \boldsymbol{\Lambda}\mathbb{E}(\mathbf{f}\mathbf{f}^T)\boldsymbol{\Lambda}^T + \mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}\tag{2.3}$$

Z 2.3 tak plynie, že

$$\begin{aligned}\text{Var}X_j &= \sigma_j^2 = \sum_{k=1}^q \lambda_{jk}^2 + \Psi_j \\ \text{Cov}(X_j, X_l) &= \sigma_{jl} = \sum_{k=1}^q \lambda_{jk}\lambda_{lk}, \text{ kde } j \neq l\end{aligned}$$

Diagonálne prvky matice $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T$

$$h_j^2 = \sum_{k=1}^q \lambda_{jk}^2 = \sigma_j^2 - \Psi_j$$

sa nazývajú *komunalita premenných* a prvky Ψ_j *špecifické rozptyly* pozorovaných premenných. Komunalita premennej X_j vyjadruje tú časť jej rozptylu, ktorá je vysvetlená pôsobením obecných faktorov a zostávajúcu časť rozptylu tvoria špecifické (chybové) rozptyly.

Pretože ďalej platí, že

$$\mathbb{E}(\mathbf{X}\mathbf{f}^T) = \mathbb{E}((\boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\epsilon})\mathbf{f}^T) = \boldsymbol{\Lambda}\mathbb{E}(\mathbf{f}\mathbf{f}^T) + \mathbb{E}(\boldsymbol{\epsilon}\mathbf{f}^T) = \boldsymbol{\Lambda}$$

tak parametre λ_{jk} predstavujú kovarianciu j-tej pozorovanej premennej s k-tým obecným faktorom.

Úlohou faktorovej analýzy je tak na základe kovariančnej matice pozorovaných premenných Σ nájsť jej rozklad daný rovnicou 2.3 s omedzením, že $\boldsymbol{\Psi}$ je diagonálna a q je malé. Pri tvorbe modelu faktorovej analýzy však vzniká niekoľko problémov:

1. identifikovateľnosť modelu
2. odhad parametrov modelu, tj. matíc $\boldsymbol{\Lambda}$ a $\boldsymbol{\Psi}$
3. odhad hodnôt jednotlivých faktorov a ich interpretácia

2.2.1 Interpretácia faktorov

Predpokladajme, že sme našli faktorový model s q obecnými faktormi, ktorý spĺňa predpoklady 2.2. Chceme tak vysvetliť, čo tieto nájdené faktory predstavujú. Ako najjednoduchší spôsob pre interpretáciu sa javí, že sa pozrieme na kovariančnú štruktúru medzi pozorovanými premennými X_j a obecnými faktormi f_k . Tým dostaneme korelačnú maticu R_{Xf} a na jej základe ich už budeme vedieť interpretovať.

Kovariančná matica medzi \mathbf{X} a \mathbf{f} má z predpokladov 2.2 tvar

$$\Sigma_{Xf} = \mathbb{E}(\mathbf{\Lambda f} + \boldsymbol{\epsilon})\mathbf{f}^T = \mathbf{\Lambda}.$$

Korelačná matica R_{Xf} je tak ďalej v tvare

$$P_{Xf} = D^{-\frac{1}{2}}\mathbf{\Lambda}$$

kde $D = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$.

Pre ilustráciu interpretácie obecných faktorov sa môžeme vrátiť k motivačnému príkladu z Hurt (2014), kde sa meralo osem morfológických premenných človeka a tieto premenné sme chceli vysvetliť menším počtom latentných premenných. Ako výsledok sme dostali nasledujúcu tabuľku.

Pozorovaná premenná	Faktorové váhy	
	1.	2.
Výška	0,86	-0,33
Rozpätie rúk	0,85	-0,41
Dĺžka predlaktí	0,81	-0,41
Dĺžka nohy	0,82	-0,34
Váha	0,75	0,56
Obvod stehna	0,64	0,51
Obvod hrudi	0,56	0,49
Šírka hrudi	0,62	0,37

Tabuľka 2.1: Faktorové váhy morfológických premenných

Vidíme, že faktorové váhy všetkých premenných u prvého faktoru sú pomerne veľké a majú rovnaké znamienko. To ukazuje na dôležitosť prvého faktoru a vzhľadom k vecnému významu ho možno považovať za obecný faktor rastu. Ak sa ďalej pozrieme na faktorové váhy u druhého faktoru, tak vidíme, že nadobúda kladné aj záporné hodnoty. Pozitívna korelácia je s váhou, obvodom stehna, obvodom hrudi a šírkou hrudi, teda s premennými, charakterizujúcimi hrúbku tela. Naopak, negatívna korelácia je s výškou, rozpätím rúk, dĺžkou predlaktí a dĺžkou nohy, teda s premennými charakterizujúcimi urastlosť človeka. Druhý obecný faktor tak vzhľadom k vecnému významu možno nazvať faktorom typu tela.

2.2.2 Nejednoznačnosť faktorových váh

Predtým ako si ukážeme problém nejednoznačnosti faktorových váh, tak si zdefinujeme pojem ortogonálnej matice.

Definícia 2. *Nech $Q_{(n \times n)}$ je reálna štvorcová matica, $n \in \mathbb{N}$. Povieme, že matica Q je ortogonálna, ak jej transponovaná matica je súčasne maticou inverznou, tj. platí $Q^T = Q^{-1}$.*

Z definície 2 tak plynie, že $Q^T Q = \mathbb{I}_n$ a pre ľubovoľný vektor $y \in \mathbb{R}^n$ platí: $(Qy)^T(Qy) = y^T y$.

Predpokladajme teda, že $Q_{(q \times q)}$ je reálna ortogonálna matica. Potom \mathbf{X} v 2.1 môžeme vyjadriť aj ako:

$$\mathbf{X} = (\Lambda Q)(Q^T \mathbf{f}) + \epsilon. \quad (2.4)$$

Z 2.4 tak vyplýva, že ak sme našli faktorový model s maticou faktorových váh Λ a q -rozmerným vektorom obecných faktorov \mathbf{f} , tak potom existuje iný faktorový model s maticou faktorových váh ΛQ a q -rozmerným vektorom obecných faktorov $Q^T \mathbf{f}$.

Definícia 3. *Nech \mathbf{f} je q -rozmerný vektor obecných faktorov ortogonálneho faktorového modelu, spĺňajúci podmienky 2.2 a nech $Q_{(q \times q)}$ je ortogonálna matica. Potom pre násobenie vektora \mathbf{f} maticou Q nazývame rotáciou systému osí.*

Tento fakt nejednoznačnosti faktorových váh je podstatný pri ich interpretácii, pretože vhodnou rotáciou vieme dostať maticu faktorových váh $\Lambda Q = (\ddot{\lambda}_{ij})$, ktoré vieme ľahšie interpretovať. Ako sme skôr ukázali, že faktorové váhy vyjadrujú koreláciu medzi faktormi a pôvodnými pozorovanými premennými, tak rotáciou chceme docieľiť, aby každý faktor vysoko pozitívne (negatívne) koreloval s istou skupinou pozorovaných premenných a so zvyšnými pozorovanými premennými nekoreloval (ideálne nulová korelácia). Možno tak povedať, že obecným účelom rotácie je väčšie faktorové váhy zväčšiť a menšie faktorové váhy zmenšiť. Existuje niekoľko rotačných metód:

- Varimax

Rotačná metóda Varimax je tzv. rozptyl maximalizujúca rotácia, ktorá minimalizuje počet premenných, ktoré majú vysoké faktorové váhy s jednotlivými faktormi. Účelová funkcia je daná súčtom mierok rozptýlenosti váh jednotlivých faktorov. Táto mierka je rovná empirickému rozptylu kvadrátov príslušných faktorových záťaží, tj. mierka rozptýlenosti váh r -tého faktoru je v tvare:

$$s_r^2 = \frac{1}{p} \sum_{j=1}^p (\ddot{\lambda}_{jr}^2)^2 - \frac{1}{p^2} \left(\sum_{j=1}^p \ddot{\lambda}_{jr}^2 \right)^2.$$

Sumárna účelová funkcia, ktorú chceme maximalizovať, je tak v tvare

$$\max_Q \sum_{r=1}^q s_r^2$$

- Quartimax

Rotácia Quartimax maximalizuje rozptyl kvadrátov faktorových váh pre jednotlivé pozorované premenné (riadky matice $\mathbf{\Lambda}Q$), tj. mierka rozptýlenosti váh pre r-tú pozorovanú premennú je v tvare

$$s_r^2 = \frac{1}{q} \sum_{j=1}^q \left(\ddot{\lambda}_{rj}^2 \right)^2.$$

Účelová funkcia je tak ďalej v tvare

$$\max_Q \sum_{r=1}^p s_r^2.$$

- Equamax

Rotácia Equamax je kombinácia metód Varimax a Quartimax. Je pri nej minimalizovaný počet premenných, ktoré majú vysoké faktorové váhy s každým spoločným faktorom a taktiež počet faktorov, ktoré sú potrebné pre vysvetlenie pozorovaných premenných.

Treba poznamenať, že vzhľadom k podmienkam 2.2, kedy uvažujeme ortogonálny faktorový model, tak rotácia faktorov, ktorá bola doposiaľ predstavená sa nazýva *ortogonálna rotácia*. Existuje totiž ešte *kosouhlá (šikmá) rotácia*, ktorá pripúšťa korelovanosť faktorov. Je menej používaná ako ortogonálna rotácia, ale jej použitie niekedy lepšie odpovedá realite. Typickými šikmými rotáciami sú Oblimin a Promax. V tejto diplomovej práci budeme predovšetkým pracovať s ortogonálnym faktorovým modelom, tak len pripomenieme, že Promax je výpočtovo rýchlejšie metóda a je tak vhodnejšia pre veľké súbory dát.

Typickým postupom pre eliminovanie nejednoznačnosti faktorových váh je zavedenie dodatočných obmedzení pre matice $\mathbf{\Lambda}$ a $\mathbf{\Psi}$. Chceme teda, aby

$$\mathbf{\Lambda}^T \mathbf{\Psi}^{-1} \mathbf{\Lambda} \text{ bola diagonálna} \quad (2.5)$$

alebo aby

$$\mathbf{\Lambda}^T D^{-1} \mathbf{\Lambda} \text{ bola diagonálna.} \quad (2.6)$$

Pretože cieľom faktorovej analýzy je odhadnúť matice $\mathbf{\Lambda}_{(p \times q)}$, ktorá má pq parametrov a $\mathbf{\Psi}_{(p \times p)}$, ktorá je diagonálna a má p parametrov, tak celkovo odhadujeme $pq + p$ parametrov. Podmienky 2.5 respektíve 2.6 na diagonálnosť matice tak prinášajú $\frac{1}{2}q(q-1)$ obmedzení. Preto počet stupňov volnosti v modeli s q faktormi je

$$\begin{aligned}
d &= (\# \text{ parametrov } \Sigma \text{ bez obmedzení}) - (\# \text{ parametrov } \Sigma \text{ s obmedzeniami}) \\
&= \frac{1}{2}p(p+1) - \left(pq + p - \frac{1}{2}q(q-1)\right) \\
&= \frac{1}{2}(p-q)^2 - \frac{1}{2}(p+q).
\end{aligned}$$

Môžu nastať tri situácie:

- $d < 0$ - model je neurčitý a existuje nekonečne veľa riešení vzťahu 2.1. Táto možnosť nastáva v prípade, že počet parametrov vo faktorovom modeli je väčší ako počet parametrov v pôvodnom modeli, alebo v prípade, že počet faktorov q nie je dostatočne malý relatívne k počtu pozorovaných premenných p
- $d = 0$ - existuje jedinečné riešenie vzťahu 2.1 (okrem rotácií)
- $d > 0$ - najčastejší prípad, ktorý vzniká. Konkrétne a jedinečné riešenie neexistuje, pretože je viac rovníc ako odhadovaných parametrov. Preto sa hľadajú približné riešenia.

Je teda vidieť, že počet stupňov voľností vo faktorovom modeli hrá dôležitú úlohu, pretože nám dáva informáciu o maximálnom počte obecných faktorov q , ktoré môžu dostatočne dobre vysvetliť pozorované premenné.

2.3 Odhad faktorového modelu

V tejto podkapitole si ukážeme jednotlivé metódy odhadu faktorového modelu. Znamená to nájsť odhad $\hat{\Lambda}$ matice faktorových váh Λ a odhad $\hat{\Psi}$ matice špecifických rozptylov Ψ a tým dostať hľadaný faktorový model. Ak označíme \mathbf{S} ako výberovú kovariančnú maticu pozorovaných premenných \mathbf{X} , tj. \mathbf{S} je odhad kovariančnej matice Σ , tak

$$\mathbf{S} = \hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi}$$

Ak teda máme odhad $\hat{\Lambda}$, tak odhad $\hat{\Psi}_j$ j-teho špecifického rozptylu môžeme položiť ako

$$\hat{\Psi}_j = s_{jj} + \sum_{k=1}^q \hat{\lambda}_{jk}^2,$$

kde s_{jk} je odhad kovariancie pozorovaných premenných X_j a X_k . V ideálnom prípade ($d = 0$) dostávame presné riešenie faktorového modelu. V modeli s viac stupňami voľnosti však toto presné riešenie nedostaneme a musíme sa pokúsiť vhodne aproximovať maticu \mathbf{S} odhadmi $\hat{\Lambda}$ a $\hat{\Psi}$.

2.3.1 Metóda hlavných faktorov

Pred predstavením metódy hlavných faktorov si zdefinujeme dôležité pojmy pre použitie tejto metódy.

Definícia 4. *Nech $A_{(n \times n)}$ je štvorcová matica, kde $n \in \mathbb{N}$. Ak existuje vektor $\mathbf{y} \in \mathbb{R}^n \neq \mathbf{0}$ spĺňajúci*

$$A\mathbf{y} = \lambda\mathbf{y},$$

kde λ je nejaký skalár. Potom skalár λ nazveme vlastným číslom matice A a vektor \mathbf{y} nazveme príslušným vlastným vektorom.

Veta 1. *Nech $A_{(n \times n)}$ je štvorcová matica s n lineárne nezávislými vlastnými vektormi \mathbf{y}_i , $i = 1, \dots, n$, kde $n \in \mathbb{N}$. Potom matica A môže byť rozložená ako*

$$A = \mathbf{Y}\mathbf{\Lambda}\mathbf{Y}^{-1}, \quad (2.7)$$

kde $\mathbf{Y}_{(n \times n)}$ je štvorcová matica, ktorej i -ty stĺpec je vlastný vektor \mathbf{y}_i a $\mathbf{\Lambda}$ je diagonálna a na diagonále má príslušné vlastné čísla matice A . Vzťah 2.7 nazývame spektrálnym rozkladom matice A .

Veta 2. *Nech $A_{(n \times n)}$ je štvorcová symetrická matica, kde $n \in \mathbb{N}$. Potom matica A môže byť rozložená ako*

$$A = \mathbf{Y}\mathbf{\Lambda}\mathbf{Y}^T, \quad (2.8)$$

kde $\mathbf{Y}_{(n \times n)}$ je ortogonálna matica, ktorej i -ty stĺpec je vlastný vektor \mathbf{y}_i matice A a $\mathbf{\Lambda}$ je diagonálna a na diagonále má príslušné vlastné čísla matice A . Vzťah 2.8 nazývame spektrálnym rozkladom symetrickej matice A .

Dôkazy viet 1 a 2 je možné nájsť v Bican (2009).

Metóda hlavných faktorov sa zameriava miesto kovariančnej matice Σ na dekompozíciu korelačnej matice \mathbf{R} pozorovaných premenných \mathbf{X} . Nech teda

$$\mathbf{R} = \mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Psi} \quad (2.9)$$

Diagonálne prvky matice \mathbf{R} tak vyzerajú ako:

$$1 = \sum_{r=1}^q \lambda_{ir}^2 + \Psi_i, \quad i = 1, \dots, p.$$

Motiváciou pre použitie tejto metódy je, že ak poznáme maticu $\mathbf{\Psi}$, tak potom z obmedzenia 2.6 na diagonalitu vyplýva, že stĺpce matice $\mathbf{\Lambda}$ tvoria ortogonálne vektory a teda, že sú vlastnými vektormi matice $\mathbf{\Lambda}\mathbf{\Lambda}^T = \mathbf{R} - \mathbf{\Psi}$. Pred tým ako sa dostaneme k algoritmu, ktorý používa táto metóda, zdefinujeme si ďalší potrebný pojem a lemmu.

Definícia 5. *Nech $A_{(n \times n)}$ je štvorcová matica, kde $n \in \mathbb{N}$. Povieme, že matica A je pozitívne semidefinitná, ak pre každý vektor $\mathbf{y} \in \mathbb{R}^n \neq \mathbf{0}$ platí, že $\mathbf{y}^T A \mathbf{y} \geq 0$.*

Lemma 3. *Nech $A_{(n \times n)}$ je pozitívne semidefinitná matica, $n \in \mathbb{N}$. Potom $\|A - \mathbf{y}\mathbf{y}^T\|^2 \geq \|A\|^2 - \lambda_1^2$, $\forall \mathbf{y} \in \mathbb{R}^n$, pričom rovnosť nastáva pre $\mathbf{y} = \sqrt{\lambda_1}\mathbf{y}_1$, kde λ_1 je najväčšie vlastné číslo matice A a \mathbf{y}_1 je príslušný vlastný vektor.*

Dôkaz tejto lemy je možné nájsť v Hurt (2014).

Dôsledkom tejto lemy je, že najlepšou aproximáciou pozitívne semidefinitnej matice A maticou hodnosti jedna je $\lambda_1\mathbf{y}_1\mathbf{y}_1^T$. resp. maticou hodnosti dva je $\lambda_1\mathbf{y}_1\mathbf{y}_1^T + \lambda_2\mathbf{y}_2\mathbf{y}_2^T$ a tak ďalej, kde λ_2 je druhé najväčšie vlastné číslo matice A .

Čiže, ak predpokladáme, že prvých q vlastných čísel je pozitívnych, tak potom môžeme určiť maticu $\mathbf{\Lambda}$ spektrálnym rozkladom v 2.7 matice $\mathbf{\Lambda}\mathbf{\Lambda}^T$ a q by bolo počet faktorov. V praxi sa často používa Kaiserovo kritérium a q sa volí tak, aby $\lambda_q \geq 1$ a $\lambda_{q+1} < 1$. Myšlienkou tohto kritéria je, že pozorované premenné sú väčšinou štandardizované a tak porovnávame rozptyl faktorov (príslušné vlastné čísla) s hodnotou 1. Do faktorovej analýzy sú tak zahrnuté tie faktory, ktorých rozptyl je väčší ako 1.

Hoci je toto kritérium veľmi jednoduché, tak pri rozhodovaní o počte faktorov je neefektívne, problematické a jeho použitie sa neodporúča. Fabrigar a kol. (1999) zhrnul tri problémy, ktoré pri používaní Kaiserovho kritéria môžu vzniknúť.

1. Kaiserovo kritérium bolo pôvodne navrhnuté pre inú mnohorozmernú štatistickú analýzu a to analýzu hlavných komponentov, kde sa brali vlastné čísla z korelačnej matice, ktorá má na diagonále jednotky. V prípade faktorovej analýzy, metódy hlavných faktorov, však na diagonále redukovanej korelačnej matice už nie sú jednotky, ale odhady komunalít pozorovaných premenných.
2. Kritérium taktiež môže viesť k ľubovoľným rozhodnutiam, pretože napríklad nedáva zmysel, ak sa rozhodneme zahrnúť faktor s vlastným číslom 1,01 a na druhej strane v inom modeli faktor s vlastným číslom 0,99 už vylúčime.
3. Posledným problémom je, že na základe simulačných štúdií, toto kritérium ukázalo, že má tendenciu nadhodnocovať počet faktorov a v niektorých prípadoch ich dokonca podhodnocovať (viď Zwick (1986)).

Populárnou súčasťou metódou pre určenie počtu faktorov je Cattellov sutinový test (viď Cattell (1966)), ktorý zahŕňa vizuálne preskúmanie grafickej reprezentácie vlastných čísel. V tejto metóde sú taktiež vlastné čísla zoradené zostupne a sú spojené čiarou. Hľadá sa bod, v ktorom dochádza k poslednému výraznému skoku, inak povedané, kde sa táto čiara ustáli. Podľa tejto metódy tento bod rozdeľuje dôležité a podstatné faktory od tých menej dôležitých. Problémom je však opäť subjektivita metódy, pretože neexistuje žiadna objektívna definícia tohto limitného bodu medzi dôležitými faktormi a faktormi, ktoré sú už bezvýznamné. Treba však poznamenať, že tento test sa ukázal na základe štúdií (Zwick (1986)), že je viac presný a menej premenlivý ako Kaiserovo kritérium.

Podľa štúdií (Humphreys a Montanelli (1975) a Zwick (1986)) je však paralelná analýza najpresnejšou metódou pre rozhodnutie počtu faktorov. Táto metóda je najmenej premenlivá a najmenej citlivá na počet faktorov. Metódu paralelnej analýzy navrhol John L. Horn (viď Horn (1965)), ktorý porovnával pozorované vlastné čísla extrahované z korelačnej matice s vlastnými číslami, ktoré

by sme dostali z náhodného výberu z normálneho rozdelenia. Z výpočtového hľadiska tak ide o Monte-Carlo simuláciu, kde očakávané vlastné čísla dostaneme simuláciou náhodného výberu z normálneho rozdelenia, ktorého veľkosť je rovnaká ako pozorované premenné v zmysle počtu pozorovaní a počtu premenných. Pôvodne sa faktor považoval za dôležitý, ak jeho korešpondujúce vlastné číslo je väčšie ako stredná hodnota vlastného čísla obdržaného z náhodného výberu. V súčasnosti sa však odporúča použiť vlastné číslo, ktoré odpovedá danému kvantilu (najčastejšie 95% kvantilu) odvodeného z náhodného výberu (Glorfeld (1995)).

Dostávame sa tak ďalej k algoritmu metódy hlavných faktorov. Tento algoritmus pozostáva z troch krokov:

1. odhad komunalít
2. odhad faktorových váh (odhad Λ)
3. prípadné iterácie krokov 1 a 2

V popise algoritmu budeme označovať vlastné čísla γ miesto zaužívaného znaku λ kvôli tomu, že už týmto znakom značíme faktorové váhy.

1. odhad komunalít

Na odhad komunalít \hat{h}_j^2 sa využívajú dva prístupy:

(a) odhad komunalít \hat{h}_j^2 je rovný druhej mocnine koeficientu mnohonásobnej korelácie pozorovanej premennej \mathbf{X}_j s pozorovanými premennými \mathbf{X}_k , kde $k \neq j$. To znamená, že $\hat{h}_j^2 = 1 - \hat{\Psi}_j = \rho^2(\mathbf{X}_j, \mathbf{W}\hat{\beta})$, kde $\mathbf{W} = (\mathbf{X}_k)_{k \neq j}$ a $\hat{\beta}$ je OLS odhad lineárnej regresie X_j na \mathbf{W} : $\mathbf{X}_j = \mathbf{W}\beta + \nu$.

(b) odhad komunalít $\hat{h}_j^2 = \max_{k \neq j} |r_{jk}|$, kde $r_{jk} = \text{corr}(X_j, X_k)$

2. odhad faktorových váh (odhad Λ)

Pri odhade faktorových váh si vezmeme redukovanú korelačnú maticu $\hat{\mathbf{R}} = \mathbf{R} - \hat{\Psi}$. Táto matica tak má na diagonále miesto jednotiek, hodnoty $1 - \hat{\Psi}_j$, tj odhady komunalít. Hľadáme tak rozklad $\hat{\mathbf{R}} = \Lambda\Lambda^T$. Z vety 2 o spektrálnom rozklade symetrickej matice tak máme

$$\mathbf{R} - \hat{\Psi} = \sum_{k=1}^p \gamma_k \mathbf{y}_k \mathbf{y}_k^T \quad (2.10)$$

s vlastnými číslami $\gamma_1 \geq \dots \geq \gamma_p$ a príslušnými vlastnými vektormi $\mathbf{y}_1, \dots, \mathbf{y}_p$. Matica $\hat{\mathbf{R}}$ nemusí byť vždy pozitívne semidefinitná a teda aspoň jedno vlastné číslo môže byť záporné. Predpokladajme, že máme q vlastných čísel, ktoré sú pozitívne. Potom z lemy 3 máme

$$\hat{\lambda}_k = \sqrt{\gamma_k} \mathbf{y}_k, \quad k = 1, \dots, q.$$

tj. maticovo

$$\hat{\Lambda} = \mathbf{Y}\mathbf{\Gamma}^{\frac{1}{2}},$$

kde

$$\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_q) \text{ a } \mathbf{\Gamma} = \text{diag}(\gamma_1, \dots, \gamma_q)$$

3. prípadné iterácie krokov 1 a 2

V tomto kroku algoritmu položíme odhad špecifických rozptylov $\hat{\Psi}_j$ ako

$$\hat{\Psi}_j = 1 - \sum_{k=1}^q \hat{\lambda}_{jk}^2, \quad j = 1, \dots, p.$$

Z tohto odhadu vieme ďalej určiť nové odhady komunalít \hat{h}_j^2 , nový odhad redukovanej korelačnej matice $\hat{\mathbf{R}}$ a nový odhad matice faktorových váh $\hat{\Lambda}$. Algoritmus ukončíme vtedy, ak hodnoty $\hat{\Psi}_j$ začnú konvergovať k istej hodnote.

2.3.2 Metóda hlavných komponentov

Zvlášť prípadom metódy hlavných faktorov je faktorové riešenie metódou hlavných komponentov. Metóda hlavných komponentov začína odhadom $\hat{\Lambda}$ matice faktorových váh Λ . Z vety 2 o spektrálnom rozklade symetrickej matice vieme rozložiť výberovú kovariančnú maticu S ako $S = \mathbf{Y}\mathbf{\Gamma}\mathbf{Y}^T$. Potom opäť môžeme odhadnúť maticu Λ pomocou prvých q vlastných vektorov

$$\hat{\Lambda} = [\sqrt{\gamma_1}\mathbf{y}_1, \dots, \sqrt{\gamma_q}\mathbf{y}_q]$$

Odhad špecifických rozptylov potom tvorí diagonála matice $S - \hat{\Lambda}\hat{\Lambda}^T$, tj. $\hat{\Psi}_i = s_{ii} - \sum_{k=1}^q \hat{\lambda}_{ik}^2$, $i = 1, \dots, q$.

2.3.3 Metóda maximálnej vierohodnosti

Použitie metódy maximálnej vierohodnosti vyžaduje od p -rozmerného vektora pozorovaných premenných mnohorozmerné pravdepodobnostné rozdelenie.

Definícia 6. *Nech $\mathbf{Y} = (Y_1, \dots, Y_p)^T$, kde $Y_i \sim \mathcal{N}(0,1)$ sú nezávislé. Nech ďalej $A_{(n \times p)}$ je matica a $\boldsymbol{\mu} \in \mathbb{R}^n$ je pevný vektor. Náhodný vektor definovaný ako $\mathbf{X} = \mathbf{A}\mathbf{Y} + \boldsymbol{\mu}$ potom má n -rozmerné normálne rozdelenie s parametrami $\boldsymbol{\mu}$ a Σ , kde $\Sigma = \mathbf{A}\mathbf{A}^T$. Značíme $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$.*

Predpokladajme tak, že $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$. Majme ďalej N pozorovaní vektora \mathbf{X} . Označme ich $\mathbf{x}_1, \dots, \mathbf{x}_N$. Ak vyjdeme zo znalosti, že hustota \mathbf{X} vzhľadom k Lebesguovej miere na \mathbb{R}^p má tvar

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})} \text{ pre } \mathbf{x} \in \mathbb{R}^p,$$

tak vierohodnostná funkcia $L(\mathbf{X}; \boldsymbol{\mu}; \Sigma)$ bude mať tvar

$$\begin{aligned} L(\mathbf{X}; \boldsymbol{\mu}; \Sigma) &= \prod_{i=1}^N f(\mathbf{x}_i; \boldsymbol{\mu}, \Sigma) \\ &= \frac{1}{(2\pi)^{Np/2}} \frac{1}{|\Sigma|^{N/2}} \exp \left(- \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) / 2 \right). \end{aligned}$$

Zlogaritmovaním dostávame logaritmickú vierohodnostnú funkciu v tvare

$$\ell(\mathbf{X}; \boldsymbol{\mu}, \Sigma) = -\frac{Np}{2} \log(2\pi) - \frac{N}{2} \log(|\Sigma|) - \frac{1}{2} \left(\sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right). \quad (2.11)$$

Pretože ďalej platí, že

$$\begin{aligned} (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) &= (\mathbf{x}_i - \bar{\mathbf{x}})^T \Sigma^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\ &\quad + 2(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}), \end{aligned}$$

tak sčítaním tohoto výrazu cez $i = 1, \dots, N$ dostávame

$$\sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) = \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})^T \Sigma^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) + N(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}). \quad (2.12)$$

Pretože zjavne platí, že

$$\begin{aligned} (\mathbf{x}_i - \bar{\mathbf{x}})^T \Sigma^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) &= \text{tr} \{ (\mathbf{x}_i - \bar{\mathbf{x}})^T \Sigma^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \} \\ &= \text{tr} \{ \Sigma^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T \} \end{aligned}$$

tak môžeme ďalej výraz 2.12 rozpísať ako

$$\begin{aligned} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) &= \text{tr} \left\{ \Sigma^{-1} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T \right\} + \\ &\quad + N(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\ &= \text{tr} \{ \Sigma^{-1} NS \} + N(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \quad (2.13) \end{aligned}$$

Dosadením 2.13 do 2.11 dostávame

$$\begin{aligned}
\ell(\mathbf{X}; \boldsymbol{\mu}, \Sigma) &= -\frac{Np}{2} \log(2\pi) - \frac{N}{2} \log(|\Sigma|) - \frac{N}{2} \text{tr}\{\Sigma^{-1}S\} + \\
&+ \frac{N}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\
&= -\frac{N}{2} [p \log(2\pi) + \log(|\Sigma|) + \text{tr}\{\Sigma^{-1}S\} - (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})].
\end{aligned}$$

Pre konzistentný a nestranný odhad $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ dostávame

$$\ell(\mathbf{X}; \hat{\boldsymbol{\mu}}, \Sigma) = -\frac{N}{2} [\log(|(2\pi)^p \Sigma|) + \text{tr}\{\Sigma^{-1}S\}]. \quad (2.14)$$

Nahradením $\Sigma = \mathbf{\Lambda}\mathbf{\Lambda}^T + \boldsymbol{\Psi}$ vo výraze 2.14 dostávame

$$\ell(\mathbf{X}; \hat{\boldsymbol{\mu}}, \mathbf{\Lambda}, \boldsymbol{\Psi}) = -\frac{N}{2} [\log\{|(2\pi)^p(\mathbf{\Lambda}\mathbf{\Lambda}^T + \boldsymbol{\Psi})|\} + \text{tr}\{(\mathbf{\Lambda}\mathbf{\Lambda}^T + \boldsymbol{\Psi})^{-1}S\}]. \quad (2.15)$$

Nájdenie matíc $\mathbf{\Lambda}$ a $\boldsymbol{\Psi}$ maximalizujúce 2.15 však vôbec nie je jednoduché a je potrebné použiť iteratívne numerické algoritmy.

Metóda maximálnej vierohodnosti taktiež umožňuje testovať hypotézu, či q -faktorový model je vhodný pre vysvetlenie napozorovanej kovariančnej štruktúry pozorovaných premenných. Testovaná hypotéza má tvar

$$H_0 : \Sigma = \mathbf{\Lambda}\mathbf{\Lambda}^T + \boldsymbol{\Psi}.$$

$$H_1 : \Sigma \text{ nie je obmedzená}$$

Označme si

$$L_0^* = \max_{\Sigma \in \mathbb{R}^{p \times p}} L(\mathbf{X}; \boldsymbol{\mu}, \Sigma) \quad (2.16)$$

$$L_1 = L(\mathbf{X}; \boldsymbol{\mu}, \Sigma), \quad (2.17)$$

kde v 2.16 je maximum vierohodnostnej funkcie pre nulovú hypotézu a v 2.17 je hodnota vierohodnostnej funkcie pre výberovú kovariančnú maticu Σ . Nech ďalej $\hat{\mathbf{\Lambda}}$ a $\hat{\boldsymbol{\Psi}}$ sú maximálne vierohodné odhady matíc $\mathbf{\Lambda}$ a $\boldsymbol{\Psi}$ a teda $L_0^* = L(\mathbf{X}; \hat{\boldsymbol{\mu}}, \hat{\mathbf{\Lambda}}, \hat{\boldsymbol{\Psi}})$.

Uvažujme ďalej vierohodnostný pomer

$$\lambda_N^* = \frac{L_0^*}{L_1}. \quad (2.18)$$

Je zjavné, že môžeme mať sklon nezamietnuť testovanú hypotézu H_0 , ak λ_N^* je veľké a naopak ju zamietnuť, ak je λ_N^* príliš malé. Test pomerom vierohodností nám hovorí, kedy presne máme uprednostniť H_0 pred H_1 .

Ak si vierohodnostný pomer 2.18 vyjadríme pomocou logaritmickej vierohodnostnej funkcie, tak máme

$$-2 \log \lambda_N^* = 2 (\ell_1 - \ell_0^*)$$

a za platnosti nulovej hypotézy platí, že

$$-2 \log \lambda_N^* \xrightarrow{D} \chi_{\frac{1}{2}\{(p-q)^2 - p - q\}}^2 \quad \text{pre } N \rightarrow \infty$$

Hypotézu H_0 tak zamietame v prospech H_1 , ak

$$2 \log \lambda_N^* \geq \chi_{\frac{1}{2}\{(p-q)^2 - p - q\}}^2 (1 - \alpha),$$

kde α je hladina testu.

2.3.4 Metóda najmenších štvorcov

V prípade, že dáta nepochádzajú z mnohorozmerného normálneho rozdelenia je metóda najmenšou štvorcov vítanou metódou, pretože dáva väčšinou rovnaké výsledky ako metóda maximálnej vierohodnosti.

Metóda najmenších štvorcov sa snaží dosiahnuť takú maticu faktorových váh $\mathbf{\Lambda}$ a maticu špecifických rozptylov $\mathbf{\Psi}$, ktoré minimalizujú súčet štvorcov reziduálnych korelačných koeficientov, ktoré neležia na diagonále (diagonálne prvky, ako súčet komunality a špecifického rozptylu, sú rovné jednej). Minimalizujeme tak funkciu

$$\phi(\mathbf{\Lambda}, \mathbf{\Psi}) = \frac{1}{2} \mathbf{S} \mathbf{S} \mathbf{Q} (\mathbf{R} - \mathbf{\Lambda} \mathbf{\Lambda}^T - \mathbf{\Psi})^2$$

cez všetky matice $\mathbf{\Lambda} \in \mathbb{R}^{p \times q}$ a $\mathbf{\Psi} \in \mathbb{D}^p$, diagonálne matice rádu p . Na výpočet tejto optimalizačnej úlohy existuje niekoľko rôznych algoritmov, ako napríklad Thompsonov algoritmus, Comreyho algoritmus a Harmanov algoritmus. Všetky tieto algoritmy sú zhrnuté v Chen (2010).

2.3.5 Zobecnená metóda najmenších štvorcov

Zobecnená metóda najmenších štvorcov je obdobná metóde najmenších štvorcov. Taktiež sú faktory extrahované tak, aby bol minimalizovaný súčet štvorcov reziduálnych korelačných koeficientov, ktoré neležia na diagonále korelačnej matice. Reziduálne korelačné koeficienty sú pritom ale prevážené tak, že korelácie premenných, ktoré majú väčší špecifický rozptyl, majú nižšiu váhu ako korelácie premenných s menšími špecifickými rozptylmi. Preto sa metóda najmenších štvorcov často v štatistických softvéroch nazýva ako nevážená metóda najmenších štvorcov.

Definícia 7. *Nech $\mathbf{X}_1, \dots, \mathbf{X}_n$ je náhodný výber z normálneho p -rozmerného rozdelenia s parametrami $\mathbf{0}$ a $\mathbf{\Sigma}$, $X_i \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$. Nech ďalej $\mathbf{X}^T = (\mathbf{X}_1, \dots, \mathbf{X}_n)$. Potom $\mathbf{W} = \mathbf{X}^T \mathbf{X}$ má Wishartovo rozdelenie s n stupňami voľnosti. Značíme $\mathbf{W} \sim \mathbf{W}_p(\mathbf{\Sigma}, n)$.*

Myšlienkou zobecnenej metódy najmenších štvorcov, že za predpokladu, že \mathbf{X} má mnohorozmerné normálne rozdelenie, $X \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$, tak $S = \mathbf{X}^T \mathbf{X}$ má Wishartovo rozdelenie, $S \sim W_p(\Sigma, N)$, kde N je veľkosť náhodného výberu. Priamym aplikovaním Aitkenovho princípu zobecnenej metódy najmenších štvorcov, tak máme účelovú funkciu, ktorú chceme minimalizovať v tvare

$$\phi(\mathbf{\Lambda}, \mathbf{\Psi}) = \frac{1}{2} S S Q [\Sigma^{-1}(S - (\mathbf{\Lambda}\mathbf{\Lambda}^T - \mathbf{\Psi}))^2 \quad (2.19)$$

V praxi avšak nepoznáme maticu Σ a tak berieme S ako jej odhad. Nahradením matice Σ v 2.19 jej odhadom S , tak máme

$$\begin{aligned} \phi(\mathbf{\Lambda}, \mathbf{\Psi}) &= \frac{1}{2} S S Q [S^{-1}(S - \mathbf{\Lambda}\mathbf{\Lambda}^T - \mathbf{\Psi})]^2 \\ &= \frac{1}{2} S S Q [\mathcal{I}_p - S^{-1}(\mathbf{\Lambda}\mathbf{\Lambda}^T - \mathbf{\Psi})]^2 \end{aligned}$$

Je treba podotknúť, že na rozdiel od metódy najmenších štvorcov, tak zobecnená metóda pracuje s kovariančnou maticou, miesto korelačnej a tak pred výpočtom kovariančnej matice musia byť pozorované premenné standardizované. Algoritmus pre výpočet tejto optimalizačnej úlohy je možné nájsť v Jöreskog a Goldberger (1972).

2.4 Faktorové skóre

Doteraz sme sa v práci venovali iba odhadu faktorových váh a komunalít, respektíve ich interpretácií. Často nás však aj zaujímajú jednotlivé hodnoty spoločných obecných faktorov. Tieto hodnoty sa nazývajú *faktorové skóre*. Konkrétnejšie, faktorové skóre sú odhady nepozorovaných náhodných vektorov \mathbf{f}_k , $k = 1, \dots, q$, pre každého jednotlivého jednotlivca \mathbf{x}_i , $i = 1, \dots, n$. Doteraz boli prezentované tri metódy odhadu, ktoré v praxi prinášajú veľmi podobné výsledky. Treba taktiež poznamenať, že faktorové skóre je možné počítat z ľubovoľného riešenia faktorových váh. Rotáciou faktorových váh zostávajú výpočetné vzorce rovnaké. V tejto práci si predstavíme najpoužívanejšiu regresnú metódu.

2.4.1 Regresná metóda

Regresná metóda je považovaná za najjednoduchšiu metódu a je nenáročnú ju implementovať. Bez ujmy na obecnosti opäť predpokladajme, že \mathbf{X} je centrovanej. Inak by sme miesto \mathbf{X} písali $\mathbf{X} - \boldsymbol{\mu}$.

Veta 4. *Nech $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$, $X_1 \in \mathbb{R}^r$, $X_2 \in \mathbb{R}^{p-r}$. Nech ďalej $X_{2.1} = X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1$, kde*

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Potom

$$\begin{aligned} X_1 &\sim \mathcal{N}_r(\mu_1, \Sigma_{11}), \\ X_{2,1} &\sim \mathcal{N}_{p-r}(\mu_{2,1}, \Sigma_{22,1}) \end{aligned}$$

sú nezávislé a

$$\mu_{2,1} = \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1, \quad \Sigma_{22,1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

Dôkaz vety je možné nájsť v Härdle (2007).

Majme teda združené rozdelenie \mathbf{X} a \mathbf{f} . Kovariančná matica má tvar

$$\text{Var} \begin{pmatrix} \mathbf{X} \\ \mathbf{f} \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Psi} & \mathbf{\Lambda} \\ \mathbf{\Lambda}^T & \mathbb{I}_q \end{pmatrix}. \quad (2.20)$$

Za predpokladu, že (\mathbf{X}, \mathbf{f}) má združené mnohorozmerné normálne rozdelenie, tak podmienené rozdelenie $\mathbf{f}|\mathbf{X}$ je taktiež mnohorozmerné normálne. Regresná metóda tak hľadá lineárnu funkciu pozorovaných hodnôt

$$\hat{\mathbf{f}}_j = \boldsymbol{\beta}_j^T \mathbf{x},$$

kde $\boldsymbol{\beta}_j$ je vektor typu $p \times 1$. Metódou najmenších štvorcov tak dostaneme odhad j-teho spoločného faktoru $\hat{\mathbf{f}}_j$ v tvare $\hat{\mathbf{f}}_j = \boldsymbol{\lambda}_j(\mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Psi})^{-1}\mathbf{x}$, kde $\boldsymbol{\lambda}_j$ označuje j-ty stĺpec matice $\mathbf{\Lambda}$. Celkovo tak vektor odhadov $\hat{\mathbf{f}}_1, \dots, \hat{\mathbf{f}}_q$ má tvar

$$\hat{\mathbf{f}} = \mathbf{\Lambda}^T(\mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Psi})^{-1}\mathbf{X}$$

V praxi však nahradzujeme neznáme $\mathbf{\Lambda}$ a $\mathbf{\Sigma}$ ich príslušnými odhadmi. Taktiež aby sme eliminovali chybu určenia počtu faktorov v modeli, sú faktorové skóre počítané priamo z výberovej kovariančnej matice \mathbf{S} . Vektor odhadu faktorových skóre q obecných faktorov pre i -teho jednotlivca x_i tak je

$$\hat{\mathbf{f}}_i = \hat{\mathbf{\Lambda}}^T \mathbf{S}^{-1} \mathbf{x}_i. \quad (2.21)$$

Ak boli odhady matíc počítané z korelačnej matice \mathbf{R} miesto kovariančnej matice \mathbf{S} , tak výraz 2.20 je platí iba ak miesto \mathbf{X} uvažujeme štandardizované premenné $\mathbf{Z} = \mathbf{D}^{-1/2}\mathbf{X}$. Potom odhady faktorov v 2.21 majú tvar

$$\hat{\mathbf{f}}_i = \hat{\mathbf{\Lambda}}^T \mathbf{R}^{-1} \mathbf{z}_i \quad (2.22)$$

Kapitola 3

Faktorová analýza EFW indexu

Táto kapitola bude venovaná faktorovej analýze EFW indexu, ktorý publikoval Fraser Institute v roku 2014 a hodnoty indexu ukazovali ekonomickú slobodu krajín za rok 2012. Predtým, ako sa dostaneme k faktorovej analýze, pomocou ktorej budeme hľadať, aké zdrojové faktory ho tvoria a koľko ich je, sa presvedčíme, či sú dáta vhodné na použitie faktorovej analýzy.

3.1 Predpoklady pre faktorovú analýzu

Vhodné predpoklady pre použitie faktorovej analýzy sú:

1. normalita

Štatistický záver je presnejší, ak dáta pochádzajú z mnohorozmerného normálneho rozdelenia. Je to taktiež nutná podmienka pre použitie metódy maximálnej vierohodnosti, ktorá bola predstavená v kapitole 2.3.3.

2. lineárny vzťah medzi pozorovanými premennými

3. dostatočne veľký počet pozorovaní

Na nájdenie spoľahlivých odhadov korelácií medzi pozorovanými premennými je potrebný dostatočný počet pozorovaní. Pomer počtu pozorovaní N k počtu obecných faktorov q by mal byť dostatočne veľký. V ideálnom prípade je $\frac{N}{q} > 20$. Faktorovú analýzu je stále rozumné použiť, ak $\frac{N}{q} > 5$.

4. faktorovateľnosť

Tento predpoklad pojednáva o tom, že medzi pozorovanými premennými by mal byť istý stupeň kolinearita, aby sme mohli identifikovať spoločné obecné faktory. Tento stupeň by však nemal byť až príliš vysoký, aby nevznikol problém singularita dát. O platnosti tohto predpokladu je možné rozhodnúť na základe Kaiser-Myer-Olkinovom (KMO) kritéria a takisto aj na základe Bartlettovho testu sféricity.

- KMO kritérium

Hodnota KMO kritéria je v tvare

$$KMO = \frac{\sum \sum_{j \neq k} r_{jk}^2}{\sum \sum_{j \neq k} r_{jk}^2 + \sum \sum_{j \neq k} g_{jk}^2}, \quad (3.1)$$

kde g_{jk} je parciálna derivácia medzi j -tou a k -tou pozorovanou premennou. Kritérium indikuje, či sú parciálne derivácie medzi pozorovanými premennými malé. KMO kritérium z definície nadobúda hodnoty od 0 do 1 a čím je bližšie k 1, tým viac indikuje vhodnosť použitia faktorovej analýzy. Hodnota KMO kritéria by tak mala byť väčšia aspoň ako 0.5. Taktiež možno na základe KMO kritéria zistiť stupeň kolinearít pre jednotlivé pozorované premenné. Pre pozorovanú premennú X_j je v tvare

$$KMO_j = \frac{\sum_{\substack{k=1 \\ k \neq j}}^p r_{jk}^2}{\sum_{\substack{k=1 \\ k \neq j}}^p r_{jk}^2 + \sum_{\substack{k=1 \\ k \neq j}}^p g_{jk}^2}.$$

Ak $KMO_j < 0.5$, tak pozorovaná premenná X_j má príliš vysoký stupeň kolienarity a môže byť vyjadrená ako lineárna kombinácia iných pozorovaných premenných.

- Bartlettov test sférickosti

Bartlettov test sférickosti porovnáva korelačnú maticu pozorovaných premenných \mathbf{R} k identickej matici. Ak sú pozorované premenné perfektne korelované, tak nám postačuje jeden faktor. Ak sú ortogonálne, tak počet faktorov je rovný počtu pozorovaných premenných. Nulová hypotéza tak je, že pozorované premenné sú ortogonálne. V prípade zamietnutia nulovej hypotézy je použitie faktorovej analýzy relevantné. Pre účely Bartlettovho testu sférickosti je dôležitý determinant matice \mathbf{R} . Za platnosti nulovej hypotézy je $|\mathbf{R}| = 1$. Ak naopak sú pozorované premenné silno korelované, tak $|\mathbf{R}| \approx 0$.

Testová štatistika tak indikuje, ako veľmi sa korelačná matica odchyľuje od referenčnej situácie $|\mathbf{R}| = 1$. Testová štatistika má tvar

$$\chi^2 = - \left(N - 1 - \frac{2p + 5}{6} \right) \cdot \ln |\mathbf{R}|.$$

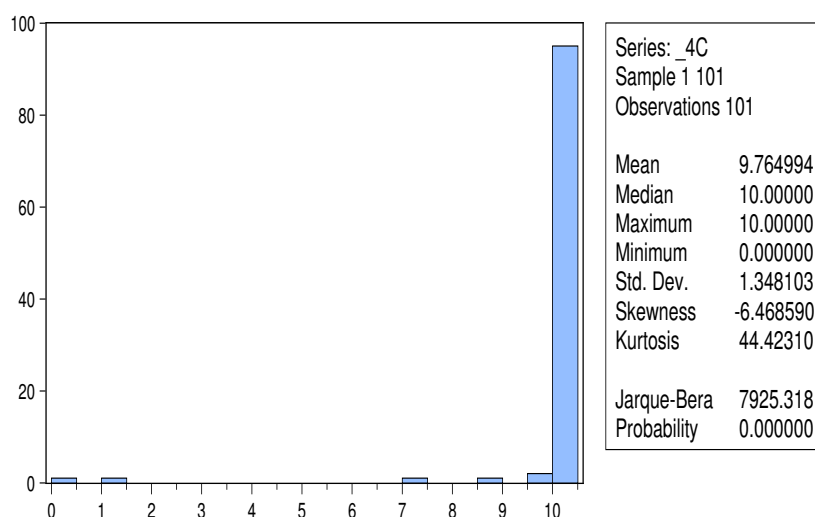
Za platnosti nulovej hypotézy má testová štatistika chí-kvadrát rozdelenie s $\frac{p(p-1)}{2}$ stupňami voľnosti.

Ako bolo už spomenuté v prvej kapitole, po očistení dát o krajiny, ktorým chýba aspoň jeden indikátor, poznáme hodnoty EFW indexu a všetkých indikátorov pre 101 krajín. Taktiež treba poznamenať, že sme sa rozhodli z pozorovaných 42 indikátorov vylúčiť tie indikátory, ktoré boli ordinálne a teda v nich už bola istá strata informácie. Bolo to týchto osem indikátorov:

- 1C štátne podniky a investície

- 1Di daň z príjmu
- 1Dii daň zo mzdy
- 3D sloboda vlastniť bankový účet v cudzej mene
- 5Ai vlastníctvo bánk
- 5Aiii kontrola úrokových sadzieb
- 5Biv regulácia pracovnej doby
- 5Bvi odvody

Taktiež sme sa rozhodli vylúčiť indikátor o cenách na čiernom trhu (4C), pretože indikátor nadobúda okrem šiestich krajín výlučne jednu hodnotu a to najvyššiu, vid' histogram 3.1. Môžeme preto aj povedať, že indikátor neprináša žiadnu novú informáciu do modelu. Pre model faktorovej analýzy tak máme celkovo 33 indikátorov (pozorovaných premenných) pre 101 krajín (počet pozorovaní).



Obr. 3.1: Histogram (4C) indikátoru - ceny na čiernom trhu

Ďalším krokom tak je overiť predpoklady pre vhodnosť použitia faktorovej analýzy.

3.1.1 Overenie predpokladov faktorovej analýzy

- Normalita

Pre otestovanie toho, či dáta pochádzajú z mnohorozmerného normálneho rozdelenia využijeme Mardia test. Testovaná hypotéza je, že dáta pochádzajú z mnohorozmerného normálneho rozdelenia. Na základe p-hodnoty Mardia testu a Q-Q grafu (obrázok 3.2) tak zamietame hypotézu a vidíme, že pozorované indikátory nepochádzajú z mnohorozmerného rozdelenia.

Mardia's Multivariate Normality Test

```

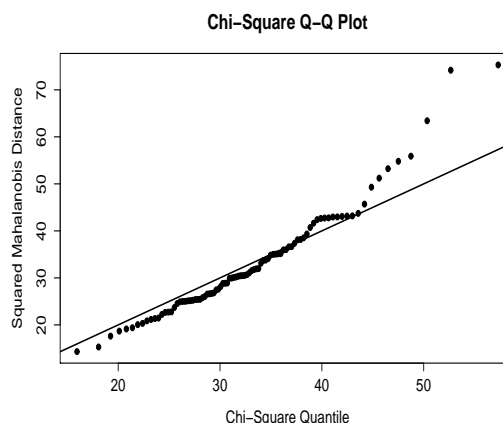
data : Indicators

g1p      : 549.1188
skew     : 8419.822
p.value.skew : 1.825455e-51

g2p      : 1218.258
kurtosis : 6.312065
p.value.kurt : 2.75336e-10

small.skew : 8710.911
p.value.small : 1.969135e-66
    
```

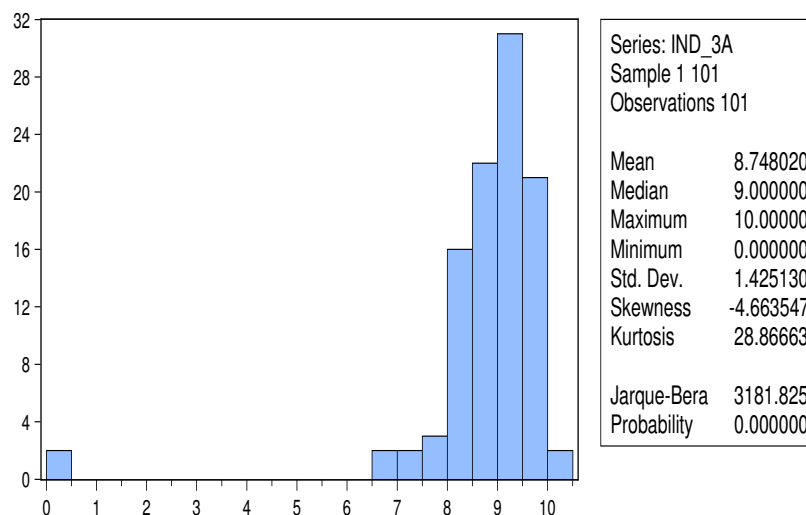
(a) Mardia test



(b) Q-Q graf

Obr. 3.2: Testovanie normality

Podobný záver sme aj očakávali, pretože pre väčšinu pozorovaných indikátorov sme zamietli hypotézu normality Jarque-Bera testom na 95% hladine kvôli vyšším hodnotám špicatosti a nižším hodnotám šikmosti oproti normálnemu rozdeleniu (viď pre príklad obrázok 3.3). Medzi pozorovanými indikátormi sa však našli aj niektoré indikátory, ktorým Jarque-Bera test hypotézu normality nezamietol, ako napríklad indikátor regulácie najímania a prepúšťania pracovnej sily (obrázok 3.4).

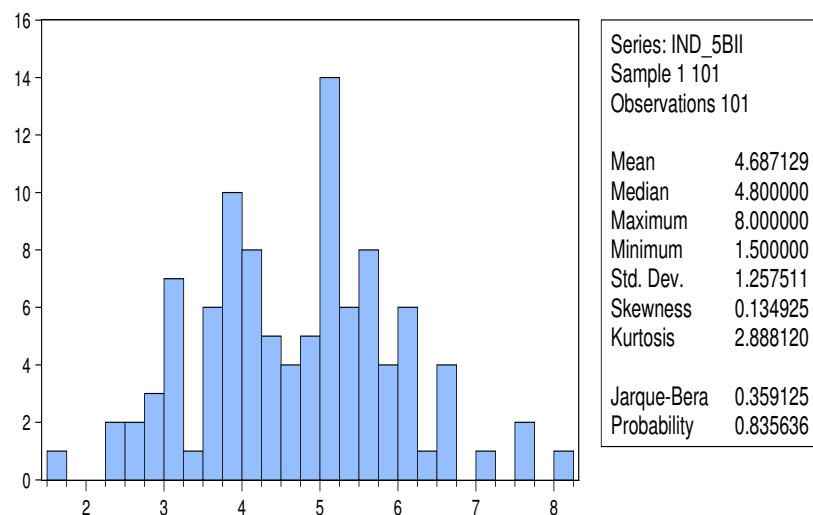


Obr. 3.3: Histogram a test normality (3A) indikátoru - rast miezd

Z nenormality dát tak vyplýva, že obecné spoločné faktory nebudeme môcť odhadovať metódou maximálnej vierohodnosti.

- Lineárne vzťahy medzi pozorovanými premennými

Na základe korelačnej matice môžeme hneď vidieť, či sú medzi pozorovanými premennými lineárne vzťahy, respektíve ako silné sú a ktoré pozo-



Obr. 3.4: Histogram a test normality (5Bii) indikátoru - regulácia najímania a prepúšťania pracovnej sily

rované premenné sa javia ako vhodní kandidáti na nahradenie spoločným faktorom. Z časti korelačnej matice, ktorá je na obrázku 3.5 vidíme, že vysokú pozitívnu koreláciu medzi sebou majú indikátory neustrannosti súdov (2B), ochrany vlastnických práv (2C) a spoľahlivosti polície (2H). Je tu teda veľký predpoklad, že tieto tri indikátory bude možné nahradiť jedným spoločným faktorom. Na druhej strane, napríklad indikátory rastu miezd (3A) a regulačných obmedzení predaja nehnuteľností (2G) nemajú výraznú koreláciu so žiadnym indikátorom a tak ich zrejme nebude možné spojiť s inými indikátormi a tým nahradiť spoločným faktorom.

	1A	1B	2A	2B	2C	2D	2E	2F	2G	2H	2I	3A
1A	1	0.56	-0.5	-0.31	-0.43	-0.53	-0.58	-0.36	-0.12	-0.5	-0.45	-0.05
1B	0.56	1	-0.38	-0.14	-0.38	-0.63	-0.62	-0.39	-0.2	-0.48	-0.48	-0.19
2A	-0.5	-0.38	1	0.87	0.9	0.57	0.67	0.4	0.17	0.83	0.5	0.18
2B	-0.31	-0.14	0.87	1	0.88	0.37	0.54	0.38	0.15	0.77	0.46	0.14
2C	-0.43	-0.38	0.9	0.88	1	0.59	0.68	0.46	0.2	0.88	0.57	0.32
2D	-0.53	-0.63	0.57	0.37	0.59	1	0.65	0.41	0.26	0.62	0.5	0.21
2E	-0.58	-0.62	0.67	0.54	0.68	0.65	1	0.57	0.2	0.78	0.75	0.21
2F	-0.36	-0.39	0.4	0.38	0.46	0.41	0.57	1	0.45	0.5	0.54	0.13
2G	-0.12	-0.2	0.17	0.15	0.2	0.26	0.2	0.45	1	0.21	0.2	0.07
2H	-0.5	-0.48	0.83	0.77	0.88	0.62	0.78	0.5	0.21	1	0.74	0.25
2I	-0.45	-0.48	0.5	0.46	0.57	0.5	0.75	0.54	0.2	0.74	1	0.12
3A	-0.05	-0.19	0.18	0.14	0.32	0.21	0.21	0.13	0.07	0.25	0.12	1

Obr. 3.5: Časť korelačnej matice indikátorov

- Dostatočne veľký počet pozorovaní

Ako bolo spomenuté v prvej kapitole, tak poznáme kompletný súbor dát EFW indexu pre 101 krajín. To značí, že pomer medzi počtom pozorovaní a počtom faktorov bude ideálny pre päť a menej faktorov. Avšak, nepredpokladáme, že budeme odhadovať model pre desať a viac faktorov, preto je tento počet pozorovaní pre faktorovú analýzu dostatočný.

- Faktorovateľnosť

Na základe hodnôt KMO kritéria, ktorého hodnota je 0,85 vidíme, že v dátach je dostatočný stupeň kolinearitu a že sú dobre faktorovateľné. Avšak, z napočítaného KMO kritéria pre jednotlivé indikátory (tabuľka 3.1) si môžeme všimnúť, že indikátor - štandardná odchýlka tarifných sadziieb (4Aiii) má KMO kritérium pod hraničnou hodnotou 0,5 čo značí, že premenná prináša do modelu privysokú kolinearitu a tak zrejme nebude vhodným kandidátom do faktorového modelu.

Indikátor	KMO
1A	0.84
1B	0.88
2A	0.89
2B	0.86
2C	0.90
2D	0.87
2E	0.94
2F	0.89
2G	0.76
2H	0.90
2I	0.90
3A	0.75
3B	0.55
3C	0.73
4Ai	0.86
4Aii	0.77
4Aiii	0.30
4Bi	0.89
4Bii	0.90
4Di	0.83
4Dii	0.81
4Diii	0.80
5Aii	0.60
5Bi	0.76
5Bii	0.67
5Biii	0.64
5Bv	0.86
5Ci	0.79
5Cii	0.90
5Ciii	0.83
5Civ	0.93
5Cv	0.89
5Cvi	0.87

Tabuľka 3.1: KMO kritérium pre jednotlivé indikátory

Taktiež náš záver o vhodnosti faktorovateľnosti potvrdzuje Bartlettov test sféricity (tabuľka 3.2), ktorý jednoznačne zamieta hypotézu ortogonalitu

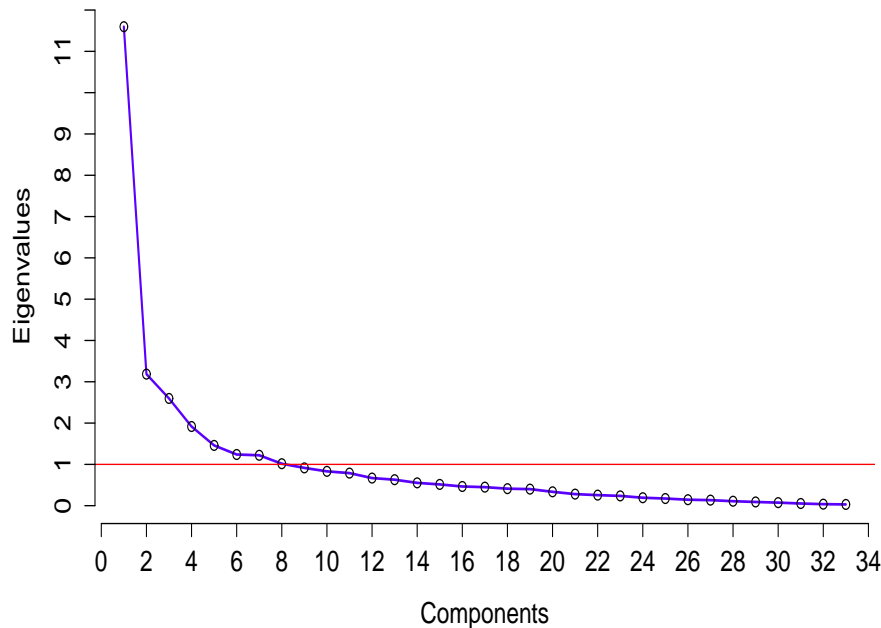
pozorovaných premenných.

Hodnota testovej štatistiky	2598.65
Počet stupňov volnosti	528
P-hodnota	0.00

Tabuľka 3.2: Bartlettov test sférickosti

3.2 Hľadanie počtu faktorov

Pri hľadaní vhodného modelu faktorovej analýzy nás najprv zaujíma, aký je optimálny počet faktorov, ktoré dostatočne dobre vysvetľujú pozorované premenné. Pre tento účel porovnáme kritéria opísané v kapitole 2.3.1.



Obr. 3.6: Kaiserovo kritérium

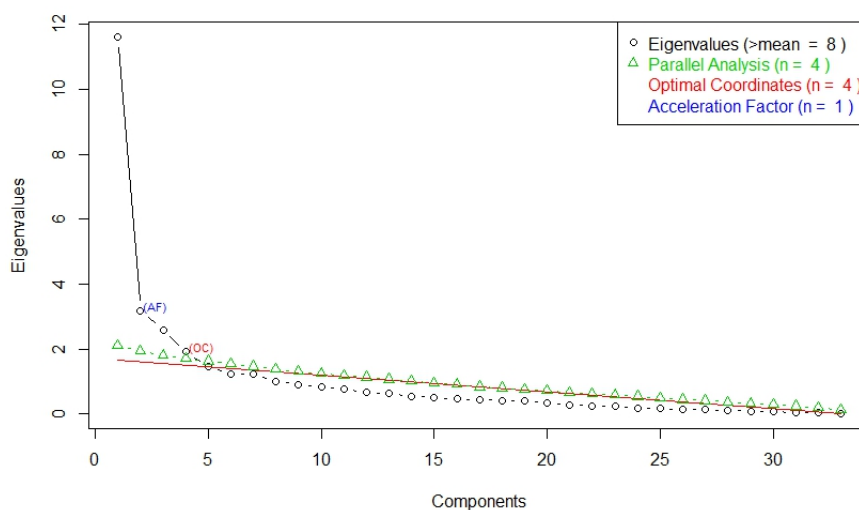
Na ako prvé sa pozrieme Kaiserovo kritérium, podľa ktorého by sme mali do modelu zahrnúť iba tie faktory, ktorých príslušné vlastné čísla sú väčšie ako 1. Z obrázku 3.6 a tabuľky 3.3 vidíme, že podľa Kaiserovho kritéria by sme mali do modelu zahrnúť osem faktorov.

Ak by sme sa rozhodli použiť Cattellov sutinový test, tak vidíme, že je ťažké určiť bod useknutia, od ktorého faktory už nie sú významné. Ako vhodný kandidát na bod useknutia sa predsa len javí skok medzi vlastným číslom siedmeho a ôsmeho faktora, od ktorého hodnoty vlastného čísla začínajú konvergovať k svojej limitnej hodnote.

Ako už však bolo spomenuté v podkapitole 2.3.1, tak najpresnejšie kritérium na určenie počtu faktorov je Hornova paralelná analýza. Z obrázku 3.7 vidíme, že

Faktor	Vlastné číslo
1	11.62
2	3.19
3	2.60
4	1.92
5	1.46
6	1.24
7	1.21
8	1.02
9	0.91
10	0.83
11	0.78
12	0.67
13	0.63
14	0.55
15	0.51
16	0.46
17	0.45
18	0.41
19	0.40
20	0.34

Tabuľka 3.3: Vlastné čísla príslušné faktorom



Obr. 3.7: Hornova paralelná analýza s hladinou spoľahlivosti 95%

toto kritérium odporúča použiť štyri faktory, pretože ešte štvrté najväčšie vlastné číslo leží nad hodnotou 95%-ného kvantilu získaného Monte-Carlo simuláciou.

3.3 Odhad faktorového modelu

Predtým ako začneme odhadovať vhodný model faktorovej analýzy, sa musíme ešte rozhodnúť, aké faktorové váhy budeme považovať za dôležité a taktiež koľko percent rozptylu pozorovaných premenných chceme, aby model vysvetľoval. Pre účely tejto diplomovej práce budeme považovať rotované faktorové váhy $|\lambda_{jk}| \geq 0,4$ za dôležité. Niektorí výskumníci používajú dokonca oveľa striktnějšíe pravidlo $|\lambda_{jk}| \geq 0,7$. Obmedzenie, ktoré sme vybrali pre túto diplomovú prácu však považujeme za dostatočne dobré.

Odhadnutý model je možné považovať za dostatočne dobrý, ak celkové percento vysvetleného rozptylu pozorovaných premenných je v intervale od 50% do 75%. Samozrejme, že pre model je prínosnejšie, ak je vysvetlené percento rozptylu čo najvyššie, avšak netreba zabudnúť na to, že cieľom faktorovej analýzy je predovšetkým zachytenie kovariančnej štruktúry pozorovaných premenných a ich vysvetlenie pomocou čo najmenšieho počtu faktorov.

Faktor	Vlastné čísla		
	Hodnota	% rozptylu	Kumulatívne %
1	11.62	35.22	35.22
2	3.12	9.65	44.86
3	2.60	7.87	52.74
4	1.92	5.82	58.55
5	1.46	4.43	62.98
6	1.24	3.74	66.73
7	1.21	3.68	70.40
8	1.02	3.08	73.48
9	0.91	2.76	76.24
10	0.83	2.50	78.75

Tabuľka 3.4: Desiat najväčších vlastných čísel a ich podiel na vysvetlenom rozptyle pozorovaných premenných

Pre ilustráciu, v tabuľke 3.4 môžeme vidieť, ako dobre aproximuje desiat najväčších vlastných čísel (vlastných vektorov) korelačnej matice \mathbf{R} pozorované premenné. Ak by sme napríklad aproximovali maticu \mathbf{R} pomocou jej desiatich najväčších vlastných čísel (vlastných vektorov), tak by nám najviac zostalo približne 21.25% nevysvetleného rozptylu pozorovaných premenných, keďže súčet druhých mocnín zvyšných vlastných čísel zhora ohraničuje súčet štvorcov elementov reziduálnej korelačnej matice, tj.

$$\sum_{i,j} (\mathbf{R} - \hat{\Lambda}\hat{\Lambda}^T - \hat{\Psi})_{ij}^2 \leq \gamma_{q+1}^2 + \dots + \gamma_p^2. \quad (3.2)$$

Výraz 3.2 tak implikuje, že podiel z celkového rozptylu vysvetlovaných premenných vzhľadom k j-temu faktoru je

- $\gamma_j / \sum_{j=1}^p s_{jj}$, ak vychádza faktorová analýza z kovariančnej matice
- γ_j / p , ak vychádza faktorová analýza z korelačnej matice

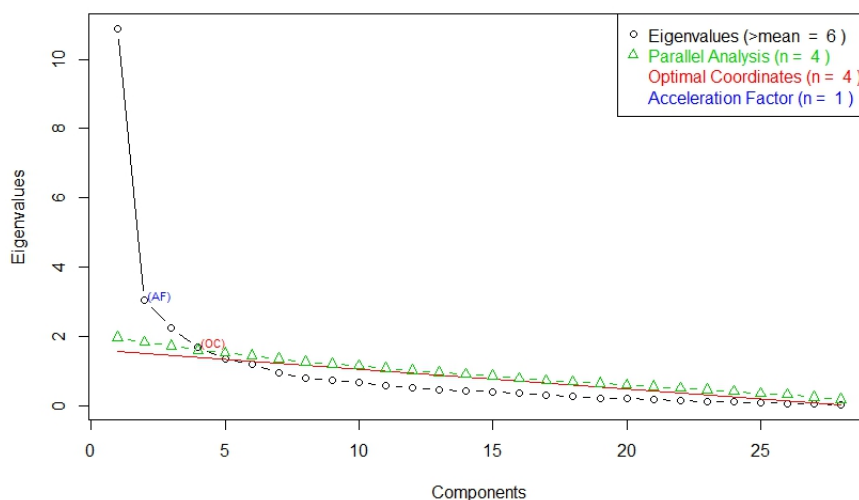
Skúsime sa tak postupne pozrieť na odhady faktorových modelov, pričom začneme od osemfaktorového, ktoré navrhuje Kaiserovo kritérium, aby sme sa presvedčili o tom, či skutočne toto kritérium nadhodnocuje počet faktorov v modeli. Navzájom budeme porovnávať modely odhadnuté jednotlivými metódami. Pretože taktiež chceme, aby jednotlivé faktory boli navzájom ortogonálne, tak budeme v modeloch používať rotáciu Varimax.

3.3.1 Metóda hlavných faktorov

Ako prvým modelom začneme osemfaktorovým modelom pre 33 pozorovaných indikátorov. Odhadnutá rotovaná matica faktorových váh je zobrazená v tabuľke 3.5. V tabuľke sú zobrazené iba faktorové váhy $|\lambda_{jk}| \geq 0,4$, ktoré považujeme za dôležité.

Môžeme vidieť, že z modelu pre ďalšie účely môžeme odstrániť indikátory 4Diii, 5Bi a 4Dii, pretože žiaden z faktorov ich nevysvetľuje dostatočne, resp. nemajú dostatočne veľkú (kladnú alebo zápornú) korelovanosť so žiadnym faktorom. Taktiež môžeme z modelu odstrániť indikátory 3A a 2G, pretože sú triviálne, tj. príslušný faktor, ktorý je s nimi najviac korelovaný, vysvetľuje iba jednu individuálnu premennú. Ďalej preto chceme odhadnúť päťfaktorový model pre zvyšných 28 indikátorov, ktoré sme ponechali v modeli.

Na obrázku 3.8 môžeme vidieť, že Hornova paralelná analýza stále odporúča použiť štyri faktory na odhad 28 indikátorov, zatiaľčo podľa Kaiserovho kritéria by sme mali použiť v modeli šesť spoločných faktorov.



Obr. 3.8: Hornova paralelná analýza 28 indikátorov s hladinou spoľahlivosti 95%

Pred odhadom nového modelu sa opäť môžeme pozrieť na Bartlettov test sféricity a KMO kritérium (viď tabuľka 3.6). Vidíme, že po odstránení piatich indikátorov hodnota KMO kritéria vzrástla a Bartlettov test sféricity opäť zamietol hypotézu ortogonalít indikátorov. Pre ďalšie modely uvedené v tejto diplomovej práci už nebudeme hodnoty KMO kritéria a Bartlettovho testu sféricity uvádzať, pretože ani v jednom modeli nezamietajú faktorovosť. Podrobné výstupy je však možné nájsť v priloženom CD.

	Faktor							
	1	2	3	4	5	6	7	8
2H	.886							
5Civ	.866							
2A	.841							
2C	.839							
2E	.829							
5Cii	-.811	-.406						
2B	.793							
2I	.769							
4Bi	.655							
2D	.623							
1A	-.600		.504					
5Cvi	.548							
5Bv	.526							.415
1B	-.508		.498					
2F	.506			.446				
4Di	.439							
4Bii		.687						
3B		.662						
5Ciii		.556						
4Ai		.471						
4Diii								
5Bii			.782					
5Ci	.435		.679					
5Biii			.608					
5Bi								
2G				.717				
4Aii					.720			
4Aiii					.646			
3C					.482			
4Dii								
5Cv							.499	
5Aii							.427	
3A								.605

Tabuľka 3.5: Metódou Varimax rotovaná matica faktorových váh 33 indikátorov osemfaktorového modelu odhadnutého metódou hlavných faktorov

KMO kritérium		.865
Bartlettov test sféricnosti	Testová štatistika	2343.945
	Stupne voľnosti	.378
	P-hodnota	.000

Tabuľka 3.6: KMO kritérium a Bartlettov test sféricnosti pre model s 28 indikátormi

Z rotovanej odhadnutej matice faktorových váh päťfaktorového modelu (viď tabuľka 3.7) vidíme, že indikátory 5Aii a 5Cv nemajú významnú koreláciu so žiadnym faktorom a tak sú v modeli nadbytočné. Taktiež sme sa rozhodli vylúčiť z modelu indikátory 5Cii, 5Civ a 2H, ktoré majú takmer rovnaké významné korelácie s viac ako jedným faktorom. Toto rozhodnutie nám zľahčuje interpretáciu spoločných faktorov, pretože chceme, aby každá pozorovaná premenná mala významnú koreláciu práve s jedným faktorom.

	Faktor				
	1	2	3	4	5
2E	.822				
2I	.751				
2D	.691				
1B	-.672			.416	
1A	-.671				
2H	.671	.604			
2F	.613				
5Cii	-.613	-.560	-.430		
5Bv	.566				
5Cvi	.486				
2B		.899			
2C	.500	.773			
2A	.507	.745			
4Di		.688			
5Ci		.685		.563	
5Civ	.624	.673			
4Bi		.561			
4Bii			.701		
5Ciii			.609		
3B			.582		
4Ai			.470		
5Cv					
5Bii				.746	
5Biii				.646	
5Aii					
4Aii					.758
4Aiii					.610
3C					.514

Tabuľka 3.7: Metódou Varimax rotovaná matica faktorových váh 28 indikátorov päťfaktorového modelu odhadnutého metódou hlavných faktorov

Po odstránení týchto premenných dostávame odhadnutý päťfaktorový model 23 indikátorov, ktorý je zobrazený v tabuľke 3.8. Z tejto tabuľky tak vidíme, že každý pozorovaný indikátor je jasne korelovaný práve s jedným faktorom a preto tento model považujeme za dobrý.

Prvý faktor tak vysvetľuje nasledujúce indikátory:

- 2E Integrita právneho systému

	Faktor				
	1	2	3	4	5
2E	.826				
2I	.739				
2D	.688				
1A	-.654			.432	
1B	-.652			.420	
2F	.621				
5Bv	.575				
5Cvi	.508				
2B		.911			
2C	.502	.770			
2A	.503	.730			
5Ci		.711		.515	
4Di		.697			
4Bi		.564			
4Bii			.712		
3B			.592		
5Ciii			.583		
4Ai			.462		
5Bii				.770	
5Biii				.652	
4Aii					.735
4Aiii					.612
3C					.519

Tabuľka 3.8: Metódou Varimax rotovaná matica faktorových váh 23 indikátorov päťfaktorového modelu odhadnutého metódou hlavných faktorov

- 2I Kriminálna činnosť v danej zemi
- 2D Vojenský zásah do súdnictva a politiky
- 1A Vládna spotreba
- 1B Transféry a dotácie
- 2F Právna vymáhateľnosť záväzkov plynúcich zo zmlúv
- 5Bv Náklady na prepustenie pracovníka
- 5Cvi Náklady na dodržiavanie daňových predpisov

Tento faktor má najsilnejšiu koreláciu (faktorovú váhu) s indikátorom integrity právneho systému (0,826). Takisto si môžeme všimnúť, že faktor nadobúda negatívne faktorové váhy s indikátormi vládnej spotreby (-0,654) a transférov a dotácií (-0,652). Tento faktor môžeme interpretovať ako obecný faktor prítomnosti vlády, ktorý vyjadruje veľkosť vládnej spotreby (indikátory 1A a 1B) a ako účinný je právny systém štátu (indikátory 2E, 2I, 2D, 2F, Bv a 5Cvi).

Indikátor vládnej spotreby je braný ako podiel na celkovej spotrebe voči HDP a preto malý rozdiel medzi ním a HDP značí nahradzovanie individuálnej voľby jednotlivých subjektov za politickú voľbu. To isté platí aj pre indikátor transférov a dotácií, ktorý je vzťahovaný k celkovému vyprodukovanému produktu v krajine. Preto tieto negatívne korelácie faktora s indikátormi 1A a 1B dokazujú, že krajiny s veľkou vládnu spotrebou a dotáciami majú nižšiu ekonomickú slobodu.

Druhý faktor vysvetľujúci indikátory

- 2B Nestrannosť súdov
- 2C Ochrana vlastníckych práv
- 2A Nezávislosť súdnictva
- 5Ci Administratívne obmedzenia
- 4Di Reštrikcie zahraničného vlastníctva a investovania
- 4Bi Netarifované obchodné prekážky

má pozitívne faktorové váhy so všetkými indikátormi, ktoré vysvetľuje. Taktiež vidíme, že indikátory ochrany vlastníckych práv a nezávislosti súdnictva majú významné korelácie aj s prvým faktorom, čiže sú komplexnejšie. Tento faktor môžeme interpretovať ako faktor vysvetľujúci právny systém (indikátory 2B, 2C a 2A) a obchodné obmedzenia (indikátory 5Ci, 4Di a 4Bi). Faktor tak vyjadruje, že neskorumpované krajiny, ktoré majú spoľahlivý právny systém (nezávislosť súdnictva, nestrannosť súdov, nestrannosť polície) a ktoré implementujú vhodné obmedzenia v obchodnej a podnikateľskej sfére, budú mať vysokú úroveň ekonomickej slobody.

Tretí faktor vysvetľujúci indikátory

- 4Bii Náklady na dodržiavanie obchodných predpisov
- 3B Smerodajná odchýlka inflácie
- 5Ciii Založenie podnikania
- 4Ai Príjem z daní zafažujúcich zahraničný obchod

má taktiež so všetkými indikátormi, ktoré vysvetľuje, pozitívnu koreláciu, pričom najväčšiu faktorovú váhu má s indikátorom nákladov na dodržanie obchodných predpisov (0,712). Tento faktor tak vysvetľuje ako vláda môže regulovať podnikanie, zahraničný obchod prípadne svoju menovú politiku. Preto možno tento faktor interpretovať ako faktor popisujúci regulácie a monetárnu politiku danej krajiny.

Štvrtý faktor popisuje síce iba dva indikátory

- 5Bii Regulácia najímania a prepúšťania pracovnej sily
- 5Biii Sila centrálného vyjednávania a odborov

ale o to ľahšia je jeho interpretácia, pretože oba indikátory spadajú do rovnakého komponentu EFW indexu. Tento faktor preto nazveme faktorom vysvetľujúcim reguláciu trhu práce. Faktor má pozitívne faktorové váhy s oboma indikátormi, preto môžeme predpokladať, že krajiny s flexibilnejším trhom práce budú mať vyššiu úroveň ekonomickej slobody.

Piaty faktor vysvetľujúci indikátory

- 4Aii Priemerná sadzba dane uvalenej na zahraničný obchod
- 4Aiii Smerodajná odchýlka daní uvalených na medzinárodný obchod
- 3C Inflácia za posledný kalendárny rok

je veľmi ťažko interpretovať, pretože indikátor popisujúci infláciu za posledný kalendárny rok je veľmi odlišný od zvyšných dvoch. Faktor je pozitívne korelovaný so všetkými tromi indikátormi a najväčšie faktorové váhy nadobúda s indikátormi popisujúcimi tarifné bariéry medzinárodného obchodu.

	Faktor				
	1	2	3	4	5
2E	.896				
2I	.775				
2D	.754				
2C	.742	.584			
2A	.708	.551			
1B	-.668		.425		
2F	.648				
1A	-.634		.443		
5Bv	.624				
5Cvi	.589				
4Bi	.571	.427			
2B	.577	.776			
5Ci		.660	.520		
4Di		.646			
5Bii			.772		
5Biii			.651		
4Bii	.449			.653	
3B				.598	
5Ciii	.461			.504	
4Ai				.402	
4Aii	.546				.685
4Aiii					.617
3C					.493

Tabuľka 3.9: Metódou Quartimax rotovaná matica faktorových váh 23 indikátorov päťfaktorového modelu odhadnutého metódou hlavných faktorov

Ak by sme miesto rotačnej metódy Varimax použili rotačnú metódu Quartimax (viď tabuľka 3.9), tak vidíme, že oproti metóde Varimax by prvý faktor

vysvetľoval o tri indikátory (2C, 2A a 4Bi). Tieto indikátory však stále majú významné korelácie aj s druhým faktorom, ktorý ich vysvetľoval pri rotácii Varimax. Interpretácia faktorov sa tak touto rotáciou nijak zásadne nezmení.

Skúsme sa ďalej takisto pozrieť na štvorfaktorový model pre 33 indikátorov, ktorý pôvodne navrhuje Hornova paralelná analýza (viď obrázok 3.7). Podobnou metodikou, kedy vyradujeme tie indikátory, ktoré majú nevýznamné korelácie so všetkými faktormi alebo ktoré majú takmer rovnaké významné korelácie s viac ako jedným faktorom, dostávame konečný model, ktorý vysvetľuje 24 indikátorov (viď tabuľka 3.10).

	Faktor			
	1	2	3	4
2E	.821			
2I	.760			
2D	.675			
1B	-.662			.407
1A	-.661			
5Bv	.577			
2F	.575			
5Cvi	.491			
2B		.912		
5Ci		.738		.501
2C	.482	.736		
2A	.493	.715		
4Di		.660	.406	
4Bi	.405	.527		
4Bii			.759	
5Ciii			.573	
4Ai			.539	
5Cv			.506	
3B			.496	
3A			.479	
4Diii			.426	
3C			.412	
5Bii				.728
5Biii				.702

Tabuľka 3.10: Metódou Varimax rotovaná matica faktorových váh 24 indikátorov štvorfaktorového modelu odhadnutého metódou hlavných faktorov

Vidíme, že prvé dva faktory sú rovnaké a vysvetľujú tie isté indikátory ako v päťfaktorovom modeli. Takisto štvrtý faktor popisujúci reguláciu trhu práce sa nezmenil. Vznikol nám však nový tretí faktor, ktorý popisuje:

- 4Bii Náklady na dodržiavanie obchodných predpisov
- 5Ciii Založenie podnikania
- 4Ai Príjem z daní zafažujúcich medzinárodný obchod

- 5Cv Licenčné zákony
- 3B Smerodajná odchýlka inflácie
- 3A Rast peňažnej zásoby
- 4Diii Možnosť zahraničných osôb navštíviť danú krajinu
- 3C Inflácia za posledný kalendárny rok

	Faktor	
	1	2
5Cii	-.924	
5Civ	.914	
2H	.883	
2C	.863	
2E	.851	
2A	.816	
2D	.773	
1B	-.697	.516
2I	.686	
2B	.679	.543
4Bi	.639	
4Bii	.597	
1A	-.594	
5Bv	.580	
2F	.576	
4Aii	.576	
5Cvi	.548	
5Ciii	.465	
4Ai	.454	
4Diii	.421	
4Dii	.420	
5Ci		.834
5Bii		.658
4Di	.474	.548
5Biii		.427

Tabuľka 3.11: Metódou Varimax rotovaná matica faktorových váh 25 indikátorov dvojfaktorového modelu odhadnutého metódou hlavných faktorov

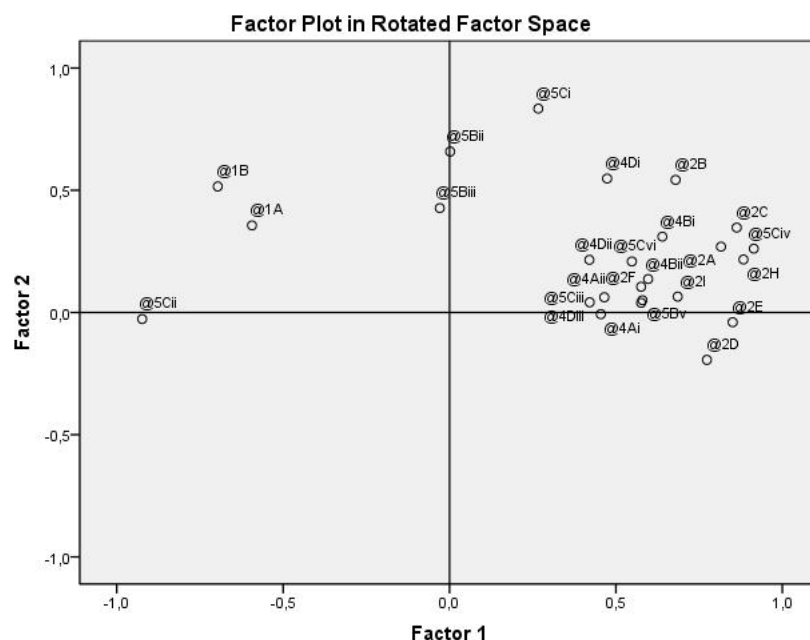
Môžeme povedať, že tento faktor lepšie popisuje reguláciu podnikania a zahraničného obchodu ako v päťfaktorovom modeli, pretože okrem indikátorov (4Bi, 5Ciii, 4Ai), ktoré už vysvetľoval, tak vysvetľuje aj indikátory (5Cv, 4Diii), ktoré úzko súvisia s touto reguláciou. Faktor okrem týchto indikátorov popisuje aj indikátory (3A, 3B, 3C), ktoré spadajú do rovnakej oblasti EFW indexu a to peňažných slobôd. Interpretácia faktora tak zostáva rovnaká v smere, že faktor stále vyjadruje, ako vláda reguluje podnikanie a zahraničný obchod, ale navyše súčasne aj dbá o to, aby bola zaistená stabilná inflácia a stabilný rast peňažnej zásoby.

Pozitívne korelácie so všetkými indikátormi značia, že krajiny s vyššími hodnotami sledovaných indikátorov, budú ekonomicky slobodnejšie. Vypadol tak piaty faktor, ktorý sme nevedeli presne definovať v predošlom modeli.

Ďalej sa ešte pozrieme na to, ako indikátory vysvetľuje najjednoduchší dvojfaktorový model. Rovnakou metodikou tak nachádzame dvojfaktorový model pre 25 indikátorov (viď tabuľka 3.11).

Vidíme, že prvý faktor vysvetľuje až 21 indikátorov a druhý faktor iba 4 indikátory. Prvý faktor tak interpretujeme ako obecný faktor ekonomickej slobody v krajine, čo značí silný nezávislý právny systém, ďalej zodpovednú vládu, ktorá dohliada na dodržiavanie základných ľudských a majetkových práv a ktorá podporuje rast podnikania a naopak nepoužíva žiadne zbytočné obchodné a iné ekonomické reštrikcie.

Druhý faktor tak popisuje menej dôležité (sekundárne) indikátory ekonomickej slobody, ako je regulácia trhu práce a administratívne obmedzenia.



Obr. 3.9: Rotované faktorové váhy dvojfaktorového modelu odhadnutého metódou hlavných faktorov

Taktiež z obrázku 3.9 je zreteľné, že faktorové váhy indikátorov sú relatívne blízko medzi sebou. Výnimku tvoria indikátory vládnej spotreby (1A), transférov a dotácií (1B) a nákladov na byrokráciu (5Cii), ktoré majú negatívnu koreláciu s ekonomickou slobodou krajiny.

3.3.2 Metóda najmenších štvorcov

Pri metóde najmenších štvorcov budeme postupovať rovnako ako pri metóde hlavných faktorov. A teda, že začneme od osemfaktorového modelu pre 33 faktorov a postupne sa dostaneme k finálnemu ortogonálnemu modelu, ktorý spĺňa naše požiadavky na model.

V tabuľke 3.12 tak vidíme odhadnutý päťfaktorový model pre 24 indikátorov. Ak ho porovnáme s odhadnutým päťfaktorovým modelom metódou hlavných fak-

	Faktor				
	1	2	3	4	5
2E	.809				
2I	.756				
2D	.652				
1B	-.644			.424	
2F	.607				
1A	-.601			.445	
5Bv	.580				
5Cvi	.468				
2B		.915			
2C	.449	.772			
2A	.456	.744			
5Ci		.713		.517	
5Civ	.569	.681			
4Di		.673			
4Bi		.575			
4Bii			.807		
4Ai			.598		
5Ciii			.555		
4Diii			.448		
5Bii				.800	
5Biii				.627	
4Aii			.469		.678
4Aiii					.631
3C					.495

Tabuľka 3.12: Metódou Varimax rotovaná matica faktorových váh 24 indikátorov päťfaktorového modelu odhadnutého metódou najmenších štvorcov

torov (tabuľka 3.8), tak vidíme, že prvý, štvrtý a piaty faktor vysvetľujú rovnaké indikátory. Druhý faktor, ktorý popisoval právny systém a obchodné obmedzenia, má takisto rovnakú interpretáciu a navyše vysvetľuje indikátor miery korupcie a extra platieb za zvýhodňovanie oproti ostatným (5Civ), ktorý priamo súvisí s danou oblasťou.

Tretí faktor tentokrát popisuje indikátory

- 4Bii Náklady na dodržiavanie obchodných predpisov
- 4Ai Príjem z daní zafažujúcich medzinárodný obchod
- 5Ciii Založenie podnikania
- 4Diii Možnosť zahraničných osôb navštíviť danú krajinu

pričom má jasne najväčšiu faktorovú váhu s indikátorom nákladov na dodržiavanie obchodných predpisov (0,807). Faktor tak interpretujeme ako reguláciu predovšetkým zahraničného obchodu, ale aj podnikateľskej sféry. Krajiny s dobrou prístupnosťou pre zahraničné osoby a jednoduchou možnosťou založenia podnikania tak sú ekonomicky slobodné.

Ďalej sa takisto pozrieme na štvorfaktorový model (viď tabuľka 3.13). Po odstránení indikátorov, už spomenutou metodikou, dostávame model pre 27 indikátorov.

	Faktor			
	1	2	3	4
2E	.792			
2I	.722			
1A	-.701			
2D	.683			
1B	-.675			
2F	.582			
5Bv	.527			
5Cvi	.493			
4Bii		.740		
4Aii		.633		
4Ai		.566		
5Ciii		.543		
5Cv		.534		
3C		.486		
4Diii		.471		
3A		.447		
3B		.443		
4Dii		.418		
2B			.894	
5Ci			.756	.481
2C	.478		.731	
2A	.508		.695	
4Di		.436	.674	
4Bi			.538	
5Biii				.704
5Bii				.664
2G				.415

Tabuľka 3.13: Metódou Varimax rotovaná matica faktorových váh 27 indikátorov štvorfaktorového modelu odhadnutého metódou najmenších štvorcov

Ak ho porovnáme so štvorfaktorovým modelom získaným metódou hlavných faktorov (tabuľka 3.10), tak vidíme, že opäť prvý faktor popisujúci prítomnosť vlády zostáva rovnaký. Taktiež faktor právneho systému a obchodných obmedzení sa nezmenil. Faktor, ktorý popisoval reguláciu podnikania, zahraničný obchod a menovú politiku štátu, vysvetľuje v tomto modeli aj indikátory

- 4Aii Priemerná sadzba dane uvalenej na zahraničný obchod
- 4Dii Kontrola kapitálu

ktoré jasne úzko súvisia s oblasťou, ktoré daný faktor popisuje, pričom indikátor priemernej sadzby dane má dokonca druhú najvýznamnejšiu faktorovú váhu (0,633) a tak interpretácia faktoru je rovnaká.

Štvrtý faktor popisuje indikátory

- 5Biii Sila centrálného vyjednávania a odborov
- 5Bii Regulácia najímania a prepúšťania pracovnej sily
- 2G Regulačné reštrikcie na nakladanie s nehnuteľnosťami.

Spája tak dva úzko súvisiace indikátory (5Bii, 5Biii) s indikátorom regulačných reštrikcií na nakladanie s nehnuteľnosťami, ktorý je od nich veľmi odlišný a tým nám sťažuje interpretovateľnosť faktora. Faktor však má jasne významnejšie faktorové váhy s indikátormi popisujúcimi reguláciu trhu práce v porovnaní s indikátorom reštrikcií s faktorovou váhou (0,415), ktorú iba tesne považujeme za významnú a pri striktnejšom kritériu by indikátor nebol v modeli.

Ďalej nás taktiež bude zaujímať najjednoduchší dvojfaktorový model, ktorý vieme odhadnúť metódou najmenších štvorcov. Tento model je zhodný s dvojfaktorovým modelom odhadnutým metódou hlavných faktorov (tabuľka 3.11) a tak ho ponechávame bez komentára.

3.3.3 Zobecnená metóda najmenších štvorcov

Pri zobecnenej metóde najmenších štvorcov taktiež postupujeme ako doposiaľ. Ako prvý tak dostávame päťfaktorový model popisujúci 24 indikátorov (viď tabuľka 3.14).

Ak porovnáme získaný model s päťfaktorovým modelom odhadnutým metódou najmenších štvorcov, tak vidíme, že prvý faktor (faktor právneho systému a obchodných obmedzení) popisuje aj indikátor nákladov na byrokráciu (5Cii). S týmto jediným indikátorom má faktor negatívnu koreláciu (-0,654), čo značí, že čím vyššie náklady má krajina na byrokráciu, tým je menej ekonomicky slobodnejšia.

Druhý faktor (faktor prítomnosti vlády) opäť popisuje osem rovnakých indikátorov. Tretí faktor vysvetľujúci indikátory

- 5Ci Administratívne obmedzenia
- 5Bii Regulácia najímania a prepúšťania pracovnej sily
- 5Biii Sila centrálného vyjednávania a odborov

má predovšetkým významné váhy s prvými dvomi indikátormi. Indikátory majú medzi sebou úzky súvis a tak faktor interpretujeme ako faktor regulácie trhu práce.

Štvrtý faktor popisujúci indikátory

- 5Ciii Založenie podnikania
- 4Bii Náklady na dodržiavanie obchodných predpisov
- 3A Rast peňažnej zásoby
- 5Cv Licenčné zákony

	Faktor				
	1	2	3	4	5
2B	.932				
2A	.871				
2C	.841				
5Civ	.750	.532			
5Cii	-.654	-.533			
4Di	.617				
4Bi	.513				
2I		.807			
2E	.401	.790			
1B		-.663	.470		
2D		.605			
2F		.598			
5Bv		.555			
1A		-.536	.436		
5Cvi		.446			
5Ci	.588		.730		
5Bii			.729		
5Biii			.566		
5Ciii				.626	
4Bii	.408			.617	
3A				.607	
5Cv				.467	
4Aii					.919
3C					.541

Tabuľka 3.14: Metódou Varimax rotovaná matica faktorových váh 24 indikátorov päťfaktorového modelu odhadnutého zobecnenou metódou najmenších štvorcov

má približne rovnaké pozitívne faktorové váhy u prvých troch indikátorov a o niečo menšiu u štvrtého. Faktor preto vysvetľuje rovnako reguláciu podnikateľskej sféry ako aj monetárnu politiku štátu.

Piaty faktor modelu nechávame bez interpretácie, pretože zlučuje indikátory (4Aii, 3C), medzi ktorými nie je zjavný žiaden vzťah.

Vhodný štvorfaktorový model sa nám zobecnenou metódou najmenších štvorcov nepodarilo nájsť, pretože algoritmus pre výpočet matice faktorových váh, ohraničený na 10 000 iterácií, nenašiel žiadne lokálne minimum účelovej funkcie 2.19.

Preto sa pokúsime odhadnúť získaný päťfaktorový model štyrmi faktormi (viď tabuľka 3.15). Vidíme, že faktory prítomnosti vlády, právneho systému a obchodných obmedzení a takisto aj regulácie trhu práce zostali rovnaké. Zvyšné dva faktory sa spojili do jedného spoločného. Faktor tak interpretujeme opäť ako reguláciu podnikateľskej sféry a monetárnej politiky štátu, pretože dva indikátory - priemerná sadzba dane uvalenej na medzinárodný obchod (4Aii) a inflácia za posledný kalendárny rok (3A), ktoré štvrtý faktor pribral, popisujú taktiež túto oblasť.

	Faktor			
	1	2	3	4
2B	.922			
2A	.872			
2C	.845			
5Civ	.747	.546		
5Cii	-.660	-.544		
4Di	.610		.410	
4Bi	.499	.402		
2I		.805		
2E	.403	.792		
1B		-.675		.467
2D		.642		
2F		.593		
5Bv		.558		
1A		-.557		
5Cvi		.481		
4Aii			.695	
4Bii	.426		.624	
3C			.570	
5Cv			.563	
5Ciii			.543	
3A			.497	
5Ci	.568			.751
5Bii				.714
5Biii				.545

Tabuľka 3.15: Metódou Varimax rotovaná matica faktorových váh 24 indikátorov štvorfaktorového modelu odhadnutého zobecnenou metódou najmenších štvorcov

3.4 Výber faktorového modelu

V predošlej podkapitole sme sa pozreli na odhady jednotlivých faktorových modelov pomocou rôznych metód a interpretáciu faktorov. V tejto podkapitole porovnáme tieto modely z pohľadu toho, ako dobre zachytávajú kovariančnú štruktúru pozorovaných indikátorov, tj. ako dobre odhadli modely korelačnú (kovariančnú) maticu a následne vyberieme ten najlepší.

Bude nás zaujímať, koľko prvkov (resp. relatívny počet prvkov) odhadnutej reprodukovanej korelačnej (kovariančnej) matice sa významne líši od pôvodnej korelačnej (kovariančnej) matice. Konkrétne, prvok s_{ij} odhadnutej kovariančnej matice \mathbf{S} , $\mathbf{S} = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}^T + \hat{\mathbf{\Psi}}$ tak považujeme za významne sa odlišujúci od pôvodnej kovariančnej matice Σ , ak $|s_{ij} - \sigma_{ij}| > 0,05$ (pre korelačnú maticu podobne). Pre ilustráciu, z reziduálnej korelačnej matice päťfaktorového modelu odhadnutého metódou hlavných faktorov (viď tabuľka 3.16) vidíme, že 82 prvkov, tj. približne 16% prvkov, sa významne odlišuje od odhadovanej korelačnej matice.

	1A	1B	2A	2B	2C	2D	2E	2F	2I	3B	3C	4Ai	4Aii	4Aiii	4Bi
1A		-.009	-.047	-.009	-.008	.002	.001	-.026	.036	.112	-.057	-.012	-.001	.018	.069
1B	-.009		.018	.007	.027	.006	-.009	.002	-.044	-.078	-.036	.035	.038	.007	-.059
2A	-.047	.018		.029	.026	.036	-.012	-.027	-.053	-.022	-.039	-.003	.033	.004	-.023
2B	-.009	.007	.029		-.003	-.016	-.018	.027	-.021	.013	.022	-.028	.037	-.005	-.043
2C	-.008	.027	.026	-.003		.026	-.024	-.010	.002	-.020	-.007	.013	.015	-.031	.002
2D	.002	.006	.036	-.016	.026		-.033	-.047	-.047	.007	-.117	-.038	.040	.010	.025
2E	.001	-.009	-.012	-.018	-.024	-.033		-.016	.066	.049	.012	.048	-.032	-.005	.013
2F	-.026	.002	-.027	.027	-.010	-.047	-.016		.041	.025	-.002	-.008	.033	-.011	-.018
2I	.036	-.044	-.053	-.021	.002	-.047	.066	.041		.032	.059	-.025	-.050	.010	.057
3B	.112	-.078	-.022	.013	-.020	.007	.049	.025	.032		.056	-.104	-.023	-.012	-.005
3C	-.057	-.036	-.039	.022	-.007	-.117	.012	-.002	.059	.056		.012	.051	-.051	-.005
4Ai	-.012	.035	-.003	-.028	.013	-.038	.048	-.008	-.025	-.104	.012		.056	-.038	.011
4Aii	-.001	.038	.033	.037	.015	.040	-.032	.033	-.050	-.023	.051	.056		.021	-.064
4Aiii	.018	.007	.004	-.005	-.031	.010	-.005	-.011	.010	-.012	-.051	-.038	.021		.027
4Bi	.069	-.059	-.023	-.043	.002	.025	.013	-.018	.057	-.005	-.005	.011	-.064	.027	

Tabuľka 3.16: Časť reziduálnej korelačnej matice päťfaktorového modelu odhadnutého metódou hlavných faktorov

Hľadáme medzi odhadnutými modelmi taký model, ktorý má najmenej významných korelačných (kovariančných) reziduí. Na tomto mieste môžeme pripomenúť, že rotáciou sa počet významných reziduí nezmení.

Metóda odhadu	Počet faktorov	Významné reziduá
Metóda hlavných faktorov	5	16%
Metóda hlavných faktorov	4	25%
Metóda hlavných faktorov	2	47%
Metóda najmenších štvorcov	5	11%
Metóda najmenších štvorcov	4	25%
Metóda najmenších štvorcov	2	47%
Zobecnená metóda najmenších štvorcov	5	18%
Zobecnená metóda najmenších štvorcov	4	22%

Tabuľka 3.17: Porovnanie reziduálnych korelačných matíc jednotlivých odhadnutých modelov

Z tabuľky 3.17 tak vidíme, že najlepšie kovariančnú štruktúru medzi pozorovanými indikátormi zachytáva päťfaktorový model odhadnutý metódou najmenších štvorcov, v ktorom iba 11% prvkov reziduálnej korelačnej matice presahuje hranicu významnosti. Naopak, dvojfaktorové modely zachytili kovariančnú štruktúru najhoršie.

Pomocou štvorfaktorových modelov sme tak síce popísali viac indikátorov ako pomocou päťfaktorových podľa dopredu daných kritérií, avšak vidíme, že kovariančnú štruktúru sa nám podarilo zachytiť menej.

Skúsme sa teraz detailne pozrieť na tento päťfaktorový model z pohľadu komunalít a faktorového skóre.

3.4.1 Analýza komunalít

Ako už bolo spomenuté v kapitole 2, tak pomocou komunalít vieme zistiť, ako dobre model popisuje rozptyl pozorovaných premenných, resp. ktoré pozorované premenné popisuje veľmi dobre a ktoré nie. V tabuľke 3.18 tak môžeme vidieť, aké odhady komunalít sme dostali po extrakcii päťfaktorového modelu metódou najmenších štvorcov.

Vstupné odhady komunalít sa určili, podobne ako pri metóde hlavných faktorov, ako druhá mocnina koeficientu mnohonásobnej korelácie daného indikátoru s ostatnými indikátormi. Po extrakcii faktorov tak vidíme, že pri deviatich indikátoroch zostáva po odhade modelu väčšia časť rozptylu nevysvetlená, tj. $\hat{h}_j^2 < 0,5$ pre j -ty indikátor.

Najlepšie model zachytil indikátor nestrannosti súdov (2B), kedy až 95,1% rozptylu indikátoru je vysvetlená. Najmenej naopak indikátor možnosti zahraničných osôb navštíviť krajinu (4Diii), kedy model zachytáva iba 28,1% jeho rozptylu a až 71,9% necháva v špecifickom rozptyle. Tieto rozdiely sú spôsobené predovšetkým tým, že faktor, ktorý vysvetľuje indikátor 2B má s daným indikátorom najvyššiu faktorovú váhu (0,915) v celom modeli. Na druhej strane faktor vysvetľujúci indikátor 4Diii má s danými indikátorom najnižšiu faktorovú váhu (0,448) v celom modeli a tým ponecháva značnú časť rozptylu indikátora nevysvetlenú.

	Odhad komunalít		Špecifický rozptyl
	Vstupné	Po extrakcii	
1A	.601	.587	.413
1B	.724	.736	.264
2A	.918	.843	.157
2B	.938	.951	.049
2C	.930	.931	.069
2D	.761	.634	.366
2E	.796	.817	.183
2F	.491	.480	.520
2I	.703	.667	.333
3C	.554	.379	.621
4Bi	.654	.593	.407
4Bii	.670	.768	.232
4Di	.668	.640	.360
5Bii	.624	.687	.313
5Biii	.477	.427	.573
5Bv	.527	.432	.568
5Ci	.820	.781	.219
5Ciii	.549	.457	.543
5Cvi	.552	.384	.616
4Diii	.426	.281	.719
4Aii	.797	.830	.170
4Aiii	.524	.416	.584
5Civ	.921	.917	.083
4Ai	.478	.427	.573

Tabuľka 3.18: Odhad komunalít päťfaktorového modelu odhadnutého metódou najmenších štvorcov

3.4.2 Odhad faktorového skóre

Ďalej sa takisto pozrieme na odhady faktorového skóre pre jednotlivé krajiny regresnou metódou. Ako už bolo spomenuté v podkapitole 2.3.4, tak pri odhade metódou najmenších štvorcov pracujeme s korelačnou maticou pozorovaných premenných a preto pri odhade faktorového skóre regresnou metódou pracujeme so štandardizovanými pozorovanými premennými.

V tabuľke 3.19 vidíme ako sa jednotlivé pozorované indikátory podieľajú na hodnotách faktorov. V regresnej metóde to je matica β , kde $\hat{\mathbf{f}} = \beta^T \mathbf{X}$ a z 2.22 plynie, že

$$\hat{\beta} = \hat{\Lambda}^T \mathbf{R}^{-1}.$$

Výsledné odhadnuté faktorové skóre pre jednotlivé krajiny je možné nájsť na priloženom CD.

Na záver sa na základe odhadnutých faktorových skóre pokúsime zostaviť vlastné poradie ekonomickej slobody krajín. Vlastný EFW index získame váženým priemerom faktorových skóre. Váhy pre jednotlivé faktory určíme podľa

	Faktor				
	1	2	3	4	5
1A	-.117	.017	.114	.116	-.089
1B	-.135	.087	-.082	.175	-.035
2A	-.011	-.071	-.107	-.130	.037
2B	-.176	.848	-.431	-.177	-.042
2C	.007	.132	.353	-.052	-.295
2D	.125	.012	-.007	-.053	-.173
2E	.415	-.099	-.063	.093	-.060
2F	.125	-.116	.063	.112	-.167
2I	.193	-.049	-.223	.084	.153
3C	-.014	.012	-.004	-.025	-.042
4Bi	-.028	.088	-.080	-.022	.256
4Bii	-.224	-.051	.601	.030	-.124
4Di	-.128	.086	.044	.017	.308
5Bii	.098	-.092	.057	.467	-.073
5Biii	.033	.024	-.013	.162	-.049
5Bv	.098	-.078	-.050	.070	.154
5Ci	-.012	.032	-.093	.278	.011
5Ciii	.064	-.061	.071	.126	-.028
5Cvi	.043	-.040	.063	.046	-.167
4Diii	-.030	-.005	.054	-.014	-.072
4Aii	-.001	-.118	.139	.070	.981
4Aiii	.008	.060	-.106	-.034	.053
5Civ	.171	.197	.210	-.080	-.296
4Ai	-.059	.025	.112	-.021	-.061

Tabuľka 3.19: Regresnou metódou odhadnuté koeficienty faktorového skóre päť-faktorového modelu

Faktor	Vysvetlený rozptyl	Váha
1	19.32%	0.31
2	18.36%	0.29
3	11.11%	0.18
4	7.95%	0.13
5	6.03%	0.10

Tabuľka 3.20: Vysvetlený rozptyl indikátorov rotovaným riešením faktorových váh

toho, ako sa jednotlivé faktory podieľajú na percente vysvetleného rozptylu pozorovaných indikátorov. Podiel celkového rozptylu vyčerpaného j-tým faktorom je daná ako

- $\sum_{j=1}^p \hat{\lambda}_{jk}^2 / \sum_j s_{jj}$, ak vychádza faktorová analýza z kovariančnej matice
- $\sum_{j=1}^p \hat{\lambda}_{jk}^2 / p$, ak vychádza faktorová analýza z korelačnej matice

Z tabuľky je zrejmé, že faktory dokopy vysvetľujú 62,77% rozptylu indiká-

torov v modeli. Najviac sa na vysvetlenom rozptyle podiela prvý faktor (faktor prítomnosti vlády a účinnosti právneho systému) a najmenej piaty faktor (faktor, ktorý sme ponechali bez interpretácie).

V tabuľke 3.21 tak máme desať krajín s najvyššou hodnotou skóre pôvodného EFW indexu a nami odhadnutého skóre EFW indexu. Vidíme, že krajiny na prvých troch miestach Singapúr, Hong Kong a Nový Zéland zostali, iba sa zmenilo poradie na prvých dvoch miestach. V prvej desiatke takisto zostalo Fínsko, ktoré sa z desiateho miesta posunulo na štvrté, ďalej takisto Švajčiarsko, ktoré pokleslo o jedno miesto, respektíve Spojené arabské emiráty, ktoré poklesli o dve miesta. Najväčší pozitívny skok v poradí zaznamenalo Luxembursko, ktoré sa v pôvodnom poradí EFW indexu objavilo na tridsiatomdeviatom mieste a v nami odhadnutom poradí sa objavilo už na šiestom mieste. Na druhej strane, Jordánsko pokleslo z ôsmej priečky v pôvodnom poradí až na tridsiatetretie miesto.

Poradie	EFW index		Pôvodné poradie	Odhadnuté poradie
	Pôvodný	Odhadnutý		
1	Hong Kong	Singapúr	2	2
2	Singapúr	Hong Kong	1	1
3	Nový Zéland	Nový Zéland	3	3
4	Švajčiarsko	Fínsko	10	5
5	SAE	Švajčiarsko	4	7
6	Kanada	Luxembursko	39	9
7	Austrália	SAE	6	22
8	Jordánsko	Veľká Británia	12	33
9	Čile	Kanada	7	24
10	Fínsko	Holandsko	34	4

Tabuľka 3.21: Porovnanie TOP desiatich krajín pôvodného EFW indexu s odhadnutým EFW indexom

Záver

Cieľom práce bola detekcia latentných premenných (faktorov) podieľajúcich sa na tvorbe indexu ekonomickej slobody pomocou faktorovej analýzy. Chceli sme zistiť, ako sa vhodná mnohorozmerná štatistická metóda líši od logického a intuitívneho prístupu autorov indexu pri zoskupovaní vstupných indikátorov do výsledného skóre.

V prvej kapitole sme sa venovali predstaveniu v súčasnosti dvoch najznámejších indexov ekonomickej slobody (IEF, EFW). V krátkosti sme si ukázali štruktúru tvorby týchto indexov a na základe akých sledovaných oblastí sa tvorí výsledný index. Taktiež sme porovnali vývoj indexov ekonomickej slobody pre Slovensko a Českú republiku.

Druhú kapitolu sme venovali charakterizácii faktorovej analýzy. Ukázali sme aká motivácia stojí za použitím faktorovej analýzy. Ďalej sme popísali jednotlivé metódy odhadov faktorov, ktoré sa v súčasnosti využívajú. Detailne sme sa pozreli na metódu hlavných faktorov a metódu maximálnej vierohodnosti. Taktiež sme vysvetlili ako funguje regresná metóda pri odhade faktorového skóre.

Tretiu kapitolu sme už venovali aplikácii faktorovej analýzy. Hľadali sme aké latentné premenné sa podieľajú na tvorbe EFW indexu. Najprv sme overovali predpoklady pre použitie faktorovej analýzy, kedy sme museli zamietnuť použitie metódy maximálnej vierohodnosti. Potom sme odhadovali, na základe definovaných kritérií (Kaiserovo pravidlo, Cattellov sutinový test, Hornova paralelná analýza), počet faktorov, ktoré vysvetľujú sledované indikátory. Následne sme odhadovali rôznymi metódami vhodný model faktorovej analýzy, ktorý najlepšie odpovedá kovariančnej štruktúre indikátorov. Zistené dvojfaktorové, štvorfaktorové a päťfaktorové modely sme medzi sebou porovnali a snažili sa interpretovať faktory podľa toho, aké indikátory vysvetľujú.

Zistili sme tak, že najlepšie kovariančnú štruktúru zachytil päťfaktorový model odhadnutý metódou najmenších štvorcov, pri ktorej zostal najmenší počet významných reziduí korelačnej matice. Odhadli sme ním, že na tvorbe EFW indexu sa podieľajú faktory prítomnosti vlády a účinnosti právneho systému, obchodných obmedzení, regulácie zahraničného obchodu a podnikateľskej sféry, regulácie trhu práce a faktor, ktorý sa nám nepodarilo zinterpretovať.

Treba podotknúť, že pri odhade jednotlivých modelov sme zistili, že tri indikátory (vládna spotreba, transféry a dotácie, náklady na byrokráciu) boli ako jediné negatívne korelované s faktorom, ktorý ich vysvetľoval. To značí, že čím má daná krajina vyššiu hodnotu týchto indikátorov, tým je menej ekonomicky slobodnejšia. Je to tak v rozpore s tým, ako je konštruovaný EFW index. Autori indexu totiž tvrdia, že čím má daná krajina vyššiu hodnotu indikátora, tým je ekonomicky slobodnejšia v sledovanej oblasti. Kvôli tomuto zisteniu sa výsledná hodnota EFW indexu (priemer hodnôt všetkých indikátorov) stáva nepresnou a

skresľujúcou.

Index ekonomickej slobody by preto nemal byť považovaný za smerodajný, ale mal by slúžiť len ako predstava o ekonomickej slobode v jednotlivých krajinách vzhľadom na zistenia tejto diplomovej práce.

Literatúra

- BICAN, L. (2009). *Lineární algebra a geometrie*. Academia, Praha. ISBN 978-80-200-1707-9.
- CATTELL, R. (1966). The scree test for the number of factors. *Heritage Foundation*, **1**, 245–276.
- CHEN, J. (2010). Least squares method for factor analysis. Masters's thesis, University of California, Los Angeles.
- FABRIGAR, L., WEGENER, D., MACCALLUM, R. a STRAHAN, E. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, **3**, 272–299.
- FRASER INSTITUTE (2015). Economic freedom of the world index. <http://www.freetheworld.com/>. [Online].
- GLORFELD, L. (1995). An improvement on horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, **55**, 377–393.
- HEBÁK, P., HUSTOPECKÝ, J. a JAROŠOVÁ, E. (2004). *Vícerozměrné statistické metody*. První díl. Informatorium, Praha. ISBN 80-7333-025-3.
- HERITAGE FOUNDATION (2015). Index of economic freedom. <http://www.heritage.org/index/>. [Online].
- HORN, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, **30**, 179–185.
- HÄRDLE, W. a SIMAR, L. (2007). *Applied Multivariate Statistical Analysis*. Second Edition. Springer, Berlín a Louvain-la-Neuve. ISBN 978-3-642-17228-1.
- HUMPHREYS, L. a MONTANELLI, R. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, **10**, 193–206.
- HURT, J. (2014). Mnohorozměrná analýza - přednáška. Univerzita Karlova v Praze.
- JÖRESKOG, K. G. a GOLDBERGER, A. S. (1972). Factor analysis by generalized least squares. *Psychometrika*, **37**, 243–260.
- LAWLEY, D. N. a MAXWELL, A. E. (1962). Factor analysis as a statistical method. *Journal of the Royal Statistical Society*, **12**, 209–229.

- MILLER, T. (2009). *Index of Economic Freedom*. Heritage Foundation, Washington. ISBN 978-0-89195-280-0.
- SPEARMAN, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology*, **15**, 201–292.
- THURSTONE, L. L. (1947). *Multiple Factor Analysis*. University of Chicago Press. ISBN 978-0226801094.
- ZWICK, W.R. & VELICER, W. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, pages 432–442.

Zoznam obrázkov

1.1	Vývoj IEF indexu Českej republiky a Slovenskej republiky	4
1.2	Mapa IEF indexu 2015	5
1.3	Mapa EFW indexu 2014	8
1.4	Vývoj EFW indexu Českej republiky a Slovenskej republiky	8
3.1	Histogram (4C) indikátoru - ceny na čiernom trhu	27
3.2	Testovanie normality	28
3.3	Histogram a test normality (3A) indikátoru - rast miezd	28
3.4	Histogram a test normality (5Bii) indikátoru - regulácia najímania a prepúšťania pracovnej sily	29
3.5	Časť korelačnej matice indikátorov	29
3.6	Kaiserovo kritérium	31
3.7	Hornova paralelná analýza s hladinou spoľahlivosti 95%	32
3.8	Hornova paralelná analýza 28 indikátorov s hladinou spoľahlivosti 95%	34
3.9	Rotované faktorové váhy dvojfaktorového modelu odhadnutého me- tódou hlavných faktorov	42

Zoznam tabuliek

2.1	Faktorové váhy morfológických premenných	12
3.1	KMO kritérium pre jednotlivé indikátory	30
3.2	Bartlettov test sférickosti	31
3.3	Vlastné čísla príslušné faktorom	32
3.4	Desať najväčších vlastných čísel a ich podiel na vysvetlenom rozptyle pozorovaných premenných	33
3.5	Metódou Varimax rotovaná matica faktorových váh 33 indikátorov osemfaktorového modelu odhadnutého metódou hlavných faktorov	35
3.6	KMO kritérium a Bartlettov test sférickosti pre model s 28 indikátormi	35
3.7	Metódou Varimax rotovaná matica faktorových váh 28 indikátorov päťfaktorového modelu odhadnutého metódou hlavných faktorov	36
3.8	Metódou Varimax rotovaná matica faktorových váh 23 indikátorov päťfaktorového modelu odhadnutého metódou hlavných faktorov	37
3.9	Metódou Quartimax rotovaná matica faktorových váh 23 indikátorov päťfaktorového modelu odhadnutého metódou hlavných faktorov	39
3.10	Metódou Varimax rotovaná matica faktorových váh 24 indikátorov štvorfaktorového modelu odhadnutého metódou hlavných faktorov	40
3.11	Metódou Varimax rotovaná matica faktorových váh 25 indikátorov dvojfaktorového modelu odhadnutého metódou hlavných faktorov	41
3.12	Metódou Varimax rotovaná matica faktorových váh 24 indikátorov päťfaktorového modelu odhadnutého metódou najmenších štvorcov	43
3.13	Metódou Varimax rotovaná matica faktorových váh 27 indikátorov štvorfaktorového modelu odhadnutého metódou najmenších štvorcov	44
3.14	Metódou Varimax rotovaná matica faktorových váh 24 indikátorov päťfaktorového modelu odhadnutého zobecnenou metódou najmenších štvorcov	46
3.15	Metódou Varimax rotovaná matica faktorových váh 24 indikátorov štvorfaktorového modelu odhadnutého zobecnenou metódou najmenších štvorcov	47
3.16	Časť reziduálnej korelačnej matice päťfaktorového modelu odhadnutého metódou hlavných faktorov	48
3.17	Porovnanie reziduálnych korelačných matíc jednotlivých odhadnutých modelov	49
3.18	Odhad komunalít päťfaktorového modelu odhadnutého metódou najmenších štvorcov	50

3.19	Regresnou metódou odhadnuté koeficienty faktorového skóre päťfaktorového modelu	51
3.20	Vysvetlený rozptyl indikátorov rotovaným riešením faktorových váh	51
3.21	Porovnanie TOP desiatich krajín pôvodného EFW indexu s odhadnutým EFW indexom	52