UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

BAKALÁRSKA PRÁCA

Meranie rizika portfólia akcií

Bratislava 2011 Zuzana Kuižová

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
KATEDRA APLIKOVANEJ MATEMATIKY A ŠTATISTIKY

MERANIE RIZIKA PORTFÓLIA AKCIÍ

BAKALÁRSKA PRÁCA

Zuzana Kuižová

9.1.9 aplikovaná matematika
ekonomická a finančná matematika
79a47b8c-8f55-4278-82ac-a3e143fdd49c

Vedúci bakalárskej práce:
Mgr. Ing. Pavol Jurča, PhD.

BRATISLAVA 2011

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Zuzana Kuižová
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: bakalárska
Jazyk záverečnej práce: slovenský

Názov : Meranie rizika portfólia akcií

Cieľ : Cieľom práce je modelovanie výnosovosti a rizika portfólia väčšieho počtu akcií pomocou pomocou zníženia dimenzie.

Anotácia : Práca bude spočívať v porovnaní aplikácie rôznych štatistických metód (napr. analýza hlavných komponentov, faktorová analýza) na zníženie dimenzie problému modelovania výnosovosti a rizika (pomocou Value at Risk) pri portfóliu obsahujúcom väčší počet akciových titulov. Metódy by mali byť otestované na príklade fondov kolektívneho investovania.

Vedúci : Mgr. Ing. Pavol Jurča, PhD.

Dátum zadania: 27.10.2010

Dátum schválenia: 08.11.2010

....
doc. RNDr. Margaréta Halická, CSc.
garant študijného programu

Kuižová
....
študent

....
vedúci práce

Dátum potvrdenia finálnej verzie práce, súhlas s jej odovzdaním (vrátane spôsobu sprístupnenia)

....
vedúci práce

Čestné prehlásenie

Vyhlasujem, že som bakalársku prácu vypracovala samostatne s použitím teoretických vedomostí a uvedenej odbornej literatúry.

Bratislava 2.6.2011 ....

Vlastnoručný podpis

Poďakovanie

Rada by som poďakovala vedúcemu práce Mgr. Ing. Pavlovi Jurčovi, PhD. za odborné rady, ochotu a čas pri vypracovávaní tejto práce. Chcela by som sa poďakovať aj mojej rodine a priateľom za ich podporu počas doterajšieho štúdia.

Abstrakt

Táto práca sa zaoberá vysvetlením a aplikáciou analýzy hlavných komponentov a faktorovej analýzy. Obe metódy sú vhodné na zjednodušenie štatistických analýz. Ak je prvotný počet premenných veľmi vysoký a existuje medzi nimi lineárna závislosť, tak pomocou týchto metód môžeme získať menší počet nových nezávislých premenných. Existuje veľa skúmaných problémov, ktorých riešenie by bolo jednoduchšie, ak by bol počet pozorovaných znakov nižší. Jedným z nich je určenie miery akciového rizika na slovenskom finančnom trhu, resp. identifikácia jeho systémového komponentu. Pomocou dostupných informácií o fondoch kolektívneho investovania a fondoch doplnkových dôchodkových spoločností sme sa snažili zistiť, či je možné zjednodušenie tohto problému a identifikácia spoločných faktorov pomocou analýzy hlavných komponentov alebo faktorovej analýzy. Výsledok práce ukazuje, že je možné znížiť dimenziu.

Abstract

This paper provides explanation and aplication of Principal Components Analysis and Factor Analysis. Both methods are useful for simplification of statistical analyses. As long as the original number of variables is very high and there is a linear dependance it is possible to obtain smaller number of new variables by using these methods. There are many research areas where the solution would be easier with smaller number of observations. One of them is the determination of equity risk on the slovak financial market, respectively identification its system component. Through public information about collective investments funds and pension fund management companies we try to find out if it is possible to simplify this problem and identify common factors with principal components analysis or factor analysis. Result of this paper illustrates that it is possible to reduce the dimension.

Obsah

1 Úvod

Jednou z možností ako zhodnotiť nadobudnuté finančné prostriedky je investovanie do akcií. Vývoj na finančných trhoch je nepredvídateľný. Kvôli snahe znížiť riziko a zvýšiť výnos je výhodnejšie investovať do viacerých rôznych akcií (portfólia akcií), čím dochádza k diverzifikácii rizika. Na Slovensku takúto možnosť ponúka kolektívne investovanie a III. pilier dôchodkového sporenia (fondy doplnkových dôchodkových spoločností). Podielové fondy, ktoré vytvára a spravuje správcovská spoločnosť, sa stávajú stále viac atraktívnejšou alternatívou zhodnocovania financií.

Pri investovaní do veľkého počtu rôznych akcií by bolo vhodné, aby dané portfólio akcií bolo reprezentované menšou vzorkou. Analýza menšieho počtu premenných je vždy jednoduchšia, obzvlášť čo sa týka akcií, ktorých vývoj nie je jednoznačný. Hlavným cieľom zníženia dimenzie je uľahčiť identifikáciu systémového akciového rizika, ktorému je vystavený slovenský finančný sektor ako celok. Práve identifikácii rizika vo finančnom sektore zo systémového pohľadu sa v súčasnosti venuje zvýšená pozornosť. To potvrdzuje aj vznik Európskeho výboru pre systémové riziká dňa 16. decembra 2010, ktorého poslaním je prispievať k predchádzaniu alebo zmierňovaniu systémových rizík pre finančnú stabilitu v EÚ, ktoré vyplývajú z vývoja v rámci finančného systému.

Na zníženie dimenzie pozorovaní slúžia dve príbuzné metódy:

1. Metóda hlavných komponentov

2. Faktorová analýza.

Obe analýzy patria do skupiny metód, ktoré slúžia na odhalenie skrytých vzťahov medzi premennými. Najdôležitejším predpokladom použitia oboch metód je vzájomná korelácia premenných. Výsledkom použitia oboch analýz je menšia skupina nezávislých premenných bez straty väčšej časti informácie.

Metóda hlavných komponentov je označovaná za jednu z najlepších aplikácií lineárnej algebry a jej použitie je veľmi rozmanité (dokonca aj v počítačovej grafike). Vznik tejto metódy siaha do začiatku 20. storočia, navrhol ju K. Pearson (1901) a o jej ďalší rozvoj sa zaslúžil H. Hotelling (1933). Opisu metódy hlavných komponentov je venovaná druhá kapitola práce. Začína sa základnou charakteristikou metódy a pokračuje vysvetlením princípu jej fungovania a použitia na skupinu premenných, pričom dôraz je kladený najmä na interpretáciu umožňujúcu pochopenie najdôležitejších pojmov a vzťahov, nie na jej teoretické odvodenie.

Tretia kapitola sa zaoberá o čosi zložitejšou faktorovou analýzou. Podstatou tejto metódy je taktiež analyzovať vzťahy medzi premennými a určiť nové premenné, ktoré umožňujú lepšie pochopiť analyzované dáta. Faktorová analýza je v podstate rozšírenie metódy hlavných komponentov. Jej základný rozdiel spočíva v tom, že je všeobecnejšia a počet nových premenných je potrebné určiť pred vykonaním faktorovej analýzy. Kvôli zložitejšiemu modelu a viacerým možnostiam riešenia je pri popise faktorovej analýzy použitej viac matematiky. Avšak nezachádza sa do detailov a dôležité je pochopiť princíp a zmysel jej použitia. Faktorová analýza je časťou skupiny všeobecných lineárnych modelov. Vzájomnému porovnaniu PCA a FA je venovaná štvrtá kapitola.

Analyzovaním konkrétnych dát pomocou oboch metód sa zaoberám v piatej kapitole. Ako už bolo spomínané, dobrou možnosťou investovania na slovenskom finančnom trhu je investovanie do podielových fondov. Analýza zahŕňa aj fondy doplnkových dôchodkových spoločností. Niektoré takéto fondy majú vo svojom portfóliu akciovú zložku, ktorej percentuálne zastúpenie je rôzne a s časom sa mení. Spolu tvoria tieto akcie všetkých fondov na slovenskom trhu jedno portfólio, ktorého výnosnosť sa budeme snažiť modelovať znížením dimenzie. Hlavným cieľom tejto práce je teda zistiť, či je možné toto portfólio popísať niektorými vybranými akciami, resp. fondami.

2 Analýza hlavných komponentov

Popis Analýzy hlavných komponentov je urobený hlavne na základe kníh [1] a [4] a internetových dokumentov [2] a [3] v zozname literatúry.

2.1 Základná charakteristika

Analýza hlavných komponentov (ďalej PCA - *Principal Components Analysis*) je neparametrická metóda, ktorej hlavným cieľom je zníženie dimenzie, čiže počtu začiatočných premenných. Deje sa to bez straty podstatnej časti informácie. Táto metóda zvýrazní podobnosti a odlišnosti medzi jednotlivými premennými. Dimenziu môžeme znížiť len vtedy, ak sú premenné korelované, to znamená, že medzi niektorými premennými existuje lineárna závislosť. V takom prípade môžeme podstatnú časť informácie vyjadriť menším počtom premenných.

Ak v praxi študujeme viacrozmerné údaje a chceme ich čo najlepšie analyzovať, vysoký počet premenných túto úlohu sťažuje. Veľmi často sa stáva, že sledujeme viac faktorov a na popísanie daného problému ich všetky nepotrebujeme. Ktoré premenné, resp. kombinácie premenných sú pre nás kľúčové, by malo byť rozpoznateľné po použití PCA. Keďže sa jedná o mnohorozmerné analýzy, grafická interpretácia nie je možná na odhalenie závislosti medzi nimi.

V tejto práci sa nebudem do hĺbky venovať matematickému vysvetleniu. Opis tejto metódy bude spočívať hlavne v tom, aby čitateľ pochopil, aká je hlavná podstata jej fungovania a prečo je dobré ju použiť vo viacrozmerných analýzach.

V najjednoduchšom dvojrozmernom prípade (ak skúmame dve premenné) môžeme ich závislosť zakresliť do grafu. Ak sa údaje zobrazia približne do priamky, tak sú zavislé a na popis nášho problému postačí jediná premenná. Preto má zmysel PCA používať minimálne v trojrozmerných analýzach.

Východiskom PCA je kovariančná, resp. korelačná matica, pôvodných premenných. Výsledkom sú nové latentné premenné, ktoré síce nie sú priamo merateľné, ale majú určitú schopnosť vecnej interpretácie vzťahov medzi pôvodnými premennými. Nové premenné, ktoré sú nekorelované, čiže lineárne nezávislé, sa nazývajú hlavné komponenty. Tieto hlavné komponenty sú vlastne lineárnymi kombináciami pôvodných premenných.

Z geometrického hľadiska ide o otočenie pôvodného súradnicového systému. Analýza hlavných komponentov nám už sama o sebe dáva výsledky, ktoré sú priamo analyzovateľné, ale môžeme ich ešte ďalej použiť v iných analýzach.

2.2 Transformácia údajov

Analýza hlavných komponentov sa používa na nájdenie skutočného rozmeru. Z nameraných údajov vytvoríme $m \times n$ maticu tak, že každý riadok je jeden n-rozmerný časový rad (napríklad denné hodnoty akcie za jeden rok). Podľa toho, koľko premenných skúmame, toľko riadkov bude mať matica pôvodných údajov. Pôvodné riadkové premenné nemusia byť len časové rady, ale môžeme skúmať rôzne vlastnosti na n vzorkách výrobkov. Tieto vlastnosti budú riadkové premenné. Ďalším príkladom je zisťovanie spokojnosti v rôznych oblastiach záujmu testovania (výška platu, pracovné prostredie, zamestnanecké výhody, ...) u n zamestnancov firmy. Nové premenné sú lineárnou kombináciou pôvodných, čo môžme zapísať takto:

$$Y = PX \tag{1}$$

Teda X je $m \times n$ matica pôvodných premenných, Y je $m \times n$ matica nových latentných premenných a riadky $P(m \times m$ matice) určujú už spomínané lineárne transformácie. Kvôli správnemu fungovaniu PCA musíme ale maticu X ešte trochu upraviť. Pre každú riadkovú premennú vypočítame strednú hodnotu, prípadne jej nevychýlený odhad, a potom ju od každého prvku odčítame. Z geometrického hľadiska to nie je už len otočenie, ale aj posunutie začiatočného bodu, pričom euklidovské vzdialenosti medzi objektmi sú zachovávané. Takto vznikne matica s riadkovými premennými, ktoré budú mať nulovú strednú hodnotu. Ak by sme tak neurobili, nedosiahli by sme potrebnú ortogonálnu transformáciu. Čiže ak $\widetilde{X}_1, \widetilde{X}_2, ..., \widetilde{X}_m$ je našich m pozorovaných premenných už s odčítanými strednými hodnotami a koeficienty a_{ij}, $i = 1, ...m, j = 1, ...m$, sú prvkami matice P, potom súbor transformovaných premenných $Y_1, Y_2, ..., Y_m$ vyzerá takto:

$$\begin{aligned}
Y_1 &= a_{11}\widetilde{X}_1 + a_{12}\widetilde{X}_2 + ...a_{1m}\widetilde{X}_m \\
Y_2 &= a_{21}\widetilde{X}_1 + a_{22}\widetilde{X}_2 + ...a_{2m}\widetilde{X}_m \\
&... \\
Y_m &= a_{m1}\widetilde{X}_1 + a_{m2}\widetilde{X}_2 + ...a_{mm}\widetilde{X}_m
\end{aligned} \tag{2}$$

Tieto nové transformované premenné sa nazývajú hlavné komponenty. Takto sú hlavné komponenty definované v knihe [1] v zozname literatúry, ale v internetových dokumentoch [2] a [3] sú za hlavné komponenty považované riadky matice P. Nastáva tu akási nekonzistencia. V podstate to ale nie je úplne odlišný prístup. Vieme, že riadky matice P určujú lineárnu transformáciu pôvodných premenných a riadky matice Y predstavujú už tieto konkrétne lineárne transformácie. Teda oba tieto prístupy definujú matice P a Y podľa vzťahu (1) rovnako, len ich inak pomenúvajú. V tejto práci budeme za hlavné komponenty považovať riadky matice Y, keďže už od začiatku chceme hľadať nové premenné, pomocou ktorých je možné znížiť dimenziu. PCA je nastavená tak, že hlavné komponenty sú navzájom (po dvojiciach) nekorelované. Korelácia je veličina, ktorou vieme určiť, či je medzi dvoma premennými lineárny vzťah. Čím je jej absolútna hodnota vyššia, tým je lineárna závislosť silnejšia (v kladnom alebo zápornom zmysle). Nulová hodnota korelácie, ale aj kovariancie, značí lineárnu nezávislosť. To znamená, že kovariančná matica hlavných komponentov je diagonálna. Zložky mimo diagonály (kovariancie medzi jednotlivými hlavnými komponentami) sú nulové, teda premenné sú po dvojiciach lineárne nezávislé. Diagonála obsahuje variancie jednotlivých nových premenných.

Zabezpečenie nezávislosti a ostatných vlastností hlavných komponentov vychádza z riešenia PCA. Známe sú dve možnosti:

- riešenie cez vlastné čísla a vlastné vektory kovariančnej, resp. korelačnej matice pôvodných premenných už s odčítanými strednými hodnotami
- riešenie cez singulárny rozklad kovariančnej, resp. korelačnej matice pôvodných premenných už s odčítanými strednými hodnotami

Ale podstatou tejto práce nie je matematické detailné vysvetlenie PCA. Vo všeobecnosti môžeme definovať vzťahy, ktoré musia platiť pre koeficienty matice P:

$$a_{i1}^2 + a_{i2}^2 + ... + a_{im}^2 = 1 \quad \text{pre každé } i = 1, 2, ..., m, \tag{3}$$

čo znamená, že rozptyl nových a pôvodných premenných sa rovná 1. Vlastnosť, že nové premenné sú navzájom nezávislé zaručuje vzťah:

$$a_{i1}a_{j1} + a_{i2}a_{j2} + ... + a_{im}a_{jm} = 0 \quad \text{pre všetky } i \neq j \text{ a } i, j = 1, 2, ..., m \tag{4}$$

Je to vlastne súčin 2 vektorov, ktorý keď sa rovná nule, znamená kolmosť týchto vektorov.

Teda zo vzťahov (3) a (4) vyplýva, že matica P je maticou ortonormálnych vlastných vektorov kovariančnej, resp. korelačnej matice pôvodných premenných.

2.3 Vlastnosti hlavných komponentov

V procese analýzy hlavných komponentov je dôležitým krokom zostrojenie kovariančnej matice pôvodných premenných. Keďže je táto matica štvorcová a symetrická, môžeme vypočítať jej vlastné čísla a vlastné vektory. Tieto vlastné vektory sa usporiadajú podľa prislúchajúcej vlastnej hodnoty od najväčšej po najmenšiu a tvoria riadky matice P. Tým sa dosiahne, že prvý hlavný komponent popisuje najväčšiu časť informácie. Pričom celkový objem informácie získame súčtom rozptylov jednotlivých premenných. Postupne sa množstvo tejto informácie znižuje, až na posledný hlavný komponent ostane len malá časť.

Vlastnosti hlavných komponentov:

1. $E(Y_i) = 0, i = 1, 2, ..., m,$
2. $D(Y_i) = \lambda_i, i = 1, 2, ..., m,$ kde $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m \geq 0$ sú vlastné čísla kovariančnej matice, resp. korelačnej matice,
3. $cov(Y_i, Y_j) = 0, i \neq j,$
4. $D(Y_1) \geq D(Y_2) \geq ... \geq D(Y_m) \geq 0,$

2.4 Štandardizácia premenných

Pri skúmaní rôznych premených sa stáva, že nie sú všetky v rovnakých meracích jednotkách. V takom prípade musíme údaje štandardizovať, to znamená odčítať strednú hodnotu a vydeliť disperziou, teda vychádzať z korelačnej matice. Štandardizácia je potrebná, ak sa disperzie premenných výrazne líšia. Napríklad ak sú premenné viaceré akcie, tak je pravdepodobné, že ich rozptyl je rôzny. Ak vychádzame z korelačnej matice, ďalší postup je taký istý ako pri kovariančnej matici. Vypočítame jej vlastné čísla a vlastné vektory, ktoré určujú lineárnu kombináciu pre nové premenné.

Ak však nemusíme údaje štandardizovať, tak je zo štatistického hľadiska lepšie vychádzať z kovariančnej matice, lebo k-ty hlavný komponent je taká lineárna kombinácia premenných, ktorý vysvetľuje k-tu najväčšiu časť celkového rozptylu premenných. Maximalizácia tohto rozptylu pri normovaných premenných má umelý charakter. V teórii je dokázané, že vlastnosti hlavných komponentov odvodené z korelačnej matice sú omnoho komplikovanejšie, ako je to pri hlavných komponentoch z kovariančnej matice.[1]

2.5 Počet hlavných komponentov

V tejto kapitole sa konečne dostaneme k tomu, ako znížiť dimenziu. Počet doteraz spomínaných hlavných komponentov je po aplikovaní PCA m, čo je takisto ako počet pôvodných premenných. Takže našou ďalšou úlohou je vyriešiť ako vybrať niekoľko k prvých hlavných komponentov. Vieme, že tieto nové latentné premenné sú usporiadané podľa objemu informácie, ktorý popisujú. Existuje viacero pravidiel na určenie počtu hlavných komponentov, ktoré nám postačia na vysvetlenie celkového rozptylu pôvodných údajov. Avšak tieto spôsoby nie sú vždy jednoznačné a výber daného počtu závisí aj od konkrétneho skúmaného problému. Pred konkrétnym popísaním daných pravidiel si pripomenieme pár faktov:

- Kovariančná matica hlavných komponentov je diagonálna.

[1] Stankovičová, I. - Vojtková, M. 2007. Viacrozmerné štatistické metódy s aplikáciami. Bratislava: Iura Edition, 2007. ISBN 978-80-8078-152-1.

- Keďže je to kovariančná matica, diagonála obsahuje disperzie hlavných komponentov (štandardné odchýlky umocnené na druhú) usporiadané od najväčšej po najmenšiu.
- Keďže je to diagonálna štvorcová matica, jej vlastné čísla sú priamo prvky diagonály. Z predchádzajúceho bodu vyplýva, že sú kladné a takisto sa rovnajú vlastným číslam hlavných komponentov.

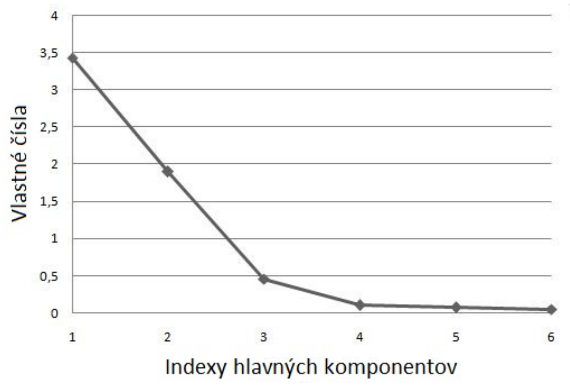
Spôsoby na určenie počtu hlavných komponentov:

1. Kaiserovo pravidlo: Vyberieme toľko prvých k hlavných komponentov, koľko vlastných hodnôt je väčších ako priemer všetkých vlastných čísel. Ak sme údaje štandardizovali, tak priemer vlastných čísel je rovný 1.
2. Môžeme sa rozhodnúť aj na základe grafického zobrazenia. Os x budú indexy hlavných komponentov a os y budú prislúchajúce vlastné hodnoty. Dôležitým krokom je nájsť zlom v tomto grafe a použiť prvých k hlavných komponentov po tento zlomový bod. Na vysvetlenie nám poslúži obrázok (1). Vidíme, že zlom nastal počnúc 3. hlavným komponentom, to znamená, že na analýzu vyberáme prvé 2 nové premenné.
3. Použijeme tie hlavné komponenty, ktoré popisujú aspoň napr. 80% celkového rozptylu údajov.

Percento určujúce celkový rozptyl údajov vyjadruje, akú časť z celkovej variability dát vysvetľuje jeden alebo viacero hlavných komponentov. Vypočítame ho pomocou daných disperzií hlavných komponentov. Konkrétne vezmeme ich odmocniny (štandardné odchýlky) a označíme si ich σ_i, $i = 1, 2, ...m$. *Individuálne percento* $\sigma_j / \sum_{i=1}^{m} \sigma_i$ určuje, akou mierou prispieva j-ty hlavný komponent k celkovej variabilite údajov. *Kumulatívne percento* $\sum_{j=1}^{k} \sigma_j / \sum_{i=1}^{m} \sigma_i$ vyjadruje, aký objem informácie popisuje prvých k hlavných komponentov.

2.6 Komponentné váhy

Komponentné váhy sú vlastne korelácie medzi hlavnými komponentmi a centrovanými premennými. Určíme ich vypočítaním korelačnej matice medzi nimi, čiže určením



Obr. 1: Indexový graf vlastných čísel

koeficientov korelácie medzi danými dvoma premennými, ktoré sú určené vzťahom:

$$\rho_{X_i,Y_j} = \frac{\lambda_j}{\sigma_i} \quad \text{pre každé} \quad i, j = 1, 2, ..., m \tag{5}$$

Pričom zachovávame symboliku, takže $X_i, i = 1, 2, ..., m$, sú pôvodné premenné a $Y_j, j = 1, 2, ..., m$, sú hlavné komponenty. Teda λ_j sa rovná rozptylu j-teho hlavného komponentu a σ_i je štandardná odchýlka i-tej pôvodnej premennej. Aby malo zmysel použiť PCA, každý z prvých k hlavných komponentov by mal korelovať aspoň s 2 pôvodnými premennými kvôli možnosti zníženia dimenzie. To znamená, že koeficient korelácie by mal byť v absolútnej hodnote aspoň $0,5$. Ak by každý hlavný komponent koreloval iba s jednou pôvodnou premennou, znamenalo by to, že medzi premennými $X_i, i = 1, 2, ..., m$, nie je lineárna závislosť. V takom prípade nie je možné znížiť dimenziu. Ak hlavný komponent koreluje s viacerými pôvodnými premennými, tak určuje aj vzťah medzi nimi. Komponentné váhy nám môžu slúžiť aj na kontrolu, či počet vybraných hlavných komponentov je dostatočný, aby sme zachovali čo najväčší objem informácie.

3 Faktorová analýza

Táto kapitola je spracovaná hlavne na základe knihy [1] a internetového dokumentu [5] v zozname literatúry.

3.1 Základná charakteristika

Faktorová analýza (ďalej FA - *Factor Analysis*) je štatistická metóda, ktorá je používaná na odhalenie vzájomných vzťahov medzi veľkým množstvom premenných. Kvôli zjednodušeniu štatistických analýz je častokrát potrebné a veľmi vhodné zníženie dimenzie pozorovaných premenných. Základným cieľom FA je nájdenie spôsobu ako zhustiť informácie obsiahnuté vo veľkom počte pôvodných premenných do menej rozmernej množiny nových premenných (tzv. *faktorov*) pri minimálnej strate informácie. FA sa snaží zistiť, či možno sledované premenné rozdeliť do skupín, v ktorých by ich vzájomné korelácie boli významné a medzi týmito skupinami by zase významné neboli. Dôležitým predpokladom FA je preto závislosť pôvodných premenných. Kým PCA sa snaží nájsť hlavné komponenty tak, aby bola vysvetlená čo najväčšia časť variability pôvodných premenných, pričom nevysvetlenou časťou variancie sa ďalej nezaoberá. FA nevysvetlenú varianciu popisuje tzv. *individuálnym faktorom* vstupujúcim do vývoja jednotlivých premenných. Faktory, nové vytvorené premenné, sú v praxi priamo nemerateľné, ale umožnujú lepšie pochopiť analyzované dáta. Tak ako aj pri PCA môžeme tieto nové premenné použiť v ďalších štatistických analýzach.

Začiatky vývoja faktorovej analýzy siahajú až do začiatku 20. storočia. Vznikla v psychológii v roku 1904 zásluhou Ch. Spearmana, kde sa doteraz výrazne používa. O jej rozvoj sa v spoločenskovedných disciplínach zaslúžili aj vedci L.L. Thurstone, R.B. Cattel, C. Burt, G. Thomson a iní. Napriek jej dlhej histórii má FA v štatistickej literatúre malé pokrytie a mnohí štatistici jej pripisujú nízku štatistickú vážnosť. Časť tejto kontroverznosti sa začala okolo roku 1950 a pokračuje až dodnes v dôsledku chýbajúcej dohody týkajúcej sa významu FA. Štatistici ju kritizujú pre jej nejednoznačné riešenie, subjektivitu v niektorých jej cieľoch a krokoch, hmlistú interpretáciu a približnosť výsledkov.[2] Faktorová analýza je v niektorých textoch zamieňaná s metódou

[2]Stankovičová, I. - Vojtková, M. 2007. Viacrozmerné štatistické metódy s aplikáciami. Bratislava:

hlavných komponentov, čo prispieva k jej kontroverznosti. FA sa najviac v štatistike zaoberali vedci D.N. Lawley, M.S. Bartlett, C.R. Rao a ďalší.

Faktorová analýza ako pojem sa vzťahuje aj na všetky metódy analýzy dát používajúce matice faktorov, vrátane PCA a FA. Teda existujú tu nejasnosti ohľadom vymedzenia tohto pojmu. Názov faktorová analýza buď pomenúvava určitú skupinu štatistických metód alebo konkrétnu analýzu, ktorú budem popisovať v tejto kapitole. Preto sa na označenie nepozorovateľnej, hypotetickej premennej používa skôr pojem *spoločný faktor (common factor)* ako len faktor, ktorý prispieva k vysvetleniu aspoň dvoch pôvodných premenných.

Druhy faktorovej analýzy:

1. **prieskumná (exploratory)** - používa sa, ak máme k dispozícii málo znalostí o faktorovej štruktúre. Prieskumná FA sa snaží nájsť túto skrytú východiskovú štruktúru pomerne veľkej množiny premenných. Patrí medzi najbežnejšie formy FA.

2. **potvrdzujúca (confirmatory)** - používa sa na overenie nejakej hypozézy o faktorovej štruktúre, ktorá je známa. Premenné sa vyberajú na základe tejto teórie a potvrdzujúca FA sa používa na zistenie očakávaného počtu faktorov. Výskumnícka hypotéza zahrňuje predpoklady asociácie každého faktora s konkrétnou podmnožinou premenných.

V tejto kapitole sa budem zaoberať prieskumnou FA, keďže podstatou mojej práce je odhalenie vzťahou medzi premennými (časovými radmi portfólia akcií).

3.2 Matematický model

Faktorová analýza je metóda, ktorá slúži na zníženie počtu premenných. Prvým krokom je tak ako pri PCA zostavenie $m \times n$ matice X, ktorej každý riadok je jedna pozorovateľná premenná (náhodná veličina). Pripomínam, že každá premenná je v našom prípade časový rad akcie alebo portfólia akcií. Matici X prislúcha m-rozmerný vektor stredných hodnôt μ_X a kovariančná matica Σ_m hodnosti m. Výsledkom FA je k v pozadí

Iura Edition, 2007. ISBN 978-80-8078-152-1.

stojacich *spoločných faktorov* $F_1, F_2, ..., F_k$, ktorých by malo byť výrazne menej ako m. Potom môžeme j-tu pozorovateľnú náhodnú premennú $X_j, j = 1, 2, ...m$, zapísať v tvare:

$$X_j = \mu_j + a_{j1}F_1 + a_{j2}F_2 + ... + a_{jk}F_k + \epsilon_j \tag{6}$$

kde

$\epsilon_j, j = 1, 2, ...m$, sú *špecifické faktory* (náhodná (chybová) zložka), čo sú nepozorovateľné, hypotetické premenné. Každý prispieva k vysvetleniu jednej pôvodnej premennej.

$a_{jl}, l = 1, 2, ...k$, sú faktorové váhy (saturácie), ktoré vyjadrujú vplyv l-tého spoločného faktora na premennú X_j.

V maticovom tvare môžeme FA vyjadriť takto:

$$X = \mu_X + AF + \epsilon, \quad resp. \quad X - \mu_X = AF + \epsilon \tag{7}$$

kde
A je matica faktorových váh typu $m \times k$,
F je matica spoločných faktorov typu $k \times n$,
ϵ je matica špecifických faktorov,
X je matica pôvodných merateľných premenných, nazývaných aj *indikátory*.

Faktorové váhy (saturácie) sú pri splnení určitých podmienok vlastne kovariancie medzi indikátormi a nepozorovateľnými faktormi. Tak ako aj pri PCA môžeme bez straty všeobecnosti vychádzať z normovaných premenných Z_j. Maticový model tohto bežnejšieho prípadu FA vyzerá nasledovne:

$$\begin{aligned} \frac{X - \mu_X}{\sigma_X} &= A^*F^* + \epsilon^* \\ Z &= A^*F^* + \epsilon^* \end{aligned} \tag{8}$$

kde prvky matice $A^*(a^*_{jl})$ predstavujú korelačné koeficienty medzi indikátormi X_j a faktormi F_l.

Vzhľadom na nezávislosť spoločných faktorov rozlišujeme dva základné modely FA:

- *kosouhlý (oblique)* - model s navzájom korelovanými faktormi.
- *ortogonálny* - model FA, ktorý budem opisovať v nasledujúcich častiach tejto kapitoly.

Ortogonálny model FA musí spĺňať nasledujúce predpoklady:

1. **Spoločné faktory** $F_l, l = 1, 2, ...k$:
   - lineárna nezávislosť,
   - rovnako rozdelené náhodné veličiny,
   - $E(F_l) = 0$,
   - $D(F_l) = 1$.
2. **Špecifické faktory** $\epsilon_j, j = 1, 2, ...m$:
   - náhodné premenné,
   - $E(\epsilon_j) = 0$,
   - $D(\epsilon_j) = e_j$,
   - $cov(\epsilon_i, \epsilon_j) = 0$, čo značí nezávislosť po dvojiciach.
3. $cov(F_l, \epsilon_j) = 0$, nezávislosť F_l a ϵ_j pre každú dvojicu $l = 1, 2, ...k$ a $j = 1, 2, ...m$.

Z týchto predpokladov vyplýva nasledujúce:

$$\Sigma_X = cov(X) = cov(\mu_X + AF + \epsilon) = Acov(F)A^T + cov(\epsilon) = AI_kA^T + E = AA^T + E \quad (9)$$

V matematickom modeli FA je známa iba matica X pôvodných premenných, preto nebude existovať len jedno riešenie tohto problému. Nájdením správneho riešenia sa budeme zaoberať v nasledujúcich podkapitolách. Teraz sa ešte zamerajme na kovariančnú maticu Σ_X, ktorá na základe vzťahu (9) slúži na nájdenie matice faktorových váh matice A a kovariančnej matice špecifických faktorov E, a dôsledky, ktoré vyplývajú z daných predpokladov:

- **Rozptyl j-tej premennej:**

$$D(X_j) = s_j^2 = (a_{j1}^2 + a_{j2}^2 + ... + a_{jk}^2) + e_j^2 = h_j^2 + u_j^2 \quad (10)$$

kde h_j^2 je tzv.*komunalita*, čo je rozptyl j-tej pozorovanej premennej X_j vysvetlený pôsobením k spoločných faktorov a u_j^2 je tzv.*unicita*, čo je reziduálny rozptyl j-tej pozorovanej premennej.[3] Komunality môžu byť interpretované ako hodnovernosť daného indikátora. Vo všeobecnosti ukazujú, pre ktorú meranú premennú faktorová analýza pracuje lepšie a pre ktorú menej dobre. Ak je komunalita nízka, tak model faktorovej analýzy nie je pre danú premennú správny a ak je možné, mala by byť z modelu odstránená. Nízke komunality v určitej množine premenných znamenajú ich malý vzájomný vzťah. Avšak v modeli nie je dôležitá len výška komunality, ale aj interpretovateľnosť faktora, ktorý vplýva na danú premennú. Niekedy môže byť aj nízka komunalita, napr. 0,25, významná, ak prispieva k dobrej definovanosti faktora.

- **Kovariancia medzi pôvodnými premennými** sa dá vyjadriť aj pôsobením spoločných faktorov:

$$cov(X_i, X_j) = \sum_{h=1}^{k} a_{ih}a_{jh} \tag{11}$$

- **Rozptyl j-tej normovanej premennej Z_j:**

$$D(Z_j) = 1 = (a_{j1}^{*2} + a_{j2}^{*2} + ... + a_{jk}^{*2}) + e_j^{*2} = h_j^{*2} + u_j^{*2} \tag{12}$$

Model faktorovej analýzy vychádza častejšie z korelačnej matice indikátorov (matica R) a môžeme ho zapísať v tvare:

$$R = A^*(A^*)^T + E^* \tag{13}$$

kde $A^* = \rho(X, F)$ je korelačná matica obsahujúca koeficienty v intervale $\langle -1; 1\rangle$.

3.3 Štatistické problémy FA

Medzi dôležité predpoklady faktorového modelu kvôli jednoduchosti patrí lineárny vzťah medzi pôvodnými premennými a faktormi, počet $k < m$ spoločných faktorov a komunality by mali byť blízke 1. Avšak pri odhade parametrov vznikajú štatistické

[3] Stankovičová, I. - Vojtková, M. 2007. Viacrozmerné štatistické metódy s aplikáciami. Bratislava: Iura Edition, 2007. ISBN 978-80-8078-152-1.

problémy. Odhad parametrov modelu FA vychádza z kovariančnej matice Σ_X, resp. korelačnej matice R, ktorá je častejšie tou východiskovou maticou. Preto sa v tejto časti zameriame na vysvetlenie problému cez korelačnú maticu, pričom to ekvivalentne platí aj pre kovariančnú maticu. Podľa vzťahu (13) musíme z jednej matice odhadnúť dve, a to matice A^* a E^*. Ak by sa počet faktorov rovnal počtu pôvodných premenných, tak bez väčších problémov by sme jednoznačne rozložili maticu R na $A^*(A^*)^T$, pričom $E^* = 0$. Ale faktorová analýza je metóda, ktorá slúži na zníženie dimenzie. Preto musí byť počet faktorov výrazne menší ako počet indikátorov a vtedy existuje nekonečne veľa riešení (keďže matica $E^* \neq 0$). Táto *nejednoznačnosť faktorového modelu* vyvoláva možnosť rotácie faktorov a hľadanie, čo najlepšieho riešenia. Metódy rotácie faktorov budú popísané v osobitnej podkapitole.

Pri hľadaní vhodného riešenia je hlavne podstatné, aby sa faktory ľahko interpretovali. Spoločné faktory musia aj čo najlepšie vysvetľovať koreláciu medzi indikátormi. Určenie ich čo najmenšieho počtu patrí tiež medzi štatistické problémy FA. Jednoduchý nie je ani odhad riešenia FA, aj keď máme vopred zadaný počet spoločných faktorov. Problémom môže byť určenie lineárne nezávislých faktorov z danej korelačnej matice R.

3.4 Metódy odhadu parametrov

Existuje mnoho metód odhadu parametrov FA. Môžeme ich rozdeliť na neiteračné a iteračné.

Neiteračné metódy - metóda hlavných faktorov (PFA), metóda hlavných komponentov (PCA), Harrisova neiteračná kanonická faktorová analýza a imidž-analýza.

Iteračné metódy - metóda nevážených najmenších štvorcov (ULS), metóda maximálnej vierohodnosti (ML), alfa-faktorová analýza (ALPHA) a iteračná metóda hlavných faktorov (PRINIT).

Vo všeobecnosti dáva každá metóda rôzne riešenie, ale ak je počet indikátorov a rozsah výberu vysoký, tak výsledky sú skoro rovnaké. Bližšie si popíšeme metódu hlavných faktorov a metódu maximálnej vierohodnosti.

3.4.1 Metóda maximálnej vierohodnosti

Hlavnou výhodou metódy maximálnej vierohodnosti (ďalej ML - *Maximum Likelihood*) je nezávislosť od použitých meracích jednotiek. Na jej vývoji sa významne podieľal D. N. Lawleym (1940) a K. G. Jöreskogom (1968, 1975). Táto metóda je na odhad parametrov FA vhodná, ak pozorované údaje pochádzajú z náhodného výberu a majú viacrozmerné normálne rozdelenie, čo je aj predpokladom použitia metódy ML. Logaritmus funkcie vierohodnosti má tvar:

$$\ln[L(A,E)] = -\frac{1}{2}(n-1)[\ln(det(AA^T+E)) + st(AA^T+E)^{-1}S] \tag{14}$$

kde

S je výberová kovariančná matica, ktorej prvky majú Wishartovo rozdelenie,

$st(\cdot)$ je stopa matice.

Wishartovo rozdelenie je viacrozmerným zovšeobecnením chí-kvadrát rozdelenia. Ak $p \times n$ matica X pochádza z viacrozmerného normálneho rozdelenia ($N_p(0,V)$), tak matica X^TX má Wishartovo rozdelenie s n stupňami voľnosti.

Maximalizáciou logaritmu zo vzťahu (14) a pridaním dodatočného predpokladu, že matica $A^TE^{-1}A$ je diagonálna, dostaneme jednoznačné riešenie pre matice A a E. Matica $A^TE^{-1}A$ je diagonálna, ak platí:

- $F \sim N_k(0, I_k)$,
- podmienené rozdelenie $X - \mu$, dané F, je $N_m(AF, E)$,

potom podmienené rozdelenie F, dané X, je $N_k(A^T\Sigma_X(X-\mu), [A^TE^{-1}A + I_k]^{-1})$.

ML poskytuje určité kritériá pre odhad počtu spoločných faktorov a umožňuje aj testovať hypotézu o jeho vhodnosti.

3.4.2 Metóda hlavných faktorov

Metóda hlavných faktorov (ďalej PFA - *Principal factor analysis*) je vlastne metóda hlavných komponentov aplikovaná na *redukovanú* kovariančnú resp. korelačnú maticu indikátorov. Redukovaná kovariančná matica je daná vzťahom:

$$\Sigma_r = \Sigma_X - E \tag{15}$$

resp. redukovaná korelačná matica:

$$R_r = R - E^* \tag{16}$$

kde E a E^* sú matice špecifických faktorov definované presne tak, ako v časti 3.3. Na základe predpokladov FA vieme, že tieto matice majú nenulové len prvky diagonály, ktoré predstavujú rozptyly špecifických faktorov. Podľa vzťahov (10) a (12) vidíme, že diagonálne prvky redukovanej matice sú príslušné komunality.

Dôležitým krokom PFA je teda určenie hodnôt komunalít (aspoň približne). K dispozícii máme viacero možností. Ak je východiskovou korelačná matica sú to napríklad:

- *Squared Multiple Correlation* - štvorec koeficientu mnohonásobnej korelácie premennej X_j s ostatnými premennými,
- *Proportional to Squared Multiple Correlation* - priemerný štvorec koeficientu korelácie premennej X_j s ostatnými premennými,
- *Maximum Absolute Correlation* - maximálny koeficient korelácie premennej X_j s ostatnými premennými,
- *Set all priors to one* - všetky komunality zvolíme 1, vtedy sa PFA mení na PCA, čiže rozptyly špecifických faktorov sa rovnajú nule.[4]

Koeficient mnohonásobnej korelácie, spomínaný v prvom bode, vyjadruje lineárny vzťah medzi viac ako dvoma premennými. Meria sa koeficientom determinácie R^2, ktorý určuje vhodnosť lineárnej regresie. Matematický model FA sa podobá na model lineárnej regresie, avšak pri FA poznáme len maticu X, resp. Z pôvodných premenných.

PFA neumožňuje testovať hypotézu o počte spoločných faktorov, ktoré sú potrebné na opísanie vzťahov medzi indikátormi a nie je ani invariantná k zmenám mierky premenných.

3.5 Heywoodov prípad

Pri použití FA nastávajú prípady, keď riešenie nie je správne (prípustné). Vo FA nie je prípustné, aby komunality boli väčšie ako 1 a zároveň matica špecifických fak-

[4] Odstavec spracovaný na základe: Stankovičová, I. - Vojtková, M. 2007. Viacrozmerné štatistické metódy s aplikáciami. Bratislava: Iura Edition, 2007. ISBN 978-80-8078-152-1.

torov (matica E) musí byť pozitívne definitná. Stačí, aby jeden prvok matice E bol nekladný a vedie to k nevhodnému riešeniu. Takýto prípad sa nazýva Heywoodov prípad. Príčin takého výsledku je viacero, napríklad nízky počet indikátorov, malý rozsah vzorky, vysoký počet spoločných faktorov, zlý odhad komunalít, či prítomnosť outlierov. Ak dokážeme odhaliť príčinu neprípustného riešenia, môžeme sa ju pokúsť odstrániť. Niekedy je potrebné zvýšiť počet premenných alebo odstrániť problematické premenné.

3.6 Počet spoločných faktorov

Pred určením kritérií počtu spoločných faktorov si pripomenieme, že faktorová analýza spočíva vo vyriešení rovnice $\Sigma_X = AA^T + E$, ak vychádzame z kovariančnej matice Σ_X alebo $R = A^*(A^*)^T + E^*$, ak vychádzame z korelačnej matice R. Potrebujeme odhadnúť faktorové váhy, čo predstavuje matica A, resp. A^* a koeficienty matice E, resp. E^*. Aby bola FA užitočná, chceme, aby k spoločných faktorov bolo výrazne menej ako indikátorov. Od hodnoty k závisí, či budeme riešiť problém s viacerými neznámymi ako rovniciami alebo neznámych bude nanajvýš toľko, koľko rovníc. Vo všeobecnosti však ani v druhom prípade nebude existovať jednoznačné riešenie, pretože každé riešenie môže byť nekonečne veľakrát rotované.

Na určenie približného počtu spoločných faktorov existuje tak, ako v metóde hlavných komponentov viacero objektívnych kritérií, ale i subjektívnych rád:

- Počet spoločných faktorov môže byť rovnaký ako počet vlastných čísel redukovanej korelačnej matice väčších ako jedna.
- Celková komunalita je daná ako súčet komunalít všetkých m indikátorov. Vybraný počet spoločných faktorov by mal reprezentovať viac ako 90% tejto celkovej komunality.
- Ďalším kritériom môže byť tak, ako v PCA graf (tzv. *scree plot*) vlastných čísel redukovanej kovariančnej resp. korelačnej matice v závislosti od indexu spoločného faktora. Dôležitým je už spomínaný zlom na krivke ako na obrázku 1. Za optimálny počet spoločných faktorov budeme považovať index pred bodom zlomu.

- Naopak vynechať by sme mali tzv. *triviálne* faktory. Takéto spoločné faktory nie sú vhodné pre našu analýzu, pretože významne korelujú len s jedným indikátorom.

V niektorých prípadoch sa stane, že počet spoločných faktorov je známy dopredu z iných analýz. Avšak objektívne kritérium na určenie počtu spoločných faktorov poskytuje metóda maximálnej vierohodnosti. Test pomerom vierohodností vyžaduje predpoklad viacrozmerného normálneho rozdelenia pôvodných premenných a dostatočne veľký rozsah vzorky n. Pre zvolený počet spoločných faktorov testujeme nulovú hypotézu:

$$H_0 : R = A^*(A^*)^T + E^* \tag{17}$$

pričom testovacia štatistika V má pri dostatočne veľkom n približne chí-kvadrát rozdelenie s počtom stupňov voľnosti $p = [(m-k)^2 - (m+k)]/2$.Počet stupňov voľnosti p musí byť kladné číslo, inak je model faktorovej analýzy zle definovaný. Teda pre počet spoločných faktorov k musí platiť:

$$k < \frac{1}{2}(2m + 1 - \sqrt{8m+1}) \tag{18}$$

3.7 Metódy rotácie faktorov

Rotácia spoločných faktorov umožňuje vytvoriť zrozumiteľnejší výstup a väčšinou je nutná kvôli uľahčeniu interpretácie faktorov. Pre získanie čo najlepšieho riešenia by každá korelácia dvoch premenných mala byť vysvetlená čo najmenším počtom faktorov. Znamená to, že pre každú dvojicu stĺpcov matice faktorových váh by malo byť málo premenných, ktoré majú vysoké váhy v obidvoch stĺpcoch.[5] Existuje viacero spôsobov rotácie faktorov, ale väčšina z nich sa snaží získať čo najviac faktorových váh blízkych nule a tých zvyšných čo najviac blízkych jednej. Matica faktorových váh A, resp. A^* nie je určená jednoznačne a rotácia faktorov nám poskytuje možnosť hľadať takú maticu saturácií, ktorá bude zmysluplnejšia. Ortogonálnou rotáciou matice A získame nové riešenie. Teda ak je matica A riešením faktorového modelu $\Sigma_X = AA^T + E$, tak je

[5] Odstavec spracovaný na základe: Stankovičová, I. - Vojtková, M. 2007. Viacrozmerné štatistické metódy s aplikáciami. Bratislava: Iura Edition, 2007. ISBN 978-80-8078-152-1.

riešením aj matica $B = AT$, kde T je ortogonálna matica. Platí to aj pre maticu A^*, ak vychádzame z korelačnej matice indikátorov.

Metódy rotácie faktorov sa rozdeľujú na dve skupiny:

1. **Ortogonálne (pravouhlé)** — zachovávajú nekorelované spoločné faktory. Suma vlastných čísel redukovanej matice zostáva rovnaká, ale mení sa prerozdelenie vysvetlenej variability medzi faktormi.

2. **Kosouhlé (šikmé)** — vedú k riešeniu so závislými spoločnými faktormi, takže navyše poskytujú korelačnú maticu medzi nimi. Komunality sa zachovávajú, ale faktorové váhy sa nerovnajú kovarianciám resp. koreláciam medzi indikátormi a faktormi. Tieto metódy sú niektorými autormi odmietané a inými vítané. Pre prax sú niekedy reálnejšie závislé spoločné faktory. Do tejto skupiny patria: *oblimin, oblimax, maxplane, promax, orthoblique,...* .

Ortogonálne rotácie:

- *varimaxná* — ortogonálna rotácia faktorov maximalizáciou súčtu rozptylov druhých mocnín faktorových váh v stĺpcoch. Ide o zjednodušenie stĺpcov matice A minimalizáciou počtu premenných vysoko korelovaných s jednotlivými faktormi. Autorom tejto metódy je H. F. Kaiser (1958).

- *quartimaxná* — ortogonálna alternatíva, ktorá minimalizuje počet faktorov potrebných na vysvetlenie každej premennej. Ide o zjednodušenie riadkov matice A.

- *orthomaxná* — kombinácia varimaxnej a quartimaxnej transformácie. Existujú ešte modifikácie tejto metódy: biquartimax a equamax.

3.8 Faktorové skóre

Faktorové skóre sú hodnoty priamo nepozorovateľných premenných. Ich určenie nie je jednoduché, pretože pre maticu indikátorov X nie je možné explicitne určiť maticu spoločných faktorov F. Pri rozsahu vzorky n dostávame maticu $n \times k$ nových hodnôt. Pre odhad faktorového skóre bolo vyvinutých viacero metód, napr. *viacnásobná re-*

gresná metóda, Bartlettova metóda (vážená metóda najmenších štvorcov), Harmanova metóda a iné.[6]

3.9 Kritéria na určenie štatistickej významnosti faktorových váh

Pri vecnej interpretácii faktora uvažujeme len s významnými premennými. Predtým však musíme určiť významnosť váh indikátorov. Predpokladajme, že vychádzame z korelačnej matice R, teda pôvodné premenné sú normované a matica faktorových váh A^* je korelačnou maticou medzi pôvodnými premennými a faktormi s hodnotami z intervalu $\langle -1, 1\rangle$. Čím vyššia je faktorová saturácia, tým majú faktor a indikátor viac spoločného. Významné premenné sú väčšinou tie, ktorých váha je mimo intervalu $(-0,5; 0,5)$. Niekedy sa za už významné považujú aj faktorové váhy v absolútnej hodnote viac ako $0,3$. Nezohľadňuje to však rozsah vzorky n, preto sa odporúča použiť nasledujúci test:

$$|a_{jl}| > \frac{t_{(1-\alpha/2,n-2)}}{\sqrt{t_{(1-\alpha/2,n-2)} + (n-2)}} \sqrt{\frac{k}{k-1-l}} \tag{19}$$

kde

$j = 1, 2, ...m$ a $l = 1, 2, ...k$,

t označuje hodnotu príslušného kvantilu Studentovho rozdelenia s počtom stupňov voľnosti $(n-2)$.

Faktorová váha a_{jl} spĺňajúca danú podmienku sa považuje za štatisticky významnú na hladine významnosti α. Zlomok $\sqrt{\frac{k}{k-1-l}}$ sa nazýva tzv. *Burt-Banksova oprava*, používa sa, aby uvedená štatistika bola aj funkciou k spoločných faktorov. Zohľadňuje aj poradie spoločného faktora.

3.10 Zhrnutie krokov aplikovania faktorovej analýzy

1. **Výber premenných a zozbieranie dát**

   Pri výbere premenných je dôležitá vecná analýza problému, pri ktorej treba brať do úvahy veľkosť vzorky. Rozsah premenných n (počet štatistických jednotiek)

[6] Stankovičová, I. - Vojtková, M. 2007. Viacrozmerné štatistické metódy s aplikáciami. Bratislava: Iura Edition, 2007. ISBN 978-80-8078-152-1.

musí byť pri počte m zvolených indikátorov dostatočný. Existujú len nejaké empirické pravidlá na určenie výšky n, napr. $n > 10m$.

2. **Odhad korelačnej matice a posúdenie vhodnosti dát pre faktorovú analýzu**

Faktorová analýza je metóda na zníženie dimenzie, preto musia byť vybrané premenné závislé. Zistíme to vypočítaním ich výberovej kovariančnej, resp. korelačnej matice (ak sú premenné vyjadrené v rôznych meracích jednotkách). Z tejto matice sa odvíjajú aj ďalšie odhady parametrov.

Test, či korelačná matica sa rovná jednotkovej matici (premenné sú nezávislé), sa nazýva Bartlettov test sférickosti. Tento test však vyžaduje predpoklad o viacrozmernom normálnom rozdelení matice X. Ak zamietneme nulovú hypotézu, tak dáta sú závislé a vhodné pre FA.

Na posúdenie vhodnosti dát sa častejšie používa štatistika KMO (Kaise-Meyer-Olkin), ktorá je daná vzťahom:

$$KMO = \frac{\sum_{i=1}^{m} \sum_{j=1, j \neq i}^{m} r_{ij}^2}{\sum_{i=1}^{m} \sum_{j=1, j \neq i}^{m} r_{ij}^2 + \sum_{i=1}^{m} \sum_{j=1, j \neq i}^{m} r_{parc.,ij}^2} \tag{20}$$

kde r_{ij} je párový koeficient korelácie medzi X_i a X_j a $r_{parc.,ij}$ je parciálny koeficient korelácie.

KMO štatistika sa počíta ako celková miera adekvátnosti (vhodnosti) výberových dát pre FA a aj ako čiastková miera adekvátnosti pre jednotlivé indikátory. Je to miera homogenity premenných. Hodnoty KMO miery sa netestujú, ale používajú sa nasledujúce odporúčania podľa Kaisera a Ricea (1974)[7]:

$KMO \geq 0,9$ - vynikajúca adekvátnosť,

$KMO \in \langle 0,8; 0,9)$ - chvályhodná adekvátnosť,

$KMO \in \langle 0,7; 0,8)$ - stredne užitočné dáta,

$KMO \in \langle 0,6; 0,7)$ - priemerná adekvátnosť,

$KMO \in \langle 0,5; 0,6)$ - slabá adekvátnosť,

$KMO < 0,5$ - nedostatočná adekvátnosť.

[7]Stankovičová, I. - Vojtková, M. 2007. Viacrozmerné štatistické metódy s aplikáciami. Bratislava: Iura Edition, 2007. ISBN 978-80-8078-152-1.

Takže vidíme, že sa odporúčajú vyššie hodnoty KMO ako $0,5$. Najvhodnejšie sú hodnoty od $0,8$, ale miera KMO okolo 0,6 je ešte v tolerancii. Zvýšiť hodnotu miery KMO môžeme, ak vo FA nepoužijeme premennú s nízkou individuálnou hodnotou miery KMO.

3. **Odhad parametrov faktorového modelu**

   Ak sú vybrané dáta vhodné, ďalším krokom je rozhodnúť sa, ktorú metódu použijeme na odhadovanie parametrov (faktorových váh) modelu FA. Ich opis je v časti 3.5.

4. **Rotácia a interpretácia**

   Začiatočný odhad matice faktorových váh nie je často dobre interpretovateľný, preto sa odporúča použiť niektoré z techník rotácie spoločných faktorov a výsledky porovnať. Vyberieme ten výsledok, ktorý spĺňa teoretické požiadavky, je logický a dobre vysvetliteľný.

5. **Odhad faktorových skóre a aplikácia výsledkov faktorovej analýzy**

   Faktorová analýza je častokrát používaná na získanie nekorelovaných premenných pre ďalšie analýzy (regresné modely, zhluková analýza, diskriminačná analýza,...). Preto je potrebné odhadnúť faktorové skóre.

4 Porovnanie faktorovej analýzy a metódy hlavných komponentov

Hlavnou podstatou metódy hlavných komponentov a faktorovej analýzy je vytváranie nových premenných a zníženie dimenzie dát pri minimálnej strate informácie. FA môžeme do určitej miery považovať za rozšírenie PCA. Pre obe metódy je potrebné, aby pôvodné premenné boli lineárne závislé.

PCA je jednoduchšia a priamočiarejšia metóda a vysvetľuje len rozptyl premenných. Výpočty FA sú oveľa náročnejšie a zložitejšie.

PCA je závislá od meracích jednotiek premenných a nevieme ani jednoznačne rozhodnúť, či zvolený počet hlavných komponentov je správny. FA riešená metódou maximálnej vierohodnosti je invariantná k zmenám meracích jednotiek a poskytuje test počtu spoločných faktorov.

Faktorová analýza je síce všeobecnejšia ako PCA, ale má veľa predpokladov a subjektívnych aspektov.

5 Analýza dát

Ako už bolo spomínané v úvode práce, PCA a FA použijeme na testovanie portfólia akcií z vybraných fondov kolektívneho investovania a fondov doplnkových dôchodkových spoločností na slovenskom finančnom trhu. Podielovým fondom sa rozumie spoločný majetok podielnikov, ktorý je zhromaždený správcovskou spoločnosťou. Problematiku kolektívneho investovania upravuje zákon č. 594/2003 Z.z. o kolektívnom investovaní a povolenie na vznik a činnosť udeľuje správcovskej spoločnosti Národná banka Slovenska. Správcovská spoločnosť je vlastne akciová spoločnosť zapísaná do obchodného registra, ktorej predmetom činnosti je investovanie získaných prostriedkov, z predaja podielových listov, do cenných papierov a nástrojov peňažného trhu.[8]

Na základe dostupných informácií na internete sme z vybraných fondov zostavili portfólio akciových investícií (akcií a podielových listov fondov kolektívneho investovania), do ktorých dané fondy investujú. Budeme predpokladať, že toto portfólio je reprezentatívnou vzorkou slovenského finančného trhu. Do analýzy bola zahrnutá väčšina z fondov, ktoré majú v portfóliu akciové investície. Podiel vzorky akciových investícií, o ktorých sme zozbierali podrobné údaje, na celkovom objeme akciových investícií v analyzovaných fondoch je $58,55\%$. Na obrázku 2 je znázornený histogram rozdelenia pokrytia akciových investícií jednotlivých fondov uvedenou vzorkou (v koľkých fondoch vzorka pokrýva daný interval všetkých akciových investícií). Budeme testovať, či je možné znížiť jeho dimenziu pre lepšie modelovanie výnosnosti a rizika finančného trhu. Informácie o fondoch (čistá hodnota majetku (NAV), podiel akciových investícií na NAV, zoznam akciových investícií a ich podiel na NAV) sme čerpali z mesačných správ k 31.12.2010. Na základe získaných dát sme zostavili 129 vážených časových radov výnosov akciových investícií v intervale od 24.4.2009 do 31.12.2010, ktoré budú premennými pre našu analýzu. Váhy jednotlivých akciových investícií sú vypočítané na základe podielu na celkovom NAV všetkých fondov, čiže súčet všetkých váh sa rovná jednej a výnosy sú vypočítané na základe vzťahu:

$$y_i = \frac{x_i}{x_{i-1}} - 1$$

[8]Spracované na základe dokumentov dostupných na interenete [6] a [7] v zozname literatúry.

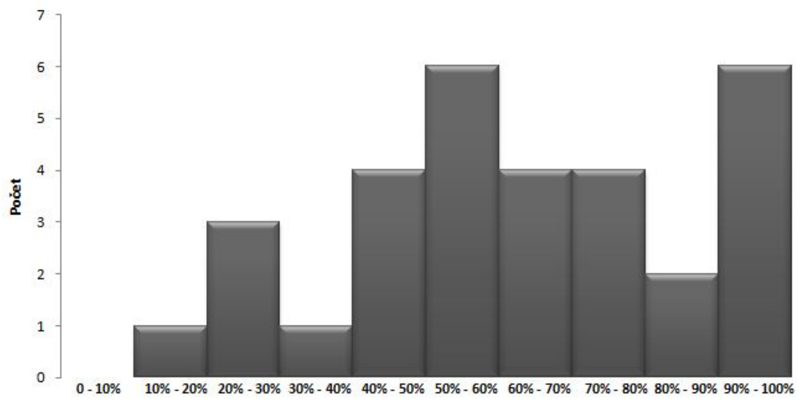
kde

x_i je cena akciovej investície na konci obchodovacieho dňa,

x_{i-1} je cena akciovej investície na konci predchádzajúceho obchodovacieho dňa,

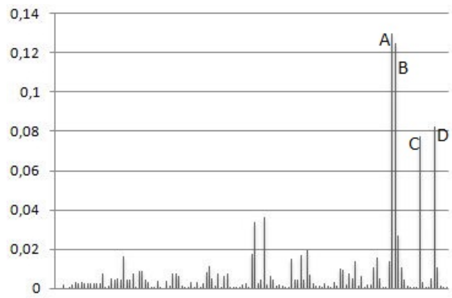
y_i je výška výnosu akciovej investície medzi danými obdobiami.

V našom prípade sú teda vypočítané denné výnosy akciových investícií. Vykonanie oboch analýz zahŕňa aj centrovanie premenných. Hodnoty váh jednotlivých akcií zobrazuje graf na obrázku 3. Výrazne vyššie hodnoty váh majú označené akciové investície (A,B,C,D), takže ich význam bude v modelovaní v porovnaní s ostatnými vysoký. Najvyššie váhy (A,B) majú podielové listy fondov správcovskej spoločnosti Tatra Asset Management (Európsky akciový fond a Americký akciový fond) a každá je prítomná v ôsmych rovnakých fondoch. Váha C prislúcha indexu (SPDR S&P 500 TRUST) a nachádza sa v siedmich fondoch. Posledná označená váha patrí akciovej investícií (VANGUARD EMERGING MARKET FUND) obsiahnutej v siedmich fondoch.



Obr. 2: Histogram rozdelenia pokrytia vzorkou akciových investícií

Potrebné informácie sa nám podarilo získať o fondoch v prílohe v tabuľke 4 a doplňujúce informácie o vybraných fondoch sa nachádzajú v prílohe v tabuľke 5. Metódu hlavných komponentov a faktorovú analýzu sme na dáta aplikovali v programe *EViews*, ktorý ponúka, hlavne čo sa týka FA, dostatok možností ich riešenia.



Obr. 3: Hodnoty váh jednotlivých akciových investícií

5.1 Analýza hlavných komponentov

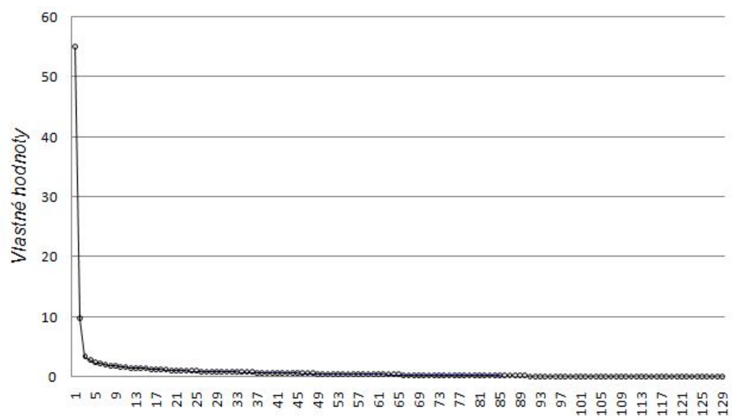
Ako prvú sme na dáta aplikovali PCA, ktorá je tou jednoduchšou metódou. Jej riešenie je jediné a našou úlohou bude určiť, či je možné znížiť dimenziu. Výsledkom PCA je znova 129 nových premenných, v našom prípade časových radov, ktoré sú lineárnymi kombináciami pôvodných vážených časových radov výnosov akciových investícií. Ak sú niektoré pôvodné premenné lineárne závislé, tak budeme môcť vybrať niekoľko prvých hlavných komponentov, ktoré popisujú rovnakú mieru rizika ako pôvodných 129 centrovaných časových radov výnosov akciových investícií. Možnosť zníženia dimenzie budeme posudzovať na základe vlastných čísel korelačnej matice pôvodných premenných, ich grafického zobrazenia, individuálneho a kumulatívneho percenta.

Východiskovou maticou bola pri riešení analýzy hlavných komponentov korelačná matica pôvodných premených, keďže štandardné odchýlky vážených výnosov akcií sa líšia. V tabuľke 1 sú zobrazené spomínané hodnoty, potrebné na určenie možnosti zníženia dimenzie, pre prvých 17 hlavných komponentov. Vidíme, že vlastná hodnota 1. hlavného komponentu je o dosť vyššia ako vlastná hodnota 2. hlavného komponentu. Z toho vyplýva, že medzi niektorými pôvodnými premennými existuje lineárna závislosť a bude možné znížiť dimenziu pozorovaného problému. Prvý hlavný komponent popisuje aj dosť vysoký rozptyl údajov, až $42,63\%$, čo ukazuje hodnota individuálneho percenta. Avšak dôležité je určiť, koľko prvých hlavných komponentov bude dostatočne vysvetľovať celkový rozptyl pôvodných časových radov. Keďže východiskovou je korelačná matica, tak priemer vlastných hodnôt je jedna a jedným z kritérií určenia počtu hlavných komponentov je počet vlastných hodnôt väčších ako jedna. Rozdiely

Hlavný komponent	Vlastné hodnoty	Individuálne percento	Kumulatívne percento
1	54.99477	0.4263	0.4263
2	9.672046	0.0750	0.5013
3	3.287380	0.0255	0.5268
4	2.806300	0.0218	0.5485
5	2.294458	0.0178	0.5663
6	2.165178	0.0168	0.5831
7	2.009638	0.0156	0.5987
8	1.834419	0.0142	0.6129
9	1.788149	0.0139	0.6268
10	1.620603	0.0126	0.6393
11	1.612395	0.0125	0.6518
12	1.485118	0.0115	0.6633
13	1.437294	0.0111	0.6745
14	1.358119	0.0105	0.6850
15	1.346962	0.0104	0.6954
16	1.265326	0.0098	0.7053
17	1.240465	0.0096	0.7149

Tabuľka 1: Výsledok analýzy hlavných komponentov

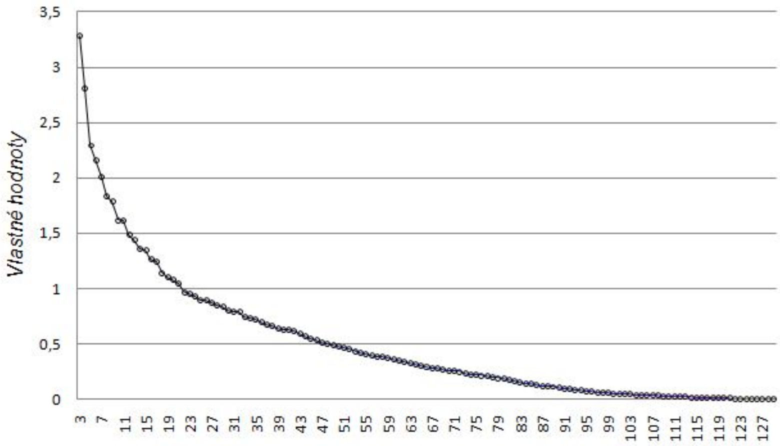
medzi vlastnými hodnotami po 7. hlavný komponent sú viditeľné a dá sa povedať, že podstatné. Počnúc 8. hlavným komponentom hodnoty vlastných čisiel klesajú približne rovnakým tempom. Takže na základe veľkosti vlastných hodnôt sa nedá jednoznačne určiť počet hlavných komponentov. Nevieme povedať, či ešte zahrnúť hlavné komponenty s vlastným číslom v intervale $\langle 0,8; 0,99 \rangle$, ktorých je deväť (33. - 41. hlavný komponent), alebo zahrnúť iba hlavné komponenty, ktorých vlastné čísla sú väčšie ako $1,2$ (prvých 17 hlavných komponentov). Po zobrazení grafu vlastných hodnôt od indexu hlavného komponentu (obrázok 4) sa nedá jednoznačne určiť bod zlomu. Keďže vlastné hodnoty 1. a 2. hlavného komponentu sú výraznejšie vyššie ako ostatné, kvôli lepšej možnosti určenia zlomu sme si zobrazili indexový graf vlastných čísel od 3. hlavného komponentu (obrázok 5). Ani z tohto grafu to však nevieme presne určiť. Čo sa týka kumulatívneho percenta, tak prvých sedem hlavných komponentov popisuje skoro 60% celkového rozptylu údajov a postupne sa jeho hodnota zvyšuje každým hlavným komponentom približne o 1%. Z týchto dôvodov nie je jednoduché určiť počet hlavných komponentov tak, aby malo zníženie dimenzie zmysel. Podľa teórie by mal vybraný počet hlavných komponentov popisovať aspoň 80% celkového rozptylu. Ale pravidlá určenia počtu hlavných komponentov nie sú jednoznačné a môžu sa líšiť v závislosti od skúmaného problému. V našom prípade sme sa rozhodli vybrať prvých 17 hlavných komponentov, ktorých vlastné hodnoty sú vyššie ako $1,2$.



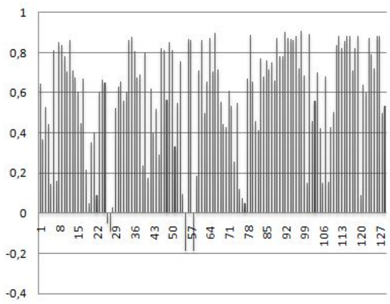
Obr. 4: Indexový graf vlastných čísel

Dôležité ešte je, aby sme sa pozreli na to, s koľkými, prípadne s ktorými akciovými investíciami významne koreluje prvý hlavný komponent. Zistíme to vypočítaním komponentných váh a budeme hľadať hodnoty približne vyššie ako $0,5$. Po uskutočnení výpočtu komponentných váh sme zistili, že 1. hlavný komponent významne koreluje až so 101 akciovými investíciami (komponentné váhy vyššie ako $0,5$). Preto sme sa zamerali na hodnoty vyššie ako $0,85$. Takto vysoko koreluje prvý hlavný komponent s 27 akciovými investíciami. Výšky komponentných váh pre 1. hlavný komponent zobrazuje graf na obrázku 6. Stále je to dosť vysoký počet a nenašli sme ani žiadnu spoločnú charakteristiku týchto akcií. Keďže už u 1. hlavného komponentu sme nenašli žiadnu možnú dobrú interpretáciu, nemá význam hľadať ju u ďalších.

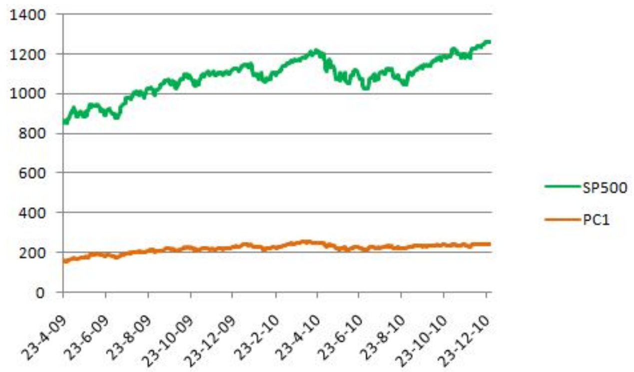
Kvôli lepšej ilustrácii 1. hlavného komponentu (PC1), ktorý popisuje značnú časť celkového rozptylu údajov, sme sa rozhodli jeho vývoj cien porovnať s vývojom akciového indexu $SP500$ (graf na obrázku 7). Ich korelácia výnosov je rovná $0,90422$, čo je dosť vysoká hodnota.



Obr. 5: Indexový graf vlastných čísel od 3.hlavného komponentu



Obr. 6: Komponentné váhy pre 1. hlavný komponent



Obr. 7: Porovnanie SP500 a 1. hlavného komponentu

5.2 Faktorová analýza

Riešenie faktorovej analýzy je síce zložitejšie ako PCA, ale poskytuje aj viac možností získania spoločných faktorov, ktoré môžu byť ešte rotované. Rotácia spoločných faktorov umožňuje nájdenie takých nových premenných, ktoré vieme dobre interpretovať. V tejto práci sú z viacerých metód odhadu parametrov faktorového modelu popísané dve (podkapitola 3.4), a to metóda hlavných faktorov a metóda maximálnej vierohodnosti. Keďže metóda maximálnej vierohodnosti vyžaduje normalitu dát, ktorú naše vážené časové rady výnosov akcií nesplĺňajú, tak na odhad parametrov použijeme metódu hlavných faktorov.

Metóda hlavných faktorov vyžaduje, aby sme na začiatku určili počet spoločných faktorov a spôsob odhadu komunalít. Východiskovou je takisto ako pri PCA korelačná matica. EViews ponúka veľa možností určenia počtu spoločných faktorov, vrátane možnosti zadania konkrétneho počtu, odhadu komunalít a následnej rotácie spoločných faktorov. Ako kritérium na určenie počtu spoločných faktorov sme vybrali "minimum eigenvalue"(minimálna vlastná hodnota), a to konkrétne hodnotu $1,2$ na základe výsledkov PCA, pri iterovanej metóde spoločných faktorov a za odhad komunalít sme použili "squared multiple correlation". Na odhad komunalít existuje viacero možností, ale keďže vykonávame iteračnú metódu hlavných faktorov, tak začiatočný výber odhadu určenia východiskových hodnôt komunalít nie je dôležitý, pretože komunality sa pri každej iterácii znova prepočítavajú na základe odhadu matice faktorových váh A.

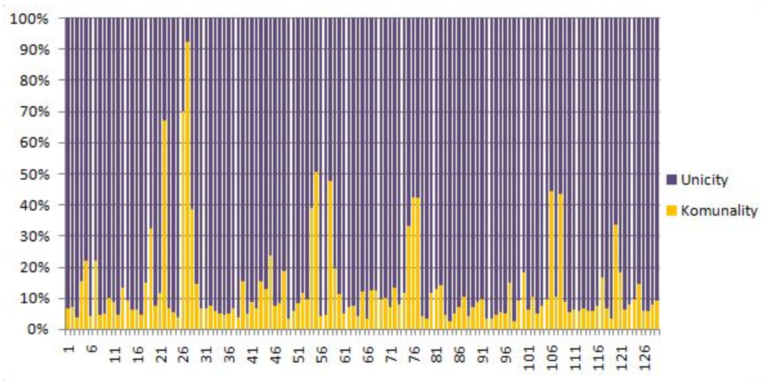
Po odhade parametrov je výsledkom 17 spoločných faktorov. Tento počet splĺňa vzťah (18), podľa ktorého musí byť počet spoločných faktorov menší ako 113. Dôležité je teraz určiť, či sú faktorové váhy významné a či je možná interpretácia spoločných faktorov. Významnosť faktorových váh určíme na základe vzťahu (19) z kapitoly 3. Hraničné hodnoty, ktoré musí váha v absolútnej hodnote spĺňať pre 17 spoločných faktorov sú vypočítané v tabuľke 2 na hladine významnosti 5%.

Poradie spoločného faktora (l)	$\|a_{jl}\| >$
1	0,1004419
2	0,1038477
3	0,1077678
4	0,1121683
5	0,117156
6	0,1228742
7	0,1295208
8	0,1373776
9	0,1468628
10	0,15863
11	0,1737704
12	0,1942812
13	0,2243366
14	0,2747552
15	0,3885625

Tabuľka 2: Vypočítané kritériá významnosti faktorových váh pre 17 spoločných faktorov

Prvým výsledkom faktorovej analýzy sú nerotované spoločné faktory, kde sme si vyznačili významné faktorové váhy. Hneď pri prvých faktoroch sme si všimli, že váhy niektorých akcií sú významné pri viacerých faktoroch. Preto sme sa rozhodli použiť ortogonálnu rotáciu typu "varimax", aby sme odstránili tento nedostatok. Samozrejme podstatná je interpretácia jednotlivých spoločných faktorov. Po určení významných váh u prvého rotovaného faktora sa nám nepodarilo odhaliť spoločné charakteristické znaky.

Akciové investície s významnými váhami sú veľmi rozmanité a nedá sa ani vybrať niekoľko fondov, ktoré by reprezentovali daný spoločný faktor. Poznamenajme, že prvý spoločný faktor má 24 významných váh. Celková variancia pôvodných centrovaných časových radov je podľa vzťahu (10) v kapitole 3 daná súčtom celkovej komunality a unicity, pričom daný počet spoločných faktorov popisuje 100% celkovej komunality. Avšak v našom výsledku sú jednotlivé unicity oproti jednotlivým komunalitám dosť vysoké (znázorňuje to graf na obrázku 8). Tento výsledok nie je veľmi priaznivý, keďže spoločné faktory popisujú potom malú časť celkovej variancie.

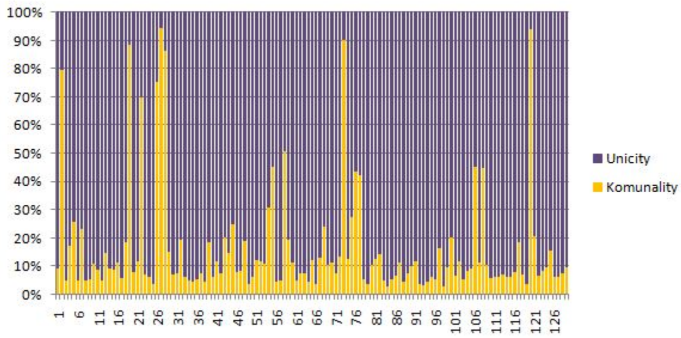


Obr. 8: Porovnanie výšky komunalít a unicít pri 17 spoločných faktoroch

Na základe doterajších výsledkov sme sa rozhodli vyskúšať rotáciu typu "quartimax", či nedá lepšie výsledky. Avšak výsledkom tejto rotácie sú významné váhy pre rovnaké akciové investície ako v predchádzajúcom prípade. Výsledkom je teda to, že pri danom výbere kritéria o počte spoločných faktorov je možné znížiť dimenziu, ale jednotlivé nové premenné nie sú interpretovateľné.

Určenie minimálnej hodnoty vlastných čísel v predchádzajúcom prípade nemusí byť správne, preto vykonáme faktorovú analýzu ešte raz. Ako minimálnu vlastnú hodnotu zadáme 1, čo je podložené teóriou. Na výsledné spoločné faktory použijeme znova rotáciu typu "varimax", keďže nastal znova ten istý problém (váhy niektorých akcií sú významné pri viacerých prvých faktoroch). Nasleduje ten istý postup zistenia možnosti interpretácie rotovaných spoločných faktorov. Keďže je teraz počet spoločných faktorov určený na 21, je potrebné prepočítať hraničné hodnoty pre významnosť jednotlivých váh (tabuľka 3).

Ani zvýšenie počtu spoločných faktorov nepriniesIo zlepšenie výsledkov, ani čo sa týka výšky vysvetlenej variancie spoločnými faktormi (obrázok 9). Prvý spoločný faktor má významnú váhu len jednu, druhý spoločný faktor ich má dve. Dané dve akciové investície sú vlastne fondy, ktoré majú minimálne investície do akcií. Takže prvé dva spoločné faktory dávajú málo informácií, skúmajme preto významné váhy u ďalších spoločných faktorov. Tretí spoločný faktor má 15 významných váh a štvrtý spoločný faktor 24. Aj v tomto riešení faktorovej analýzy majú jednotlivé spoločné faktory významné váhy pre rozdielne akciové investície bez spoločnej charakteristiky a výška vysvetlenej variancie spoločnými faktormi je nízka.



Obr. 9: Porovnanie výšky komunalít a unicít pri 21 spoločných faktoroch

Možnosť zníženia dimenzie sme potvrdili obidvoma metódami (PCA a FA), ale nepodarilo sa nám nájsť žiadne vhodné interpretácie nových premenných. Dôvodom je veľký počet pôvodných premenných, čo sťažuje aj následný rozbor výsledkov.

Poradie spoločného faktora (l)	$\lvert a_{jl} \rvert >$
1	0,0990762
2	0,1017911
3	0,1047422
4	0,1079658
5	0,1115066
6	0,1154203
7	0,1197773
8	0,1246681
9	0,1302116
10	0,1365671
11	0,1439543
12	0,1526866
13	0,1632289
14	0,1763073
15	0,193135
16	0,2159315
17	0,2493362
18	0,3053733
19	0,431863

Tabuľka 3: Vypočítané kritériá významnosti faktorových váh pre 21 spoločných faktorov

6 Záver

Teoretická časť práce sa zaoberá vysvetlením dvoch metód, ktoré umožňujú zníženie dimenzie skúmaného problému, ak sú niektoré prvotné znaky lineárne závislé, a to analýzou hlavných komponentov a faktorovou analýzou. Analýza hlavných komponentov je jednoduchšia metóda, ktorá je založená na hľadaní nových premenných, ktoré sú lineárnymi kombináciami pôvodných premenných. Výsledkom je jedno riešenie, z ktorého je ešte potrebné na základe uvedených kritérií určiť potrebný počet hlavných komponentov. Druhej metóde (faktorovej analýze) je v práci venovaný väčší priestor, pretože podstata tejto analýzy je zložitejšia, ale ponúka aj širšie možnosti jej riešenia a za splnenia predpokladu o normálnom rozdelení umožňuje testovať hypotézu o správnom počte spoločných faktorov (nových premenných). Ak dáta nepochádzajú z normálneho rozdelenia, tak je potrebné pred samotnou faktorovou analýzou jednoznačne určiť počet spoločných faktorov, čo sa u analýzy hlavných komponentov robí až na konci. Faktorová analýza ponúka ešte možnosť rotácie výsledných spoločných faktorov. Častokrát to umožní lepšiu interpretáciu nových premenných. Výsledkom oboch metód je menší počet nezávislých premenných. Ale ak vo faktorovej analýze vykonáme kosouhlú rotáciu, namiesto ortogonálnej, konečné spoločné faktory budú závislé.

Cieľom praktickej časti práce bolo určenie systémového akciového rizika, ktorému je vystavený slovenský finančný sektor. Do analýzy boli vybrané fondy kolektívneho investovania a fondy dôchodkových správcovských spoločností, ktoré majú významnú akciovú zložku. Na základe dostupných informácií sme zostavili vzorku portfólia akciových investícií, do ktorých tieto fondy investujú. Na centrované vážené časové rady denných výnosov akciových investícií sme aplikovali obe analýzy a skúmali výsledky. Zistili sme, že medzi niektorými časovými radmi je určite lineárna závislosť a bude možné vybrať menší počet nových premenných. Ale pri skúmaní spoločných znakov u pôvodných časových radov, s ktorými nové premenné korelujú sa nám ani po rotácií nepodarilo nájsť možnú interpretáciu. Výsledkom tejto práce je, že existuje možnosť zjednodušenia určenia rizika tohto portfólia akciových investícií, ale nie je možné toto portfólio reprezentovať len niektorými fondami alebo významnými akciami. Napriek tomu považujeme výsledok za dosť pozitívny, pretože vzorku portfólia vieme dobre reprezentovať pomocou malého počtu hlavných komponentov. Hlavný cieľ práce sa nám teda podarilo splniť. Možno by bolo vhodné dáta ďalej analyzovať. Jednou z možností je použitie lineárnej regresie na vybraný menší počet nových nezávislých premenných a ich porovnanie s nejakým akciovým indexom.

Literatúra

[1] Stankovičová, I. - Vojtková, M. 2007. Viacrozmerné štatistické metódy s aplikáciami. Bratislava: Iura Edition, 2007. ISBN 978-80-8078-152-1.

[2] Shlens, J. 2005. A tutorial on Principal Component Analysis. Dostupné na internete: https://dspace.ist.utl.pt/bitstream/2295/56448/1/pca.pdf

[3] Smith, L. I. 2002. A tutorial on Principal Component Analysis. Dostupné na internete: http://users.ecs.soton.ac.uk/hbr03r/pa037042.pdf

[4] Seber, G. A. F. 2004. Multivariate Observation. New Jersey: Wiley - interscience, 2004. ISBN 0-471-69121-6.

[5] Factor Analysis. Dostupné na internete: http://faculty.chass.ncsu.edu/garson/PA765/factor.htm

[6] Fondy. Dostupné na internete: http://www.finance.sk/investovanie/informacie/fondy/

[7] Základné informácie o kolektívnom investovaní. Dostupné na in ternete: http://www.nbs.sk/sk/dohlad-nad-financnym-trhom/dohlad-nad-trhom-cennych-papierov/kolektivne-investovanie/zakladne-informacie

Príloha

Kód fondu	Názov fondu	Názov správcovskej spoločnosti
ALICOGI	Global Index	Alico Funds Central Europe
ALICOOB	Optimal Balanced	Alico Funds Central Europe
AMSSGAFF	Globálny akciový fond fondov	Asset Management Slovenskej sporiteľne
AMSSFMV	Fond maximalizovaných výnosov	Asset Management Slovenskej sporiteľne
AMSSRF	Realitný fond	Asset Management Slovenskej sporiteľne
CSOBAMPROP	ČSOB Property	ČSOB Asset Management
CSOBAMRAST	ČSOB Rastový	ČSOB Asset Management
CSOBAMVYV	ČSOB Vyvážený	ČSOB Asset Management
CSOBAMEK	EURO konvergentný	ČSOB Asset Management
CSOBAMSA	Svetový akciový fond	ČSOB Asset Management
DDSTATRARP	Rastový príspevkový fond	Doplnková dôchodková spoločnosť Tatra banky
DDSTATRAVP	Vyvážený príspevkový fond	Doplnková dôchodková spoločnosť Tatra banky
AXAAPDF	Príspevkový fond	AXA
ALLIANZAMGO	Growth Opportunities	Allianz Asset Management
IADKDPROS	KD Prosperita	IAD Investments
IADKDRUSIA	KD Russia	IAD Investments
AEGONPF	Príspevkový fond	AEGON
INGTATRYRPF	Rastový príspevkový fond	ING Tatry - Sympatia
TATRAAMAA	Americký akciový fond	Tatra Asset Management
TATRAAMCF	ConservativeFund	Tatra Asset Management
TATRAAMEA	Európsky akciový fond	Tatra Asset Management
TATRAAMHF	HarmonicFund	Tatra Asset Management
TATRAAMPV1	Private V1	Tatra Asset Management
TATRAAMPV2	Private V2	Tatra Asset Management
TATRAAMPV3	Private V3	Tatra Asset Management
TATRAAMSMART	SmartFund	Tatra Asset Management
TATRAAMSF	StrategicFund	Tatra Asset Management
TATRAAMSEA	Stredoeurópsky akciový fond	Tatra Asset Management
VUBAMDP	VÚB AM Dynamické portfólio	VÚB Asset Management
VUBAMPPM30	VÚB AM Privátne portfólio MIX 30	VÚB Asset Management
VUBAMVRF	VÚB AM Vyvážený rastový fond	VÚB Asset Management

Tabuľka 4: Zoznam vybraných fondov

Kód fondu	Veľkosť v % z celkového objemu NAV všetkých analyzovaných fondov	%-ny podiel akciovej zložky na NAV fondu	%-na časť akciovej vzorky fondu	Sektor
ALICOGI	2,28%	93%	49,46%	kolektívne investovanie
ALICOOB	0,78%	51%	50,72%	kolektívne investovanie
AMSSGAFF	3,35%	70%	15,92%	kolektívne investovanie
AMSSFMV	0,45%	86,91%	37,58%	kolektívne investovanie
AMSSRF	1,5%	66%	55,66%	kolektívne investovanie
CSOBAMPROP	0,56%	100%	49,66%	kolektívne investovanie
CSOBAMRAST	0,38%	77,83%	55,76%	kolektívne investovanie
CSOBAMVYV	0,53%	45,89%	42,62%	kolektívne investovanie
CSOBAMEK	0,37%	21,82%	21,43%	kolektívne investovanie
CSOBAMSA	0,21%	78,33%	55,92%	kolektívne investovanie
DDSTATRARP	3,61%	56,78%	56,78%	fondy DDS
DDSTATRAVP	24,08%	16,93%	15,89%	fondy DDS
AXAAPDF	0,75%	41,3%	27,4%	fondy DDS
ALLIANZAMGO	1,08%	42,1%	26,62%	kolektívne investovanie
IADKDPROS	0,74%	52,61%	18,31%	kolektívne investovanie
IADKDRUSIA	10,14%	89,09%	35,94%	kolektívne investovanie
AEGONPF	0,28%	5,68%	5,67%	fondy DDS
INGTATRYRPF	0,75%	36,8%	25,65%	fondy DDS
TATRAAMAA	4,8%	99,16%	25,3%	kolektívne investovanie
TATRAAMCF	2,1%	29,03%	19,23%	kolektívne investovanie
TATRAAMEA	6,37%	98,94%	11,5%	kolektívne investovanie
TATRAAMHF	7,53%	53,65%	27,83%	kolektívne investovanie
TATRAAMPV1	1,59%	25,31%	13,45%	kolektívne investovanie
TATRAAMPV2	3,03%	53,63%	40,53%	kolektívne investovanie
TATRAAMPV3	1,43%	72,17%	55,73%	kolektívne investovanie
TATRAAMSMART	2,66%	29,33%	26,11%	kolektívne investovanie
TATRAAMSF	5,45%	71,8%	39,77%	kolektívne investovanie
TATRAAMSEA	1,24%	97,66%	52,17%	kolektívne investovanie
VUBAMDP	4,53%	43,32%	21,22%	kolektívne investovanie
VUBAMPPM30	1,62%	24,07%	12,62%	kolektívne investovanie
VUBAMVRF	5,79%	57,04%	17,02%	kolektívne investovanie

Tabuľka 5: Doplňujúce informácie vybraných fondov