UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



Využitie metódy hlavných komponentov v úlohe optimálneho portfólia

DILPOMOVÁ PRÁCA

2021 Filip Májek

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

Využitie metódy hlavných komponentov v úlohe optimálneho portfólia

DIPLOMOVÁ PRÁCA

Študijný program:	Ekonomicko-finančná matematika a modelovanie
Študijný odbor:	1113 Matematika
Školiace pracovisko:	Katedra aplikovanej matematiky a štatistiky
Vedúci práce:	Mgr. Katarína Lučivjanská, PhD.

Bratislava 2021 **Filip Májek**





Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Filip Májek
Študijný program: ekonomicko-finančná matematika a modelovanie (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Využitie metódy hlavných komponentov v úlohe optimálneho portfólia
Principal component analysis in the optimal portfolio problem

Anotácia: V úlohe optimálneho portfólia (mean-variance alebo minimum-variance) je dôležitým vstupom kovariančná matica. Pri vyššom počte aktív (10, 50, 100, ..) je však odhad tejto matice citlivejší na vstupné dáta, a preto sa často predpokladá určitá štruktúra, ktorá úspornejšie opisuje jej komponenty.

Vedúci: Mgr. Katarína Lučivjanská, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Marek Fila, DrSc.

Dátum zadania: 17.12.2019

Dátum schválenia: 21.12.2019 prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
vedúci práce

Poďakovanie: Ďakujem mojej školiteľke diplomovej práce Mgr. Kataríne Lučivjanskej, PhD. za veľkú odbornosť v tematike a tiež za pozitívny prístup pri konzultáciách. Vďaka nej je práca na odbornejšej úrovni a formálna úprava práce je oveľa akademickejšia. Okrem toho ďakujem mojim blízkym za podporu v osobnom živote.

Abstrakt v štátnom jazyku

Májek, Filip: Využitie metódy hlavných komponentov v úlohe optimálneho portfólia [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: Mgr. Katarína Lučivjanská, PhD., Bratislava, 2021, 82 s.

V našej diplomovej práci sa venujeme úprave Markowitzovej teórie portfólia, aby odhadnuté optimálne portfólio dosahovalo lepšie výsledky. Parametre pri optimalizačnej úlohe sú stredná hodnota a kovariančná matica výnosov. Z viacerých štúdií vyplýva, že pre veľký počet aktív je potrebná veľmi veľká sada historických údajov, aby boli momenty výnosov dobre odhadnuté. Problém je v tom, že väčšina aktív nemá takú dlhú históriu, aká je pre odhad momentov potrebná. V diplomovej práci vysvetľujeme podpriestorovú Markowitzovu teóriu portfólia od Chena a Yuana [6], ktorá sa zaoberá zmenšením prípustnej množiny v úlohe optimálneho portfólia. Optimalizačná úloha sa mení tak, že investovať nie je možné do ľubovoľného aktíva, ale do kombinácií aktív, vytvorených pomocou analýzy hlavných komponentov. Tým sa zmenšujú rozmery strednej hodnoty a kovariančnej matice výnosov a na ich presný odhad nám stačí menšie množstvo historických údajov. Cieľom tejto práce je v empirickej časti ukázať na zlú výkonnosť odhadnutého Markowitzovho optimálneho portfólia pre veľký počet aktív a pre momenty výnosov odhadnuté na relatívne malom počte historických údajov. Ďalším cieľom je porovnať výsledky odhadnutého Markowitzovho optimálneho portfólia s výsledkami odhadnutého podpriestorového optimálneho portfólia a s výsledkami nami navrhnutých rozšírení pre výpočet podpriestorového optimálneho portfólia.

Kľúčové slová: Markowitzova teória portfólia, optimálne portfólio, stredná hodnota výnosov, kovariančná matica výnosov, podpriestorová Markowitzova teória portfólia, analýza hlavných komponentov, prípustná množina optimalizačnej úlohy

Abstract

Májek, Filip: Principal component analysis in mean-variance portfolio [Master Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Katarína Lučivjanská, PhD., Bratislava, 2021, 82 p.

In our master thesis, we focus on the modification of Modern Portfolio Theory to gain better results for the estimated mean-variance portfolio. Parameters during optimization are the mean value and the covariance matrix of the returns. Many studies show that for a large number of assets a very large set of historical data is needed for good estimations of moments of returns. The problem is that most assets do not have the long history required to estimate the moments. In the thesis, we explain Subspace Mean-Variance Analysis by Chen and Yuan [6], which deals with restricting an investment in a carefully chosen linear subspace for optimal portfolio. The optimization task is amended so that it is not possible to invest in any asset but in a combination of assets, created by Principal Component Analysis. This reduces the dimensions of the mean value and covariance matrix of returns and for their accurate estimation is needed a smaller amount of historical data. The aim of this paper in the empirical part is to point out the poor performance of the estimated Markowitz mean-variance portfolio for a large number of assets and moments of returns estimated on a relatively small number of historical data. Another aim is to compare the results of the estimated Markowitz mean-variance portfolio with the results of the estimated subspace mean-variance portfolio and the results produced by our extensions for the calculation of the subspace mean-variance portfolio.

Keywords: Modern Portfolio Theory, optimal portfolio, mean-variance portfolio, mean value of returns, covariance matrix of returns, Subspace Mean-Variance Analysis, Principal Component Analysis

Obsah

Úvod **9**

1 Markowitzova teória portfólia **12**

1.1 Alokácia do rizikových aktív 12

1.2 Kombinácia bezrizikového aktíva s rizikovými aktívami 14

1.3 Úloha optimálneho portfólia pre nadmerné výnosy 15

2 Oceňovanie aktív **17**

2.1 Model oceňovania kapitálových aktív 17

2.2 Teória arbitrážneho oceňovania 20

3 Metóda hlavných komponentov **22**

3.1 Výber počtu hlavných komponentov 25

4 Znižovanie dimenzie prípustnej množiny v úlohe optimálneho portfólia **28**

4.1 Podpriestorová Markowitzova teória portfólia 28

4.2 Rozšírenie podpriestorovej Markowitzovej teórie portfólia 32

5 Testovanie na dátach **35**

5.1 Stratégie 36

5.2 Spôsob vyhodnotenia stratégií 37

5.3 Markowitzova teória portfólia v rámci a mimo vzorky 40

5.3.1 Dáta 40

5.3.2 Použitie Stratégií 41

5.3.3 Vyhodnotenie 42

5.4 Znižovanie dimenzie prípustnej množiny v úlohe optimálneho portfólia - analýza na simulovaných dátach 44

5.4.1 Generovanie dát 44

5.4.2 Použitie Stratégií 45

5.4.3 Vyhodnotenie 46

5.5 Znižovanie dimenzie prípustnej množiny v úlohe optimálneho portfólia - analýza na reálnych dátach 48
5.5.1 Dáta 48
5.5.2 Použitie Stratégií 50
5.5.3 Výsledky pre rôzne voľby d 51
5.5.4 Výsledky pre optimálne d 53

Záver **63**

Zoznam použitej literatúry **66**

Príloha 1 - prehľad o použitých dátach **69**

Príloha 2 - kódy diplomovej práce **70**

Úvod

Teóriou portfólia sa ako prvý zaoberal vo svojej práci H.M. Markowitz v roku 1952 [20]. Odvtedy bolo napísaných veľa štúdií, ktoré teóriu portfólia skúmali ďalej. V našej diplomovej práci riešime práve jedno zo slabých miest a kritík Markowitzovej teórie portfólia, a to problém odhadu očakávanej strednej hodnoty a očakávanej kovariančnej matice. Medzi prvými, ktorí problém uviedli a kvantifikovali, boli Chopra a Ziemba [7]. Vo svojej štúdii ukazujú na zlú výkonnosť portfólia pri aplikovaní teórie na reálne dáta, hlavne kvôli zlému odhadu strednej hodnoty a tiež kvôli nepresnému odhadu kovariančnej matice na základe historických dát.

Slabým stránkam pri aplikácii na reálnych dátach sa tiež venovali Miguel et al. [9], ktorí vo svojej práci porovnali teóriu portfólia a jej viaceré rozšírenia s rovnako váženým portfóliom na viacerých dátových sadách. V práci ukázali, že rovnako vážené portfólio vykazovalo lepšie výsledky ako optimalizované portfóliá a žiadne z portfólií nedokázalo konzistentne porážať rovnako vážené portfólio. Reakciou na túto prácu bola práca od Kritzman et al. [16], kde ukázali, že optimálne portóliá z teórie portfólia konzistentne porážajú rovnako vážené portfólio, ale na odhad strednej hodnoty a kovariančnej matice používali väčší počet historických dát ako Miguel et al.

Miguel et al. navyše ukázali, že pre portfólio s viac ako 50 aktívami by bolo treba tisíce mesačných výnosov na to, aby teória portfólia prekonala rovnako vážené portfólio. Problém chyby odhadu parametrov sa snažili riešiť viacerí autori. Keďže chyba odhadu strednej hodnoty výnosov má väčší vplyv na výkonnosť optimálneho portfólia, tak niektorí autori (napr. Scherer [26]) skúmali stav, keď očakávaná stredná hodnota je 0 a teda úloha sa mení na úlohu minimalizácie rozptylu výnosov. Jorion [14] používal zmrštenie strednej hodnoty výnosov na zníženie chyby odhadu. Ledoit a Wolf v prácach [17] a [18] robili úpravu kovariančnej matice výnosov zmrštením jej extrémnych hodnôt. Tu a Zhou [28] skúšali namiesto len teórie portfólia kombinovať optimálne váhy z teórie portfólia s rovnako váženými váhami. Bjerring et al. [3] riešia problém veľkého počtu aktív vytvorením zhlukov na základe korelácie výnosov a následnou optimalizáciou nie na výnosoch všetkých aktívach, ale na výnosoch stredných bodov z vytvorených zhlukov, čím sa zníži rozmer úlohy. Branger et al. [4] tiež redukujú dimenziu úlohy na skupiny aktív, ktoré majú podobné očakávané výnosy, majú podobné optimálne váhy v

prípade optimalizácii na všetkých aktívach, a ktoré majú podobné α a β voči trhovému portfóliu. V rámci skupín používajú rovnaké váženie pre aktíva a optimalizáciu tvoria na skupinách.

V našej diplomovej práci nadväzujeme na prácu od Chen a Yuan [6], ktorí navrhujú vysporiadať sa s chybou odhadu pomocou ohraničenia prípustnej množiny optimalizačnej úlohy na hlavné komponenty výnosov všetkých aktív. Takýto prístup zakladajú na myšlienke, že výnosy všetkých aktív sú tvorené niekoľkými trhovými faktormi. Preto ohraničením prísпustnej množiny do týchto trhových faktorov môžeme znížiť chybu odhadu (lebo odhadujeme momenty pre menej aktív), pričom ale vytvoríme systematickú chybu, ktorá označuje ohraničenie investičných príležitostí. Ich zámerom je nájdenie bodu, v ktorom je súčet systematickej chyby a chyby odhadu minimálny.

Navrhujeme 2 rozšírenia práce [6], ktorými meníme výpočet podpriestorového optimálneho portfólia. V prvom rozšírení tvoríme podpriestor pre optimalizačnú úlohu pomocou korelačnej matice namiesto kovariančnej matice. Pri použití kovariančnej matice majú vplyv na analýzu hlavných komponentov rozptyly výnosov jednotlivých aktív. Použitie korelačnej matice je ekvivalentné štandardizácii výnosov aktív na rozptyl rovný jednej pred analýzou hlavných komponentov. V druhom rozšírení sme sa zamerali na vzorec pre výpočet podpriestorového optimálneho portfólia a navrhli sme pre ňho jednu technickú zmenu.

Práca je rozdelená na 5 kapitol. V prvej kapitole uvádzame teóriu Markowitzovho portfólia a odvádzame ako sa optimálne portfólio mení, keď ho tvoríme na nadmerných výnosoch. V druhej kapitole sa zaoberáme modelmi oceňovania aktív, ktorých teória je podstatná pre hlavnú časť práce. V tretej kapitole opisujeme metódu hlavných komponentov, ktorú používame pre hľadanie faktorov na trhu. Štvrtá kapitola je hlavná teoretická časť tejto práce a opisujeme v nej teóriu podpriestorovej Markowitzovej teórie portfólia, ktorú definovali Chen a Yuan [6]. Okrem toho v štvrtej kapitole definujeme aj naše 2 rozšírenia pre danú teóriu. V piatej kapitole uvádzame praktickú časť našej diplomovej práce. Najprv poukazujeme na zlú výkonnosť odhadnutého Markowitzovho optimálneho portfólia na reálnych dátach pre veľký počet aktív a pre momenty výnosov odhadnuté na relatívne malom počte historických údajov. Potom tvoríme a hodnotíme odhadnuté podpriestorové optimálne portfólio na simulovaných dátach. Nakoniec na

reálnych dátach porovnávame výkonnosť odhadnutého podpriestorového optimálneho portfólia oproti odhadnutému Markowitzovmu optimálnemu portfóliu, naivnému portfóliu a oproti nami navrhnutými rozšíreniami pre výpočet podpriestorového optimálneho portfólia.

1 Markowitzova teória portfólia

Kapitola je rozdelená na 3 časti. V prvej časti uvádzame Markowitzovu teóriu portfólia, ktorá je definovaná pre investíciu do rizikových aktív. V druhej časti sa zaoberáme priamkou kapitálových aktív, ktorá je odpoveďou na otázku, ako vyzerá efektívna hranica pri kombinácii rizikových aktív s bezrizikovým aktívom. V tretej časti odvádzame úlohu optimálneho portfólia, keď ju tvoríme nie na výnosoch rizikových aktív, ale na nadmerných výnosoch rizikových aktív.

1.1 Alokácia do rizikových aktív

Teóriu portfólia uviedol vo svojej práci H. M. Markowitz [20] v roku 1952. V práci sa zaoberá problematikou, ako z možných investičných aktív vytvoriť optimálne portfólio. V tejto kapitole čerpáme najmä z poznatkov z [20], [19] a [15].

Dva základné predpoklady Markowitzovej teórie portfólia sú, že investori sú rizikovo averzní a chcú čo najvyššie výnosy. Riziková averznosť znamená, že pre investora je pri fixnom výnose vždy výhodnejšie portfólio s nižším rizikom a druhý predpoklad zaručuje, že medzi dvomi aktívami s rovnakým rizikom si investor vyberie to, ktoré má vyšší výnos. Markowitz taktiež predpokladá, že výber portfólia je čisto založený na vzťahu medzi očakávaným výnosom a očakávaným rozptylom, kde vyšší rozptyl znamená vyššie riziko. Ostatné predpoklady môžu byť vyjadrené nasledovne:

(a) akciové trhy existujú bez daní a transakčných nákladov

(b) uvažované aktíva sú ľubovoľne deliteľné na menšie časti

(c) všetky aktíva na trhu sú likvidné

(d) uvažujeme stabilnú alokáciu počas celej časovej jednotky

Podľa Markowitza by mal racionálny investor zvažovať iba efektívne portfóliá. Za efektívne portfóliá považuje tie, ktoré majú maximálny výnos pre určený rozptyl alebo portfóliá, ktoré majú minimálny rozptyl pre určený výnos. Pre voľbu efektívnych portfólií používa Markowitz nasledovnú účelovú funkciu:

$$w^\top \mu - \frac{\gamma}{2} w^\top \Sigma w, \tag{1}$$

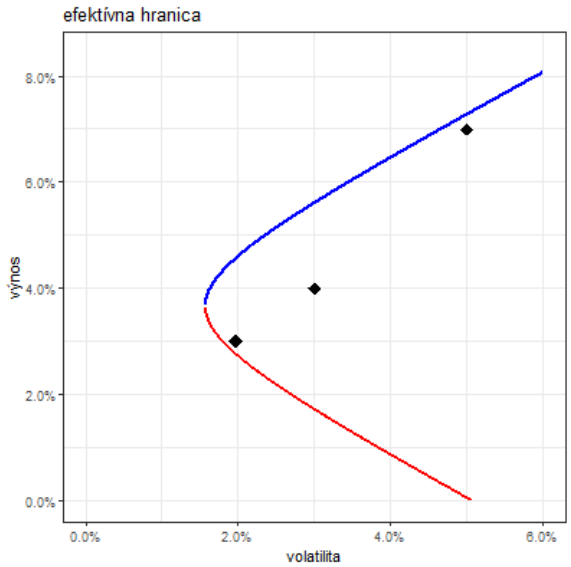
kde w je vektor váh, μ je vektor očakávaných výnosov vybraných aktív (odhadovaný pomocou strednej hodnoty výnosov) a Σ je prislúchajúca očakávaná kovariančná matica výnosov aktív. Ďalej γ označuje parameter pre rizikovú averziu, ktorý penalizuje pridané riziko do portfólia. Vyššia γ vyjadruje vyššiu averznosť a nižšia γ vyjadruje nižšiu averznosť voči riziku. Stručne popísané, účelová funkcia (1) vyjadruje kompromis medzi očakávaným výnosom portfólia a očakávaným rizikom.

S pridaním obmedzení pre váhy vyzerá Markowitzom popísaná optimalizačná úloha nasledovne:

$$\begin{aligned} &\max_w w^\top \mu - \frac{\gamma}{2} w^\top \Sigma w, \\ &\text{pri podmienkach } \sum_{i=1}^{n} w_i = 1, \end{aligned} \tag{2}$$

kde w_i je i-ta hodnota vektora w. Iné ohraničnia pre váhy nie sú, a preto sú povolené aj záporné váhy (krátke pozície). Riešením úlohy (2) dostávame optimálne váhy pre investovanie do rizikových aktív.

Pre rôzne voľby parametra γ investor získava efektívne portfóliá, ktoré spolu tvoria efektívnu hranicu (*efficient frontier*). Na obr. 1 vidíme ukážku, ako vyzerá efektívna hranica pre 3 aktíva (modrá čiara). Obr. 1 sme vytvorili postupom popísaným v [24].



Obr. 1: Efektívna hranica portfólia - modrou farbou zvýraznená efektívna hranica pri investícii do 3 aktív.

Čierne kosoštvorce znázorňujú 3 aktíva, do ktorých je možné investovať. Celá krivka

vyšla riešením úlohy (2) pre rôzne voľby γ. Modrá krivka označuje efektívnu hranicu a červená krivka označuje neefektívnu hranicu, lebo pre takú istú volatilitu sa dá získať portfólio s vyšším výnosom na modrej krivke.

1.2 Kombinácia bezrizikového aktíva s rizikovými aktívami

Kombináciu optimálneho portfólia s bezrizikovým aktívom predstavil vo svojom článku W. F. Sharpe [25] v rámci modelu oceňovania kapitálových aktív (*Capital asset pricing model* - CAPM), ktorý podrobnejšie opisujeme v podkapitole 2.1. Bezrizikové aktívum, je také aktívum, ktorého výnos r_f je nenáhodný a ktorého rozptyl a kovariancie s inými aktívami sú nulové. Aj keď v realite toto aktívum neexistuje, klasicky sa za bezrizikové aktíva považujú krátkodobé dlhopisy stabilných štátov, lebo je veľmi nízka pravdepodobnosť ich nesplatenia.

Pri zapojení bezrizikového aktíva sa efektívna hranica mení z hyperboly na polpriamku. Všetky efektívne portfóliá (portfóliá z efektívnej hranice - polpriamky) sú kombináciou bezrizikového aktíva a jediného portfólia rizikových aktív. Očakavaný výnos a rozptyl efektívneho portfólia sú potom:

$$\mu_p = w_0 r_f + w^\top \mu, \tag{3}$$

$$\sigma_p = \sqrt{w^\top \Sigma w}, \tag{4}$$

kde

w_0 - váha pre bezrizikový výnos,

w - vektor váh pre výnosy rizikových aktív,

μ - stredná hodnota výnosov rizikových aktív,

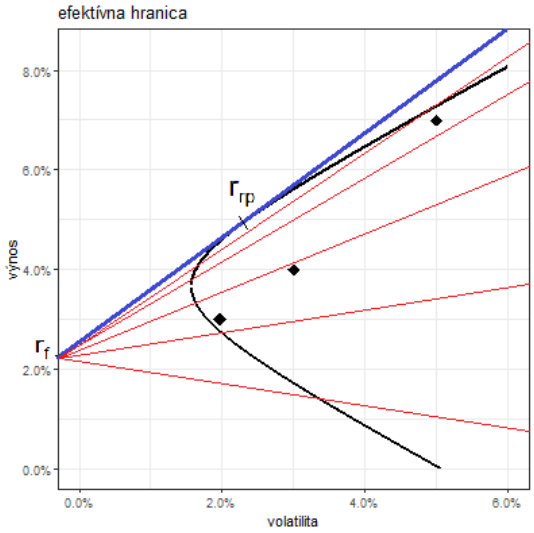
μ_p - stredná hodnota výnosu celého portfólia,

σ_p - volatilita výsledneho portfólia,

Σ - kovariančná matica výnosov rizikových aktív.

Pre kombináciu s bezrizikovým aktívom je najefektívnejšie zobrať také portfólio rizikových aktív, ktoré je v bode spojenia dotyčnice vychádzajúcej z bezrizikového výnosu a hyperboly, ktorá je efektívnou hranicou iba rizikových aktív. Toto portfólio sa nazýva dotyčnicové portfólio (*tangency portfolio*). Na obr. 2 vidíme modrou čiarou zvýraznenú efektívnu hranicu, ktorá je kombináciou výnosu bezrizikového aktíva r_f a

výnosu dotyčnicového portfólia r_{rp}. Vidíme, že modrá polpriamka je optimálna voľba, pretože pri kombinácii bezrizikového aktíva s ľubovoľnými inými prípustnými portfóliami rizikových aktív by vznikli polpriamky (na obr. zvýraznené červenou farbou), ktoré by dosahovali nižší výnos ako modrá polpriamka pri takej istej volatilite. Modrá polpriamka sa nazýva priamka kapitálového trhu (CML - *Capital market line*).



Obr. 2: Priamka kapitálového trhu - modrou farbou zvýraznená efektívna hranica pri investícii do 3 rizikových aktív a bezrizikového aktíva.

1.3 Úloha optimálneho portfólia pre nadmerné výnosy

Nadmerný výnos (*excess return*) R_i pre i-te aktívum je celkový výnos r_i očistený o bezrizikový výnos r_f:

$$R_i = r_i - r_f. \tag{5}$$

Nadmerný výnos vyjadruje prémiu za podstúpenie rizika, kedže v ňom nie je zahrnutý výnos bez rizika.

Nasledujúcim postupom odvodíme definovanie Markowitzovej úlohy pre nadmerné výnosy. Úloha (2) tvorená na nA rizikových aktívach a jednom bezrizikovom aktíve má

nasledovnú formuláciu:

$$\max_{w,w_0} w^\top \mu + w_0 r_f - \frac{\gamma}{2} \begin{bmatrix} w_0 & w \end{bmatrix}^\top \begin{bmatrix} 0 & 0 \\ 0 & \Sigma \end{bmatrix} \begin{bmatrix} w_0 \\ w \end{bmatrix} =$$

$$\max_{w,w_0} w^\top \mu + w_0 r_f - \frac{\gamma}{2} w^\top \Sigma w, \tag{6}$$

$$\text{pri podmienkach } \sum_{i=0}^{n} w_i = w_0 + w^\top \mathbb{1}_{nA} = 1.$$

V (6) vidíme, že rozptyl portfólia s bezrizikovým aktívom závisí len od rizikových aktív, pretože pre bezrizikové aktívum platí, že jeho rozptyl a kovariancie s rizikovými aktívami sú nulové. Ďalej upravujeme (6) na:

$$\max_{w,w_0} w^\top \mu + w_0 r_f - \frac{\gamma}{2} w^\top \Sigma w =$$

$$\max_{w,w_0} w^\top (\mu - \mathbb{1}_{nA} r_f + \mathbb{1}_{nA} r_f) + w_0 r_f - \frac{\gamma}{2} w^\top \Sigma w =$$

$$\max_{w,w_0} w^\top (\mu - \mathbb{1}_{nA} r_f) + (w_0 + w^\top \mathbb{1}_{nA}) r_f - \frac{\gamma}{2} w^\top \Sigma w, \tag{7}$$

$$\text{pri podmienkach } w_0 + w^\top \mathbb{1}_{nA} = 1.$$

V (7) vidíme, že $(\mu - \mathbb{1}_{nA} r_f)$ je stredná hodnota pre nadmerné výnosy. Kvôli podmienke platí $w_0 + w^\top \mathbb{1}_{nA} = 1$, a preto výraz $1 * r_f$ v optimalizačnej funkcii môžeme vynechať, lebo je to konštanta, ktorá neovplyvňuje voľbu optimálnych váh. Navyše kovariančná matica výnosov aktív sa rovná kovariančnej matici nadmerných výnosov aktív, lebo posun o konštantu nemení rozptyl a ani kovariancie medzi rizikovými aktívami: $\text{cov}(r) = \text{cov}(r - \mathbb{1}_{nA} r_f) = \text{cov}(R)$. Preto je riešenie (6) rovné riešeniu (8).

$$\max_{w} w^\top \mu^* - \frac{\gamma}{2} w^\top \Sigma^* w, \tag{8}$$

kde μ^* a Σ^* sú momenty pre nadmerné výnosy. Optimalizačná úloha už nezavisí od w_0, lebo w_0 sa už nenachádza v optimalizačnej funkcii, a úloha nemá obmedzenie pre váhy, lebo pri ľubovoľnej hodnote $\sum_{i=1}^{nA} w_i$ vieme zvoliť w_0 tak, aby $\sum_{i=0}^{nA} w_i = 1$. Našim odvodením sme ukázali, že optimalizačná úloha pre výnosy rizikových aktív a výnos bezrizikového aktíva s obmedzením pre sumu váh je ekvivalentná optimalizačnej úlohe pre nadmerné výnosy rizikových aktív bez obmedzenia.

Keďže v Markowitzovej teórii portfólia minimalizujeme riziko a maximalizuje zisk a v úlohe (6) používame váhy pre rizikové aktíva a váhu pre bezrizikové aktívum, tak intuitívne nám vyplýva, že optimálne riešenia úlohy (6) a tým pádom aj úlohy (8) budú ležať na priamke kapitálového trhu (viď obr. 2). Dôkaz tejto predstavy uvádza napr. Campbell a Viceira v [5, str. 13-15].

2 Oceňovanie aktív

V tejto kapitole sa zaoberáme modelmi pre oceňovanie aktív, kvôli súvislosti medzi podpriestorovou Markowitzovou teóriou portfólia a teóriou arbitrážneho oceňovania. Predpoklad platnosti teórie arbitrážneho oceňovania je použitý v podkapitole 4.1 v tvrdení 2 na dokázanie rovnosti Sharpovho pomeru odhadnutého podpriestorového optimálneho portfólia a Sharpovho pomeru pravého optimálneho portfólia pre počet aktív a veľkosť trénovacieho okna rastúce do nekonečna. Model oceňovania kapitálových aktív úvadzame tiež, lebo je považovaný za predchodcu teórie arbitrážneho oceňovania a je medzi nimi určitá podobnosť.

Obidva modely oceňovania aktív riešia správne ocenenie aktíva, ktoré je pridané do dobre diverzifikovaného portfólia. Diverzifikácia portfólia je nutná pre to, aby individuálne chyby aktív (*specific risk*) mali zanedbateľný vplyv na riziko portfólia. Tým pádom je riziko portfólia vyjadrené len systematickým rizikom (*systamatic risk*).

2.1 Model oceňovania kapitálových aktív

Jeden z prvých významných modelov na oceňovanie aktív bol model oceňovania kapitálových aktív (CAPM) predstavený W. F. Sharpom a J. Lintnerom v roku 1960. Model oceňovania kapitálových aktív vyjadruje, že očakávaný výnos aktív (zodpovedajúci systematickému riziku) závisí len na vzťahu s trhovým portfóliom. Tento model je jednofaktorový, keďže na vyjadrenie očakávaných výnosov aktív nám stačí jediný faktor - trhové portfólio. Pojem trhové portfólio je vysvetlené napríklad v [15, str. 67]. Predpoklady pre tento model preberáme z [27, str. 19]:

- investori majú rovnaké informácie a z nich robia rovnaké predikcie o budúcich výnosoch aktív
- všetci investori majú rovnaký horizont jedného obdobia
- investori ohodnocujú svoje portfóliá podľa očakávanej výnosnosti a volatility pri horizonte jedného obdobia
- neberú sa do úvahy transakčné náklady a ani zdaňovanie

- existuje bezrizikové aktívum a investori môžu nakúpiť alebo predať ľubovoľné množstvo tohto aktíva za tzv. bezrizikový výnos (bezrizikovú výnosovú mieru)
- jednotlivé aktíva sú nekonečne deliteľné
- keď si investori môžu vybrať medzi dvoma portfóliami s rovnakou volatilitou, tak si vyberú to s vyšším výnosom a keď si vyberajú medzi portfóliami s rovnakým výnosom, vždy si vyberú portfólio s nižšou volatilitou

Z predpokladov je vidno, že je v nich vysvetlený ideálny svet odchýlený od reality. Ku každému predpokladu by sa dalo nájsť kritiku, v čom je nereálny, ale za najdôležitejšie považujeme upozorniť na najsilnejší predpoklad. Najsilnejší predpoklad je, že investori na základe rovnakých informácií robia rovnaké predikcie o budúcich výnosoch aktív. Tento predpoklad súvisí s predpokladom, že investori ohodnocujú portfóliá len podľa očakávanej výnosnosti a volatility pri horizonte jedného obdobia. V realite sa investori zameriavajú na rôzne informácie o aktívach a ich predikcie do budúcna sa líšia.

Vzťah pre model CAPM je nasledovný:

$$\begin{aligned} \mu_i = r_f &+ \beta_i(\mu_m - r_f), \\ \beta_i &= \frac{\sigma_{im}}{\sigma_m^2}, \end{aligned} \tag{9}$$

kde

μ_i - stredná hodnota pre výnos i-teho aktíva,

μ_m - stredná hodnota pre výnos trhového portfólia,

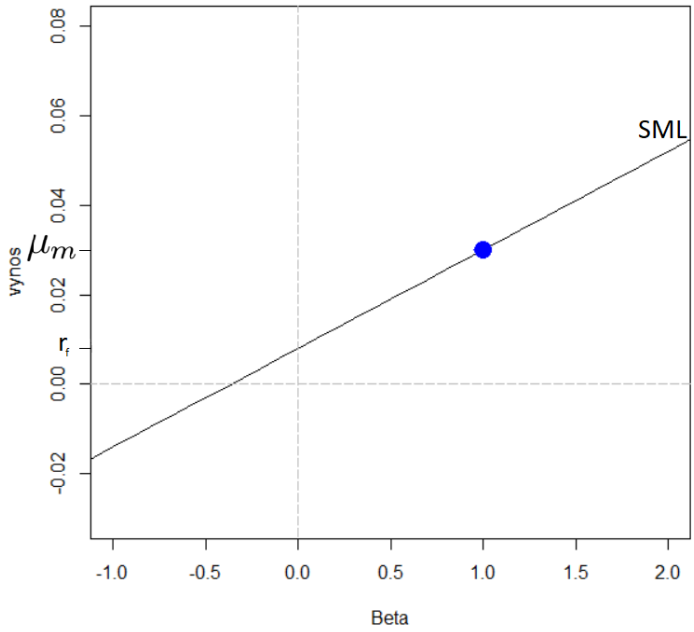
σ_{im} - kovariancia i-teho aktíva a trhového portfólia,

σ_m^2 - rozptyl trhového portfólia.

Časť rovnice $(\mu_m - r_f)$ označuje trhovú rizikovú prémiu, teda aký máme nadvýnos oproti bezrizikovému aktívu, keď podstupujeme riziko vlastnenia trhového portfólia. Časť $\beta_i(\mu_m - r_f)$ označuje rizikovú prémiu pre i-te aktívum.

Vzťah (9) sa dá interpretovať, že výnos každého aktíva je zložený z bezrizikovej výnosovej miery a svojej rizikovej prémie. Riziko sa v tomto modeli meria relatívne k riziku trhového portfólia. Pre $(\mid \beta_i \mid = 1)$ je riziko i-teho aktíva rovné riziku trhového portfólia. Aktíva s $(\mid \beta_i \mid < 1)$ majú menšie riziko a aktíva s $(\mid \beta_i \mid > 1)$ majú väčšie riziko ako trhové portfólio.

Priamka, ktorá vznikne zo vzťahu (9) pre $\beta_i \in \mathbb{R}$ sa nazýva priamka trhu cenných papierov (*Security market line* - SML). Na obr. 3 vidíme priamku trhu cenných papierov, na ktorej podľa CAPM ležia v ekvilibriu všetky aktíva na trhu.



Obr. 3: Priamka trhu cenných papierov - SML. Modrou bodkou je znázornené trhové portfólio, $\mu_m = 0.03$ a $r_f = 0.008$.

2.2 Teória arbitrážneho oceňovania

Poznatky v tejto podkapitole čerpáme z [23], [6, str. 9-13], [12] a [27]. Teória arbitrážneho oceňovania (*arbitrage pricing theory* - APT) je na rozdiel od CAPM založená na tom, že systematické riziko na trhu ovplyvňuje nie jeden, ale viacero spoločných faktorov. Tak isto APT používa iné predpoklady ako CAPM pre funkčnosť teórie.

Z [12, str. 2]: „Matematický dôkaz APT požaduje ohraničenie na β a kovariančnú maticu $\Sigma = E(\varepsilon\varepsilon^\top)$. Častý dodatočný predpoklad je, že $E(\varepsilon \mid f) = 0$, ale tento predpoklad nie je nutný v niektorých APT rozšíreniach. Predpokladá sa tiež, že počet aktív n je oveľa väčší ako počet faktorov k.“

Pred definovaním APT modelu najprv vyjadrujeme nadmerné výnosy individuálnych aktív pomocou viacfaktorového modelu:

$$\begin{aligned} R_{i,t} = r_{i,t} - r_f = \alpha_i \quad &+ \quad \beta_{i,1}(f_{1,t} - r_f) + ... + \beta_{i,K}(f_{K,t} - r_f) + \varepsilon_{i,t}, \\ \text{pre } i \quad &= \quad 1, 2, ...nA, \ t = 1, 2, ...T, \end{aligned} \tag{10}$$

kde

$R_{i,t}$ - nadmerný výnos i-teho aktíva v mesiaci t,

α_i - špecifický výnos pre i-te aktívum nezahrnutý vo faktorových zložkách,

$\beta_{i,j}$ - koeficient prislúchajúci k i-temu aktívu a j-temu faktoru,

$f_{j,t}$ - hodnota j-teho faktora v mesiaci t,

$\varepsilon_{i,t}$ - individuálna chyba pre i-te aktívum v mesiaci t.

Za podmienky, že systematické riziko ovplyvňuje len niekoľko faktorov na trhu, APT model určuje nasledovný vzťah pre výpočet strednej hodnoty nadmerného výnosu:

$$\mu_i^* = \alpha_i + \beta_{i,1}(\mu_{f_1} - r_f) + ... + \beta_{i,K}(\mu_{f_K} - r_f), \quad \text{pre } i = 1, 2, ...nA, \tag{11}$$

kde

μ_i^* - stredná hodnota nadmerného výnosu pre i-te aktívum,

μ_{f_j} - stredná hodnota j-teho faktora a

α_i - nulová pri exaktnom APT modeli.

V (11) vidíme, že pre exaktný APT model sa očakávaný nadmerný výnos všetkých aktív dá vyjadriť pomocou kombinácie očakávaných hodnôt faktorov očistených o bezrizikový výnos.

Kovariančná matica pre nadmerné výnosy R vyzerá nasledovne:

$$\Sigma^* = B\Sigma_f B^\top + \Sigma_\varepsilon, \tag{12}$$

kde

B - matica rozmeru $nA \times K$, ktorej i-ty riadok je $\beta_i^\top$ (vektor s koeficientami pre i-te aktívum),

Σ_f - kovariančná matica faktorov a

Σ_ε - kovariančná matica individuálnych chýb.

APT model je založený na myšlienke, že na trhu neexistuje arbitráž. Arbitráž nastáva, keď sa dá získať nadvýnos bez dodatočného rizika. V prípade, že by bolo nadhodnotené aktívum, tak by investor mohol kúpiť krátku pozíciu tohto aktíva a k tomu kúpiť takú dlhú pozíciu z ostatných správne nacenených aktív, ktorá by mala takú istú expozíciu (také isté koeficienty β) do faktorov na trhu. Tým pádom by kupoval nejaké riziko v dlhej pozícii a také isté riziko by v krátkej pozícii predal, ale za vyššiu cenu, čím mu vznikla arbitrážna situácia. Preto v prípade bezarbitrážneho trhu nenastane nadhodnotenie žiadneho aktíva a bude platiť APT model.

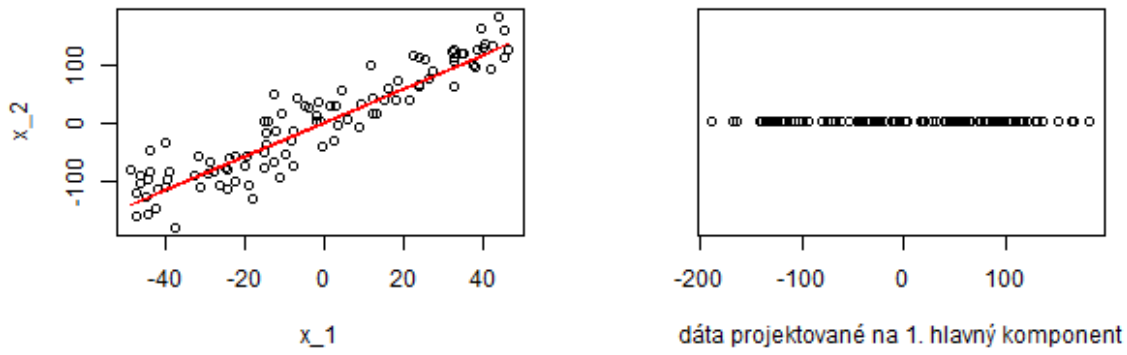
Nevýhodou APT modelu je, že neurčuje koľko spoločných faktorov na trhu je a aké konkrétne faktory to sú, pričom CAPM model jednoznačne určuje, že sa jedná o jediný faktor a to trhové portfólio.

3 Metóda hlavných komponentov

Metóda hlavných komponentov sa používa na redukciu dimenzie dát. Pri hľadaní trhových faktorov sa snažíme práve o redukciu počtu aktív len na niekoľko trhových faktorov. Metóda hlavných komponentov je jednou z technických metód, ktoré sa používajú na tento účel. Poznatky v tejto kapitole sme získali na predmete [22] a z knihy [13].

Majme náhodný vektor x, ktorý má p parametrov. Metóda hlavných komponentov sa zameriava na nájdenie takých lineárnych závislostí medzi parametrami, ktoré zachovávajú čo najviac rozptylu z prvotných dát, a pritom sú medzi sebou lineárne nezávislé. Vybratím m takých lineárnych kombinácií sa dá redukovať dimenzia dát, ak zvolíme $m < p$. Týmto spôsobom vieme zredukovať dimenziu dát a súčasne zachovať veľkú časť rozptylu.

Pre znázornenie tejto redukcie zobrazujeme na obr. 4 redukciu z 2-rozmerného priestoru do 1-rozmerného priestoru.



Obr. 4: Redukcia dát z 2-rozmerného priestoru do 1-rozmerného priestoru

V prvej časti obr. 4 je červenou úsečkou znázornený smer, ktorý zachováva maximum rozptylu dát. V druhej časti je už znázornená len úsečka a na nej vyznačené body, ktoré vznikli kolmou projekciou dát na červenú úsečku v prvej časti. V tomto prípade zachytávame veľké množstvo rozptylu prvotných dát, lebo parametre x_1 a x_2 sú medzi sebou veľmi korelované, ako je vidno na obr. 4. Metóda hlavných komponentov je vhodná

hlavne v prípade, ak sú dáta medzi sebou aspoň trochu korelované a tým pádom, síce znížime dimenziu, ale neznížime o veľa rozptyl dát. To by pre úplne nekorelované dáta nebolo možné.

Hlavné komponenty sa získavajú pomocou maximalizácie rozptylu lineárnej kombinácie základných parametrov. Označme Σ ako kovariančnú maticu rozmeru $p \times p$, kde na i-tej pozícii na diagonále sú rozptyly i-teho parametra a na pozícii (i, j) je kovariancia i-teho a j-teho parametra, pre $i \neq j$. Hlavný komponent k označme $z_k = \alpha_k^\top x$, kde α_k je vektor s príspevkami pre náhodný vektor x.

Pre získanie hlavných komponentov treba začať maximalizáciou rozptylu prvého hlavného komponentu. Teda optimalizačná úloha je nasledovná:

$$\max_{\alpha_k} \operatorname{var}(z_k) = \max_{\alpha_k} \operatorname{var}(\alpha_k^\top x) = \max_{\alpha_k} \alpha_k^\top \Sigma \alpha_k, \qquad (13)$$
$$\text{pri podmienkach } \alpha_k^\top \alpha_k = 1,$$

kde k volíme 1 a $\operatorname{var}(x) = \Sigma$.

V podmienke je ohraničenie na α tak, aby súčet štvorcov príspevkov v α bol rovný 1. Bez ohraničenia pre α by výsledok optimalizácie mohol byť rovný ∞, lebo prvky vektororu α by mohli byť ∞. Preto je nejaké ohraničenie nutné a toto konkrétne ohraničenie sa volí kvôli dobre riešiteľnej optimalizácii. Jolliffe podotýka ([13, str. 5]), že v niektorých konkrétnych prípadoch by mohlo byť vhodnejšie iné ohraničenie.

Optimalizáciu (13) môžeme riešiť pomocou Langrangeových multiplikátorov:

$$\max_{\alpha_1} \alpha_1^\top \Sigma \alpha_1 - \lambda(\alpha_1^\top \alpha_1 - 1),$$

kde λ je Lagrangeov miltiplikátor.

Po derivácii podľa α_1 dostávame:

$$\begin{aligned} \Sigma\alpha_1 - \lambda\alpha_1 &= \\ (\Sigma - \lambda I_{p\times p})\alpha_1 &= 0, \end{aligned} \qquad (14)$$

kde $I_{p\times p}$ je matica identity rozmeru $p \times p$.

Zo vzťahu (14) vidíme, že α_1 je vlastný vektor kovariančnej matice Σ a λ je prislúchajúce vlastné číslo. Navyše po dosadení (14) do (13) máme:

$$\max_{\alpha_k} \alpha_k^\top \Sigma \alpha_k = \max_{\alpha_k, \lambda} \alpha_k^\top \lambda \, \alpha_k = \max_{\alpha_k, \lambda} \lambda \, \alpha_k^\top \alpha_k = \max_{\lambda} \lambda. \qquad (15)$$

Preto je λ najväčšie vlastné číslo matice Σ. Po získaní prvého hlavného komponentu získame ostatné hlavné komponenty pomocou vzorca (13), pričom pridávame podmienky, nech $z_k \perp\!\!\!\perp z_i$, kde $i = 1, \ldots k-1$. Tým dostaneme, že α_k pre k-ty hlavný komponent sa rovná vlastnému vektoru matice Σ prislúchajúcemu ku k-tej najväčšej vlastnej hodnote matice Σ.

3.1 Výber počtu hlavných komponentov

Pri metóde hlavných komponentov sa volí, koľko hlavných komponentov sa použije. V tejto kapitole neuvádzame prehľad metód pre výber počtu hlavných komponentov, ale zameriavame sa na 2 metódy, ktoré v práci používame. Prvá z metód je jednoduchšia a veľmi intuitívna metóda a druhá metóda je zložitejšia a asymptoticky významná.

Výber pomocou percentilu celkového rozptylu dát

Pri tejto metóde volíme hodnotu h, ktorá vyjadruje aký percentil celkového rozptylu dát musí súčet rozptylov vybratých hlavných komponentov dosiahnuť. Celkový rozptyl dát označujeme ako súčet rozptylov všetkých parametrov dát. Ak je Σ regulárna matica, tak súčet rozptylov všetkých parametrov dát sa tiež rovná súčtu rozptylov všetkých hlavných komponentov:

$$\text{celkový rozptyl } = \sum_{i=1}^{p} \Sigma_{i,i} = \sum_{k=1}^{p} \text{var}(z_k), \tag{16}$$

kde $\Sigma_{i,i}$ je rozptyl i-teho parametra dát.

Pri metóde hlavných komponentov predpokladáme, že rozptyl parametrov obsahuje informácie o dátach. Preto výberom počtu hlavných komponentov pomocou percentilu celkového rozptylu dát určujeme, koľko percent z prvotných informácií chceme zachovať pri znížení dimenzie dát. Počet hlavných komponentov pomocou percentilu celkového rozptylu dát je najmenšie m, pre ktoré platí:

$$h < 100 \ \frac{\sum_{k=1}^{m} \text{var}(z_k)}{\sum_{i=1}^{p} \Sigma_{i,i}}, \tag{17}$$

kde h je hodnota percentilu, ktorú musíme zvoliť.

Jolliffe [13, str. 113] uvádza, že klasicky sa h volí medzi 70 a 90, ale záleží na konkrétnych dátach, na ktoré metódu používame. My v našej práci tvoríme analýzu hlavných komponentov pre výnosy, pri ktorých záleží aj na veľmi malých odchýlkach, a preto volíme 92. a 96. percentil celkového rozptylu dát.

Výber optimálneho počtu hlavných komponentov

Bai a Ng [1] sa zaoberajú nájdením vzťahu pre optimálny počet hlavných komponentov vo viacfaktorovom modeli pre počet aktív a veľkosť trénovacieho okna rastúce

do nekonečna. Vyjadrujú, že klasické Akaikeho informačné kritérium (AIC) a Bayesovo informačné kritérium (BIC), ktoré sú funkcie iba počtu aktív nA alebo iba veľkosti trénovacieho okna T, nefungujú dobre, keď obidve premenné rastú do nekonečna. Preto navrhujú také vlastné kritériá, že za istých predpokladov regularity pre nA, $T \to \infty$ platí $P(\widehat{d} = K) \to 1$, kde K je pravý počet faktorov a $\widehat{d}$ je odhadnutý optimálny počet faktorov.

Chybu pre počet faktorov d a odhadnuté faktory $\widehat{F}^d$ označujú ako:

$$V(d, \widehat{F}^d) = \min_{\Lambda} \frac{1}{nA\ T} \sum_{i=1}^{nA} \sum_{t=1}^{T} (X_{it} - \lambda_i^{d\top} \widehat{F}_t^d)^2, \tag{18}$$

kde

nA - počet všetkých aktív,

T - počet zaznamenaných výnosov,

X_{it} - výnosy aktíva i v mesiaci t,

λ_i^d - koeficienty faktorov pre i-te aktívum pri použití d faktorov,

$\widehat{F}_t^d$ - výnosy d odhadnutých faktorov v mesiaci t.

Na nájdenie optimálneho počtu faktorov uvažujú funkciu:

$$PC(d) = V(d, \widehat{F}^d) + d\ g(nA, T), \tag{19}$$

kde $d\ g(nA, T)$ vyjadruje penalizáciu za množstvo použitých faktorov. Pri týchto nastaveniach formulujú vetu 2.

Veta 2 [1, str. 199]

Predpokladajme, že sú splnené predpoklady A-D [1] a že faktory sú odhadnuté pomocou analýzy hlavných komponentov. Nech d_{max} je ohraničené celé číslo také, že $K \leq d_{max}$. Nech $\widehat{d} = argmin_{0 \leq d \leq d_{max}} PC(d)$. Potom $\lim_{nA,T \to \infty} P(\widehat{d} = K) = 1$, ak sú splnené tieto 2 podmienky:

i) $g(nA, T) \to 0$

ii) $C_{nA,T}^2\ g(nA, T) \to \infty$ pre $nA, T \to \infty$, kde $C_{nA,T} = min(\sqrt{nA}, \sqrt{T})$

Navyše v dôsledku 1 uvádzajú, že namiesto $PC(d)$ môžeme použiť funkciu $IC(d) = \ln(V(d, \widehat{F}^d)) + d\ g(nA, T)$ a potom optimálny počet hlavných komponentov z vety 2 bude konzistentne odhadovať pravý počet faktorov. Pri aplikácii použili Bai a

[1] technické predpoklady definované pre platnosť vety 2 z [1, str. 196]

Ng viacero funkcií $g(nA, T)$ spĺňajúcich vetu 2 a my v našej práci aplikujeme tú, ktorú použili Chen a Yuan [6]:

$$IC_{p1}(d) = \ln(V(d, \widehat{F}^d)) + d\left(\frac{nA+T}{nA\ T}\right) ln\left(\frac{nA\ T}{nA+T}\right). \tag{20}$$

4 Znižovanie dimenzie prípustnej množiny v úlohe optimálneho portfólia

V tejto kapitole rozoberáme hlavnú metódu našej práce a to znižovanie dimenzie prípustnej množiny v úlohe optimálneho portfólia. Ako sme uviedli v úvode, tak motiváciou tejto metódy je pomocou zníženia dimenzie prípustnej množiny dosiahnuť nižšiu chybu odhadu strednej hodnoty a kovariančnej matice nadmerných výnosov a tým zlepšiť výkonnnosť odhadnutého optimálneho portfólia. V podkapitole 4.1 opisujeme teóriu, ktorú vypracovali Chen a Yuan [6] a v 4.2 navrhujeme naše rozšírenia tejto teórie.

4.1 Podpriestorová Markowitzova teória portfólia

Chen a Yuan [6] predstavujú vylepšenie odhadnutého optimálneho portfólia z Markowitzovej teórie portfólia, ktoré má zlú výkonnosť pre veľké počty aktív a malé trénovacie okná. Uvádzajú v [6, tvrdenie 1, str. 6], že Sharpov pomer pre odhadnuté optimálne portfólio predelený Sharpovym pomerom pravého optimálneho portfólia klesá k nule s rastúcim počtom aktív pre fixné trénovacie okno. To znamená, že odhadnutie strednej hodnoty a kovariančnej matice nadmerných výnosov veľkého počtu aktív spôsobujú zlú výkonnosť odhadnutého optimálneho portfólia oproti pravému optimálnemu portfóliu, keď máme k dispozícii málo historických údajov.

Tento problém sa rozhodli Chen a Yuan riešiť ohraničením prípustnej množiny optimalizačnej úlohy. Danou zmenou sa snažia nájsť také optimálne portfólio, ktorému redukujú prípustnú množinu, ale popri tom nestratia dôležité informácie o výnosoch všetkých aktív. Ohraničenie tvoria obmedzením pre váhy w, aby patrili do lineárneho podpriestoru $\mathcal{P} \subset \mathbb{R}^{nA}$. To zmení úlohu (8) definovanú na nadmerných výnosoch aktív na:

$$\max_{w \in P} \left(w^\top \mu^* - \frac{\delta}{2} w^\top \Sigma^* w \right). \tag{21}$$

Táto úloha je explicitne riešiteľná a jej optimálne riešenie je:

$$w_{mv}^{\mathcal{P}} := \frac{1}{\gamma} P_{\mathcal{P}} (P_{\mathcal{P}}^\top \Sigma^* P_{\mathcal{P}})^{-1} P_{\mathcal{P}}^\top \mu^*, \tag{22}$$

kde $P_{\mathcal{P}}$ je $nA \times d$ matica, ktorej stĺpce sú ortogonálne bázové vektory podpriestoru $\mathcal{P}$ a d je dimenzia podpriestoru $\mathcal{P}$.

Preto váhy pre odhadnuté podpriestorové optimálne portfólio sú:

$$\widehat{w}_{mv}^{\mathcal{P}} := \frac{1}{\gamma} P_{\mathcal{P}} (P_{\mathcal{P}}^{\top} \widehat{\Sigma^*} P_{\mathcal{P}})^{-1} P_{\mathcal{P}}^{\top} \widehat{\mu^*}. \tag{23}$$

Neoptimálnosť odhadnutého podpriestorového optimálneho portfólia oproti pravému Markowitzovmu optimálnemu portfóliu môže byť vyjadrená ako:

$$s(w_{mv}) - s(\widehat{w}_{mv}^{\mathcal{P}}) = \left[s(w_{mv}^{\mathcal{P}}) - s(\widehat{w}_{mv}^{\mathcal{P}}) \right] + \left[s(w_{mv}) - s(w_{mv}^{\mathcal{P}}) \right], \tag{24}$$

kde $s()$ je Sharpov pomer, ktorý vyjaduje výkonnosť portfólia [6, str. 5].

Prvá zátvorka na pravej strane rovnice vyjadruje chybu odhadu, lebo odrátavame od výkonnosti pravého podpriestorového optimálneho portfólia výkonnosť odhadnutého podpriestorového optimálneho portfólia. Druhá zátvorka vyjadruje chybu vzniknutú zúžením investičného priestoru (označujeme ju systematická chyba), lebo od výkonnosti pravého Markowitzovho optimálneho portfólia sa odrátava výkonnosť pravého podpriestorového optimálneho portfólia. Typicky pri náraste dimenzie podpriestoru $\mathcal{P}$ rastie chyba odhadu, pretože narastá počet parametrov, ktoré odhadujeme a systematická chyba klesá, lebo menej obmedzujeme investičné možnosti. Na základe tejto úvahy by mohli optimálne váhy pre portfólio vzniknúť pomocou kompromisu medzi chybou odhadu a systematickou chybou. Chen a Yuan ilustrujú názornú ukážku na [6, obr. 1, str. 8].

Výber lineárneho podriestoru

Pri úlohe (21) je kľúčový výber lineárneho podriestoru $\mathcal{P}$. Chen a Yuan navrhujú vytvoriť $\mathcal{P}$ pomocou analýzy hlavných komponentov. Stĺpce $\mathcal{P}$ sú tvorené vlastnými vektormi matice Σ^* prislúchajúcich k d najväčším vlastným číslam. Vytvorené hlavné komponenty dokážu zachytiť veľkú časť informácie obsiahnutej vo výnosoch všetkých aktív. Pri tomto výbere lineárneho podriestoru je riešenie (22) rovné riešeniu:

$$w_d = \frac{1}{\gamma} \sum_{k=1}^{d} \theta_k^{-1} \eta_k \eta_k^{\top} \mu^*, \tag{25}$$

kde

θ_k - k-ta najväčšia vlastná hodnota matice Σ^* a

η_k - vlastný vektor matice Σ^* prislúchajúci k θ_k.

Hlavné komponenty nie sú vopred známe, keďže nepoznáme pravé Σ^* a preto môžu byť odhadnuté pomocou minulých výnosov aktív a teda pomocou prvých d vlastných vektorov matice $\widehat{\Sigma^*}$:

$$\widehat{w}_d = \frac{1}{\gamma} \sum_{k=1}^{d} \widehat{\theta}_k^{-1} \widehat{\eta}_k \widehat{\eta}_k^\top \widehat{\mu^*}. \tag{26}$$

Za predpokladu platnosti viacfaktorového modelu (10) Chen a Yuan uvádzajú predpoklady A-F pre platnosť tvrdenia 2. My ich tu pre ich technickosť nebudeme uvádzať, ale len v krátkosti prediskutujeme ich zmysel. Technicky zameranejší čitateľ ich môže nájsť v [6, str. 10, 12]. Predpoklady A-C sú o faktoroch, faktorových koeficientoch a individuálnych chybách a sú to všeobecné predpoklady používané pri viacfaktorových modeloch. Popri tom sa v obmedzeniach nepredpokladá normalita faktorov a individuálnych chýb. V predpokladoch D-F sú špecifikované obmedzenia pre maximálne slabú závislosť medzi faktormi, medzi individuálnymi chybami rôznych aktív a tiež medzi faktormi a individuálnymi chybami.

V tvrdení 2 je vyjadrený vzťah medzi výkonnosťou váh pravého Markowitzovho optimálneho portfólia a výkonnosťou váh odhadnutého podpriestorového optimálneho portfólia.

Tvrdenie 2: [6, str. 12]
Nech $d = K$ je fixné a konečné. Potom za predpokladu asymptotického APT modelu a predpokladov A-F platí:

$$s^2(\widehat{w}_d) = s^2(w_{mv}) + O_{\mathcal{P}}(T^{-\frac{1}{2}} + nA^{-\frac{1}{2}}), \tag{27}$$

kde
d - počet použitých hlavných komponentov,
K - pravý počet faktorov platný pre APT model,
$\widehat{w}_d$ - váhy odhadnutého podpriestorového Markowitzovho optimálneho portfólia,
w_{mv} - váhy pravého Markowitzovho optimálneho portfólia.

Z tvrdenia 2 vyplýva, že ak je veľký počet aktív nA a veľké množstvo historických údajov v trénovacom okne T, tak Sharpov pomer pre $\widehat{w}_d$ bude približne rovný ako Sharpov pomer pre w_{mv}, ak sa počet hlavných komponentov rovná počtu pravých faktorov na trhu. Chen a Yuan pripomienkujú, že vo všeobecnosti sa hlavné komponenty líšia

od všeobecných faktorov, ale tvrdenie 2 ukazuje, že pri podpriestorovej Markowitzovej teórii portfólia je analýza hlavných komponentov dobrým odhadom faktorov pri veľkom počte aktív.

Pravý počet faktorov K je neznámy, ale v podkapitole 3.1 sme uviedli, že Bai a Ng vyjadrili vzťah pre nájdenie počtu hlavných komponentov $\widehat{d}$, taký že pod istými predpokladmi regularity pre nA, $T \to \infty$ platí $P(\widehat{d} = K) \to 1$ pre viacfaktorový model.

Z dôsledku 3 [6, str. 13] vyplýva, že tvrdenie 2 platí aj pre odhadnuté $\widehat{d}$ také, že $P(\widehat{d} = K) \to 1$ pre nA, $T \to \infty$. Preto dostávame, že $s(\widehat{w}_{\widehat{d}})$ je asymptoticky rovné $s(w_{mv})$.

4.2 Rozšírenie podpriestorovej Markowitzovej teórie portfólia

V našej diplomovej práci priamo nadväzujeme na spôsob redukcie prípustnej množiny v úlohe optimálneho portfólia od Chena a Yuana. Podpriestorovú Markowitzovu téoriu rozširujeme 2 spôsobmi. V prvom spôsobe obmieňame podpriestor $\mathcal{P}$ tak, že jeho stĺpce tvoria vlastné vektory z korelačnej matice namiesto vlastných vektorov kovariančnej matice. V druhom spôsobe sme sa zamerali na vzorec (23) a navrhli sme pre ňho jednu technickú zmenu.

Prvé rozšírenie

Analýza hlavných komponentov sa tvorí na kovariančnej matici, keď majú parametre dát rovnaké jednotky, alebo približne rovnaké rozptyly. My používame výnosy aktív, ktoré sú v percentách a tým pádom sa javí použitie kovariančnej matice ako adekvátne. Napriek tomu za istých okolností, ktoré uvádzame nižšie by mohlo byť použitie korelačnej matice výhodnejšie.

Ako sme uviedli v podkapitole 2.2, pre vývoj výnosov jednotlivých aktív uvažujeme faktorový model (10). Ak je individuálna chyba aktív v priebehu času nízka v porovnaní s faktorovými zložkami $(\beta_{j,1}f_{1,t} + ... + \beta_{j,K}f_{K,t})$, tak väčšina rozptylu na trénovacom okne vznikne pôsobením faktorov. Tým pádom pomocou analýzy hlavných komponentov na kovariančnej matici získavame najviac rozptylu spôsobeného faktormi, čo je žiadaný stav. Ale v prípade, že by bol rozptyl niektorých aktív silne ťahaný individuálnymi chybami a nie faktorovými zložkami, tak vyšší rozptyl týchto aktív na trénovacom období by oproti iným aktívam tvoril nežiadúci stav. Stav je nežiadúci, lebo rozptyl nenesie viac informácie o faktoroch, ale o svojej individuálnej chybe. V takom prípade by sme chceli rozptyl týchto aktív ponížiť, lebo kvôli ich rozptylu by sme nachádzali nesprávne faktory. Problém je, že nevieme jasne rozlíšiť aktíva, ktoré sú viac ťahané faktorovými zložkami oproti aktívam ťahanými individuálnymi chybami. Môžeme urobiť úpravu, ktorou štandardizujeme rozptyly všetkých aktív, čo je ekvivalentné tvoreniu analýzy hlavných komponentov na korelačnej matici. Pri štandardizácii neponížime len najväčšie rozptyly aktív, ale tiež povýšime najnižšie rozptyly aktív oproti rozptylom ostatných aktív. Ak nízky rozptyl aktíva je tvorený prevažne individuálnou chybou, tak nám to bude kaziť analýzu hlavných komponentov, pričom ak je jeho nízky rozptyl

tvorený prevažne faktorovými zložkami, tak to zlepší analýzu hlavných komponentov. Vidíme, že analýza hlavných komponentov na korelačnej matici môže uškodiť, ale aj pomôcť či už pre aktíva s vyšším alebo nižším rozptylom oproti iným aktívam.

V teoretických dôkazoch naše rozšírenie 1 nemá zmysel, lebo sa predpokladá, že individuálne chyby aktív sú zanedbateľné. Napriek tomu to v realite nemusí platiť a naše rozšírenie môže mať lepšie výsledky ako odhadnuté podpriestorové optimálne portfólio. Myslíme si, že použitie korelačnej matice môže pomôcť hlavne v prípade, keď jednotlivé aktíva majú špecifické problémy ako problémy s manažmentom, extrémne oslabený dopyt po produktoch/ službách alebo škandál suvisiaci s firemnou činnosťou. Naopak naše rozšírenie bude fungovať horšie, ak pre aktíva platí, že ich individuálne chyby sú minimálne.

Preto na základe (26) budú optimálne váhy prvého rozšírenia podpriestorového optimálneho portfólia nasledovné:

$$\widehat{w}_d^{r1} = \frac{1}{\gamma} \sum_{k=1}^{d} \widehat{\theta}_k^{C^{-1}} \widehat{\eta}_k^C \widehat{\eta}_k^{C^\top} \widehat{\mu}, \tag{28}$$

kde

$\widehat{\theta}_k^C$ - k-ta najväčšia vlastná hodnota korelačnej matice výnosov $\widehat{C}$ a

$\widehat{\eta}_k^C$ - je vlastný vektor korelačnej matice výnosov $\widehat{C}$ prislúchajúci k $\widehat{\theta}_k^C$.

Druhé rozšírenie

V druhom rozšírení riešime technickú zmenu výpočtu podpriestorového optimálneho portfólia. Najprv uvádzame technické úpravy, ktoré nastávajú pri zmene odhadnutého Markowitzovho optimálneho portfólia na odhadnuté podpriestorové optimálne portfólio.

Pri úlohe Markowitzovho optimálneho portfólia sú odhadnuté optimálne váhy funkciou od odhadnutej strednej hodnoty a odhadnutej kovariančnej matice výnosov:

$$\widehat{w}_{mv} = \frac{1}{\gamma} \widehat{\Sigma}^{*^{-1}} \widehat{\mu^*} \tag{29}$$

V úlohe podpriestorového Markowitzovho portfólia môžeme prepísať výpočet optimálneho portfólia na vzťah, v ktorom vystupuje funkcia od $\widehat{\Sigma}^*$:

$$\widehat{w}_{mv}^{\mathcal{P}} := \frac{1}{\gamma} P_{\mathcal{P}} (P_{\mathcal{P}}^{\top} \widehat{\Sigma}^{*} P_{\mathcal{P}})^{-1} P_{\mathcal{P}}^{\top} \widehat{\mu^{*}} = \frac{1}{\gamma} func(\widehat{\Sigma}^{*}) \widehat{\mu^{*}},$$
$$func(\widehat{\Sigma}^{*}) = P_{\mathcal{P}} (P_{\mathcal{P}}^{\top} \widehat{\Sigma}^{*} P_{\mathcal{P}})^{-1} P_{\mathcal{P}}^{\top}, \tag{30}$$

kde

$func(\widehat{\Sigma}^{*})$ - fukncia, ktorá transformuje maticu $\widehat{\Sigma}^{*}$ a

$P_{\mathcal{P}}$ - $nA \times d$ matica, ktorej stĺpce sú vlastné vektory matice $\widehat{\Sigma}^{*}$ prislúchajúce k d najväčším vlastným hodnotám.

Funkcia $func()$ nieje funkcia, ktorá len transformuje kovarinčnú maticu do podpriestoru $\mathcal{P}$, ale daná transformácia je komplexnejšia a nie teoreticky interpretovateľná.

V našom druhom rozšírení sa zameriavame namiesto transformácie kovariančnej matice $\widehat{\Sigma}^{*}$ na transformáciu korelačnej matice $\widehat{C}^{*}$ pomocou funkcie $func()$.

Ako prvý krok prepisujeme v (29) kovariančnú maticu na maticový súčin korelačnej matice a diagonálnych matíc s rozptylmi na diagonále:

$$\widehat{w}_{mv} = \frac{1}{\gamma} (D_{\widehat{\sigma^{*}}} \widehat{C}^{*} D_{\widehat{\sigma^{*}}}^{\top})^{-1} \widehat{\mu^{*}} = \frac{1}{\gamma} D_{\widehat{\sigma^{*}}}^{-1} \widehat{C^{*}}^{-1} D_{\widehat{\sigma^{*}}}^{\top^{-1}} \widehat{\mu^{*}} \tag{31}$$

Potom definujeme naše druhé rozšírenie podpriestorového optimálneho portfólia ako:

$$\widehat{w}_{d}^{r2} := \frac{1}{\gamma} D_{\widehat{\sigma^{*}}}^{-1} func(\widehat{C}^{*}) D_{\widehat{\sigma^{*}}}^{\top^{-1}} \widehat{\mu^{*}}, \tag{32}$$

kde $func()$ závisí od $\widehat{C}^{*}$, a preto matica $P_{\mathcal{P}}$ je tvorená vlastnými vektormi matice $\widehat{C}^{*}$.

5 Testovanie na dátach

V tejto kapitole tvoríme praktickú časť práce. Najprv opisujeme, aké všetky stratégie používame a ako ich vyhodnocujeme. Potom v 3 podkapitolách vyhodnocujeme Markowitzovo optimálne portfólio na reálnych dátach a podpriestorové optimálne portfólio aj na simulovaných a aj na reálnych dátach. Keďže v častiach používame rôzne sady dát, tak prehľad o použitých dátach prikladáme v tab. 9 v prílohe 1.

Ešte pred opísaním samotných stratégií definujeme najdôležitejšie premenné, na ktoré sa často-krát odkazujeme:
n - dĺžka celého testovacieho obdobia, konkrétnejšie počet mesiacov za celé testovacie obdobie,
t - konkrétny mesiac z celého testovacieho obdobia, $t = 1, ...n$,
$R_{t,i}$ - mesačné výnosy realizované od mesiaca $t-1$ do mesiaca t pre i-te aktívum, ktoré sú očistené od bezrizikového výnosu. Teda sa jedná o tzv. nadmerné výnosy,
$w_{t,i}$ - váha v mesiaci t pre aktívum i,
T - trénovacie okno, je to počet mesiacov použitých pri metóde mimo vzorky na odhad parametrov,
nA - počet aktív,
R - matica výnosov s rozmermi $n \times nA$ s $R_{t,i}$ v stĺpci t a riadku i.

Celý proces tvorenia portfólia na základe optimálnych váh podľa Markowitzovej teórie portfólia označujeme v našej práci ako Markowitovu stratégiu. Pre analýzu je dôležité podrobne pochopiť rozdiel medzi teoretickou Markowitzovou stratégiou, jej aplikáciou v rámci vzorky (*in sample*) a aplikáciou mimo vzorky (*out of sample*). Teoretická Markowitzova stratégia berie do úvahy pravú strednú hodnotu a pravú kovariančnú maticu, ktoré sú vyrátané z pravého rozdelenia výnosov aktív. Toto pravé rozdelenie aktív ale v realite nepoznáme.

Markowitzova stratégia v rámci vzorky je stratégia, ktorej strednú hodnotu a kovariančnú maticu odhadujeme na celej vzorke. Investor nevie zrealizovať túto stratégiu, pretože na to aby určil optimálne váhy potrebuje poznať, aké výnosy sa zrealitujú. Je to zodpovedanie otázky, aké konštantné váhy by boli optimálne ak by sme poznali informáciu o výnosoch aktív na celom období.

Pri metóde mimo vzorky sa už jedná o predikciu výnosov do budúcna, teda zrealizovateľnú stratégiu. Na jej vytvorenie potrebujeme určiť trénovacie okno (*training window*) a testovacie okno (*testing window*). To sa dá urobiť viacerými spôsobmi. Klasicky je zaužívaný spôsob rolovacie trénovacie okno s veľkosťou T (*rolling training window*) [29, str. 307] a pre testovacie okno volíme veľkosť 1. Rolovacie trénovacie okno znamená, že vždy pri predikcii výnosu v mesiaci t, zoberieme posledných T známych minulých výnosov: r_{t-T}, ...,r_{t-1} a na nich vyrátame optimálne váhy, ktoré použijeme pre investovanie v mesiaci t. Spôsob mimo vzorky je založený na tom, že predpokladáme, že na historických údajoch vieme dobre odhadnúť parametre pre stratégiu do budúcna. Tu vzniká chyba odhadu parametrov, kvôli ktorej sú vyhotovenia mimo vzorky horšie ako vyhotovenia v rámci vzorky. Pri Markowitzovej stratégii mimo vzorky sa jedná o chybu odhadu strednej hodnoty a kovariančnej matice.

5.1 Stratégie

V celej empirickej časti formulujeme optimalizačnú úlohu na nadmerných výnosoch aktív. Najprv sa pozrieme na stratégiu odhadnutú v rámci vzorky. Optimálne váhy pre túto Markowitzovu stratégiu sú: [9, str. 1922]

$$\widehat{w}_M = \frac{1}{\gamma}\widehat{\Sigma}^{*^{-1}}\widehat{\mu}^* \tag{33}$$

kde

γ - riziková averzia investora,

$\widehat{\mu}^*$ - vektor rozmeru $nA \times 1$. Stredná hodnota nadmerných výnosov aktív za celé zvolené obdobie,

$\widehat{\Sigma}^{*^{-1}}$ - matica rozmeru $nA \times nA$. Kovariančná matica nadmerných výnosov aktív za celé zvolené obdobie.

Ďalej v práci vyhodnocujeme Markowitzovu stratégia mimo vzorky. Markowitzova stratégia mimo vzorky vznikne zmenou (33) na:

$$\widehat{w}_{M_t} = \frac{1}{\gamma}\widehat{\Sigma}_t^{*^{-1}}\widehat{\mu}_t^* \tag{34}$$

kde

$\widehat{\mu}_t^*$ - vektor rozmeru $nA \times 1$. Stredná hodnota nadmerných výnosov aktív v mesiaci t

výpočítaná z posledných minulých T mesačných nadmerných výnosov,
$\widehat{\Sigma}_t^{*^{-1}}$ - matica rozmeru $nA \times nA$. Kovariančná matica nadmerných výnosov aktív v mesiaci t výpočítaná z posledných minulých T mesačných nadmerných výnosov.

Vo výsledkoch porovnávame výkonnosť dvoch hore uvedených stratégií s podpriestorovou Markowitzovou stratégiou mimo vzorky. Túto stratégiu mimo vzorky rátame vzorcom, ktorý sme uviedli v (26):

$$\widehat{w}_d = \frac{1}{\gamma} \sum_{k=1}^{d} \widehat{\theta}_k^{-1} \widehat{\eta}_k \widehat{\eta}_k^{\top} \widehat{\mu}$$

Ako posledné tiež robíme porovnanie s naivnou stratégiou. Naivná stratégia priraďuje rovnaké váhy jednotlivým aktívam:

$$w_N = \frac{\mathbb{1}_{nA}}{nA}, \tag{35}$$

kde $\mathbb{1}_{nA}$ označuje jednotkový vektor rozmeru $nA \times 1$.

Pre Markowitzovu stratégiu v rámci vzorky a pre naivnú stratégiu získame optimálne váhy jediným výpočtom. Pre Markowitzovu stratégiu mimo vzorky a podpriestorovú Markowitzovu stratégiu mimo vzorky rátame optimálne váhy pre každý mesiac t od $T + 1$ po n.

5.2 Spôsob vyhodnotenia stratégií

V empirickej časti používame pre vyhodnotenie výkonnosti stratégií Sharpov pomer (SR) a *Certainty equivalent return* (CER). Pri vyhodnoteniach vždy vynechávame prvých T mesačných výnosov, aby sme mohli porovnávať stratégiu v rámci vzorky a mimo vzorky, lebo pri stratégii mimo vzorky má trénovacie okno veľkosť T údajov, a tým pádom prvé možné hodnotenie výkonnosti nastáva až pre $T + 1$ výnos.

Následne sme nadmerné výnosy stratégií v masiaci t vyrátali ako:

$$R_{S_t} = \sum_{i=1}^{nA} w_{ti} R_{ti}, \tag{36}$$

kde R_{S_t} je nadmerný výnos stratégie v mesiaci t.

Výpočet pre Sharpov pomer uvádzame dvomi spôsobmi podľa príkladu z [9, str. 1928]. Prvý spôsob (37) je univerzálnejší, vieme ho použiť vždy keď máme zaznamenané

výnosy stratégie a v našich vyhodnoteniach ho používame pre výkonnosť stratégií mimo vzorky. Konkrétnejšie počítame Sharpov pomer z priemerného výnosu a rozptylu na základe pozorovaných výnosov portfólia.

$$SR = \frac{\mu(R_S)}{\sqrt{\text{var}(R_S)}}, \tag{37}$$

kde R_S je vektor nadmerných výnosov stratégie za celý časový úsek od $T + 1$ po n, teda vektor s hodnotou R_{S_t} na pozícii t.

Druhý spôsob (38) sa dá použiť iba v prípade, ak má stratégia celý čas stabilné váhy pre aktíva, čo platí pre Markowitzovu stratégiu v rámci vzorky. V tomto prípade vo výpočte používame priemerné výnosy a kovariančnú maticu z aktív a nie z celého portfólia.

$$SR_{in_sample} = \frac{w^\top \widehat{\mu}^*}{\sqrt{w^\top \widehat{\Sigma}^* w}}, \tag{38}$$

kde $\widehat{\mu}$ a $\widehat{\Sigma}$ sú odhadnuté na nadmerných výnosoch aktív za celé obdobie, na ktorom sme tvorili stratégiu.

Vo vzorcoch pre Sharpov pomer neodrátavame bezrizikový výnos, pretože v obidvoch prípadoch sú výpočty tvorené na nadmerných výnosoch.

Alternatívnou mierou výkonnosti stratégií, ktorú používame popri Sharpovom pomere v empirickej časti je *Certainty equivalent return*(CER):

$$CER = \mu(R_S) - \frac{\gamma}{2}\text{var}(R_S) \tag{39}$$

kde γ je v každom mesiaci taká istá a zvolili sme ju 1, lebo pri tejto voľbe sa naše výsledky najviac zhodovali s [9].

V hlavnej časti testovania sme sa zamerali aj na meranie obratu aktív (*turnover*). Toto vyhodnotenie je dôležité z 2 dôvodov. Prvý dôvod je zistenie, či je stratégia zrealizovateľná v reálnom svete. Extrémne hodnoty obratu aktív nám napovedajú, že niektoré aktíva nakupujeme v tak veľkých krátkych alebo dlhých pozíciách, že by to v realite nebolo možné. Druhý dôvod je zohľadnenie poplatkov za rebalansovanie portfólia, ktoré vznikajú kvôli transakčným nákladom a tiež kvôli rozdielu medzi nákupnou

a predajnou cenou (*bid-ask spread*) aktív. Obrat aktív meriame sumou absolútnych rozdielov medzi váhami v mesiaci t a $t-1$.

$$turnover_t = \sum_{i=1}^{nA} \mid w_{t,i} - w_{t-1,i} \mid \tag{40}$$

5.3 Markowitzova teória portfólia v rámci a mimo vzorky

V tejto podkapitole ukazujeme na reálnych dátach zlú výkonnosť Markowitzovej stratégie mimo vzorky v porovnaní s Markowitzovou stratégiou v rámci vzorky keď je počet aktív relatívne veľký oproti počtu historických údajov. Okrem toho porovnávame výkonnosť Markowitzovej stratégie mimo vzorky aj s naivným portfóliom, aby sme videli rozdiel oproti jednoduchej, ale zato účinnej investičnej stratégii. V analýze používame nastavenia a aj dáta podľa príkladu z [9].

5.3.1 Dáta

V empirickej časti pracujeme s voľne dostupnými dátami zo stránky **??**. Konkrétne, používame 10 portfolií vytvorených na základe priemyselného odvetvia, do ktorého patria (nazývame premyselné portfóliá), 20 portfólií vytvorených na základe veľkosti a pomeru účtovnej a trhovej hodnoty (nazývame portfóliá$_{20}$) a 3 faktorové portfóliá - tzv. SMB, HML a trhové portfólio. Portfóliá$_{20}$ sú vytvorené z dát 25 portfólií vytvorených na základe veľkosti a pomeru účtovnej a trhovej hodnoty, pričom 5 najväčších portfólií je vynechaných. Pri všetkých portfóliách pracujeme s výnosmi na mesačnej frekvencii, v období 07/1963 až 11/2004 ($n = 497$).

Prvá dôležitá informácia pre stiahnutie správnych dát je, že pre priemyselné portfóliá používame trhovo vážené mesačné výnosy a pre portfóliá$_{20}$ rovnako vážené mesačné výnosy. Pre faktorové portfóliá nie je uvedené, ktorý spôsob bol použitý pre dostupné dáta, pričom pre trhové portfólio je jasné, že sa jedná o trhovo vážené mesačné výnosy, lebo to správnejšie odzrkadľuje vývoj amerického trhu.

Trhovo vážené výnosy spôsobia to, že výnosy menších firiem majú menšiu váhu ako väčších firiem. Zoberme si príklad, keď chceme získať výnos z technologického priemyslu. Máme údaje o Apple a potom o nejakej menšej technologickej firme. Keď tvoríme výnos z technologického priemyslu, tak chceme čo najpresnejšie odhadnúť výnos v tomto priemysle. Teda nechceme dať takú istú váhu veľkej aj malej firme, ale keďže Apple tvorí veľkú časť z technologického priemyslu dáme mu veľkú váhu. Túto váhu určujeme na základe trhovej hodnoty firmy. Trhové váženie má tiež praktický zmysel pre investorov obchodujúcich s veľkými objemami. Trhové váženie zabezpečí, že do veľkých firiem investor zainvestuje veľké množstvá a do menších firiem menšie množstvá, a

preto budú jeho investície zrealizovateľné a jeho rozdiely medzi nákupnou a predajnou cenou by mali byť nižšie ako pri rovnakom vážení. V prípade rovnakého váženia, by investor musel nakupovať veľké objemy malých firiem, čo by buď nebolo možné, alebo by mal nevýhodnú situáciu oproti predávajúcim, ktorí by mohli požadovať vyššiu cenu, lebo sa silne zvýšil dopyt po ich aktívach.

Pre portfóliá$_{20}$ volíme rovnako vážené výnosy, lebo pomocou daných výnosov nám vychádzajú podobnejšie výsledky z naivnej stratégie pri porovnaní s výsledkami z [9] a našou motivácou je dosiahnuť rovnaké výsledky ako v danej práci. Tento postup sme museli zvoliť, pretože v [9] nie je presnejšie uvedené, ktorý spôsob použili. Navyše výnosy portfólií nie sú tak citlivé na metódu váženia v prípade, že sú tvorené na základe veľkostí.

Pred prácou s dátami tvoríme nadmerné výnosy odčítaním bezrizikovej úrokovej miery od mesačných výnosov. Ako bezrizikovú úrokovú mieru používame mesačný výnos 3-mesačných amerických vládnych dlhopisov. Z [10] máme ročný výnos týchto dlhopisov, a preto sme zobrali dvanástinu z neho, čo je pri jednoduchom úročení mesačný výnos.

V prípade SMB *a* HML portfólií neodrátavame bezrizikový výnos a to kvôli tomu, že sú to portfóliá s nulovým zainvestovaním. To znamená, že takú istú hodnotu, ktorou akcie kupujú (dlhá pozícia), používajú aj na predaj akcií (krátka pozícia). Inak povedané, bezrizikový výnos tam už je zahrnutý, pretože pri dlhej pozícii sa odrátava a pri krátkej pozícii sa prirátava, a teda sa nuluje.

Tak ako Miguel et al. sme priemyselné portfóliá a portfóliá$_{20}$ rozšírili o americké trhové portfólio. To znamená, že v prvom prípade uvažujeme množinu 11 rôznych aktív a v druhom prípade množinu 21 rôznych aktív.

5.3.2 Použitie Stratégií

Na základe [9, str. 1922] volíme pre Markowitzovu stratégiu v rámci vzorky hodnotu rizikovej averzie γ rovnú $\frac{1}{\mathbb{1}_{nA}\widehat{\Sigma}^{-1}\widehat{\mu}}$ a pre Markowitzovu stratégiu mimo vzorky $\frac{1}{\mathbb{1}_{nA}\widehat{\Sigma}_t^{-1}\widehat{\mu}_t}$ pre každé t. Táto voľba γ nám zaručuje, že sumy váh pre investovanie do aktív sú v každom mesiaci t rovné 1. Samozrejme daná rovnosť platí aj pre naivné portfólio. Upozorňujeme, že pri výpočte CER sa používa iná γ ako γ pre Markowitzovu stratégiu.

5.3.3 Vyhodnotenie

V tabuľke 1 uvádzame výsledky pre Markowitzovu stratégiu v rámci vzorky, Markowitzovu stratégiu mimo vzorky a naivné portfólio pre 3 rôzne sady aktív.

Sharpov pomer	Priemyselné portfóliá[1] + 1	Faktorové portfóliá	Portfóliá$_{20}$ + 1
Markowitzova stratégia v rámci vzorky	0,21	0,28	0,53
Markowitzova stratégia mimo vzorky	0,04	0,21	0,05
Naivné portfólio	0,13	0,23	0,17
CER			
Markowitzova stratégia v rámci vzorky	0,0097	0,0045	0,0283
Markowitzova stratégia mimo vzorky	-0.01	0,004	-0,06
Naivné portfólio	0,0049	0,0042	0,0078

Tabuľka 1: Porovnanie výkonností na základe Sharpovho pomeru a CER. Stratégie mimo vzorky sú vytvorené rolovacím trénovacím oknom s veľkosťou $T = 60$ a testovacím oknom veľkosti 1. Parameter γ pre CER sa rovná 1. Celé testovacie obdobie je 07/1963 až 11/2004 (n = 497).

[1] Portfóliá s trhovo váženými výnosmi

[2] Portfóliá s rovnako váženými výnosmi

Na výsledkoch vidíme, že Markowizova stratégia mimo vzorky má dobrú relatívnu výkonnosť oproti aplikácii v rámci vzorky len pre faktorové portfóliá. Práve pre túto sadu je najnižší počet aktív (nA=3). Ostatné 2 sady majú 11 a 21 aktív a ich relatívna výkonnosť oproti aplikácii v rámci vzorky je zlá, pričom najhoršia je práve pre sadu s najväčším počtom aktív. Tiež vidíme, že naivné portfólio má pre všetky sady lepšiu výkonnosť ako Markowizova stratégia mimo vzorky. Táto podvýkonnosť vzniká zo zlých

predikcií budúcich výnosov a kovariancií výnosov na základe historických údajov.

V tabuľke 1 vidíme, že sme dospeli k veľmi podobným výsledkom ako Miguel et al. [9, str. 1931 a 1934]. Rozdiely sa vyskytli len pre Markowizovu stratégiu mimo vzorky. Vo vyhotovení Sharpovho pomeru boli výsledky podobné, ale vo vyhodnotení CER sú pre priemyselné portfóliá a portfóliá$_{20}$ veľmi odlišné výsledky oproti Miguelovým výsledkom. Napriek tomu poznatky získané z výsledkov sa zhodujú.

Miguel et al. uvádzajú, že pri vyššom počte aktív je potrebný vyšší počet historických údajov pre lepšiu výkonnosť Markowizovej stratégie mimo vzorky. Uvádzajú, že pre sadu s 25 aktívami z amerického trhu treba viac ako 3000 historických mesačných výnosov pre prekonanie naivného portfólia.

5.4 Znižovanie dimenzie prípustnej množiny v úlohe optimálneho portfólia - analýza na simulovaných dátach

V tejto podkapitole tvoríme simuláciu nadmerných výnosov aktív, pre ktoré platí viacfaktorový model a rozptyl individuálnych chýb nie je veľký. Simulovanie tvoríme so zámerom porovnať podpriestorovú (*subspace*) Markowitzovú stratégiu mimo vzorky s pravou Markowitzovou stratégiou. Pravú Markowitzovu stratégiu vieme vyhodnotiť, pretože tvoríme simulovanie a v ňom volíme aká je pravá stredná hodnota a pravá kovariančná matica. Analýzou vyhodnocujeme ako ovplyvní výsledky to, že odhadujeme faktory pomocou analýzy hlavných komponentov a ako sa mení výkonnosť pre rôzne voľby počtu hlavných komponentov.

5.4.1 Generovanie dát

Najprv vysvetľujeme, ktoré premenné získame pomocou generovania a aké predpoklady pri generovaní využívame. Nastavenia pre generovanie výnosov preberáme z [6, str. 13 až 14]. Generujeme mesačné nadmerné výnosy nA rizikových aktív z trojfaktorového modelu. Na zhodnotenie vplyvu veľkosti portfólia uvažujeme $nA = 25$ a 100. Tri faktory simulujeme z viacrozmerného normálneho rozdelenia so strednou hodnotou a kovariančnou maticou získanými z reálnych nadmerných výnosov faktorov za obdobie 7/1963 až 8/2007 zo stránky [21]. Použité faktory sú trhový (*market*) faktor, faktor veľkosti (*size*) a faktor pomeru účtovnej a trhovej hodnoty ($book - to - market$). Trhový faktor sme upravili o bezrizikovú úrokovu mieru tak ako v podkapitole 5.3. Bety pre rizikové aktíva sme náhodne generovali z rovnomerného rozdelenia medzi 0,9 až 1,2 pre trhový faktor, medzi -0,3 až 1,4 pre faktor veľkosti, medzi -0,5 až 0,9 pre faktor pomeru účtovnej a trhovej hodnoty. Pre porovnanie dĺžky trénovacieho okna T berieme do úvahy T=60, 120 alebo 240.

Predpokladáme exaktný APT model, pre ktorý platí, že epsilony sú z normálneho rozdelenia so strednou hodnotou 0_{nA} a kovariančnou maticou Σ_ε. Rozdiel pri generovaní oproti Chenovi a Yuanovi robíme pri voľbe kovariančnej matice Σ_ε, ktorú tiež berieme diagonálnu, ale generovanú z iného rozsahu. V [6, str. 14] uvádzajú, že prvky diagonály Σ_ε sú generované z rovnomerného rozdelenia z rozsahu 0,1 až 0,3. Chen a Yuan robili generovanie na základe Tu a Zhou [28, str. 208], kde uvádzajú že generujú z

rozsahu 0,1 až 0,3 pre diagonálu kovariančnej matice ročných výnosov a nie mesačných výnosov. Preto podobné generovanie dostaneme tým, že ročné hodnoty preškálujeme koeficientom $\sqrt{12}$ nasledovne: $\frac{0.1}{\sqrt{12}} = 0.029$ až $\frac{0.3}{\sqrt{12}} = 0.087$. My sme sa však rozhodli zvoliť rozsah pre volatility na základe reálnych dát. Preto sme stiahli 25 portfólií vytvorených na základe veľkosti a pomeru účtovnej a trhovej hodnoty zo stránky [21] a urobili sme regresiu tak, aby výnos každého portfólia zo súboru (vybrali sme rovnako vážené výnosy) bol v každom mesiaci t súčtom lineárnej kombinácie faktorov popísaných vyššie a individuálnej chyby. Tým sme dostali rady individuálnych chýb pre každé aktívum a pre každý mesiac. Potom sme získali štandardnú odchýlku (*volatilitu*) individuálnych chýb pre každé aktívum. Minimálna volatilita individuálnych chýb bola 0.014 a maximálna volatilita bola 0.032. Preto sme volatility σ_i generovali z rovnomerného rozdelenia z rozsahu 0.01 až 0.03 a následne vytvorili kovariančnú maticu s týmito volatilitami na diagonále a nulovými hodnotami mimo diagonály.

Pri generovaniach z viacrozmerného normálneho rozdelenia sa určuje, či sa jedná o empirickú, alebo o populačnú strednú hodnotu a kovariančnú maticu. Faktory generujeme s empirickými momentami, a preto pri generovaní faktorov vždy vyjdú na vygenerovaných nadmerných výnosoch faktorov stredná hodnota a kovariančná matica presne tie, ktoré sme zadali do generovania. Epsilony generujeme s populačnými momentami, čo spôsobí že ich stredná hodnota a kovariančná matica budú vždy na základe náhody trochu rozdielne od tých nastavených v generovaní. Týmto nastavením spôsobujeme, že vygenerovaná stredná hodnota a kovariančná matica všetkých aktív sú trochu odlišné od pravej strednej hodnoty a kovariančnej matice aktív.

Jedno generovanie robíme pre T nadmerných mesačných výnosov pre konkrétny počet aktív nA a veľkosť trénovacieho okna T. Z vygenerovaných T nadmerných mesačných výnosov rátame optimálne váhy pre pravú Markowitzovu stratégiu a pre podpriestorovú Markowitzovu stratégiu mimo vzorky pre všetky možné počty použitých hlavných komponentov $d = 1, ...nA$.

5.4.2 Použitie Stratégií

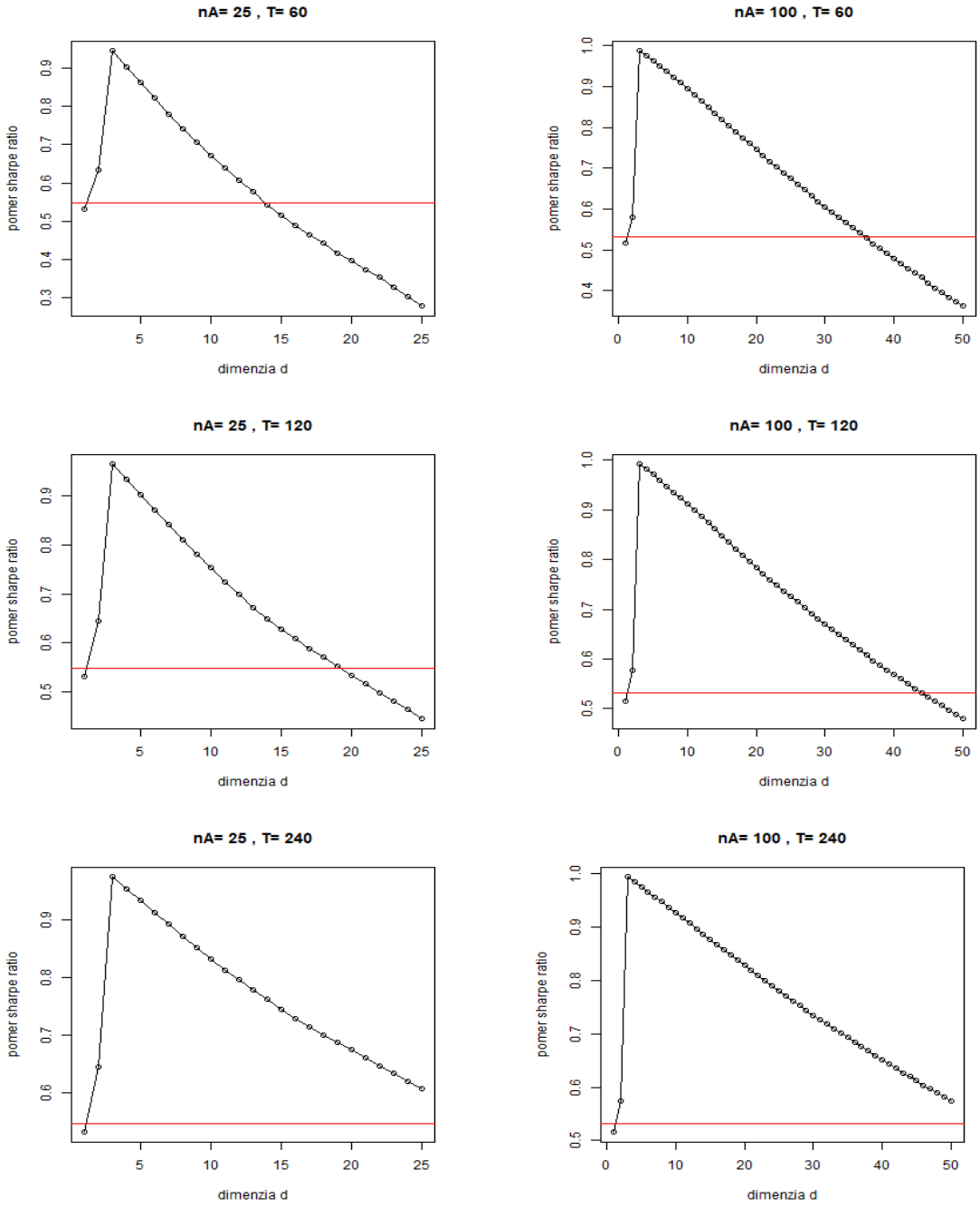
Výpočet optimálnych váh pravej Markowitzovej stratégie robíme s pravými momentami μ^* a Σ^* (používanými pre generovanie). Výpočet optimálnych váh podpriestorovej Mar-

kowitzovej stratégie mimo vzorky pomocou $\widehat{\mu}^*$ a vlastných čisel a vlastných vektorov z $\widehat{\Sigma}^*$ (momenty z vygenerovaných nadmerných výnosov).

Výnos pre vyhodnotenie Sharpovho pomeru negenerujeme, ale namiesto toho pre výpočet Sharpovho pomeru obidvoch stratégií používame μ^* a Σ^*. Považujeme to za vhodnejší prístup, lebo pravé μ^* a Σ^* určujú aký ďalší výnos by v priemere nasledoval, pričom jeden vygenerovaný výnos by bol ovplyvnený náhodou, ktorá kvôli ε nastala. Premennú γ vo výpočtoch volíme $\frac{1}{\mathbb{1}_{nA}\Sigma^{-1}\mu}$.

5.4.3 Vyhodnotenie

Na obr. 5 ukazujeme priemery pre Sharpov pomer z 1000 generovaní pre každú možnú kombináciu parametrov nA a T predelené Sharpovým pomerom pravej Markowitzovej stratégie. Na simulovaných dátach vidíme hlavnú myšlienku, s ktorou Chen a Yuan vo svojej práci prišli. Ak sa výnosy všetkých aktív dajú určiť pomocou len niekoľkých trhových faktorov, tak je oveľa výhodnejšie robiť Markowitzovu stratégiu len na týchto faktoroch ako na všetkých aktívach kvôli silne narastajúcej chybe odhadu. To je na obr. 5, v ľavých podobrázkoch vidno v tom, že pre 3 hlavné komponenty je relatívna výkonnosť podpriestorovej Markowitzovej stratégie lepšia ako pre 25 hlavných komponetov, kedy je podpriestorová Markowitzová stratégia mimo vzorky rovná Markowitzovej stratégii mimo vzorky. V pravých podobrázkoch obr. 5 nevidíme relatívnu výkonnosť Markowitzovej stratégie mimo vzorky, lebo tá nastáva až pre počet hlavných komponentov rovných 100 a my vykresľujeme x-ovú os len do hodnoty 50. Napriek tomu vidíme, že od počtu hlavných komponentov d=3, relatívna výkonnosť podpriestorovej Markowitzovej stratégie mimo vzorky systematicky klesá. Okrem toho aj pri zlom odhade počtu pravých faktorov dostávame stále dobré výsledky a pre veľa volieb d dosahujeme lepšie výsledky ako naivná stratégia.



Obr. 5: Sharpov pomer (priemer z 1000 simulácií) pre jednotlivé podpriestorové Markowitzove stratégie mimo vzorky predelený Sharpovým pomerom pravej Markowitzovej stratégie. Červenou čiarou znázornený pomer pre naivnú stratégiu.

5.5 Znižovanie dimenzie prípustnej množiny v úlohe optimálneho portfólia - analýza na reálnych dátach

V tejto podkapitole robíme analýzu na reálnych nadmerných výnosoch, aby sme overili či má podpriestorová Markowitzova stratégia zmysel, aj keď výnosy nie sú vygenerované tak, aby boli splnené predpoklady stratégie. Do analýzy tiež zapájame výpočet optimálneho počtu hlavných komponentov d, lebo nevieme koľko pravých faktorov na trhu je. Optimálny počet hlavných komponentov rátame každý mesiac t na základe posledných minulých T nadmerných výnosov, lebo nemusí platiť, že počas celého sledovaného obdobia je na trhu stále fixný počet pravých faktorov. Preto pre každý mesiac t môžeme dostať rôzny optimálny počet hlavných komponentov a na základe nich vyrátať optimálne váhy pre podpriestorovú Markowitzovu stratégiu mimo vzorky.

Okrem podpriestorovej Markowitzovej stratégie mimo vzorky prezentujeme výsledky aj pre naše návrhy na rozšírenie stratégie. Našim zámerom je na výsledkoch vidieť, či majú naše návrhy na rozšírenie stratégie zmysel pre aplikáciu na reálnych dátach.

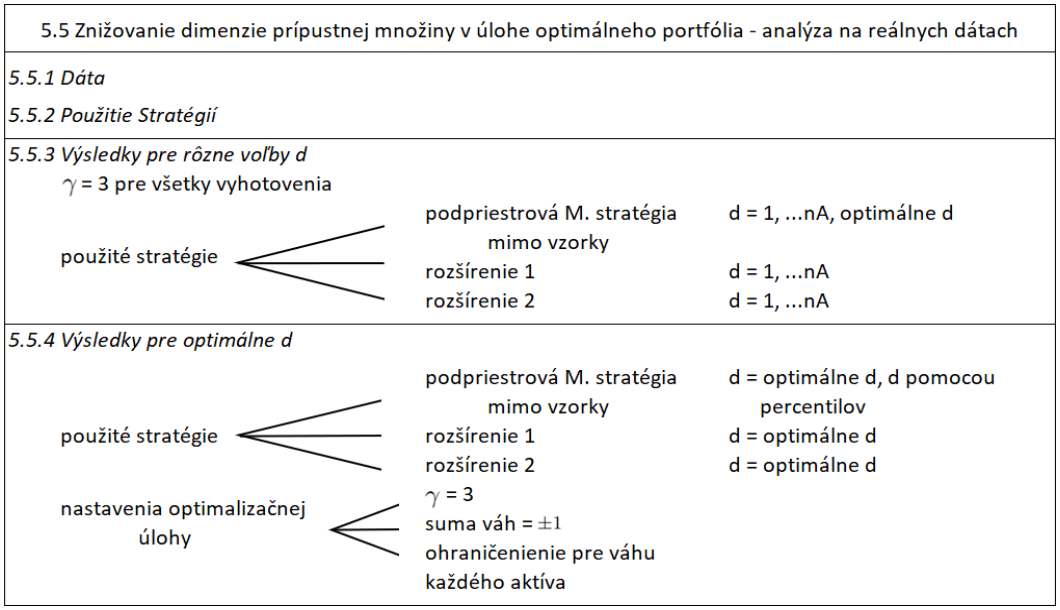
Podkapitolu 5.5.3 tvoríme pre zobrazenie výsledkov pre všetky možné fixné počty hlavných komponentov. Podkapitola tiež slúži pre porovnanie výsledkov na reálnych výnosoch s výsledkami na simulovaných výnosoch z podkapitoly 5.4.

V 5.5.4 sa už venujeme len podpriestorovým strategiám s optimálnym počtom hlavných komponentov. Optimálny počet hlavných komponentov od Bai a Ng porovnávame aj s voľbou počtu hlavných komponentov pomocou percentilov. Túto podkapitolu rozdeľujeme na viac častí, v ktorých volíme rôzne nastavenia optimalizačnej úlohy. Tieto časti rozoberáme, aby sme ukázali výsledky aj pre základné optimalizačné úlohy, a aj pre upravené optimalizačné úlohy, ktoré sú na rozdiel od tých základných uskutočniteľné z hľadiska obratu aktív.

Pre lepšie pochopenie štruktúry analýzy na reálnych dátach prikladáme v tab. 9 štruktúru kapitoly 5.5.

5.5.1 Dáta

Podľa príkladu [6] v analýze používame sady Fama-French 25 a 100 portfólií z [21], vytvorených na základe veľkostí a pomeru účtovnej a trhovej hodnoty, každé používajúce rovnako vážené (*equally weighted*) mesačné výnosy v rámci jedného portfólia. Obdobie



Tabuľka 2: Štruktúra kapitoly 5.5 - analýza na reálnych dátach.

pre testovanie stratégií je 1/1961 až 12/2010 ($n = 600$).

Sady Fama-French portfólií nie sú tvorené konkrétnymi firmami, ale už portfóliami obsahujúcimi veľa firiem. Výhodou týchto portfólií je, že sú veľmi dobre diverzifikované a preto by mali obsahovať hlavne systematické riziko a minimum individuálneho rizika. Použitie Fama-French portfólií je vhodné pre podpriestorvú Markowitzovu stratégiu, lebo v nej odhadujeme riziko všetkých aktív pomocou len niekoľkých hlavných komponentov (faktorov), ktoré dokážu zachytiť maximálne systematické riziko a nie individuálne riziko všetkých firiem.

Analýzu rozširujeme o ďalšiu sadu, ktorá obsahuje firmy z indexu S&P 500. Pre túto sadu sme zvolili obdobie 1/2000 až 2/2020 ($n = 241$), lebo firmy zo sady nemajú takú dlhú históriu dát ako je použitá pre Fama-French portfóliá. Konkrétne firmy sme vybrali tak, že sme získali zoznam firiem z indexu S&P 500 k 28.2.2020 z [30] a do sady sme zobrali všetky firmy, ktorých ceny sú dostupné na období 1/2000 až 2/2020. Tým nám vznikla naša sada 341 S&P firiem. Aktíva v tejto sade obsahujú väčšie individuálne riziká ako aktíva z Fama-French portfólií, lebo sa jedná o jednotlivé firmy. Preto je pre použitie podpriestorovej Markowitzovej stratégie sada 341 S&P firiem menej vhodná ako Fama-French portfóliá. Našim zámerom je vidieť ako sa menia výkonnosti podpriestorových Markowitzovych stratégií mimo vzorky pre túto sadu.

Tiež chceme porovnať výsledky pre naše 2 rozšírenia, ktoré robia operácie s korelačnou maticou, a preto by mohli byť pre sadu S&P 341 firiem lepšie ako podpriestorová Markowitzova stratégia mimo vzorky.

Z mesačných výnosov portfólií a S&P firiem sme odrátavali bezrizikový výnos aby sme vytvorili mesačné nadmerné výnosy. Tak isto ako v podkapitole 5.3 sme od mesačných výnosov portfólií a firiem odrátavali dvanástinu výnosu 3-mesačných amerických vládnych dlhopisov.

5.5.2 Použitie Stratégií

Na testovanie sme použili tak isto ako Chen a Yuan Markowitzovu stratégiu mimo vzorky, podpriestorovú Markowitzovu stratégiu mimo vzorky a naivnú stratégiu. K podpriestorovej Markowitzovej stratégii mimo vzorky sme priložili 2 nami navrhnuté rozšírenia z podkapitoly 4.2.

Klasicky používame na výber optimálneho počtu hlavných komponentov metódu od Bai a Ng (podkapitola 3.1). Bai a Ng dokázali, že pre ich metódu platí $P(\widehat{d} = K) \to 1$ pre nA, $T \to \infty$. Chen a Yuan ukázali, že vďaka takému výberu počtu hlavných komponentov sa Sharpov pomer pre podpriestorovú Markowitzovu stratégiu mimo vzorky bude asymptoticky rovnať hodnote Sharpovho pomeru pravej Markowitzovej stratégie. Ako alternatívu k výberu optimálneho počtu hlavných komponentov používame jednoduchú metódu, ktorá sa zameriava na množstvo obsiahnutého celkového rozptylu dát pomocou prvých d hlavných komponentov. V diplomovej práci nazývame túto metódu výber počtu hlavných komponentov pomocou percentilov, pretože sa v nej určuje aký percentil celkového rozptylu dát má byť zachovaný. Asymptoticky by mala táto zmena výberu počtu hlavných komponentov spôsobiť horšie výsledky, ale chceme porovnať výsledky na reálnych dátach, keď nA a T nerastú do ∞.

Pre 25 a 100 portfólií používame trénovacie okno T= 60, 120 a 240. Pre 341 S&P firiem meníme veľkosť najväčšieho trénovacieho okna na T = 200, lebo dáta nie sú dostatočne dlhé pre túto sadu (n=241 mesačných výnosov).

Sharpov pomer sme rátali pomocou (37), lebo rátame výkonnosti stratégií mimo vzorky. Premennú γ sme nevolili v ďalších častiach vždy tú istú a preto ju špecifikujeme v konkrétnych vyhotoveniach.

5.5.3 Výsledky pre rôzne voľby d

V tejto celej časti volíme koeficient rizikovej averzie konštantný, rovný 3. Zdôrazňujeme, že konštantná γ nemení hodnotu Sharpovho pomeru, ale ovplyvňuje obrat aktív v stratégii (*turnover*).

Na obr. 6 je čiernou krivkou zobrazená podpriestorová Markowitzova stratégia mimo vzorky pre rôzne fixné voľby počtu hlavných komponentov d. Podobrázky v každom riadku znázorňujú výsledky pre inú sadu aktív. Poznatky z výsledkov sa medzi sadami líšia. Na podobrázkoch v prvom riadku má Markowitzova stratégia mimo vzorky jednu z najlepších hodnôt Sharpovho pomeru oproti podpriestorovým Markowitzovym stratégiám. Použitie podpriestorovej Markowitzovej stratégie mimo vzorky nezlepšilo výsledky ani pre väčšinu fixných voľieb počtu hlavných komponentov a ani pre optimálny počet hlavných komponentov. Z podobrázkov sa zdá, že zvyšovanie dimenzie prípustnej množiny v úlohe optimálneho portfólia systematicky zlepšuje výsledky, a preto sa ohraničovať prípustnú množinu neoplatí.

V podobrázkoch v druhom riadku obr. 6 sú relatívne výkonnosti pre podpriestorovú Markowitzovu stratégiu mimo vzorky lepšie ako pre podobrázky v prvom riadku. Síce na podobrázku s veľkosťou trénovacieho okna $T = 60$ má prevažne lepšiu výkonnosť Markowitzova stratégia mimo vzorky oproti podpriestorovým aplikáciám stratégie, na podobrázkoch s veľkosťami trénovacích okien $T = 120$ a 240 dosahujú podpriestorové Markowitzove stratégie mimo vzorky lepšie výkonnosti ako Markowitzova stratégia mimo vzorky pre veľké rozsahy voľby fixného počtu hlavných komponentov a aj pre optimálny počet hlavných komponentov. Vidíme, že pre podpriestorvú Markowitzovu stratégiu mimo vzorky nevyšiel konkrétny najlepší fixný počet hlavných komponentov, od ktorého by hodnoty Sharpovho pomeru pre stratégie s viac hlavnými komponentami boli horšie. Tiež Sharpov pomer pre podpriestorovú stratégiu s optimálnym počtom hlavných komponentov nevychádza lepšie ako hodnoty Sharpovho pomeru pre veľa fixných voľieb počtu hlavných komponentov. Z toho sa zdá, že na trhu nie je len niekoľko faktorov, ktoré by ovplyvňovali výnosy všetkých aktív na trhu, alebo že tieto faktory nevieme nájsť pomocou analýzy hlavných komponentov. Napriek tomu, že nie je jeden výrazne najlepší počet hlavných komponentov pre podpriestorovú Markowitzovu stratégiu mimo vzorky tak ako pri simulovaných výnosoch, vidíme že sa dá získať

nadvýkonnosť oproti Markowitzovej stratégii mimo vzorky pre veľa možností počtu hlavných komponentov.

V podobrázkoch v treťom riadku obr. 6 vidíme veľmi podobné výsledky ako v podobrázkoch v druhom riadku. V podobrázku s veľkosťou trénovacieho okna $T = 60$ je opäť výkonnosť Markowitzovej stratégie mimo vzorky vo väčšine prípadoch lepšia ako výkonnosti podpriestorových stratégií mimo vzorky. Pre $T = 120$ a 200 sú znova prevažne výkonnejšie podpriestorové Markowitzove stratégie mimo vzorky oproti Markowitzovej stratégii mimo vzorky. Napriek tomu vidíme malé oslabenie podpriestorových Markowitzovych stratégií mimo vzorky pre $T = 200$, keď pre viaceré fixné voľby počtu hlavných komponentov mali aj horšiu výkonnosť ako Markowitzova stratégia mimo vzorky.

Upozorňujeme, že v podobrázkoch v prvom riadku a v druhých 2 podobrázkoch druhého riadku obr. 6 sa posledná hodnota Sharpovho pomeru rovná hodnote Sharpovho pomeru Markowitzovej stratégie mimo vzorky. V ostatných podobrázkoch to nenastáva, pretože množstvo hlavných komponentov je obmedzené tým, že veľkosť trénovacieho okna T je menšia ako počet aktív nA. Keďže hlavné komponenty sú medzi sebou nezávislé bázové vektory a z kovariančnej matice rozmeru $T \times nA$ vieme dostať maximálne $min(T, nA)$ nezávislých bázových vektorov, tak sme aj s použitím počtu všetkých hlavných komponentov nedokázali získať celú prípustnú množinu $\mathbb{R}^{nA}$.

Nami vypracované výsledky pre podpriestorovú Markowitzovu stratégiu mimo vzorky vyzerajú podobne ako na obr. 6 od Chena a Yuana [6, str. 21], ale vidíme aj výraznejšie rozdiely. Vizuálnym porovnaním sa výsledky líšia hlavne pre trénovacie okná veľkosti $T = 60$. Okrem toho, na výsledkoch pre 100 portfólií u Chena a Yuana pozorujeme, že hodnoty Sharpovho pomeru dosiahli najlepšie hodnoty v rozmedzí $d = 10$ až 20 a potom sa zdá, že systematicky klesali, pričom na našich výsledkoch daný systematický pokles nevidíme. Upozorňujeme, že pri porovnávaní grafov si treba dať pozor, lebo naše posledné 2 podobrázky v druhom riadku majú na x-ovej osi počty hlavných komponentov do hodnoty 100 (v daných testovaniach maximálny počet hlavných komponentov), pričom Chen a Yuan uvádzajú počty hlavných komponentov len do hodnoty 60.

Poznatky nadobudnuté z výsledkov z obr. 6 sú, že podpriestorová Markowitzova stratégia mimo vzorky má pre niektoré nastavenia lepšie výsledky a pre iné nastavenia

horšie výsledky ako Markowitzova stratégia mimo vzorky. Podpriestorová Markowitzova stratégia mimo vzorky vyzerá mať horšiu výkonnosť oproti Markowitzovej stratégii mimo vzorky hlavne pre malý počet aktív nA a pre malú veľkosť trénovacieho okna T.

Na obr. 7 ukazujeme pre všetky sady a pre T=120 rozdiely medzi podpriestorovou Markowitzovou stratégiou mimo vzorky a našimi dvomi navrhnutými rozšíreniami stratégie. Iné voľby trénovacích okien vynechávame, lebo výsledky sú veľmi podobné.

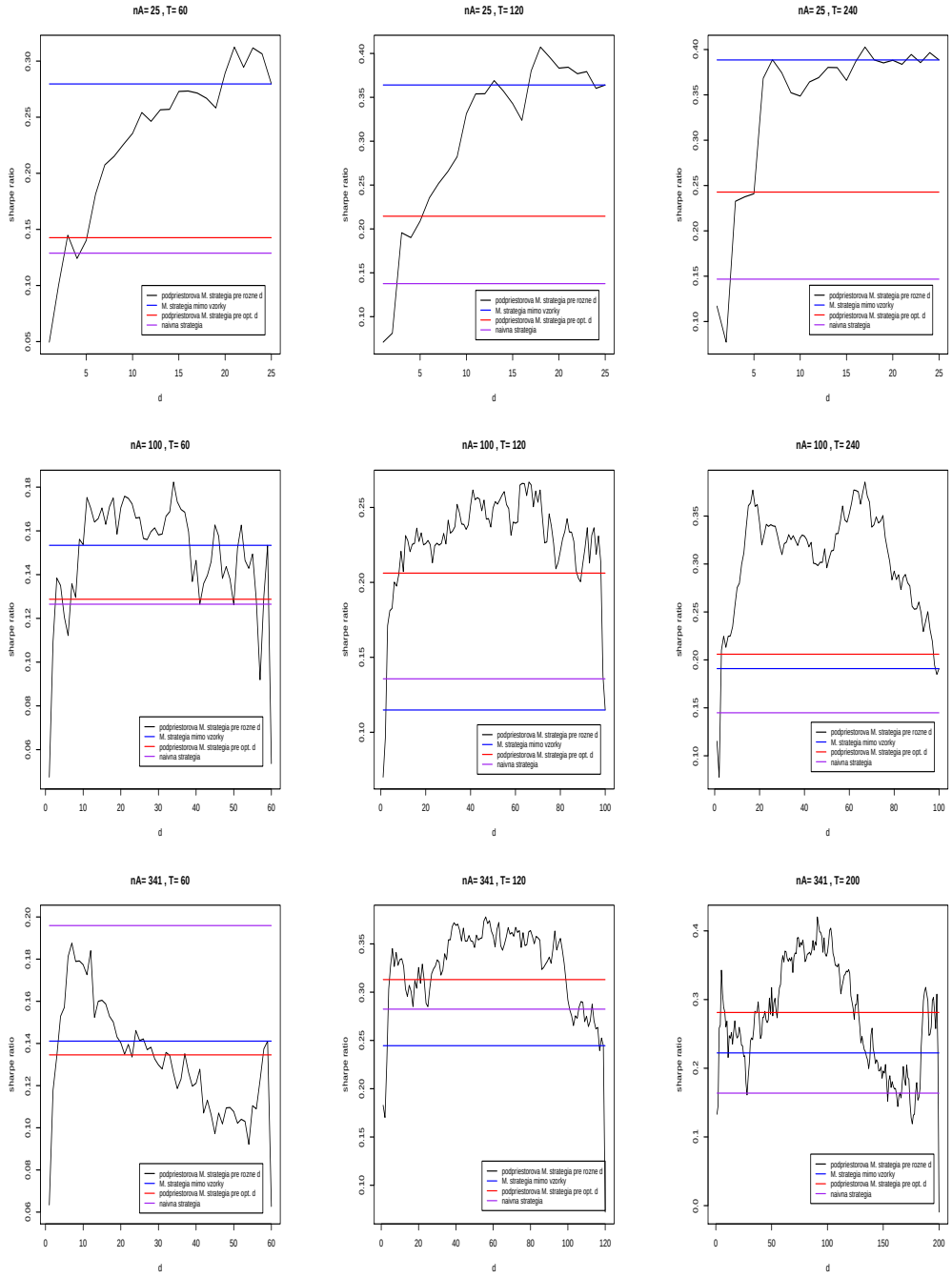
Na obr. 7 vidíme, že pre 25 a 100 portfólií ani jedno naše rozšírenie nedokázalo poraziť podpriestorovú Markowitzovu stratégiu mimo vzorky pre rôzne fixné počty hlavných komponentov d. Rozšírenie 2 malo konzistentne pre všetky fixné počty hlavných komponentov okrem niekoľkých výnimiek o trochu horšiu výkonnosť. Rozšírenie 1 malo lepšiu výkonnosť len pre d=1 a d=2. S narastajúcim d sa výkonnosť rozšírenia 1 voči výkonnosti podpriestorovej Markowitzovej stratégii mimo vzorky znižovala, čo znamená že ohraničením prípustnej množiny optimalizačnej úlohy do vlastných vektorov korelačnej matice, namiesto kovariančnej matice výnosov, sme viac dôležitej informácie pre stratégiu stratili ako získali.

Na obr. 7 vidíme, že pre 341 S&P firiem vychádzajú iné výsledky ako pre 25 a 100 portfólií. Všetky aplikácie podpriestorovej Markowitzovej stratégie mimo vzorky sú veľmi podobné, kde každá z nich má isté obdobie lepšie výsledky ako ostatné stratégie a iné obdobie horšie výsledky ako ostatné stratégie. Toto je pre nás dôležitý poznatok, lebo sme zistili, že keď sme dobre diverzifikované portfóliá nahradili reálnymi firmami, tak sa relatívne výkonnosti našich rozšírení oproti podpriestorovej Markowitzovej stratégii mimo vzorky zlepšili a už nie je ľahké určiť, ktorá aplikácia je najlepšia. Preto by naše rozšírenia mohli byť prínosom v prípade, že robíme so sadou, ktorej aktíva majú extrémne individuálne chyby.

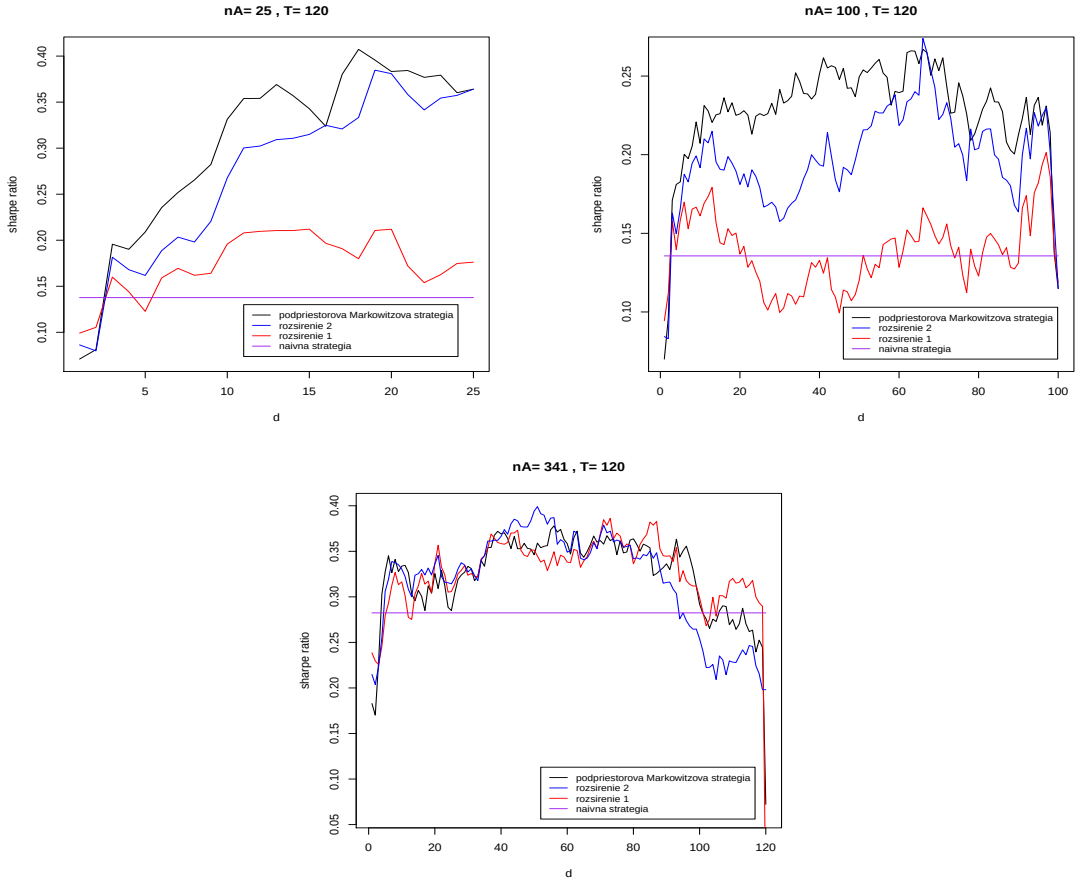
5.5.4 Výsledky pre optimálne d

Podkapitolu 5.5.4 rozdeľujeme na viac častí. V úvodnej časti úvádzame rozdiely pre voľbu počtu hlavných komponentov. V prvom vyhodnotení ukazujeme výsledky základnej optimalizačnej úlohy. Následne upravujeme optimalizačnú úlohu tak, aby obraty aktív stratégií boli uskutočniteľné v reálnom svete.

Na výpočet optimálneho počtu hlavných komponentov používame vzťah od Bai a



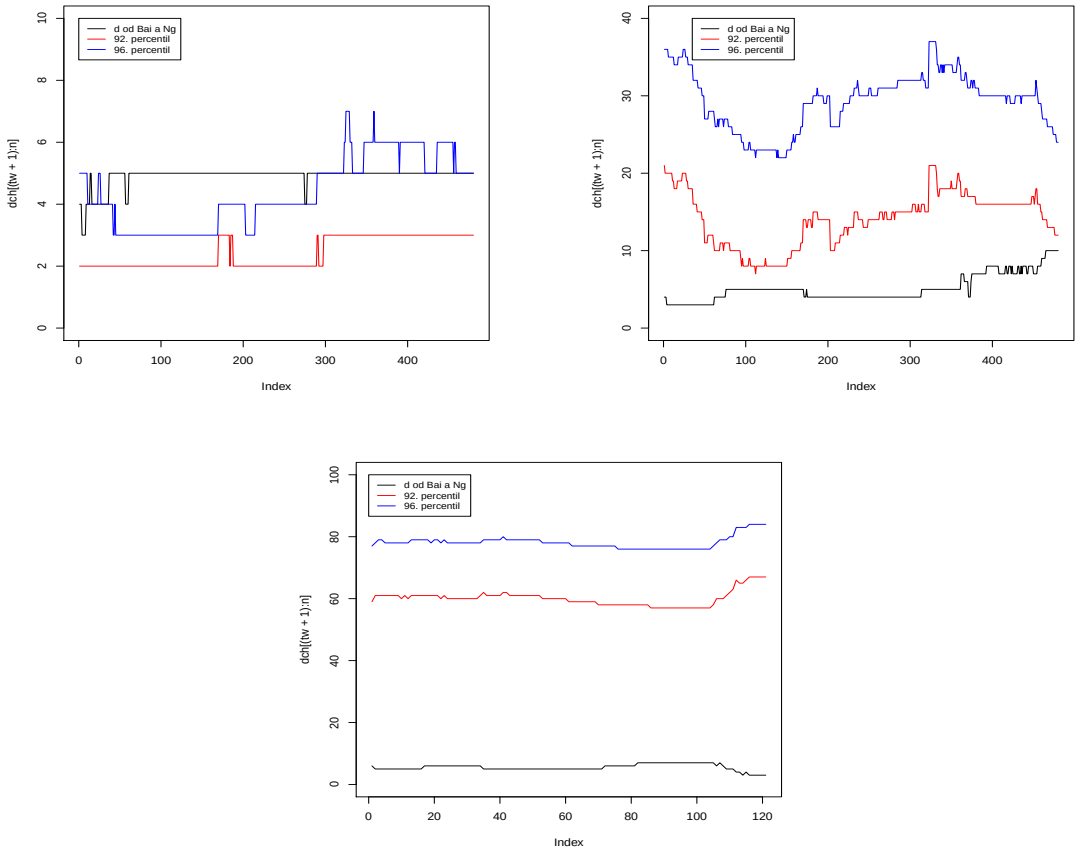
Obr. 6: Sharpov pomer pre podpriestorové Markowitzove stratégie mimo vzorky pre všetky možné fixné *d*. Podobrázky v riadkoch patria k tým istým sadám s inou voľbou trénovacieho okna. Pre porovnanie pridané horizontálne čiary označujúce Sharpov pomer pre Markowitzovu stratégiu mimo vzorky, podpriestorovú Markowitzovu stratégiu mimo vzorky s optimálnym počtom hlavných komponentov a naivnú stratégiu. Celé testovacie obdobie je 1/1961 až 12/2010.



Obr. 7: Sharpov pomer pre podpriestorové Markowitzove stratégie mimo vzorky a pre nami navrhnuté rozšírenia stratégie pre všetky možné fixné d. Pre porovnanie pridaná horizontálna čiara označujúca Sharpov pomer pre naivnú stratégiu. Celé testovacie obdobie je 1/1961 až 12/2010.

Ng (20). Tento optimálny počet hlavných komponentov volíme aj pre podpriestorovú Markowitzovu stratégiu mimo vzorky a aj pre naše dve rozšírenia stratégie. Okrem toho ukazujeme výsledky pre podpriestorovú Markowitzovu stratégiu mimo vzorky s metódou voľby počtu hlavných komponentov pomocou percentilov namiesto optimálneho počtu hlavných komponentov.

Na obr. 8 vidíme, že optimálny počet hlavných komponentov d je bez ohľadu na počet použitých aktív v priemere okolo hodnoty 6. Takýto stav je celkom žiadaný, pretože všetky 3 sady sú dobrým odhadom pre americký trh. V prípade, že na americkom trhu existuje niekoľko faktorov (dobre zachytiteľných hlavnými komponentmi nadmerných výnosov), tak by ich malo výjsť približne tak isto veľa aj pre väčší a aj pre menší počet



Obr. 8: Porovnanie hodnoty d v mesiacoch $t = T + 1$ až n. T sme zvolili 120. Čiernou krivkou zobrazené d od Bai a Ng. Červenou a modrou krivkou sú zobrazené hodnoty d, ktoré sú volené tak, aby d hlavných komponentov zachovávalo aspoň 92. alebo 96. percentil celého rozptylu dát. Celé testovacie obdobie je 1/1961 až 12/2010.

aktív z trhu. Špeciálne 2 sady: 25 portfólií a 100 portfólií označujú tie isté aktíva, len rozdelené do iného počtu portfólií. Pri metóde voľby počtu hlavných komponentov d pomocou percentilov vidíme, že pre 25 portfólií určovala metóda menšie hodnoty d a naopak pre 100 portfólií a 341 S&P firiem určovala metóda väčšie hodnoty d oproti optimálnemu počtu hlavných komponentov. V priebehu testovacieho obdobia je metóda voľby počtu hlavných komponentov pomocou percentilov rozkolísanejšia ako metóda pre voľbu optimálneho počtu hlavných komponentov.

Výsledky pre konštantnú γ

Premennú γ sme zvolili rovnú 3.

V tabuľke 3 vidíme, že pre 25 portfólií má najlepšie výsledky Markowitzova stratégia

nA	T	M. strat. mimo vzorky	naivná strat.	podpriest. M. strat. mimo vzorky	rozšírenie 1	rozšírenie 2	metóda voľby $d_{92.\ pct}$	metóda voľby $d_{96.\ pct}$
25	60	0.28	0.13	0.14	0.11	0.09	0.15	0.16
25	120	0.36	0.14	0.21	0.13	0.17	0.15	0.20
25	240	0.39	0.15	0.24	0.18	0.19	0.13	0.24
100	60	0.15	0.13	0.13	0.10	0.10	0.18	0.18
100	120	0.11	0.14	0.21	0.14	0.19	0.23	0.25
100	240	0.19	0.14	0.21	0.19	0.20	0.33	0.33
341	60	0.14	0.20	0.13	0.12	0.12	0.11	0.11
341	120	0.24	0.28	0.31	0.28	0.31	0.38	0.36
341	200	0.22	0.16	0.28	0.31	0.25	0.35	0.34

Tabuľka 3: Sharpov pomer pre stratégie. Stratégie mimo vzorky sú vytvorené na mesačných nadmerných výnosoch s rolovacím trénovacím oknom s veľkosťou T a testovacím oknom veľkosti 1. Celé testovacie obdobie je 1/1961 až 12/2010.

mimo vzorky a ostatné vyhotovenia dosahujú približne polovičný Sharpov pomer oproti nej. Pre 100 portfólií a 341 S&P firiem sú pre $T = 60$ výsledky pre všetky vyhotovenia veľmi podobné medzi sebou v rámci rovnakej sady. Pre trénovacie okná $T = 120$ a 240 (resp. 200) nastala zmena a podpriestorové stratégie prekonávali Markowitzovu stratégiu mimo vzorky.

K tomu, podobne ako v obr. 7, aj v tab. 3 vidíme, že naše rozšírenia 1 a 2 dosahujú pre optimálny počet hlavných komponentov nižšiu výkonnosť ako podpriestorová Markowitzova stratégia mimo vzorky.

Keď voľbu počtu hlavných komponentov robíme pomocou percentilov, tak veľa krát dosahujeme lepšiu výkonnosť ako pri voľbe optimálneho počtu hlavných komponentov. Táto nadvýkonnosť pravdepodobne súvisí s tým, že voľby počtu hlavných komponentov $d_{92.\ pct}$ a $d_{96.\ pct}$ majú na testovacom období v priemere násobne väčšie d ako priemerné optimálne $d_{Bai\ a\ Ng}$, čo je vidno na obr. 8 a v tab. 4.

V tab. 4 vidíme, že optimálny počet hlavných komponentov d je v priemere väčší ako priemerné d pomocou percentilov jedine pre 25 portfólií. V sadách 100 portfólií a 341 S&P firiem sú násobne väčšie priemerné d pomocou percentilov ako optimálny počet hlavných komponentov. Sharpov pomer pre optimálny počet hlavných komponentov v tab. 3 je lepší jedine vo vyhotoveniach pre 25 portfólií a naopak pre ostatné sady

nA	T	optimálne $d_{Bai\ a\ Ng}$	metóda voľby $d_{92.\ pct}$	metóda voľby $d_{96.\ pct}$
25	60	4.7	2.5	4.3
25	120	4.9	2.4	4.4
25	240	5.0	2.5	4.6
100	60	4.6	11.1	21.2
100	120	5.1	14.1	29.3
100	240	4.9	17.3	37.7
341	60	3.3	36.8	45.3
341	120	5.5	60.1	78.2
341	200	6.2	85.0	112.9

Tabuľka 4: Priemerné d z výpočtov na testovacom období. Celé testovacie obdobie je 1/1961 až 12/2010.

má horšie výsledky ako výsledky pre voľbu počtu hlavných komponentov pomocou percentilov. Z toho nám vyplýva, že nadvýkonnosť by mohla súvisieť s voľbou väčšieho počtu hlavných komponentov. Tiež ale vidíme, že pre sady 100 portfólií a 341 S&P firiem vychádzajú pre obidve voľby počtu hlavných komponentov pomocou percentilov skoro rovnaké hodnoty Sharpovho pomeru napriek tomu, že $d_{96.\ pct}$ má priemerné d o desiatky väčšie ako $d_{96.\ pct}$.

K predošlým výsledkom považujeme za užitočné pridať informácie o obrate aktív, ktoré dosahujú pri vyhotoveniach extrémne hodnoty. V nasledujúcej tab. 5 uvádzame obraty aktív len pre Markowitzovu stratégiu mimo vzorky a podpriestorovú Markowitzovu stratégiu mimo vzorky. Vyhotovenia nie sú porovnateľné z hľadiska obratu aktív, lebo systematicky majú veľmi odlišnú sumu váh napriek tomu, že parameter pre rizikovú averziu investora γ=3 je rovnaký. Suma váh aj pre zobrazené 2 stratégie je odlišná, kde Markowitzova stratégia mimo vzorky má v priemere sumu váh nad 100% a podpriestorová Markowitzova stratégia má sumu váh v priemere okolo 0.5%.

Z tab. 5 vyplýva, že obraty aktív pre Markowitzovu stratégiu mimo vzorky sú nerealisticky vysoké. V realite by sa pravdepodobne nedalo nakupovať také veľké množstvá aktív z našich sád, aké vyšli z riešení optimalizačných úloh. Ak by bolo možné robiť také veľké nákupy, tak jedine s extrémnymi príplatkami za kupovanie veľkých krátkych pozícií a tiež s veľkými stratami z rozdielu nákupnej a predajnej ceny aktíva. K obra-

nA	T	M. strat. mimo vzorky	naivná strat.	podpriest. M. strat. mimo vzorky
25	60	5230%	0%	310%
25	120	1180%	0%	150%
25	240	460%	0%	60%
100	60	18470%	0%	180%
100	120	67250%	0%	90%
100	240	2730%	0%	50%
341	60	990%	0%	60%
341	120	2040%	0%	50%
341	200	5260%	0%	50%

Tabuľka 5: Priemer mesačných obratov aktív pre $\gamma = 3$. Celé testovacie obdobie je 1/1961 až 12/2010.

tom aktív podpriestorovej Markowitzovej stratégie sa v tejto časti nevyjadrujeme, lebo považujeme za nerealistické, že suma rizikových aktív vychádza v priemere len 0.5%.

Výsledky pre $\sum_{i=1}^{nA} w_{ti} = \pm 1$

Pre problém s vysokými obratmi aktív, a tiež aby boli obraty aktív medzi stratégiami porovnateľné, meníme váhy stratégií v každom mesiaci t. Najprv rátame optimálne váhy pre stratégie a potom ich preškálujeme tak, aby $\sum_{i=1}^{nA} w_{ti} = \pm 1$ podľa toho, či je suma váh kladná alebo záporná pred škálovaním. To deformuje premennú rizikovej averzie γ , ktorá pri optimalizačnej úlohe na nadmerných výnosoch ovplyvňuje sumu váh rizikových aktív (viď podkapitola 1.3). Preto riziková averzia investora už len pôsobí na určenie, či sumu váh rizikových aktív držíme dokopy s váhou $+1$ alebo -1.

V tab. 6 vidíme, že úpravou sumy váh na ± 1 nedosahujeme reálnosť priemerných mesačných obratov aktív a v niektorých prípadoch sa práve naopak obraty aktív o veľa navyšujú. Problém vysokého obratu aktív v stratégiách je kvôli vysokým krátkym a dlhým pozíciám konkrétnych aktív v rámci tých istých mesiacov. Pri porovnávaní podpriestorovej Markowitzovej stratégie a našich rozšírení vidíme, že pre 25 portfólií majú naše rozšírenia prevažne nižší priemerný mesačný obrat aktív. Pre 100 portfólií a 341 S&P firiem majú naopak naše rozšírenia väčšinou vyšší priemerný mesačný obrat aktív. Pre 25 a 100 portfólií vidíme, že podpriestorvá Markowitzova stratégia mimo

nA	T	M. strat. mimo vzorky	naivná strat.	podpriest. M. strat. mimo vzorky	rozšírenie 1	rozšírenie 2	metóda voľby $d_{92.\ pct}$	metóda voľby $d_{96.\ pct}$
25	60	18630%	0%	7600%	1350%	3950%	4310%	3860%
25	120	2890%	0%	820%	590%	470%	1040%	330%
25	240	150%	0%	50%	90%	80%	2320%	50%
100	60	52880%	0%	2640%	980%	2910%	15040%	13430%
100	120	92890%	0%	940%	5360%	1030%	510%	570%
100	240	1780%	0%	50%	80%	60%	140%	240%
341	60	280%	0%	60%	110%	6010%	230%	250%
341	120	290%	0%	20%	50%	50%	170%	230%
341	200	380%	0%	20%	50%	50%	160%	200%

Tabuľka 6: Priemer mesačných obratov aktív pri škálovaní váh na $\sum_{i=1}^{nA} w_{ti} = \pm 1$. Celé testovacie obdobie je 1/1961 až 12/2010.

vzorky s voľbou počtu hlavných komponentov pomocou percentilov má aj vyššie a aj nižšie obraty aktív v porovnaní s voľbou optimálneho počtu hlavných komponentov. Pre 341 S&P firiem majú vyhotovenia pomocou percentilov systematicky vyšší obrat aktív ako obrat aktív optimálneho počtu hlavných komponentov.

Keďže ani úpravou sumy váh na ± 1 sme nedosiahli reálnosť mesačných obratov aktív stratégií, tak k vyhotoveniu neprikladáme ani výkonnosti Sharpovho pomeru, ale pokračujeme v nasledujúcom vyhotovení.

Výsledky pre ohraničenie váh pre každé aktívum

V tejto časti meníme váhy pre každé aktívum po vyrátaní optimálnych váh. Pred úpravou každej jednej váhy potrebujeme, aby boli váhy medzi stratégiami približne v podobných veľkostiach. Totižto ak by jedna stratégia volila extrémne vysoké váhy, tak by naše ohraničenia robili veľké zmeny pre stratégiu a ak by stratégia volila extrémne nízke váhy, tak by ju naše ohraničenia ani neovplyvnili. Preto najprv škálujeme sumu váh na $\sum_{i=1}^{nA} w_{ti} = \pm 1$. Následne váhy upravíme tak, aby váha pre každé aktívum mala veľkosť minimálne spodnej hranice pre váhu a maximálne vrchnej hranice pre váhu. Formálne sa váhy zmenia na:

$$\widehat{w}_{ti}^{new} = \min(\max(b_l, \widehat{w}_{ti}), b_u),$$
$$\text{pre } i = 1, 2, ...nA,\ t = T+1, T+2, ...n,$$

kde

b_l - spodná hranica pre váhu každého aktíva, volíme $b_l = -5\%$,

b_u - vrchná hranica pre váhu každého aktíva, volíme $b_u = 5\%$.

Tiež tu nastáva deformácia významu premennej rizikovej averzie γ tak, ako v predošlej časti. Podobný spôsob ohraničenia váh po vypočítaní optimálnych váh používajú v literatúre napríklad [8, tab. 7, str. 67], kde ohraničujú investovanie jediného aktíva medzi 0% a 150%. Dangl a Halling netvoria optimalizačnú úlohu pomocou Markowitzovej teórie portfólia, ale pomocou prediktívneho regresného modelu na základe viacerých vysvetľujúcich premenných.

nA	T	M. strat. mimo vzorky	naivná strat.	podpriest. M. strat. mimo vzorky	rozšírenie 1	rozšírenie 2	metóda voľby $d_{92.\ pct}$	metóda voľby $d_{96.\ pct}$
25	60	34%	0%	29%	31%	29%	24%	25%
25	120	20%	0%	17%	19%	17%	11%	14%
25	240	11%	0%	7%	11%	8%	5%	9%
100	60	133%	0%	52%	58%	56%	89%	113%
100	120	86%	0%	27%	35%	32%	54%	77%
100	240	35%	0%	17%	24%	20%	40%	50%
341	60	207%	0%	44%	66%	71%	187%	202%
341	120	199%	0%	23%	51%	48%	154%	204%
341	200	200%	0%	21%	53%	46%	149%	180%

Tabuľka 7: Priemer mesačných obratov aktív pri ohraničení váh všetkých aktív pomocou spodnej a vrchnej hranice. Celé testovacie obdobie je 1/1961 až 12/2010.

V tab. 7 vidíme, že po ohraničení váh pre každé aktívum získavame konečne v reálnom svete uskutočniteľné priemerné mesačné obraty aktív pre stratégie. Pre 25 portfólií vychádzajú veľmi podobné priemerné mesačné obraty aktív pre všetky stratégie. Pre 100 portfólií a 341 S&P firiem dosahuje Markowitzova stratégia mimo vzorky a podpriestorové Markowitzove stratégie mimo vzorky s voľbou počtu hlavných komponentov pomocou percentilov násobne väčšie obraty aktív ako podpriestorová Markowiztova stratégia mimo vzorky s optimálnym počtom hlavných komponentov. Naše 2 rozšírenia majú podobné obraty aktív ako podpriestorová Markowitzova stratégia mimo vzorky s optimálnym počtom hlavných komponentov. Sady s viac aktívami majú systematicky väčší priemerný obrat aktív ako sady s menej aktívami. To má logické zdôvodnenie,

nA	T	M. strat. mimo vzorky	naivná strat.	podpriest. M. strat. mimo vzorky	rozšírenie 1	rozšírenie 2	metóda voľby $d_{92.\ pct}$	metóda voľby $d_{96.\ pct}$
25	60	0.07	0.13	0.12	0.08	0.13	0.12	0.13
25	120	0.14	0.14	0.17	0.14	0.16	0.11	0.21
25	240	0.23	0.15	0.24	0.16	0.24	0.11	0.24
100	60	-0.01	0.13	0.14	0.11	0.10	0.2	0.22
100	120	-0.01	0.14	0.21	0.17	0.21	0.24	0.17
100	240	0.16	0.14	0.25	0.21	0.26	0.25	0.32
341	60	0.08	0.20	0.12	0.12	0.14	0.09	0.07
341	120	0.29	0.28	0.35	0.27	0.32	0.39	0.4
341	200	0.3	0.16	0.27	0.31	0.24	0.34	0.38

Tabuľka 8: Sharpov pomer pre stratégie pri ohraničení váh všetkých aktív pomocou spodnej a vrchnej hranice. Stratégie mimo vzorky sú vytvorené na mesačných nadmerných výnosoch s rolovacím trénovacím oknom s veľkosťou T a testovacím oknom veľkosti 1. Celé testovacie obdobie je 1/1961 až 12/2010.

pretože obrat aktív sa ráta pre celú stratégiu a nie je spriemerovaný pre jedno aktívum zo stratégie. Pre sady s viac aktívami by sme mohli znížiť priemerný mesačný obrat aktív navýšením spodnej hranice a ponížením vrchnej hranice pre váhy.

V tab. 8 sú výkonnosti Sharpovho pomeru stratégií. Markowitzova stratégia mimo vzorky má pre 4 nastavenia lepšie výsledky a pre 4 nastavenia horšie výsledky ako naivná stratégia. Podpriestorová Markowitzova stratégia mimo vzorky má lepšie výsledky v 8 prípadoch z 9 prípadov oproti Markowitzovej stratégii mimo vzorky a v 7 prípadoch z 9 má lepšie výsledky ako naivná stratégia. Rozšírenie 1 má systmaticky o niekoľko desatinných miest menší Sharpov pomer oproti podpriestorovej Markowitzovej stratégii mimo vzorky a rozšírenie 2 má aj lepšie hodnoty, rovnaké, a aj horšie hodnoty Sharpovho pomeru ako podpriestorová Markowitzova stratégia mimo vzorky. Voľba počtu hlavných komponentov pomocou percentilov väčšinou zlepšuje Sharpov pomer pre podpriestorovú Markowitzovu stratégiu.

Záver

V diplomovej práci sme riešili problém zlej výkonnosti odhadnutého Markowitzovho optimálneho portfólia pri veľkom počte aktív a pre momenty nadmerných výnosov odhadnuté na relatívne malom počte historických údajov. V teoretickej časti sme najprv odvodili Markowitzovu teóriu portfólia pre nadmerné výnosy aktív. Potom sme v teoretických kapitolách vysvetlovali predpoklady a teóriu pre podpriestorovú Markowitzovu teóriu portfólia od Chena a Yuana [6]. V nej riešia chybu odhadu momentov nadmerných výnosov pomocou ohraničenia prípustnej množiny optimalizačnej úlohy na hlavné komponenty nadmerných výnosov všetkých aktív. V práci sme navrhli 2 rozšírenia výpočtu podpriestorového optimálneho portfólia. V prvom rozšírení tvoríme podpriestor pre optimalizačnú úlohu pomocou korelačnej matice namiesto kovariančnej matice. V druhom spôsobe sme sa zamerali na vzorec pre podpriestorové optimálne portfólio a navrhli sme preňho jednu technickú zmenu.

Hlavnými praktickými podkapitolami 5.3, 5.4 a 5.5 sme výpočtami overovali teoretické informácie z diplomovej práce. Pre lepšie pochopenie problematiky a zvýšenie dôveryhodnosti našich výsledkov sme sa rozhodli vo všetkých vyhodnoteniach používať skoro také isté nastavenia ako v prácach [9] a [6]. Našu prácu sme navyše rozšírili o novú sadu dát 341 S&P firiem (viď podkapitola 5.5.1). Nami vypracované výsledky sú veľmi podobné s výsledkami dvoch vyššie referovaných prác. Napriek tomu vznikajú malé rozdiely, ktoré sú spôsobené buď miernou odlišnosťou v postupoch, alebo miernou odlišnosťou použitých dát.

V podkapitole 5.3 sme na základe tab. 1 dospeli k záveru, že naozaj platí, že pre sady s väčším počtom aktív dosahuje odhadnuté Markowitzovo optimálne portfólio horšie výsledky ako pre sady s menším počtom aktív pri takej istej veľkosti trénovacieho okna. Hodnotenie odhadnutého optimálneho portfólia sme robili relatívne voči Markowiztovmu optimálnemu portfóliu vyrátanému v rámci vzorky. Navyše nám vyšlo, že naivné portfólio má pre všetky sady lepšiu výkonnosť ako odhadnuté Markowitzovo optimálne portfólio. To je dôležitá informácia pre portfólio manažéra, lebo vidí, že jednoducho vytvorené portfólio fungovalo na historických údajoch lepšie ako zložitá stratégia, ktorej snahou bolo optimalizovať výkonnosť.

V podkapitole 5.4 máme na simulovaných dátach vytvorené grafy (obr. 5), ktoré

poukazujú na hlavnú myšlienku podpriestorovej Markowitzovej teórie portfólia a to na výhodu použitia len niekoľkých hlavných komponentov nadmerných výnosov aktív namiesto výnosov všetkých aktív, keď nadmerné výnosy aktív ovplyvňuje len niekoľko faktorov na trhu. Na simulovaných dátach nám vyšlo, že zvyšovanie počtu hlavných komponentov nad počet pravého počtu faktorov zhoršovalo výsledky odhadnutého podpriestorového optimálneho portfólia.

V podkapitole 5.5.3 uvádzame výsledky pre odhadnuté podpriestorové optimálne portfólio pre rôzne fixné voľby počtu hlavných komponentov. Narozdiel od simulovaných dát, na reálnych dátach nám nevyšiel pre odhadnuté podpriestorové optimálne portfólio konkrétny najlepší počet hlavných komponentov, od ktorého by hodnoty Sharpovho pomeru pre odhadnuté podpriestorové optimálne portfólio s viac hlavnými komponentami boli systematicky horšie. Napriek tomu nám vyšlo, že pre veľa volieb počtu hlavných komponentov má odhadnuté podpriestorové optimálne portfólio lepšie výsledky ako odhadnuté Markowitzovo optimálne portfólio a to hlavne pre sadu 100 Fama-French portfólií a sadu 341 S&P firiem.

V podkapitole 5.5.4 uvádzame výsledky odhadnutého podpriestorového optimálneho portfólia pre optimálny počet hlavných komponentov, prípadne pre počet hlavných komponentov, vyrátaný pomocou percentilov. V rámci podkapitoly sme vytvorili viac častí, v ktorých ukazujeme výsledky aj pre optimálne váhy zo základných optimalizačných úloh a aj pre tak upravené váhy, aby boli mesačné obraty aktív uskutočniteľné v reálnom svete a aby neboli príliš veľké. Totižto pre základnú úlohu odhadnutého Markowitzovho optimálneho portfólia vyšli priemerné mesačné obraty aktív približne v tisícoch percent. Aj pre základné optimalizačné úlohy nám vyšli dobré výsledky pre odhadnuté podpriestorové optimálne portfólio, ale bližšie interpretujeme len výsledky upravených váh, lebo je to užitočnejšia informácia pre portfólio manažéra. Pre upravené váhy nám z tab. 7 a tab. 8 vyplýva, že odhadnuté podpriestorové optimálne portfólio dosahuje prevažne lepšie výsledky ako odhadnuté Markowitzovo optimálne portfólio a pritom má násobne nižšie priemerné mesačné obraty aktív. Z nižších obratov aktív môže vyplynúť ďalšia nadvýkonnosť, kvôli ušetreniu na transakčných poplatkoch a rozdielu medzi nákupnými a predajnými cenami aktív.

Naše 2 rozšírenia pre výpočet odhadnutého podpriestorového optimálneho portfó-

lia mali horšie, prípadne podobné výsledky ako odhadnuté podpriestorové optimálne portfólio (viď obr. 7, tab. 7 a tab. 8). Znamená to, že operáciami s korelačnou maticou sme nedokázali dosiahnuť lepšie výsledky ako operáciami s kovariančnou maticou nadmerných výnosov. Výsledky nám vyšli iné pre sady Fama-French portfólií a pre 341 S&P firiem. Rozdiel medzi sadami je práve v individuálnych chybách ich aktív. Fama-French portfóliá sú diverzifikované portfóliá skladajúce sa z veľa aktív a preto majú menšie individuálne chyby ako jednotlivé firmy z S&P indexu. Tým, že 341 S&P firiem malo vyššie individuálne chyby, tak pre sadu nevyšla dominancia odhadnutého podpriestorvého optimálneho portfólia oproti našim rozšíreniam, ale pre všetky z nich vychádzali výsledky podobne a nedalo sa určiť, ktorá z nich je systematicky najlepšia.

Voľba počtu hlavných komponentov pomocou percentilov dokázala zlepšiť výkonnosť odhadnutého podpriestorového optimálneho portfólia, ale za cenu vyšších priemerných mesačných obratov aktív, ktoré majú podobnú veľkosť ako obraty aktív odhadnutého Markowitzovho optimálneho portfólia. Na obr. 7 a aj v tab. 3 sme ukázali, že nadvýkonosť podpriestorového optimálneho portfólia s počtom hlavných komponentov pomocou percentilov dosahuje lepšiu výkonnosť prevažne kvôli tomu, že v priemere volí vyššie počty hlavných komponentov. Vo voľbe počtu hlavných komponentov je preto priestor pre skúmanie, aká metóda by bola najlepšia na použitie pre podpriestorovú Markowitzovu teóriu portfólia.

Zoznam použitej literatúry

[1] Bai, J. and Ng, S.: *Determining the number of factors in approximate factor models*, Econometrica 70(1)(2002), 191-221

[2] Best, M. J., and Grauer, R. R.: *On the Sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results*, Review of Financial Studies 4(2)(1991), 315–342

[3] Bjerring, T., Ross, O., and Weissensteiner, A.: *Feature selection for portfolio optimization*, Annals of Operations Research 256(1)(2016), 21-40

[4] Branger, N., Lučivjanská, K., and Weissensteiner, A.: *Optimal granularity for portfolio choice*, Journal of Empirical Finance 50 (2019), 125-146

[5] Campbell, J. Y., Viceira, L. M. : *Long-Horizon Mean-Variance Analysis: A User Guide*, dodatok pre *The Term Structure of the Risk-Return Tradeoff*, 2004, dostupné na internete(04.01.2021): http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.431.1846&rep=rep1&type=pdf

[6] Chen, J., Yuan, M.: *Efficient Portfolio Selection in a Large Market*, Oxford University Press, Oxford, 2016

[7] Chopra, V. K., and Ziemba, W. T.: *The effect of errors in means, variances, and covariances on optimal portfolio choice*, The Journal of Portfolio Management 19(2)(1993), 6–11

[8] Dangl, T., Hailing, M.: *Predictive regressions with time-varying coefficients*, Journal of Financial Economics 106(2012), 157-181

[9] DeMiguel, V., Garlappi, L., Uppal, R.: *Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?*, Oxford University Press, Oxford, 2007

[10] fred.stlouisfed.org, *3-Month Treasury Bill: Secondary Market Rate*, dostupné na internete(04.01.2021): https://fred.stlouisfed.org/series/TB3MS

[11] Huberman, G.:*A simple approach to arbitrage pricing theory*, Journal of Economic Theory 28(1982), 183-191

[12] Huberman, G., Wang, Z.:*Arbitrage Pricing Theory*, Staff Reports 216, Federal Reserve Bank of New York, New York, 2005, dostupné na internete(04.01.2021): https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr216.pdf

[13] Jolliffe, I.T.: *Principal Component Analysis, Second Edition*, Springer, New York, 2002

[14] Jorion, P.: *Bayes-Stein estimation for portfolio analysis*, The Journal of Financial and Quantitative Analysis 21(3)(1986), 279–292

[15] Kašuba, J.: *Optimálne portfólio a problém pri jeho dosahovaní*, diplomová práca, Univerzita Komenského v Bratislave, Fakulta Managementu, Bratislava, 2011, dostupné na internete(04.01.2021):
https://opac.crzp.sk

[16] Kritzman, M., Page, S., and Turkington, D.: *In defense of optimization: The fallacy of 1/ N*, Financial Analysts Journal 66(2)(2010), 31–39

[17] Ledoit, O. and Wolf, M.: *A well-conditioned estimator for large-dimensional covariance matrices*, Journal of Multivariate Analysis 88(2)(2004a), 365-411

[18] Ledoit, O. and Wolf, M.: *Honey, I shrunk the sample covariance matrix: Problems in mean-variance optimization*, The Journal of Portfolio Management 30(2004b), 110-119

[19] Marakbi, Z.: *Mean-Variance Portfolio Optimization: Challenging the role of traditional covariance estimation*, diplomová práca, KTH Industrial Engineering and Management, Industrial Management, Štokholm, 2016, dostupné na internete(04.01.2021):
http://www.diva-portal.org/smash/get/diva2:1060405/FULLTEXT01.pdf

[20] Markowitz, H. M.: *Portfolio selection*, Journal of Finance 7(1952), 77-91

[21] mba.tuck.dartmouth.edu, *U.S. Research Returns Data*, dostupné na internete(04.01.2021): https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[22] Rosa S.: predmet Redukcia dimenzie dát, FMFI UK, Bratislava, 2019

[23] Ross, S. A.: *The arbitrage theory of captial asset pricing*, Journal of Economic Theory 13(1976), 341-360

[24] r-bloggers, *A gentle introduction to finance using r efficient frontier and capm part 1*, dostupné na internete(04.01.2021): https://www.r-bloggers.com/2016/05/a-gentle-introduction-to-finance-using-r-efficient-frontier-and-capm-part-1/

[25] Sharpe, W. F.: *Capital asset prices: A theory of market equilibrium under conditions of risk*, Journal of finance 19(3)(1964), 425-442

[26] Scherer, B.: *A note on the returns from minimum variance investing*, Journal of Empirical Finance 18(4)(2011), 652–660

[27] Spustová, N.: *Viacfaktorové modely výberu portfólia*, bakalárska práca, Univerzita Komenského v Bratislave FSEV, kvantitatívne metódy v ekonómii, Bratislava, 2015, dostupné na internete(04.01.2021): https://opac.crzp.sk

[28] Tu, J., Zhou, G.: *Markowitz meets Talmud: a combination of sophisticated and naive diversification strategies*, Journal of Financial Economics 99 (2011), 204-215

[29] Wang, J., Zivot, E.: *Modeling Financial Time Series with S-PLUS^R*, Springer, New York, 2002

[30] finance.yahoo.co, *Downloading historical data*, dostupné na internete(04.01.2021): https://finance.yahoo.com/

Príloha 1

		1. sada	2. sada	3. sada
5.3 Markowitzova teória portfólia v rámci a mimo vzorky	názov sady	Priemyselné portfóliá + 1	Faktorové portfóliá	Portfóliá_20 + 1
	počet všetkých mesačných údajov	497	497	497
5.4 Znižovanie dimenzie prípustnej množiny v úlohe optimálneho portfólia - analýza na simulovaných dátach	názov sady	generovanie dát		
	počet všetkých mesačných údajov	1000 simulácií		
5.5 Znižovanie dimenzie prípustnej množiny v úlohe optimálneho portfólia - analýza na reálnych dátach	názov sady	Fama-French 25 portfólií	Fama-French 100 portfólií	341 S&P firiem
	počet všetkých mesačných údajov	600	600	241

Tabuľka 9: Prehľad o použitých dátach.

Príloha 2 - kódy diplomovej práce

Regresia pre určenie individuálnych chýb pre generované aktíva

```
# data z July 1963 to August 2007 #data sme v subore predelovali 100 aby
# boli v desatinnych miestach ako percenta
mydata <- read.csv("C:/Filip/Faklas job/MAIN DEAL/diplomovka/urobene/
regresia pre FF/25_Portfolios_5x5.csv", header=T, sep=",")
attach(mydata)

mydata2 <- read.csv("C:/Filip/Faklas job/MAIN DEAL/diplomovka/urobene/
regresia pre FF/F-F_Research_Data_Factors.csv", header=T, sep=",")
attach(mydata2)

n <- length(mydata[1,])
linearMod <- lm(formula = Mkt ~ 0+as.matrix(mydata[,2:(n-1)]))
modelSummary <- summary(linearMod)
sd(modelSummary$residuals)

vol <- rep(0,n-1)
for (i in (2:n)){
linearMod <- lm(formula = Mkt ~ 0+as.matrix(mydata[,i]))
modelSummary <- summary(linearMod)
vol[i-1] <- sd(modelSummary$residuals)
}

min(vol)
max(vol)
mean(vol)
```

Analýza na simulovaných dátach

```
#install.packages("MASS")
library(MASS)

#install.packages("pspline")
library(pspline)
#install.packages("phtt", repos="http://R-Forge.R-project.org")
library(phtt)

mydata <- read.csv("C:/Filip/Faklas job/MAIN DEAL/diplomovka/data/true
data/F-F_Research_Data_Factors_3m_tbill_till_2007.csv", header=T,
sep=",")
# mesacne vynosy z July 1963 to August 2007
# mesacne vynosy sme v subore predelovali 100 aby boli v desatinnych
# miestach, teda napr. 0.01 = 1% vynos za dany mesiac

head(mydata)

# uprava
assetrets <-mydata[,2:(3+1)]/100
head(assetrets)
attach(assetrets)

# Mkt.RF ... market portfolio - riskfree urokova miera. Ako riskfree
# urokovu mieru sme pouzili 3 mesacne US dlhopisy
# SMB ... size portfolio
# HML ... B-M portfolio
# blizsie popisy su na https://mba.tuck.dartmouth.edu/pages/faculty/
# ken.french/Data_Library/f-f_factors.html

# cele simulovanie
n <- length(Mkt.RF) # pocet vsetkych mesiacov v sade
dchch <- rep(0,n) # na ukladanie optimalnej hodnoty 'd' pocas cyklu
```

```
nA <- 100 # or 100 # pocet simulovanych aktiv
tw <- 120 # or 60, or 240 # velkost traning window
sh <- matrix(0,1000,(min(nA,60)+2)) # matica na uchovavanie hodnot
# sharperatio(SR) pocas cyklu, okrem nA aktiv uchovavame este SR pre
# MVtrue a SR pre naivnu strategiu
mi <- c(mean(Mkt.RF),mean(SMB),mean(HML))
sigm <- cov(cbind((Mkt.RF),(SMB),(HML)))

for (s in (1:1000)){ # robime 1000 simulacii
# mi a sigm urcuju strednu hodnotu a kovarianciu faktorov od 1963
# do 2007
factors <- mvrnorm(n = tw, mi, sigm, tol = 1e-12, empirical = TRUE)
# vo factors sme vygenerovali 3 faktory pomocou viacrozmerneho
# normalneho rozdelenia - so strednou hodnotou mi a kovarianciou sigm
# emprirical ... If true, mu and Sigma specify the empirical not
# population mean and covariance matrix.

# generovanie Biet pomocou rovnomerneho rozdelenia
bM <- runif(nA, min = 0.9, max = 1.2) # bM ako beta_Market
bS <- runif(nA, min = -0.3, max = 1.4) # ... beta_Size
bB <- runif(nA, min = -0.5, max = 0.9) # ... beta_Book-Market

misim <- mi[1]*bM + mi[2]*bS + mi[3]*bB # stredna hodnota simulacie.
# Konkretnejsie stredna hodnota vsetkych nA aktiv
# generovanie sigmeps
sigmeps1 <- diag(runif(nA, min = 0.01, max = 0.03)) # pomocou
# rovnomerneho rozdelenia
sigmeps <- sigmeps1%*%sigmeps1 # sigmeps urcuje kovarianciu epsilonov,
# teda specifickych odchylok jednotlivych aktiv
B <- cbind(bM, bS, bB)
sigmsim <- B %*% sigm %*% t(B)+sigmeps # kovariancia simulacie.
# Konkretnejsie kovariancia vsetkych nA aktiv
```

```
# generovanie epsilonov
epss <- mvrnorm(n = tw, rep(0,nA), sigmeps, tol = 1e-12, empirical =
TRUE)
# generovanie epsilonov pomocou viacrozmerneho normalneho rozdelenia
# so strednou hodnotou 0 a kovarianciou sigmeps
# emprirical=TRUE ... vysledne epsilonu budu mat strednu hodnotu
# presne 0 a kovariancnu maticu presnu sigmeps

# vypocet vynosov aktiv
# kazde aktivum ma svoju bM, bS a bB, ktore su pevne pocas celeho
#  obdobia potrebujeme prejst kazdym casom a vyratat mesacny vynos pre
# aktivum v kazdom case ma aktivum jedinecne epsilon
assetrets <- matrix(0,tw,nA)
for (na in (1:nA)){ # na je index pre konkretne aktivum
for (tt in (1:tw)){ # tt je index pre konkretny cas v training window
assetrets[tt,na] <- (bM[na]*factors[tt,1] +
                        bS[na]*factors[tt,2] +
                        bB[na]*factors[tt,3] + epss[tt,na])
}}
factors.pca <- prcomp(assetrets)
# vyratalo hlavne komponenty pre maticu assetrets - maticu
# nasimulovanych mesacnych vynosov

v <- factors.pca$rotation
# prikaz na priradenie hlavnych vektorov do matice v

# overenie ze factors.pca$sdev^2 su vlastne cisla cov matice
# eigenvalues = svd(cov(assetrets))$d
# factors.pca$sdev[k]^4
# ano, su

# optimal 'd'
# tu to je vyratane pomocou funkcie
```

```
dchch[s] <- OptDim(assetrets, criteria = "IC1")$summary[1]

# vlastny vypocet nefunguje
#maxk<-8
#dd<- rep(0,maxk)
#choice of d
#for (k in (2:(maxk))){
#dd[k] <- log(t(rep(1,(k-1)))%*%(factors.pca$sdev[(1):(k-1)]^4))
#+k*(nA+tw)/(nA*tw)*log((nA*tw)/(nA+tw))
#dd[k]<-log(length((k+1):nA)*factors.pca$sdev[k]^2)+k*(nA+tw)/
#(nA*tw)*log((nA*tw)/(nA+tw))
#}
#dch[s]<-which.min(dd)

# odhad vynosov a sigma
eE <- colMeans(assetrets) # priemery pre aktiva pocas training window
eS <- cov(assetrets) # kovariancia pre aktiva pocas training window
ga <- 3 # gamma # taka gamma bola pouzita v Miguel

#subspace mean variance
wd <- matrix(0,min(nA,60),nA) # matica kde zaznamename vahy pre kazde
# mozne 'd'
for (d in (1:min(nA,60))){
# vzorec nizsie je z ChenYuan2016 str. 9
wd[d,] <- rep(1/(ga),nA) * t(matrix(1/(factors.pca
$sdev[1:d]^2),d,nA)*t(v[,1:d])) %*% t(v[,1:d]) %*% eE
# prikaz "factors.pca$sdev[1:d]^2" znamena prvych d vlastnych cisiel
# na vyratanie vah pouzivame vl. vektory vyratane z training window a
# tiez eE vyratane z training window

# nizsie prikladam zakomentovany alternativny vzorecvzorec zo str. 7
# wd[d,] <- rep(1/ga,nA) * v[,1:d] %*% ginv((t(v[,1:d]) %*% eS %*%
# v[,1:d]), tol = 1e-6)%*% t(v[,1:d]) %*% eE
```

```
# z obidvoch vzorcov vychadza rovnaky vysledok
sh[s,d] <- t(wd[d,])%*%misim/ sqrt(t(wd[d,])%*%sigmsim%*%wd[d,])
# vypocet a ulozenie SR
}

# true MV strategy
gaT <- ga
wT <- rep(1/gaT,length(misim)) * ginv(sigmsim, tol = 1e-12) %*% misim
# vypocet vahy pre MVtrue, z ChenYuan2016 str. 5
# vo vzorci hore pouzivame prave(v realite nezname) hodnoty misim a
# sigmsim
sh[s,(min(nA,60)+1)] <- t(wT)%*%misim/ sqrt(t(wT)%*%sigmsim%*%wT)
# vypocet a ulozenie SR

# naive strategy
gaN <- nA # gammma
wN <- rep(1/gaN,nA)
sh[s,(min(nA,60)+2)] <- t(wN[1:nA])%*%misim/ sqrt(t(wN[1:nA])%*
%sigmsim%*%wN[1:nA])
# vypocet a ulozenie SR
} # koniec celeho simulovania


sharpe <- colMeans(sh) # vytvori priemer SR pre subspace MV
# s d=1, ...25, pre MVtrue a pre naivnu strategiu
sT <- sharpe[min(60,nA)+1] # SR true MV portfolio
naive <- sharpe[min(60,nA)+2] # SR naivna strategia
```

Porovnanie stratégií na reálnych dátach pre rôzne fixné počty hlavných komponentov *d*

```
#install.packages("RiskPortfolios")
library(RiskPortfolios)

#install.packages("pspline")
#library(pspline)
#install.packages("phtt", repos="http://R-Forge.R-project.org")

library(phtt)
library(MASS)

vysledok <- NULL
# data z jan 1961 do dec 2010, data su v percentach a nie v
# desatinnych miestach
#mydata<-read.csv("C:/Filip/Faklas job/MAIN DEAL/diplomovka/data/true
#data/25_Portfolios_5x5_minus_rf.csv", header=T, sep=",")
#mydata<-read.csv("C:/Filip/Faklas job/MAIN DEAL/diplomovka/data/true
#data/100_Portfolios_10x10_minus_rf.csv", header=T, sep=",")
mydata<-read.csv("C:/Filip/Faklas␣job/MAIN␣DEAL/diplomovka/data/true
data/S&P500_companies_rets.csv", header=T, sep=",")
#attach(mydata)
#head(mydata)

nA <- dim(mydata)[2]-1
assetrets <- mydata[,2:(nA+1)]/100
n <- length(assetrets[,1])

# tw <- 60, 120, 240 (200)
tw <- 120

turnover <- matrix(0,n,9)
```

```
vynosy <- matrix(0,n,3*min(tw, nA)+3)

for (i in ((1+tw):n)){

X <- assetrets[(i-tw):(i-1),]
factors.pca <- prcomp(X)
v <- factors.pca$rotation

# corelation matrix
Y <- scale(assetrets[(i-tw):(i-1),], center=TRUE, scale=TRUE)
factors.pcaY <- prcomp(Y)
vv <- factors.pcaY$rotation
# odhad vynosov a sigmy
eE <- colMeans(assetrets[(i-tw):(i-1),])
eS <- cov(assetrets[(i-tw):(i-1),])
ga <- 3

# realne vynosy
E <- as.numeric(assetrets[i,])

# for various 'd'
for (d in (1:min(nA, tw))){ # spravne ma byt po nA

#subspace MV
## vzorec nizsie je z ChenYuan2016 str. 9
wd <- rep(1/(ga),nA) * t(matrix(1/(factors.pca
$sdev[1:d]^2),d,nA)*t(v[,1:d])) %*% t(v[,1:d]) %*% eE
vynosy[i, d] <- t(wd)%*%E[1:nA]

# vahy pre expansion 1
wdd2 <- rep(1/(ga),nA) * t(matrix(1/(factors.pcaY
$sdev[1:d]^2),d,nA)*t(vv[,1:d])) %*% t(vv[,1:d]) %*% eE
vynosy[i, min(tw, nA)+d] <- t(wdd2)%*%E[1:nA]
```

```
# vahy pre expansion 2
Dinv <- 1/sqrt(diag(eS))
Dinv <- diag(Dinv) # = D^(-1)
rho <- Dinv%*%eS%*%Dinv
wdd3 <- rep(1/ga,nA) * Dinv %*% vv[,1:d] %*% ginv((t(vv[,1:d]) %*
% rho %*% vv[,1:d]), tol = 1e-6)%*% t(vv[,1:d]) %*% Dinv %*% eE
vynosy[i, 2*min(tw, nA) + d] <- t(wdd3)%*%E[1:nA]

} # koniec for cyklu pre 'd'

# vahy pre subspace MV d estimated
dd <- OptDim(as.matrix(assetrets[(i-tw):(i-1),]), criteria = "IC1")
$summary[1]
wdd <- rep(1/(ga),nA) * t(matrix(1/(factors.pca
$sdev[1:dd]^2),dd,nA)*t(v[,1:dd])) %*% t(v[,1:dd]) %*% eE
vynosy[i, 3*min(tw, nA)+1] <- t(wdd) %*% E[1:nA]

# MV as Mean Variance sample
wMV <- rep(1/ga,length(eE)) * ginv(eS, tol = 1e-12) %*% eE
# wMV <- optimalPortfolio(mu = eE, Sigma = eS, control = list(gama =
ga))
vynosy[i, 3*min(tw, nA)+2] <- t(wMV) %*% E[1:nA]
# N as Naive
gaN <- nA
wN <- rep(1/gaN,nA)
vynosy[i, 3*min(tw, nA)+3] <- t(wN) %*% E[1:nA]
} # koniec cyklu

r <- colMeans(vynosy[(tw+1):n,])
vol <- apply(vynosy[(tw+1):n,], 2, sd)
sharpe <- r/vol
```

Porovnanie stratégií na reálnych dátach pre optimálny počet hlavných komponentov *d*

```
#install.packages("RiskPortfolios")
library(RiskPortfolios)

#install.packages("pspline")
#library(pspline)
#install.packages("phtt", repos="http://R-Forge.R-project.org")

library(phtt)
library(MASS)

vysledok <- NULL
# data z jan 1961 do dec 2010, data su v percentach a nie v
# desatinnych miestach
#mydata<-read.csv("C:/Filip/Faklas job/MAIN DEAL/diplomovka/data/true
#data/25_Portfolios_5x5_minus_rf.csv", header=T, sep=",")
mydata<-read.csv("C:/Filip/Faklas job/MAIN DEAL/diplomovka/data/true
data/100_Portfolios_10x10_minus_rf.csv", header=T, sep=",")
#mydata<-read.csv("C:/Filip/Faklas job/MAIN DEAL/diplomovka/data/true
#data/S&P500_companies_rets.csv", header=T, sep=",")
#attach(mydata)
#head(mydata)

nA <- dim(mydata)[2]-1
assetrets <- mydata[,2:(nA+1)]/100
n <- length(assetrets[,1])

# tw <- 60, 120, 240 (200)
tw <- 120

dchm <- matrix(0,n,5)
```

```
turnover <- matrix(0,n,9)
vynosy <- matrix(0,n,10)
dch <- rep(0,n)
vahy1 <- matrix(0,n,nA)
vahy2 <- matrix(0,n,nA)

for (i in ((1+tw):n)){

X <- assetrets[(i-tw):(i-1),]
factors.pca <- prcomp(X)
v <- factors.pca$rotation

helper <- rep(0,min(nA,tw))
for (j in (1:min(nA,tw))){
helper[j] <- sum(factors.pca$sdev[1:j]^2)/sum(factors.pca$sdev^2)
}

dchm[i,2] <- which.max(helper>0.92)
dchm[i,4] <- which.max(helper>0.96)

d2 <- dchm[i,2]
d4 <- dchm[i,4]

# corelation matrix
Y <- scale(assetrets[(i-tw):(i-1),], center=TRUE, scale=TRUE)
factors.pcaY <- prcomp(Y)
vv <- factors.pcaY$rotation

dch[i] <- OptDim(as.matrix(assetrets[(i-tw):(i-1),]), criteria =
"IC1")$summary[1]
d <- dch[i]

# odhad vynosov a sigmy
```

```
eE <- colMeans(assetrets[(i-tw):(i-1),])
eS <- cov(assetrets[(i-tw):(i-1),])
ga <- 3 #3

# realne vynosy
E <- as.numeric(assetrets[i,])

# vahy pre subspace MV d estimated
if (i!=(tw+1)){
wddo <- wdd
}
wdd <- rep(1/(ga),nA) * t(matrix(1/(factors.pca
$sdev[1:d]^2),d,nA)*t(v[,1:d])) %*% t(v[,1:d]) %*% eE
vahy1[i,] <- t(wdd)
vynosy[i, 1] <- t(wdd) %*% E[1:nA]
if (i!=(tw+1)){
turnover[i,1] <- sum(abs(wdd - wddo))
}

# vahy pre expansion 1
if (i!=(tw+1)){
wddo2 <- wdd2
}
wdd2 <- rep(1/(ga),nA) * t(matrix(1/(factors.pcaY
$sdev[1:d]^2),d,nA)*t(vv[,1:d])) %*% t(vv[,1:d]) %*% eE
vahy2[i,] <- t(wdd2)
vynosy[i, 2] <- t(wdd2)%*%E[1:nA]
if (i!=(tw+1)){
turnover[i,2] <- sum(abs(wdd2 - wddo2))
}

# vahy pre expansion 2
if (i!=(tw+1)){
```

```
wddo3 <- wdd3
}
Dinv <- 1/sqrt(diag(eS))
Dinv <- diag(Dinv) # = D^(-1)
rho <- Dinv%*%eS%*%Dinv
wdd3 <- rep(1/ga,nA) * Dinv %*% vv[,1:d] %*% ginv((t(vv[,1:d]) %*%
rho %*% vv[,1:d]), tol = 1e-6)%*% t(vv[,1:d]) %*% Dinv %*% eE
vynosy[i, 3] <- t(wdd3)%*%E[1:nA]
if (i!=(tw+1)){
turnover[i,3] <- sum(abs(wdd3 - wddo3))
}

#vahy pre percentily obsiahnutej variancie

#d2
if (i!=(tw+1)){
wdp2o <- wdp2
}
wdp2 <- rep(1/(ga),nA) * t(matrix(1/(factors.pca
$sdev[1:d2]^2),d2,nA)*t(v[,1:d2])) %*% t(v[,1:d2]) %*% eE
vynosy[i, 5] <- t(wdp2)%*%E[1:nA]
if (i!=(tw+1)){
turnover[i,5] <- sum(abs(wdp2 - wdp2o))
}

#d4
if (i!=(tw+1)){
wdp4o <- wdp4
}
wdp4 <- rep(1/(ga),nA) * t(matrix(1/(factors.pca
$sdev[1:d4]^2),d4,nA)*t(v[,1:d4])) %*% t(v[,1:d4]) %*% eE
vynosy[i, 7] <- t(wdp4)%*%E[1:nA]
if (i!=(tw+1)){
```

```
turnover[i,7] <- sum(abs(wdp4 - wdp4o))
}

# MV as Mean Variance sample
if (i!=(tw+1)){
wMVo <- wMV
}
wMV <- rep(1/ga,length(eE)) * ginv(eS, tol = 1e-12) %*% eE
#wMV <- optimalPortfolio(mu = eE, Sigma = eS, control = list(gama =
#ga))
vynosy[i, 9] <- t(wMV) %*% E[1:nA]
if (i!=(tw+1)){
turnover[i,9] <- sum(abs(wMV - wMVo))
}

# N as Naive
gaN <- nA
wN <- rep(1/gaN,nA)
vynosy[i, 10] <- t(wN) %*% E[1:nA]



} # koniec cyklu

r <- colMeans(vynosy[(tw+1):n,])
vol <- apply(vynosy[(tw+1):n,], 2, sd)
sharpe <- r/vol
```