

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



NÁHODNÉ MATICE: TEÓRIA A APLIKÁCIE
BAKALÁRSKA PRÁCA

2012

Martin ŠIMO

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

NÁHODNÉ MATICE: TEÓRIA A APLIKÁCIE

BAKALÁRSKA PRÁCA

Študijný program: Ekonomická a finančná matematika
Študijný odbor: 1114 Aplikovaná matematika
Školiace pracovisko: Katedra aplikovanej matematiky a štatistiky
Vedúci práce: Mgr. Martin Niepel, PhD.
Evidenčné číslo: ffb697e9-eb29-422f-879d-ab11cfb28ebb



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

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

Názov: Náhodné matice: teória a aplikácie

Cieľ: Oboznámenie sa s výsledkami teórie náhodných matíc a ich aplikáciami vo finančnej matematike..

Vedúci: Mgr. Martin Niepel, PhD.

Dátum zadania: 16.10.2011

Dátum schválenia: 27.10.2011

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

študent

vedúci

Čestné prehlásenie

Čestne prehlasujem, že na bakalárskej práci som pracoval samostatne na základe vlastných teoretických poznatkov, konzultácií a odbornej literatúry uvedenej v Zozname použitej literatúry.

Miesto, Dátum

Titul, Meno Priezvisko

Pod'akovanie Ďakujem svojmu vedúcemu práce Mgr. Martinovi Niepelovi, PhD. za ochotu, odbornú pomoc a trpezlivosť. Ďakujem aj svojej priateľke Linde Šimekovej za podporu a trpezlivosť.

Abstrakt v štátnom jazyku

ŠIMO, Martin: Náhodné matice: teória a aplikácie [Bakalárska práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; vedúci práce: Mgr. Martin Niepel, PhD., Bratislava, 2012, 34s.

V našej práci analyzujeme korelácie medzi akciami na finančnom trhu pomocou niektorých výsledkov teórie náhodných matíc. Skúmame vlastnosti empirickej korelačnej matice denných výnosov, teda rozdelenie jej prvkov a jej spektrálne vlastnosti. Tieto porovnáваме s tvrdeniami teórie náhodných matíc. Pre skúmané objekty ponúkame finančnú interpretáciu a výstupom našej analýzy sú poznatky o štruktúre trhu a možné spôsoby ako tieto poznatky využiť pri riešení úloh, v ktorých vystupuje korelačná matica denných výnosov.

Kľúčové slová: Teória náhodných matíc, Empirická korelačná matica, Marčenkovo–Pasturovo rozdelenie, Optimalizácia portfólia

Abstract

ŠIMO, Martin: Random Matrices: Theory and Applications [Bachelor Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Martin Niepel, PhD., Bratislava, 2012, 34p.

In our work we analyze cross-correlations in stock market data using results of Random Matrix Theory. We explore properties of empirical correlation matrix of daily returns, which include the distribution of entries and spectral properties. We compare these properties with the theoretical results given by Random Matrix Theory. We offer financial interpretation for these objects and the results of our analysis provide a better understanding of the stock market structure, which can be used in other applications involving correlation matrix of returns.

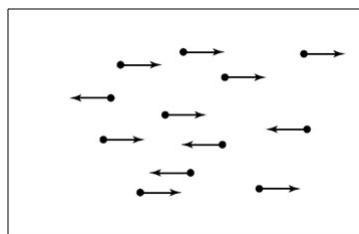
Keywords: Random Matrix Theory, Empirical Correlation Matrix, Marchenko–Pastur distribution, Portfolio Optimization

Obsah

Úvod	8
1 Východiská práce	11
1.1 Analyzované dáta	11
1.2 Predpoklady a obmedzenia analýzy	12
2 Analýza medziakciových korelácií	14
2.1 Korelačná matica denných výnosov	14
2.2 Spektrálne vlastnosti matice E	15
2.3 Vlastné hodnoty upravenej korelačnej matice E'	17
2.4 Vlastné vektory očistenej korelačnej matice E'	19
2.5 Stabilita spektrálnych vlastností matice E v čase	23
3 Ďalšie aplikácie	27
3.1 Použitie pri optimalizácií portfólia	27
3.2 Singulárny rozklad matice X	28
Záver	30
Zoznam použitej literatúry	33

Úvod

Častým objektom fyzikálneho skúmania v súčasnosti sú komplexné systémy. Štatistická fyzika skúma systémy, ktoré sa skladajú z mnohých súčastí. Správanie jednotlivých súčastí je v niektorých prípadoch známe, resp. sa dá popísať jednoduchým matematickým modelom. Skúmanie veľkých systémov sa však komplikuje kvôli veľkému množstvu interakcií medzi jednotlivými súčastami. Niektoré systémy sa dokonca správajú kvalitatívne odlišne v rôznych škálach veľkosti. Jednoduchým príkladom takéhoto systému je napríklad uzavretá nádoba v ktorej sa nachádza N častíc, ktoré sa hýbu buď doprava alebo doľava. Situáciu ilustruje Obr. 1. V tomto prípade by nás mohlo zaujímať v akom stave sa systém nachádza, napríklad aký je tlak na pravú stenu nádoby.



Obr. 1: Častice v nádobe

Klasickým riešením tohto problému je poznať polohu a rýchlosť každej častice. Potom by sme vedeli povedať, ktoré častice sú dosť blízko pri pravej stene a hýbu sa správnym smerom, teda na stenu vyvíjajú tlak. Toto riešenie však pre veľké N môže byť mimo náš dosah. V takejto situácii je veľmi náročné, niekedy priamo nemožné, odmerať polohu a rýchlosť každej častice.

Iné riešenie takýchto problémov ponúka štatistická fyzika. Takéto riešenie spočíva v zmene optiky problému. Nebudeme sa zaoberať jednotlivými časticami, ale systémom ako celkom. Určíme všetky možné stavy systému, všetky možné polohy a smery všetkých častíc, a pre každý stav zistíme aký je veľký tlak na pravú stenu. Z týchto všetkých stavov vypočítame priemerný a budeme experimentálne overovať, či takýto priemerný stav systému dobre popisuje aj náš konkrétny. Ukazuje sa, že pre veľké N v mnohých systémoch tento postup vedie k dobrému riešeniu, teda priemerný stav je dobrým odhadom konkrétneho.

Podobnú ideu využil aj Eugen Paul Wigner pri modelovaní energetických hladín

jadier ťažkých atómov. Uvažoval súbory náhodných symetrických $N \times N$ matíc, kde každá zložka pochádza z nejakého pravdepodobnostného rozdelenia. Vlastné hodnoty týchto matíc zodpovedali rôznym energetickým hladinám tohto fyzikálneho systému, teda každá matica popisovala nejaký konkrétny stav systému. Za svoj výsledok potom zobral priemerné správanie všetkých možných systémov a skúmal jeho asymptotické vlastnosti pre N idúce do nekonečna. Ukázalo sa, že takýto postup vedie k dobrým riešeniam pre tento problém a teóriou náhodných matíc sa začali zaoberať aj ďalší fyzici a matematici.

Keďže mnohé z prvých aplikácií teórie náhodných matíc vychádzali z fyzikálnych problémov, tak skúmané matice mali konkrétnu štruktúru, ktorá vyplývala z povahy riešeného problému. Jednalo sa väčšinou o reálne symetrické alebo komplexné Hermitovské matice, ktoré majú dobré spektrálne vlastnosti. Práve spektrálne vlastnosti, najmä vlastné hodnoty, týchto matíc boli medzi prvými skúmanými objektami. V súčasnosti sa teória náhodných matíc používa nielen vo fyzike, ale našla si aj mnoho ďalších uplatnení v numerike, finančníctve a neurónových sieťach.

V tejto práci sa budeme venovať aplikáciám teórie náhodných matíc na finančný trh. Oboznámime sa s niektorými teoretickými výsledkami, ktoré sú pre tieto aplikácie relevantné. Ide najmä o spektrálne vlastnosti reálnych symetrických matíc opisujúcich korelácie. Tieto teoretické vlastnosti ilustrujeme aj pomocou matematického softvéru a Monte Carlo metód. Potom tieto teoretické poznatky využijeme pri analýze reálnych trhových dát.

Hlavným objektom skúmania tejto práce budú korelácie medzi dennými výnosmi akcií na finančnom trhu. Teoretické výsledky teórie náhodných matíc používajú na analýzu medziakciových korelácií aj práce [1, 2, 3]. V našej práci sa pokúsime zopakovať experimenty popísané v týchto prácach na novších dátach. Tieto práce sa venujú analýze akcií, ktoré tvoria index S&P 500, pričom v práci [1] porovnávajú dosiahnuté výsledky aj s analýzou akcií tokijskej burzy. Popis skúmaných dát spolu s predpokladmi a obmedzeniami použitých modelov sú zhrnuté v prvej kapitole.

V druhej kapitole budeme skúmať spektrálne vlastnosti empirickej korelačnej matice denných výnosov E a porovnávať ich s teoretickými spektrálnymi vlastnosťami danými teóriou náhodných matíc. Naše výsledky budeme porovnávať aj s výsledkami

dosiahnutými v spomínaných prácach.

Tretia kapitola obsahuje možnosti použitia výsledkov druhej kapitoly pri riešení klasického Markowitzovského optimalizačného problému. Uvádzame v nej aj niektoré skutočnosti, ktoré považujeme za zaujímavé a hodné povšimnutia.

Cieľom tejto práce je oboznámiť čitateľa s vybranými tvrdeniami teórie náhodných matíc. Toto však nebudeme robiť klasickým spôsobom budovania teórie s dôkazmi. Tvrdenia teórie náhodných matíc priamo použijeme pri relevantnej analýze a ilustrujeme ich pomocou náhodne vygenerovaných dát, ktoré budeme používať ako kontrolnú vzorku. Dosiahnuté výsledky sa pokúsime vždy uviesť do reálneho kontextu.

1 Východiská práce

1.1 Analyzované dáta

Vzorka dát, ktorej sa v tejto práci budeme venovať sú údaje o 500 firmách, ktoré tvorili akciový index S&P500, v časovom období medzi 1.1.1990 a 28.12.2011. Zoznam týchto firiem spolu s príslušným sektorom a geografickou lokáciou sídla sme čerpali z [6]. Historické dáta o vývoji cien každej z týchto firiem sme získali z databázy [7]. Zo všetkých dostupných údajov nás zaujímala zatváracia cena akcií v obchodovaných dňoch. Z týchto zatváracích cien sme zostavili maticu dát, kde stĺpce predstavujú akcie a riadky obchodované dni. Dáta v tejto forme sme potom analyzovali pomocou štatistického softvéru R.

Predmetom nášho skúmania sú predovšetkým korelácie medzi jednotlivými akciami, preto sme sa rozhodli brať pri základnej analýze do úvahy posledných 10 rokov dát. Na jednej strane chceme dáť čo najviac, pretože pomer $q = N/T$, kde N je počet akcií a T je počet obchodovaných dní, je dôležitý pre relevantnosť výsledkov a je dobré, keď je čo najmenší. Na druhej strane medziakciové korelácie s veľkou pravdepodobnosťou nie sú v čase stabilné, preto nie je vhodné uvažovať ani príliš dlhé časové obdobia. V našom prípade, 10 rokov dát predstavuje približne 2500 obchodovaných dní a pomer q je približne 0.2.

Základnými údajmi, s ktorými budeme pracovať, pri hľadaní korelácií medzi akciami sú denné výnosy. Vektory denných výnosov pre akciu $i = 1, \dots, N$ vypočítame ako

$$R_i(t) \equiv \frac{S_i(t+1)}{S_i(t)} - 1, \quad (1.1)$$

kde $S_i(t)$ je zatváracia cena i -tej akcie v dni t . Z týchto denných výnosov vypočítame normalizované denné výnosy

$$X_i(t) \equiv \frac{R_i(t) - \bar{R}_i}{\sqrt{\bar{R}_i^2 - \bar{R}_i^2}}. \quad (1.2)$$

Priemerná hodnota normalizovaných výnosov X_i je nulová a smerodajná odchylka σ_i je rovná jednej. Ten istý postup je použitý aj v prácach [1, 2, 3].

Na tomto mieste považujeme za vhodné upozorniť, že kvalitu takto zostrojených dát treba dôkladne skontrolovať. Proces zostavovania pôvodnej tabuľky zatváracích cien

môže viesť k chybám. Medzi akciami existujú také, ktoré nemajú niektorý deň v dátach priradenú žiadnu hodnotu, pravdepodobne v tie dni neboli obchodované. Takýmto prázdny miestam sme priradili nulovú hodnotu. Dni, v ktorých sa nachádzalo veľa nulových cien, sme zo súboru vylúčili. Pri výpočte relatívneho výnosu však môže pri niekoľko málo akciách nastať situácia, že na mieste ceny z minulého dňa je hodnota nula. Výpočet relatívneho výnosu potom vedie k výsledku nekonečno a pokiaľ takéto dni neodstránime, vedie to k výrazne odlišným výsledkom ďalších analýz. Na túto možnosť je dôležité pamätať a dni, v ktorých sa medzi výnosmi objavuje hodnota nekonečno, je treba zo súboru vylúčiť. Rovnako ako riadky, môžu byť „zlé“ aj stĺpce. Pokiaľ ide o novú akciu, ktorá nemá potrebnú históriu, napr. je na trhu obchodovaná iba 3 roky, ukazuje sa ako vhodné vylúčiť celý stĺpec. Takéto filtrovanie dát potom vedie k tomu, že z pôvodnej matice výnosov s rozmermi 2500×500 ostane v našom prípade matica X rozmerov 2120×466 .

1.2 Predpoklady a obmedzenia analýzy

Predtým ako budeme dáta analyzovať pomocou štatistických nástrojov uvedieme predpoklady, z ktorých tieto nástroje vychádzajú. Dôležitým predpokladom je, že denné výnosy, naše vstupné dáta, považujeme za náhodnú premennú s normálnym rozdelením. Model s normálnym rozdelením používame kvôli značnému zjednodušeniu tvrdení, ktoré tento predpoklad zabezpečí. Takýto popis správania výnosov však nevychádza z praxe, ktorá je s ním skôr v rozpore. Predpoklad náhodného správania výnosov v krátkodobom horizonte, nie viac ako jeden deň, je vo finančníctve bežne využívaný a ukazuje sa ako prípustný. Rozdelenie týchto výnosov však nie je normálne. V práci [2] sa uvádza, že pravdepodobnostné rozdelenia krátkodobých výnosov majú ťažšie chvosty ako normálne rozdelenie a majú bližšie skôr k Studentovmu rozdeleniu.

Z dát o denných výnosoch sa budeme snažiť získať informácie o korelačnej matici C . Korelačná matica C v tomto prípade zodpovedá viacrozmernému normálnemu rozdeleniu, z ktorého podľa predpokladov nášho modelu pochádzajú denné výnosy. Maticu C , ktorá je parametrom predpokladaného náhodného procesu, nebudeme nikdy poznať. Budeme ju odhadovať z dát a skúmať jej vlastnosti pomocou teórie náhodných matic.

V práci budeme často využívať kontrolnú vzorku. Dáta a výsledky, ktoré pochádzajú

z kontrolnej vzorky budeme označovať pomocou dolného indexu R , napr. E_R . Objektom nášho skúmania sú korelácie medzi jednotlivými akciami. Nulová hypotéza, na ktorej je založená kontrolná vzorka, vychádza z klasického štatistického predpokladu nezávislosti denných výnosov, čo zodpovedá identickej korelačnej matici $C_R = I$. Kontrolná vzorka je tvorená náhodne vygenerovanými dátami, pričom takto zostrojené kontrolné výnosy majú normálne rozdelenie so strednou hodnotou nula a smerodajnou odchýlkou rovnou jednej. Rozmery týchto dát sú identické s rozmermi skutočných dát a v niektorých prípadoch používame niekoľko takto vygenerovaných súborov. V práci budeme skúmať odchýlky od tohto nezávislého správania, ktoré naznačujú prítomnosť netriviálnych medziakciových korelácií a deterministickej štruktúry v rámci trhu.

Všetky korelácie, ktoré budeme skúmať, sú založené na historických dátach. Takýto prístup je nutný, kvôli nedostatku akýchkoľvek iných dát. Je ale dôležité si uvedomiť, že korelácie v minulosti nemusia nutne implikovať budúce korelácie. Rovnako ako budúce výnosy nemusia byť jednoducho odhadnuteľné pomocou minulých. Tieto predpoklady sú nutným a pomerne samozrejým základom všetkých predpovedí vo finančníctve, na ich obmedzenia však treba pamätať a explicitne upozorniť.

2 Analýza medziakciových korelácií

2.1 Korelačná matica denných výnosov

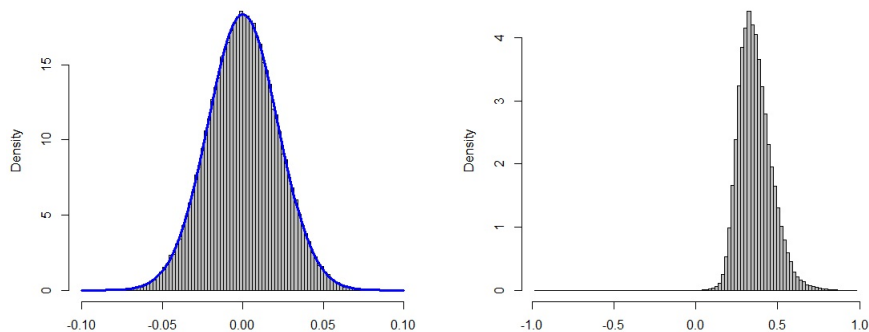
Z dát v matici X , odhadneme empirickú korelačnú maticu pomocou Pearsonovho odhadu

$$E = \frac{1}{T} X^T X. \quad (2.1)$$

Treba poznamenať, že ide o odhad z dát, teda nie o „skutočnú“ korelačnú maticu C , ktorú nepoznáme a nikdy poznať nebudeme. Intuitívne očakávame, že odhad E by sa mal od skutočnej C líšiť iba málo v prípade kedy máme k dispozícii dostatočne veľa realizácií, tj. dát. To nastáva v prípade kedy pomer $q = N/T$ ide k nule. V prípade keď $N \gg 1, T \gg 1$ a $q = O(1)$ výsledky nie sú intuitívne. Vlastnosťami empirickej korelačnej matice E v takomto prípade sa zaoberá časť teórie náhodných matíc, ktorú budeme v práci používať.

Prvou vecou, ktorú môžeme analyzovať, sú nediagonálne zložky matice E , ktoré predstavujú odhady korelačných koeficientov medzi jednotlivými párami akcií. V kontrolnom prípade, keď denné výnosy jednotlivých akcií sú nezávislé náhodné premenné, teda skutočná korelačná matica C_R je rovná identite, je skutočná hodnota korelácie medzi ľubovoľným párom $(C_R)_{i,j}$ nulová. Keďže však máme iba obmedzený počet realizácií očakávame, že v odhadnutej korelačnej matici E_R budú nediagonálne zložky blízke nule.

Pozrime sa na prvky kontrolnej korelačnej matice E_R , ktorá pochádza z náhodne vygenerovaných dát. Rozdelenie prvkov tejto matice je nulovou hypotézou s ktorou budeme porovnávať pravdepodobnostné rozdelenie prvkov empirickej korelačnej matice E , pochádzajúcej z reálnych dát. V kontrolnom prípade je priemerná hodnota korelácie blízka nule so smerodajnou odchýlkou $\sigma \approx 0.02$ a prvky matice E_R sa riadia normálnym rozdelením. S rastúcou hodnotou počtu realizácií T sa disperzia znižuje a asymptoticky je rovná nule. Zložky matice E , ktorá pochádza z reálnych dát, sú na prvý pohľad výrazne odlišné, ako je vidieť na obr. 2.1. Priemerná hodnota korelácie je odrazená od nuly a disperzia je väčšia ako v pri kontrolnej vzorke. Tak isto je z obrázku vidieť pozitívny sklon rozdelenia zložiek. Zaujímavý je aj fakt, že všetky korelácie sú kladné.

Obr. 2.1: Nediagonálne prvky matíc E_R a E

2.2 Spektrálne vlastnosti matice E

Matica E je symetrická kladne semidefinitná $N \times N$ matica. Z toho vyplýva, že existuje jej spektrálny rozklad

$$E = \sum_{\alpha=1}^N \lambda_{\alpha} V_{\alpha} V_{\alpha}^T, \quad (2.2)$$

teda existuje ortnormálna báza zložená z jej vlastných vektorov a všetky jej vlastné hodnoty sú reálne a nezáporné. Tento poznatok lineárnej algebry je základom tzv. analýzy hlavných komponentov, po anglicky Principal Component Analysis, ktorej cieľom je rozložiť korelované náhodné premenné na nekorelované smery s klesajúcou disperziou. Nekorelované smery sú vlastné vektory korelačnej matice E a vlastné hodnoty matice E zodpovedajú príslušným disperziám.

Skôr ako budeme tieto objekty skúmať je vhodné uviesť ich finančnú interpretáciu. Na ľubovoľný vektor váh w , ktorý má N zložiek sa dá pozeráť ako na portfólio Π_w , v ktorom každá zložka zodpovedá váhe danej akcie. Niektoré akcie sú v pozícií long, ak w_i je kladné, a ostatné v pozícií short, w_i je záporné, alebo sa v portfóliu nenachádzajú, kedy je w_i nulové. Realizované riziko portfólia Π_w , vypočítané na základe historických dát, je rovné

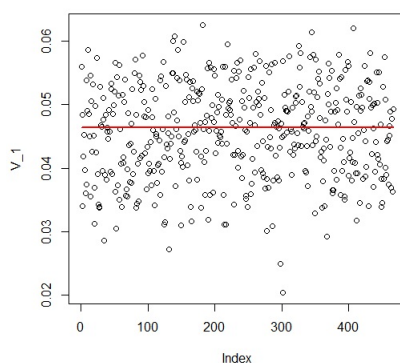
$$\sigma_w^2 = \sum w_i \sigma_i E_{i,j} \sigma_j w_j = w^T E w, \quad (2.3)$$

kde σ_i predstavuje volatilitu výnosov i -tej akcie. V našom prípade sú všetky volatility rovné jednej z čoho plynie druhá rovnosť. V prípade vlastných vektorov platí

$$\sigma_{\alpha}^2 = V_{\alpha}^T E V_{\alpha} = \lambda_{\alpha} \underbrace{V_{\alpha}^T V_{\alpha}}_1 = \lambda_{\alpha}. \quad (2.4)$$

Vďaka princípom analýzy hlavných komponentov sú portfóliá definované vlastnými vektormi navzájom nekorelované. Spektrálny rozklad matice E je teda rozkladom trhu na nezávislé portfóliá s klesajúcou volatilitou.

Ukazuje sa, že vlastný vektor zodpovedajúci najväčšej vlastnej hodnote, značíme V_1 , zodpovedá tzv. „trhovému módu“ [3]. Je to portfólio, ktoré zachytáva pohyby trhu ako celku. Vlastný vektor V_1 približne zodpovedá portfóliu bez diverzifikácie, tj. $V_{1,i} \approx 1/\sqrt{N}$. Tento výsledok sa potvrdil pri našej analýze. Zložky vektora V_1 porovnané s hodnotou $1/\sqrt{N}$ ukazuje obr. 2.2.



Obr. 2.2: Zložky prvého vlastného vektora V_1 porovnané s hodnotou $1/\sqrt{N}$

To že vektor V_1 zachytáva podstatnú časť pohybov trhu potvrdzuje výrazný rozdiel medzi prvou a druhou vlastnou hodnotou. Vlastná hodnota, zodpovedajúca vektoru V_1 , má hodnotu $\lambda_1 = 174$. Pre porovnanie, druhá najväčšia vlastná hodnota matice E je $\lambda_2 = 16$. Tento vplyv zväčšuje aj všetky medziakciové korelácie a je pravdepodobne dôvodom toho, že sú všetky pozitívne. Preto pre potreby ďalšej analýzy vplyv „trhového módu“ odstránime.

Možnosť ako modelovať vplyv, ktorý je spoločný pre všetky akcie ponúka vzorec (2.5), ktorý je vo finančníctve používaný a je napríklad základom Capital Asset Pricing Modelu.

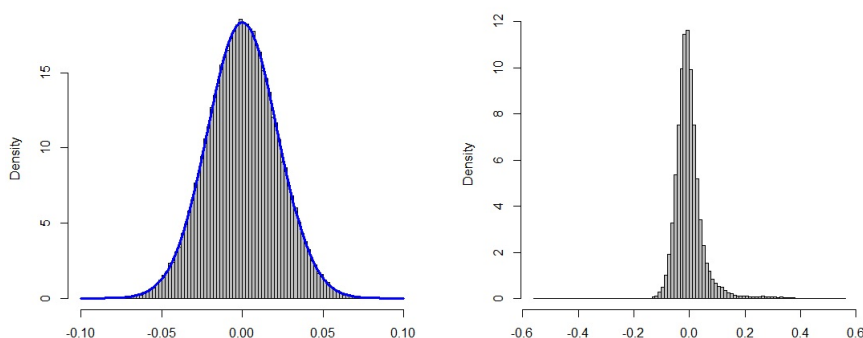
$$X_i(t) = \alpha_i + \beta_i M(t) + \epsilon_i(t) \quad (2.5)$$

Kde α_i a β_i sú parametre špecifické pre danú akciu a $M(t)$ je vplyv spoločný pre všetky. Zároveň požadujeme, aby $\overline{\epsilon(t)} = 0$ a skalárny súčin $\langle M(t)\epsilon(t) \rangle$ bol rovný nule. Spoločný

vplyv $M(t)$, ktorého presnú hodnotu nepoznáme, odhadneme pomocou

$$X^1(t) = \sum_{j=1}^N V_{1,j} X_j(t). \quad (2.6)$$

Vplyv spoločný pre všetky akcie teda definujeme ako vplyv portfólia V_1 . Novú, upravenú, korelačnú maticu E' definujeme ako korelačnú maticu odchýliek ϵ . Rozloženie prvkov korelačnej matice E' , porovnané s náhodnou korelačnou maticou E_R , zachytáva obr. 2.3. Odstránenie spoločného vplyvu trhu viedlo k výraznému zníženiu priemernej medziakciovej korelácie, nová priemerná hodnota je blízka nule. Disperzia hodnôt sa znížila, ale je stále vyššia ako v prípade kontrolnej vzorky a rozdelenie si zachovalo pozitívny sklon.



Obr. 2.3: Porovnanie nediagonálnych prvkov náhodnej a očistenej korelačnej matice E'

2.3 Vlastné hodnoty upravenej korelačnej matice E'

Vlastné hodnoty matice E' sú veľmi podobné vlastným hodnotám pôvodnej matice E . Odstránením vplyvu vektora V_1 sa najväčšia vlastná hodnota λ_1 zmenší na nulu. Ostatné vlastné hodnoty a vlastné vektory ostávajú prakticky nezmené.

Tvrdenia teórie náhodných matíc, ktoré tu spomíname, popisujú správanie empirickej korelačnej matice E , v prípade, že skutočná korelačná matica C je rovná identite. Hustota vlastných hodnôt empirickej korelačnej matice v takomto prípade je určená ako Marčenkovo–Pasturovo rozdelenie [2]

$$\rho_E(\lambda) = \frac{\sqrt{4\lambda q - (\lambda + q - 1)^2}}{2\pi\lambda q}, \quad \lambda \in [(1 - \sqrt{q})^2, (1 + \sqrt{q})^2]. \quad (2.7)$$

Iný zápis tohto rozdelenia, ktorý budeme používať je

$$\rho_E(\lambda) = \frac{1}{2\pi q} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \quad \lambda_- \leq \lambda \leq \lambda_+, \quad (2.8)$$

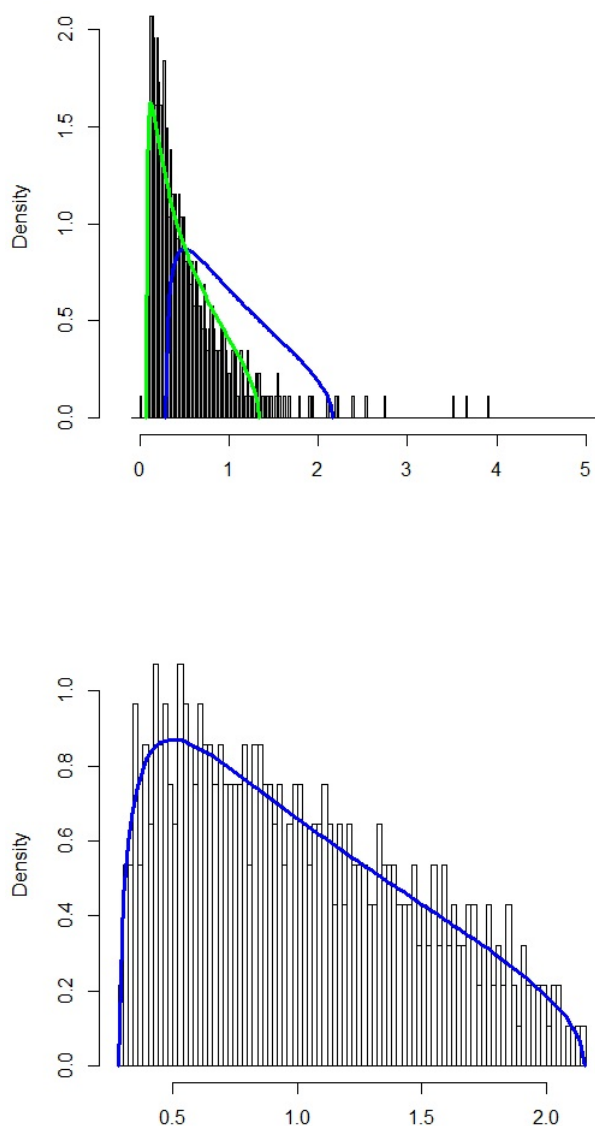
kde λ_+ a λ_- sú rovnaké hranice ako v (2.7).

Toto rozdelenie budeme, v súlade s doterajším postupom, považovať za nulovú hypotézu a budeme analyzovať odchýlky od neho v snahe odhaliť nenáhodnú štruktúru korelácií medzi akciami. Táto krivka vyjadruje „rozostrenie“ skutočného spektra matice C_R , v ktorom sú všetky vlastné hodnoty rovné jednej. Toto „rozostrenie“ závisí od pomeru q , ktorý vyjadruje koľko dát máme k dispozícii a v limite pre q idúce do nuly je toto rozdelenie totožné s Diracovým rozdelením v bode jedna. Obr. 2.4 zachytáva porovnanie hustoty vlastných hodnôt matice E' s Marčenkovým–Pasturovým rozdelením. Tmavá krivka reprezentuje hustotu rozdelenia, tak ako ho popisuje vzorec (2.8). Svetlá krivka popisuje hustotu rozdelenia, pri ktorom sme optimalizovali hornú a dolnú hranicu definičného oboru, aby sme dosiahli čo najpresnejší popis dát. Obr. 2.4 zachytáva aj porovnanie empirickej hustoty vlastných hodnôt korelačnej matice kontrolnej vzorky s teóriou predpovedaným rozdelením.

Posúvanie hraníc teóriou predpovedanej hustoty vlastných hodnôt matice E je problematický krok, pretože dolná a horná hranica v (2.7) sú jasne určené identickou korelačnou maticou a pomerom $q = N/T$. Motivácia, ktorá nás k pozmeneniu vzorca vedie, je však práve snaha zistiť rozdiel medzi „skutočnou“ korelačnou maticou C a idealizovanou identitou. Konkrétne posunutie, ktoré vedie k lepšiemu „napasovaniu“ krivky na dáta, teda môže niesť informáciu o matici C .

Štruktúra empirických vlastných hodnôt pripomína teoretické rozdelenie, najmä pri optimalizovaní hraníc intervalu, ale existujú tu odchýlky. Asi najpodstatnejšou je prítomnosť pomerne veľa vlastných hodnôt mimo definičného oboru oboch teoretických rozdelení. Tieto vlastné hodnoty naznačujú prítomnosť štruktúry a netriviálnych korelácií medzi akciami.

Vlastné hodnoty, ktoré sa nachádzajú pod teoretickou krivkou hustoty, nevieme odlíšiť od náhodného šumu. Podľa teórie náhodných matíc sú konzistentné s náhodným „rozostrením“ spektra identickej korelačnej matice a bez ďalších informácií by sme o nich mohli vyhlásiť, že zodpovedajú vlastným hodnotám rovným jednej v matici C .



Obr. 2.4: Rozloženie vlastných hodnôt empirickej a náhodnej korelačnej matice porovnané s Marčenkovým–Pasturovým rozdelením

2.4 Vlastné vektory očistenej korelačnej matice E'

Ako sme už uviedli v časti 2.2, vlastné vektory korelačnej matice reprezentujú portfóliá s nezávislými výnosmi. Vlastné vektory, ktoré prislúchajú vlastným hodnotám mimo teoretického rozdelenia (2.8), nesú informáciu o nenáhodnej štruktúre trhu. Význačné hodnoty sa nachádzajú na oboch stranách rozdelenia, teda existujú aj extrémne vysoké

aj extrémne nízke.

Vlastný vektor zodpovedajúci najväčšej vlastnej hodnote matice E , ktorý sme úpravou na maticu E' posunuli na koniec spektra, zodpovedal správaniu trhu ako celku. Ostatné vlastné vektory a im prislúchajúce vlastné hodnoty sa úpravou matice E na E' prakticky nezmenili. Pri skúmaní starších dát v prácach [2, 3] sa vlastné vektory zodpovedajúce veľkým vlastným hodnotám dali interpretovať ako jednotlivé odvetvia ekonomiky. Výsledok našej analýzy bol iný, akcie ktoré mali v týchto portfóliách najväčšie zastúpenie pochádzali z rôznych odvetví.

Vlastné vektory, zodpovedajúce najmenším vlastným hodnotám, reprezentujú portfóliá s nízkym rizikom. Sú to špecializované portfóliá typu $(0, \dots, 1/\sqrt{2}, \dots, -1/\sqrt{2}, \dots, 0)$. Pričom akcie na nenulových pozíciách sa hýbu proti sebe. Príkladom takejto dvojice by mohlo byť portfólio v pozícií long na Pepsi a short na Coca-Colu. Takéto portfólio pohltí všetky efekty trhu a odvetvia, jediné čo sa na výnosoch prejaví sú udalosti výrazne odlišujúce dané značky. Takéto vlastné vektory skutočne vznikajú [3] a to okolo akcií, ktoré majú najväčšiu mieru absolútnej korelácie $|E'_{i,j}|$. Pri našej analýze vznikajú aj portfóliá zmiešané, ktoré obsahujú priemerne 4 a viac výrazných zložiek. Tento fakt prisudzujeme tomu, že extrémne zložky matice E' , v našom prípade, sú k sebe výrazne bližšie ako v [3].

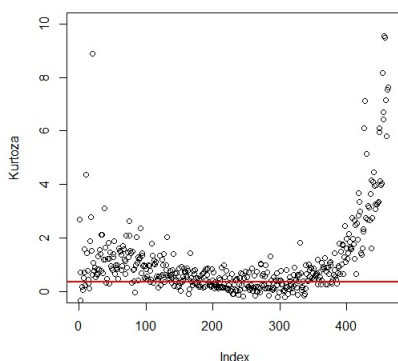
O vlastných hodnotách, ktoré sa vymykajú teoretickému rozdeleniu (2.8), vieme s istotou povedať, že nesú výpovednú hodnotu. Sú teda dôkazom prítomnosti nenáhodnej štruktúry a ukázali sme, že majú aj reálnu interpretáciu. Ďalej nás zaujíma, ako rozhodnúť či vlastné vektory tých vlastných hodnôt, ktoré nevybočujú z teoretického rozdelenia a neodlišujú sa od kontrolnej vzorky, sú len náhodné alebo nesú informačnú hodnotu. Empiricky sme overili, že zložky vlastných vektorov v kontrolnej vzorke sa riadia normálnym rozdelením s približne nulovou strednou hodnotou a odchýlkou $\sigma_{\alpha,i} \approx 0.046$. Vlastné vektory, ktoré sa od takéhoto rozdelenia odlišujú, môžu niesť informačnú hodnotu.

Ďalším spôsobom hľadania význačných vlastných vektorov, použitým v práci [2], je využitie kurtózy (anglicky kurtosis) definovanej ako

$$K_{\alpha} = \frac{\frac{1}{N} \sum_{i=1}^N (V_{\alpha,i} - \bar{V}_{\alpha})^4}{\frac{1}{N} (\sum_{i=1}^N (V_{\alpha,i} - \bar{V}_{\alpha})^2)^2} - 3. \quad (2.9)$$

Kurtóza, ako štvrtý centrálny moment, je v prípade normálneho rozdelenia rovná nule.

Kontrolné vlastné vektory majú hodnoty kurtózy blízke nule. Extrémne kladné hodnoty kurtózy vo vlastných vektoroch matice E' , teda ukazujú na vektory, ktoré majú inú štruktúru ako kontrolná vzorka. Obr 2.5 zachytáva hodnoty kurtózy nadobúdané v skúmaných vlastných vektoroch prislúchajúcich matici E' . Vodorovná hranica vyjadruje 95 percentý kvantil hodnôt kurtózy nadobúdaných v kontrolnej vzorke. Hodnoty nad hranicou sú signifikantnými odchýlkami od kontrolnej vzorky na hladine významnosti päť percent. Táto empirická hranica pochádza z viacerých simulácií s dátami rozmerov $T \times N$. Pri tomto teste vychádza skoro 30 percent vektorov nesignifikantných, teda zodpovedajúcich náhodnému šumu. Ďalších viac ako 20 percent sa nachádza v bezprostrednej blízkosti hranice signifikantnosti. Toto kritérium teda hovorí, že viac ako polovica vlastných vektorov je vlastnosťami veľmi blízka náhodnému šumu v kontrolnej vzorke.



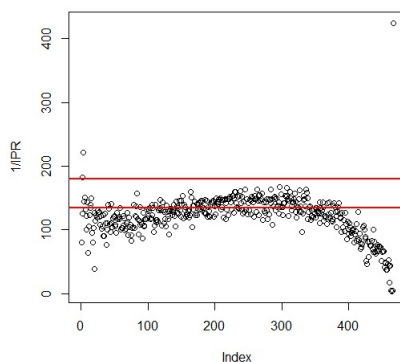
Obr. 2.5: Kurtóza vlastných vektorov matice E'

Práca [3] ponúka podobný spôsob, založený na veličine IPR, skratka z anglického Inverse Participant Ratio, definovanej ako

$$I_\alpha = \sum_{i=1}^N V_{\alpha,i}^4. \quad (2.10)$$

Dá sa vidieť, že ak je v normalizovanom portfóliu iba jedna akcia s váhou 1, tak $I_\alpha = 1$. Na druhej strane ak sú v portfóliu zastúpené všetky akcie rovnako, s váhou $1/\sqrt{N}$, bude platiť $I_\alpha = 1/N$. Táto veličina teda hovorí o obrátenej hodnote počtu výrazných prvkov vo vektore V_α . Odtiaľ pochádza jej anglický názov. Obr. 2.6 ukazuje hodnoty $1/I_\alpha$ pre naše dáta. Hranice opäť predstavujú empirickú hranicu signifikantnosti odchýlky od

kontrolnej vzorky. Podľa tohto kritéria je v súlade s kontrolnou vzorkou približne 35 percent a ďalších 20 percent je k hraniciam blízko. Výsledok je podobný ako v prípade kurtózy, polovica vlastných vektorov sa nelíši výrazne od náhodného šumu v kontrolnej vzorke.



Obr. 2.6: Hodnoty $1/I_\alpha$ pre vlastné vektory matice E'

Obe uvedené metódy vedú k podobným výsledkom. Ako je vidieť aj z obrázkov 2.5 a 2.6, medzi vlastnými vektormi matice E' sa nachádzajú také, ktoré sú v súlade s kontrolnou vzorkou. Tieto sú koncentrované v strednej časti spektra. Na okrajoch spektra sa však nachádza netriviálne množstvo vlastných vektorov, ktoré sú od kontrolnej vzorky signifikantne odlišné. Reálnu finančnú interpretáciu pre vektory na bezprostredných okrajoch spektra sme už ponúkli. Čím ďalej od krajov spektra sa však posúvame, tým ťažšie reálnu interpretáciu hľadať. Zhluky blízkych vlastných hodnôt a ich vlastných vektorov nemusia byť vôbec stabilné a to čo vidíme, môže byť s veľkou pravdepodobnosťou iba realizácia náhodného šumu. Vlastné vektory matice E' tvoria ortonormálnu bázu a to znamená, že vlastné vektory zodpovedajúce okrajom spektra ovplyvňujú aj zložky vlastných vektorov v strede spektra. Hodnoty IPR a kurtózy by preto mohli byť optimistické miery informačnej hodnoty vlastných vektorov matice E' . Najmä ak ich hodnoty porovnávame s kontrolnou vzorkou ktorá zodpovedá identickej korelačnej matici, ktorú existencia extrémnych vlastných hodnôt zamietá.

2.5 Stabilita spektrálnych vlastností matice E v čase

V predošlých častiach kapitoly 2 sme sa zaoberali spektrálnymi vlastnosťami korelačnej matice normalizovaných denných výnosov E . Počas analýzy sme niekoľkokrát upozornili na to, že závery, ku ktorým sme dospeli, nemusia byť v čase stabilné. Overovaniu časovej stability v prípade vybraných vlastností sa budeme venovať tu. Konkrétne budeme sledovať stabilitu najväčších vlastných hodnôt a im prislúchajúcich vlastných vektorov. Časovú stabilitu budeme skúmať nasledovným spôsobom.

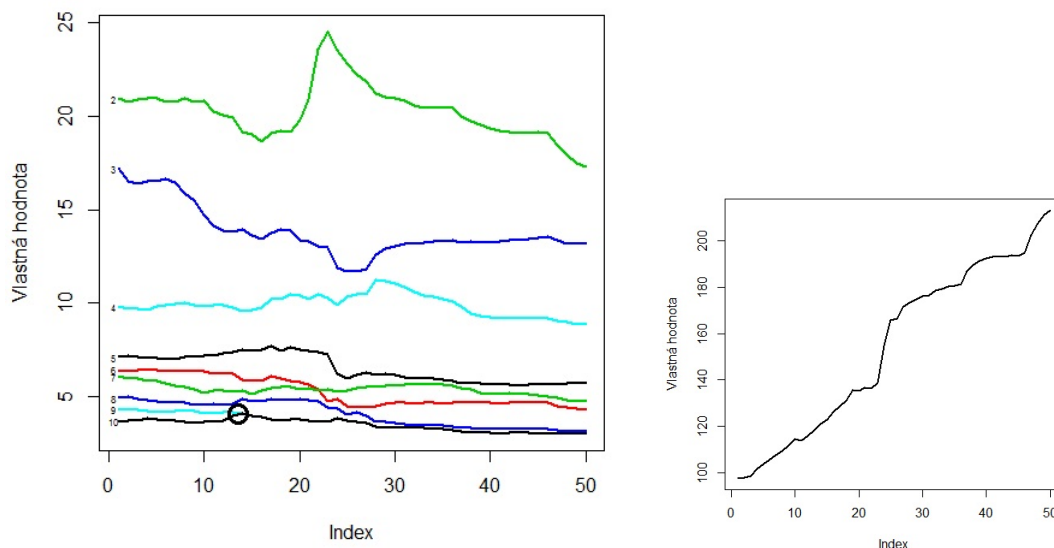
Vlastné hodnoty a vlastné vektory vypočítame pre sériu päťdesiatich empirických korelačných matíc E^τ . Matice X^τ , z ktorých korelačné matice E_τ vypočítame podľa (2.1), pochádzajú z údajov o denných výnosoch v posúvaných periódach. Maximálny počet riadkov T , je v každej z nich tisíc päťsto. Každá matica X^τ , teda zodpovedá šiestim rokom dát. Posledná empirická korelačná matica X^{50} zodpovedá najmladším dátam, tj. jej posledný riadok zodpovedá poslednému dňu v našom súbore 28.12.2011. Matice X^1, \dots, X^{49} obsahujú dáta posunuté o tridsať dní do minulosti, teda posledný riadok matice X^{49} obsahuje informáciu o dni 29.11.2011, a prvý riadok o tridsať dní menej ako v prípade matice X^{50} . Prvý riadok matice X^1 , teda najstaršie dáta, ktoré do tejto analýzy vstupujú sú z dňa 27.1.2000. Všetky takto zostrojené matice X^τ potom očistíme pomocou procesu popísaného v časti 1.1.

Vývojom vlastnej hodnoty λ_α v čase budeme ďalej rozumieť nasledovné. V prvej iterácii vyberieme 10 najväčších vlastných hodnôt matice E^1 a im prislúchajúce vlastné vektory V_1^1, \dots, V_{10}^1 . V každej ďalšej iterácii určíme novú hodnotu λ_α^τ sledovanej vlastnej hodnoty λ_α podľa

$$\lambda_\alpha^\tau = \lambda_\theta, \quad \theta = \operatorname{argmax}_\beta \underbrace{|(V_\beta^\tau)^T V_\alpha^{\tau-1}|}_{\Theta_{\alpha,\tau}}, \quad (2.11)$$

kde V_β^τ sú vlastné vektory matice E^τ . Vyberáme teda vlastnú hodnotu, tak aby vlastné vektory v iteráciách za sebou zvierali čo najmenší uhol. Keďže sú všetky vlastné vektory normalizované na dĺžku jedna, znamená to, že vyberáme najbližšie vektory. Skalárny súčin uprednostňujeme pred normou rozdielu, pretože algoritmus výpočtu vlastného vektora nie je stabilný, čo sa týka znamienka. Zmena znamienka má veľký dopad na normu rozdielu, ale na absolútnu hodnotu skalárneho súčinu nevyplýva. Na obr. 2.7 je vidieť vývoj 10 najväčších vlastných hodnôt v čase. Najväčšia vlastná hodnota je na

samostatnom obrázku, pretože je rádovo vyššia ako ostatné.

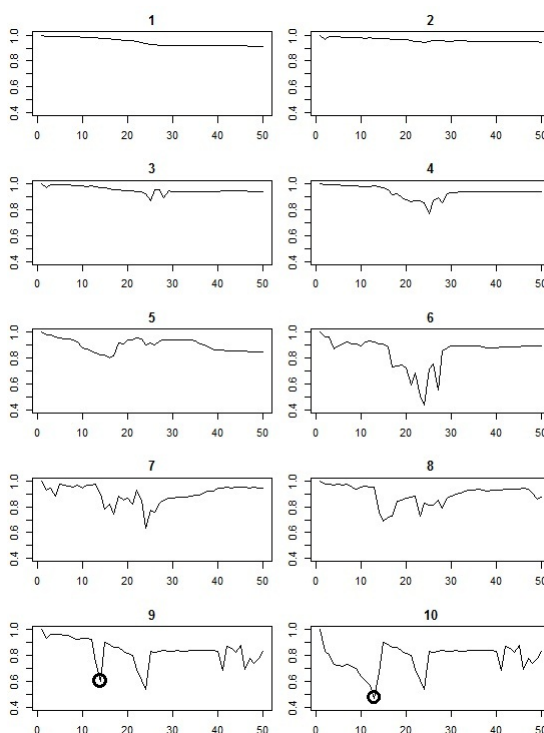


Obr. 2.7: Vývoj vlastných hodnôt $\lambda_1, \dots, \lambda_{10}$ v čase

Najväčšia vlastná hodnota empirickej korelačnej matice sa za posledných 12 rokov výrazne zväčšila na viac ako dvojnásobok, ako je vidieť na obr. 2.7. Ukazuje sa, že niekoľko ďalších vlastných hodnôt je stabilných. Obr. 2.7 však ukazuje aj prípad, kedy dve vlastné hodnoty splynuli do jednej. Konkrétne ide o hodnoty λ_9 a λ_{10} , ktorých stret je označený krúžkom.

Stabilitu vlastných vektorov, ktorým tieto vlastné hodnoty prislúchajú, sledujeme podľa absolútnej hodnoty skalárneho súčinu vektorov v nasledujúcich iteráciách, teda maxima výrazu $\Theta_{\alpha, \tau}$ v (2.11). Vývoj týchto skalárnych súčinov v čase zachytáva obr. 2.8, na ktorom kvôli prehľadnosti uvádzame vývoj pre jednotlivé vlastné vektory samostatne. Krúžky pri vývoji $\Theta_{9, \tau}$ a $\Theta_{10, \tau}$, ukazujú na moment, kedy tieto vlastné vektory splynuli do jedného.

Vlastné vektory zodpovedajúce vlastným hodnotám $\lambda_1, \dots, \lambda_5$ sa ukazujú ako pomaly meniace v čase. Vo vývoji skalárnych súčinov $\Theta_{\alpha, \tau}$ menších vlastných hodnôt sa dosahujú aj nízke hodnoty, ktoré môžu znamenať, že vlastná hodnota „preskočila“ na iný vlastný vektor. To by znamenalo, že výrazné vlastné hodnoty sa objavujú vždy, ale vlastné vektory, ktoré im zodpovedajú, sa môžu skokovo meniť vplyvom významných udalostí.



Obr. 2.8: Skalárne súčiny $\Theta_{\alpha,\tau}$ vlastných vektorov zodpovedajúcich vlastným hodnotám $\lambda_1, \dots, \lambda_{10}$ v jednotlivých iteráciách

Poznamenajme, že pri výbere vlastnej hodnoty λ_α^τ by sme mohli porovnávať vlastné vektory matice E^τ vždy s pôvodným vlastným vektorom, ktoré zodpovedá danej vlastnej hodnote λ_α pre maticu E^1 . Pri takomto postupe sa ukazuje ako extrémne stabilný iba prvý vlastný vektor. V ostatných vlastných vektoroch dochádza v priebehu dvanástich rokov k výrazným zmenám, čo sa prejavuje nízkou hodnotou Θ_α^τ , ktorá je v niektorých prípadoch iba 0.2. Takto nízka hodnota naznačuje, že pôvodný vlastný vektor pravdepodobne zanikol. Vlastné hodnoty λ_α , prislúchajúce takto vybraným vlastným vektorom, sa v čase výrazne menia a „preskakujú“.

Analyzovaním časovej stability spektrálnych vlastností empirickej korelačnej matice denných výnosov E sme dospeli k nasledujúcim záverom. Štruktúra spektra matice E je v čase stabilná v nasledujúcom zmysle. V každej skúmanej perióde existujú, pomerne izolované, veľké vlastné hodnoty, ktoré sa nachádzajú mimo teoretických hraníc daných Marčenkovým–Pasturovým rozdelením (2.7). Veľkosť týchto vlastných hodnôt sa v čase výrazne nemení, s výnimkou najväčšej. Vlastné vektory, ktoré týmto vlastným hodnotám zodpovedajú, v čase svoju štruktúru menia. Zdá sa, že ide o postupný proces,

s výnimkou krízových období, kedy dochádza k prudkým zmenám zloženia vlastných vektorov.

Vlastné hodnoty λ_α empirickej korelačnej matice E zodpovedajú volatilité portfólií definovaných jej vlastnými vektormi, ako sme ukázali v (2.4). Ukázali sme, že za posledných šesť rokov vzrástla najväčšia vlastná hodnota na viac ako dvojnásobok. Finančný trh ako celok je, pravdepodobne kvôli finančnej kríze a jej dôsledkom, dnes dva krát volatilnejší ako bol pred šiestimi rokmi. Tento poznatok vychádza zo súborov dát, ktoré každý zodpovedajú šiestim rokom obchodovaných dní. Aj v najnovších dátach je teda stále obsiahnutá kríza. Je možné, že súčasná volatilita by bola nižšia, keby sme brali do úvahy menšie súbory, pri ktorých by v posledných súboroch už nebol priamy vplyv krízy.

Vlastné vektory matice E prislúchajúce ostatným významným vlastným hodnotám sa v prácach [1, 2, 3] stotožňujú so sektormi trhu. Vo všetkých spomínaných prácach sú analyzované predkrízové dáta. V časti 2.4 sme uviedli, že v našej práci nebolo toto stotožnenie vidieť. Keď sa však pozrieme na vlastné vektory empirickej korelačnej matice zostrojenej zo starších dát pred finančnou krízou, podobný výsledok vychádza aj nám. V minulosti teda portfóliá definované prvými vlastnými vektormi korelačnej matice denných výnosov zodpovedali jednotlivým ekonomickým sektorom. Na základe časovej analýzy sme dospeli k záveru, že tieto „pôvodné“ vlastné vektory sa v čase zmenili.

3 Ďalšie aplikácie

3.1 Použitie pri optimalizácii portfólia

Optimalizácia portfólia je riešenie nasledovného problému. Zostavujeme portfólio z N akcií, tak aby malo požadovaný výnos

$$\Phi = \sum_{i=1}^N w_i g_i, \quad (3.1)$$

kde w_i sú váhy jednotlivých akcií v portfóliu a g_i sú ich očakávané výnosy, ktoré môžeme odhadovať z minulých dát. Zároveň chceme minimalizovať riziko definované ako

$$\sigma_w^2 = \sum_{i=1}^N \sum_{j=1}^N w_i \sigma_i C_{ij} \sigma_j w_j, \quad (3.2)$$

kde C je matica medziakciových korelácií a σ_i predstavuje volatilitu, smerodajnú odchylku, i -tej akcie. Keď pre jednoduchosť predpokladáme, že volatility jednotlivých akcií σ_i poznáme a predefinujeme w_i na $w_i \sigma_i$ a g_i na g_i / σ_i , môžeme vyhlásiť, že $\sigma_i \equiv 1$. Rovnica na výpočet rizika sa tým zjednoduší na

$$\sigma_w^2 = \sum_{ij} w_i C_{ij} w_j = w^T C w. \quad (3.3)$$

Takýto Markowitzovský problém sa dá vyriešiť pomocou Lagrangeových multiplikátorov a vedie k riešeniu

$$w_{Opt} = \Phi \frac{C^{-1} g}{g^T C^{-1} g}. \quad (3.4)$$

Problémom tohto riešenia je, že nepoznáme maticu C , ku ktorej inverzná v rovnici (3.4) vystupuje. Táto matica sa dá odhadnúť z minulých dát pomocou (2.1). Takto odhadnuté portfólio má však v praxi takmer vždy výrazne podcenenú volatilitu. Dôvodom je fakt, že takáto matica E v sebe obsahuje okrem informáciách o skutočných medziakciových koreláciách aj konkrétnu realizáciu náhodného šumu.

Môžeme predpokladať, že informáciu o skutočnej korelačnej matici v sebe nesú tie vlastné hodnoty matice E , ktoré sa líšia od kontrolnej vzorky. Takto inšpirovaná čistiaca metóda potom môže vyzeráť nasledovne. Z vlastných hodnôt, ktoré sa nachádzajú mimo teoretického rozdelenia (2.8), sa dá zostrojiť „očistená“ korelačná matica. Vytvoríme diagonálnu maticu s prvkami $\Lambda'_{ii} = \{\lambda_1, \lambda_2, \dots, \lambda_{10}, \bar{\lambda}, \bar{\lambda}, \bar{\lambda}\}$. Túto transformujeme pomocou ortonormálnej bázy matice E na maticu C' , pričom spoločnú vlastnú

hodnotu $\bar{\lambda}$ definujeme tak, aby platilo $Tr(C') = N$. Túto metódu používajú v práci [3] a vedie k úspešnému zníženiu rozdielu medzi predpovedaným a skutočným rizikom portfólia v nasledujúcej perióde.

Takéto čistenie však neodstráni problémy, ktoré sú prítomné v každom Markowitzovom modeli. Fakt, že budúce výnosy odhadujeme pomocou minulých je veľkým obmedzením týchto modelov, pretože nevie predpokladať „krízu“, teda náhlu zmenu v správaní akcie či trhu. Rovnako korelácie medzi akciami nemusia byť v čase stabilné ako sme už spomínali v časti 1.2.

Pre zaujímavosť spomenieme ešte čistiacu metódu spomínanú v práci [2], ktorá nesúvisí priamo s teóriou náhodných matíc, ale ilustruje tu spomínaný problém vychádzania z minulosti. Táto metóda spočíva v tom, že korelačnú maticu vstupujúcu do optimalizácie odhadne ako

$$C' = (1 - \alpha)E + \alpha I_N, \quad 0 \leq \alpha \leq 1. \quad (3.5)$$

Ukazuje sa, že táto metóda pri hodnote parametra $\alpha = 1/2$ konkuruje a v niektorých prípadoch je vhodnejšia ako uvádzaná metóda vychádzajúca z teórie náhodných matíc [2]. Pod kvalitou metódy sa v tomto prípade rozumie presnosť odhadu budúcej volatility portfólia v porovnaní so skutočnosťou. Na koeficient α sa v tejto metóde dá pozeráť ako na mieru nedôvery k matematickému modelu vychádzajúcemu z minulosti v prospech modelu vychádzajúcemu z predpokladu, že akcie sa hýbu bez vzájomných korelácií, prípadne že akékoľvek korelácie sú náhodné a v čase nestabilné. V spomínanej práci porovnávajú štyri rôzne metódy čistenia korelačnej matice.

3.2 Singulárny rozklad matice X

Singulárny rozklad obdĺžnikovej matice M je vo všeobecnosti daný ako

$$M_{T \times N} = U \Sigma V^T = \sum_i \sigma_i U_i V_i^T, \quad (3.6)$$

kde $U_{T \times T}$ aj $V_{N \times N}$ sú ortogonálne matice, σ_i sú singulárne hodnoty matice M a U_i, V_i sú i -te stĺpce príslušných matíc. Ďalej platí, že U je ortogonálna matica, tvorená bázou vlastných vektorov matice MM^T a matica V je tvorená bázou vlastných vektorov matice $M^T M$. Singulárne hodnoty matice M sú rovné odmocninám spoločných vlastných hodnôt matíc MM^T a $M^T M$, ktoré sú všetky nezáporné reálne čísla.

V našej analýze sa môžeme pozrieť na singulárny rozklad matice normalizovaných denných výnosov X , ktorých korelácie skúmame v kapitole 2. Ortonormálna báza vlastných vektorov matice $X^T X$ je totožná s ortonormálnou bázou matice E analyzovanou v spomínanej kapitole. Vlastné hodnoty sa líšia len pre násobením konštantou $1/T$ ako je vidieť z (2.1).

Nenulové vlastné hodnoty XX^T sú rovnaké ako vlastné hodnoty $X^T X$. Upozorňujeme, že matica XX^T nie je korelačnou maticou obchodovaných dní ako by sa mohlo zdať. Matica X^T totiž nemá normalizované stĺpce v zmysle uvedenom v 1.1. Vlastné vektory matice XX^T predstavujú časový vývoj výnosov v jednotlivých dňoch. Napríklad vlastný vektor zodpovedajúci najväčšej vlastnej hodnote λ_1 je násobkom vektora R_1 výnosov trhu z rovnice (2.6), ktorý sme použili pri odstraňovaní vplyvu najväčšej vlastnej hodnoty v časti 2.2. Podrobnejšia analýza a následná lepšia interpretácia finančného významu ostatných vlastných vektorov matice XX^T by mohla byť predmetom budúceho výskumu.

Záver

Cieľom tejto práce bolo oboznámiť čitateľa s niektorými výsledkami teórie náhodných matíc pomocou ich aplikácie pri analýze korelácií medzi výnosmi akcií na finančnom trhu. Analyzujeme dáta o akciách, ktoré tvorili index S&P500 s dátami za posledných dvadsať rokov. V kapitole 1 uvádzame presný popis štruktúry a vstupných úprav týchto dát. Upozorňujeme aj na obmedzenia použitého prístupu. Ide najmä o problém pravdepodobnej nestability korelácií medzi akciami počas dlhých období. V kapitole 1 sme definovali aj náhodne vygenerovanú kontrolnú vzorku nezávislých výnosov, ktorej správanie je nulovou hypotézou pri overovaní existencie signifikantných korelácií v akciových dátach.

V kapitole 2 sa venujeme skúmaniu korelačnej matice denných výnosov E . V časti 2.1 sme analyzovali jej zložky, ktoré predstavujú odhady korelačných koeficientov párov akcií. Tieto sú v porovnaní s kontrolnou vzorkou výrazne väčšie, čo poukazuje na existenciu netriviálnych korelácií v rámci trhu.

Pomocou analýzy hlavných komponentov a výsledkov teórie náhodných matíc skúmame spektrálny rozklad empirickej korelačnej matice denných výnosov E . Medzi vlastnými hodnotami empirickej korelačnej matice E existujú také, ktoré sa nachádzajú mimo Marčenkovho–Pasturovho rozdelenia (2.8), ktorým sa riadia vlastné hodnoty v kontrolnom prípade. Takéto významné vlastné hodnoty sa nachádzajú na oboch stranách spektra, teda sú aj vysoké aj nízke.

Vlastné vektory, ktoré im zodpovedajú sme analyzovali v časti 2.4. Dospeli sme k rovnakému záveru ako v [1, 2, 3], že vlastný vektor zodpovedajúci najväčšej vlastnej hodnote predstavuje „trhový mód“, tj. portfólio bez výraznejšej diverzifikácie, v ktorom sú všetky akcie zastúpené približne rovnako. Najväčšia vlastná hodnota je rádovo vyššia ako všetky ostatné a preto sme jej vplyv odstránili a vytvorili novú upravenú korelačnú maticu denných výnosov E' .

Upravená korelačná matica E' má rovnaké spektrálne vlastnosti ako pôvodná, okrem odstránenej vlastnej hodnoty, ktorá je v upravenej korelačnej matici E' rovná nule s pôvodným vlastným vektorom. Významná zmena je však v rozdelení zložiek upravenej korelačnej matice E' , ktoré predstavujú odhady korelačných koeficientov výnosov, z ktorých bol odstránený najsilnejší spoločný vplyv. Tu došlo v porovnaní s pôvod-

nou maticou E k výraznému poklesu medziakciových korelácií. Tento fakt nás vedie k záveru, že veľkú časť korelácií medzi akciami spôsobuje práve správanie trhu ako celku. Pre úplnosť dodáme, že rozdelenie zložiek upravenej korelačnej matice E' sa stále významne líšilo od normálneho rozdelenia kontrolnej vzorky.

Vlastné vektory zodpovedajúce najväčším vlastným hodnotám upravenej korelačnej matice, ktoré sa v spomínaných prácach stotožňujú s odvetviami ekonomiky, v našej práci túto štruktúru nemajú. Dochádza v nich k premiešaniu viacerých odvetví. V časti 2.5 analyzujeme stabilitu týchto vlastných vektorov v čase. Ukazuje sa, že v období pred finančnou krízou vlastné vektory zodpovedajúce vysokým vlastným hodnotám približne zodpovedali odvetviam, avšak v priebehu posledných niekoľko rokov došlo k zmene ich štruktúry, ktorú pripisujeme práve finančnej kríze.

V kapitole 2 sme uviedli, že okrem extrémne vysokých vlastných hodnôt empirickej korelačnej matice denných výnosov E existujú aj extrémne nízke. Tieto podľa prác [2, 3] zodpovedajú párom akcií, prípadne nejakým malým skupinám, ktoré sú úzko previazané. Skúmaniu tejto časti spektra sme sa v práci nevenovali podrobnejšie, ale považujeme to za jej možné pokračovanie.

Podobne sme v časti 2.4 ukázali, že aj v strednej časti spektra, anglicky „bulk“, sa nachádzajú vlastné vektory, ktoré sú významne odlišné od kontrolnej vzorky. Štúdium týchto vlastných vektorov, by mohlo priniesť viac informácií o skutočnej korelačnej matici C .

V kapitole 3 uvádzame ďalšie možné aplikácie výsledkov kapitoly 2. Výsledky teórie náhodných matíc sa dajú použiť pri čistení korelačnej matice denných výnosov E . Proces odstraňovania náhodného šumu z korelačnej matice výnosov E je dôležitý pri riešení klasického Markowitzovského optimalizačného problému. Možnému použitiu v tejto oblasti je venovaná časť 3.1.

Prínosom práce pre čitateľa by mohlo byť pomerne nenáročné oboznámenie s niektorými výsledkami teórie náhodných matíc, ktoré v práci priamo ilustrujeme pomocou kontrolnej vzorky. Tieto tvrdenia používame priamo pri skúmaní finančného trhu so snahou analyzovať jeho štruktúru.

Analýza uvádzaná v práci nie je komplexnou ani hĺbkovou analýzou finančného trhu. Je skôr len nahliadnutím do možností, ktoré teória náhodných matíc ponúka pri takejto

analýze a poukazuje na obmedzenia, ktoré z výsledkov teórie náhodných matíc plynú. V časti 2.4 sme napríklad ukázali, že podstatné množstvo vlastných vektorov, ktoré sú výsledkom analýzy hlavných komponentov zodpovedá náhodnému šumu. Možným pokračovaním analýzy je skúmanie korelácií medzi menším množstvom akcií, napríklad akcií vrámci jedného sektoru. Dalo by sa skúmať či sa spektrálny rozklad, najmä vlastné vektory, tejto menšej vzorky nejakým spôsobom prejavajú aj v spektrálnom rozklade korelačnej matice celého trhu.

Ako otvorenú otázku nechávame aj správanie spektra vlastných hodnôt empirickej korelačnej matice E pre hodnoty skutočnej korelačnej matice C odlišné od identity. Teda skúmanie toho ako sa mení hustota vlastných hodnôt matice E daná Marčenkovým–Pasturovým rozdelením, keď skutočná korelačná matica má tvar ako matica C' zo vzorca (3.5).

Teória náhodných matíc má omnoho širší záber ako tvrdenia, ktoré uvádzame v našej práci. Okrem reálnych symetrických matíc sa teória náhodných matíc zaoberá aj komplexnými Hermitovskými, reálnymi ortogonálnymi a komplexnými unitárnymi maticami, pre ktoré popis ich vlastností patrí medzi klasické výsledky matematiky a venuje sa im množstvo literatúry. Tvrdeniam tohto typu sme sa v práci nevenovali, pretože nesúvisia priamo s problematikou, ktorú sme rozoberali. Navyše plnohodnotné pochopenie týchto tvrdení spolu s ich odvodením a dôkazmi kladie veľký dôraz na mnoho odvetí matematiky, medzi ktoré okrem iných patria teória pravdepodobnosti, štatistika, komplexná analýza a lineárna algebra. Teória náhodných matíc vďaka nástrojom, ktoré ponúka, je využiteľná a využívaná v mnohých odvetviach nielen teoretickej matematiky, ale aj v aplikovaných smeroch.

Zoznam použitej literatúry

- [1] Akihiko U., Kazusumi I., Masaki O.: *Random Matrix Theory Analysis of Cross Correlations in Financial Markets*, Phys. Rev. E **70** 26110, 2004, dostupné na internete (19.5.2012):
<http://arxiv.org/pdf/cond-mat/0312643v1>
- [2] Bouchard J. P., Potters M.: *Financial Applications of Random Matrix Theory: a short review*, Science & Finance, Paris, 2009 dostupné na internete (19.5.2012):
<http://arxiv.org/pdf/0910.1205v1.pdf>
- [3] Plerou, V. et al.: *A Random Matrix Approach to Cross-Correlations in Financial Data*, Phys. Rev. E **65** 066126, 2002, dostupné na internete (19.5.2012):
<http://arxiv.org/pdf/cond-mat/0108023v1.pdf>
- [4] Potters M., Bouchard J. P.: *Financial Applications of Random Matrix Theory: Old Laces and New Pieces*, Science & Finance, Paris, 2005, dostupné na internete (19.5.2012):
<http://arxiv.org/pdf/physics/0507111v1.pdf>
- [5] Tao T.: *Topics in random matrix theory*, Department of Mathematics UCLA, Los Angeles, 2011
- [6] *Zoznam firiem v akciovom indexe S&P*, dostupné na internete (20.5.2012)
http://en.wikipedia.org/wiki/List_of_S%26P_500_companies
- [7] *Historické údaje o cenách akcií*, dostupné na internete (20.5.2012)
<http://finance.yahoo.com>