Razvoj ortogonalnih portfelja pomoću metode glavnih komponenti

Krilčić, Filip

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1559

RAZVOJ ORTOGONALNIH PORTFELJA POMOĆU METODE GLAVNIH KOMPONENTI

Filip Krilčić

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1559

RAZVOJ ORTOGONALNIH PORTFELJA POMOĆU METODE GLAVNIH KOMPONENTI

Filip Krilčić

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 4. ožujka 2024.

ZAVRŠNI ZADATAK br. 1559

Pristupnik:	**Filip Krilčić (0036540474)**
Studij:	Elektrotehnika i informacijska tehnologija i Računarstvo
Modul:	Računarstvo
Mentor:	prof. dr. sc. Zvonko Kostanjčar

Zadatak:	**Razvoj ortogonalnih portfelja pomoću metode glavnih komponenti**

Opis zadatka:

U sklopu završnog zadatka potrebno je istražiti primjenu metode glavnih komponenti (PCA) u razvoju ortogonalnih portfelja kao strategije diverzifikacije investicijskog portfelja. Prvo je potrebno opisati metodu glavnih komponenti te istražiti njezine ključne karakteristike. Zatim je potrebno pomoću metode glavnih komponenti kreirati ortogonalne portfelje na zadanom skupu financijskih vrijednosnica. Na kraju je potrebno analizirati kako metoda glavnih komponenti može identificirati nekorelirane komponente portfelja kako bi se smanjio sistemski rizik.

Rok za predaju rada: 14. lipnja 2024.

Sadržaj

Uvod

Diverzifikacija investicijskog portfelja predstavlja ključnu strategiju u upravljanju rizikom te očuvanju kapitala pri investiranju. Ulaganjem u različite financijske instrumente koji imaju malo ili nimalo korelacije međusobno, investitori mogu smanjiti volatilnost, tj. nagle promjene unutar svojih portfelja i na taj način zaštititi se od gubitaka. Niska korelacija između različitih financijskih instrumenata zapravo znači da različiti financijski instrumenti imaju različite cjenovne kretnje te da drukčije reagiraju na uvjete u tržištu. Diverzifikacija omogućava raspodjelu rizika na više različitih imovina, čime se smanjuje utjecaj negativnih događaja na cjelokupni portfelj. Ovaj pristup ne samo da pomaže u očuvanju kapitala, već također omogućava investitorima da iskoriste prilike na različitim tržištima. Smanjenje rizika, osim što služi očuvanju kapitala i dobiti, ključni je faktor za olakšavanje psihološke strane investiranja. Naime, investitori su često emocionalno povezani sa svojim novcem te zbog toga može doći do naglih i nepromišljenih odluka prilikom investiranja. Što je rizik manji, veća je sigurnost i pouzdanost investicije, a samim time se smanjuje strah te se lakše donose racionalne investicijske odluke. Osim toga, stabilniji portfelj omogućava dugoročno planiranje i postizanje financijskih ciljeva s većom sigurnošću.

Jedna od tehnika koja se koristi za postizanje diverzifikacije portfelja je metoda glavnih komponenti (Principal Component Analysis - PCA). PCA je statistički alat koji se koristi za smanjenje dimenzionalnosti podataka, čime se smanjuje kompleksnost višedimenzionalnih podataka dok se istovremeno zadržavaju bitne informacije o strukturi i trendovima podataka. Time se omogućuje olakšana analiza te identificiranje skrivenih struktura u velikim skupovima podataka. Korištenjem PCA metode, moguće je transformirati skup financijskih vrijednosnica u manje skupove nekoreliranih komponenti, što omogućava kreiranje ortogonalnih portfelja. Ortogonalni portfelji, koji se sastoje od nekoreliranih komponenti, smanjuju sistemski rizik i poboljšavaju stabilnost portfelja. Osim toga, PCA omogućava bolju interpretaciju i vizualizaciju podataka, što olakšava donošenje investicijskih odluka temeljeno na analitičkim rezultatima. Na taj način investitori mogu bolje razumjeti dinamiku tržišta i prilagoditi svoje strategije u skladu s identificiranim trendovima.

Cilj ovog istraživanja je napraviti analizu varijanci povrata portfelja, očekivanih povrata portfelja te sposobnosti predviđanja PCA metode, a detaljnije analizirati primjenu metode glavnih komponenti u razvoju ortogonalnih portfelja kao strategije diverzifikacije investicijskog portfelja. Prvo će biti opisane osnove i ključne karakteristike metode PCA. Zatim će biti demonstrirano kako se PCA može koristiti za kreiranje ortogonalnih portfelja na zadanom skupu financijskih vrijednosnica. Na kraju, analiza će pokazati kako se metoda glavnih komponenti ponaša ovisno o određenim parametrima te će se ispitati stacionarnost tržišnih uvjeta. Ova analiza pružit će uvid u učinkovitost PCA metode u različitim tržišnim uvjetima i pomoći u identifikaciji optimalnih strategija za upravljanje portfeljem.

Rad je podijeljen u 3 poglavlja. U prvom poglavlju bit će opisani financijski vremenski nizovi, njihova važnost, svojstva i metode analize. U drugom poglavlju će biti predstavljen teorijski okvir PCA metode, uključujući matematičku osnovu i ključne karakteristike. U trećem poglavlju bit će prikazana provedena analiza te dobiveni rezultati. Dokument će završiti zaključkom i pregledom glavnih nalaza istraživanja.

1. Financijski vremenski nizovi

1.1. Definicija financijskih vremenskih nizova

Financijski vremenski nizovi predstavljaju sekvencu podataka prikupljenih i zabilježenih tijekom određenog vremenskog razdoblja, koji se odnose na različite financijske instrumente i pokazatelje. Oni su temelj za mnoge vrste financijskih analiza jer omogućuju praćenje i proučavanje promjena u vrijednosti imovine te generalno u ekonomskim performansama kroz vrijeme.

Financijski vremenski nizovi uključuju širok spektar podataka, a neki od najznačajnijih su:

- **Cijene dionica**
  - Predstavljaju promjene u vrijednosti dionica pojedinih tvrtki. Ovi vremenski nizovi su ključni za analizu tržišta dionica, jer odražavaju performanse i percepciju vrijednosti tvrtke od strane investitora.
- **Valutni tečajevi**
  - Ovi nizovi prikazuju vrijednost jedne valute u odnosu na drugu, što je ključno za analizu deviznog tržišta (Forex). Valutni tečajevi odražavaju ekonomske uvjete, političke događaje i očekivanja tržišta vezana uz gospodarsku stabilnost zemalja.
- **Kamatne stope**
  - Vrijednosti kamatnih stopa, kao što su stope državnih obveznica ili međubankarske stope, ključni su pokazatelji monetarne politike i ekonomske stabilnosti. Promjene kamatnih stopa utječu na trošak kapitala i investicijske odluke.
- **Indeksi tržišta**
  - To su kompozitni pokazatelji koji prate performanse određenih segmenata tržišta, poput S&P 500 za američko tržište dionica ili MSCI World za globalno tržište. Indeksi pružaju sveobuhvatan pogled na tržišna kretanja i raspoloženje investitora.
- **Volumen trgovanja**
  - Predstavlja količinu financijskih instrumenata koji su promijenili vlasnika tijekom određenog razdoblja. Visok volumen može ukazivati na visoku razinu interesa i

aktivnosti na tržištu, dok nizak volumen može signalizirati nedostatak interesa ili nesigurnost među investitorima.

Financijski vremenski nizovi predstavljaju temeljni alat u financijskoj analizi, omogućujući detaljno praćenje i razumijevanje kretanja različitih financijskih instrumenata kroz vrijeme. Oni pružaju neophodne podatke za identifikaciju trendova, procjenu rizika i donošenje informiranih investicijskih odluka, čime se povećava efikasnost i uspješnost ulaganja.

1.2. Karakteristike financijskih vremenskih nizova

Financijski vremenski nizovi imaju nekoliko ključnih karakteristika koje analitičari moraju razumjeti kako bi učinkovito analizirali tržište i donosili informirane investicijske odluke. Neke od najpoznatijih karakteristika uključuju volatilnost, trendove, sezonalnost, stacionarnost i autokorelaciju. Razumijevanje i analiza ovih karakteristika omogućavaju investitorima da bolje predvide buduća kretanja te optimiziraju i prilagode svoje strategije ulaganja.

Volatilnost se odnosi na stupanj varijacije cijena kroz vrijeme i predstavlja mjeru rizika povezanog s određenom imovinom. Visoka volatilnost ukazuje na velike promjene cijena u kratkom vremenskom razdoblju, što može značiti veći rizik, ali i potencijalno veće prinose. Analitičari koriste različite statističke mjere, kao što su standardna devijacija i varijanca, za kvantificiranje volatilnosti.

Trendovi su dugoročne smjernice kretanja cijena, bilo prema gore (uzlazni trend) ili prema dolje (silazni trend), koji mogu trajati nekoliko mjeseci ili godina. Identifikacija trendova je ključna za donošenje investicijskih odluka jer omogućuje investitorima da se pozicioniraju u skladu s dugoročnim kretanjima na tržištu. Trendovi se često analiziraju s pomoću tehničkih indikatora, kao što su pokretni prosjeci (Moving average - MA) i trend linije.

Sezonalnost podrazumijeva ponavljajuće obrasce kretanja cijena unutar određenih vremenskih perioda, kao što su sezonske promjene potražnje ili godišnji poslovni ciklusi. Na primjer, cijene energije mogu rasti tijekom zime zbog povećane potražnje za grijanjem. Analizom sezonalnosti, investitori mogu predvidjeti kada će cijene vjerojatno porasti ili pasti, što im omogućuje da iskoriste sezonske prilike za trgovanje.

Stacionarnost se odnosi na svojstvo vremenskog niza čije statističke karakteristike (srednja vrijednost, varijanca, autokorelacija) ostaju konstantne kroz vrijeme. Mnogi financijski vremenski nizovi nisu stacionarni, što znači da njihove statističke karakteristike variraju kroz vrijeme. Za analizu i modeliranje takvih nizova često je potrebno primijeniti metode kao što su diferenciranje ili transformacije podataka kako bi se postigla stacionarnost.

Autokorelacija se odnosi na korelaciju između vrijednosti vremenskog niza u različitim vremenskim točkama. Visoka autokorelacija može ukazivati na prisutnost obrazaca ili sezonskih efekata u podacima. Autokorelacija se mjeri s pomoću autokorelacijske funkcije (ACF), koja prikazuje korelaciju između trenutne vrijednosti i nekoliko vrijednosti unazad. Identifikacija autokorelacija može pomoći u izgradnji boljih prediktivnih modela.

1.3. Metode analize financijskih vremenskih nizova

Postoji nekoliko metoda koje se koriste za analizu financijskih vremenskih nizova, svaka s jedinstvenim pristupom i ciljevima. Te metode pomažu analitičarima i investitorima da identificiraju trendove, predviđaju buduća kretanja cijena i donose informirane investicijske odluke. Glavne metode analize financijskih vremenskih nizova uključuju tehničku analizu, fundamentalnu analizu i analizu sentimenta tržišta.

Tehnička analiza koristi povijesne podatke o cijenama i volumenima trgovanja za identifikaciju obrazaca i trendova koji mogu predvidjeti buduća kretanja cijena. Temelji se na pretpostavci da su svi relevantni faktori koji mogu utjecati na cijenu već ugrađeni u cijenu imovine te da se povijest često ponavlja kroz prepoznatljive obrasce. Kod tehničke analize se koriste razni indikatori poput RSI (Relative Strength Index) i MACD (Moving Average Convergence Divergence) te razni grafovi i interpretacije obrazaca istih.

Fundamentalna analiza fokusira se na procjenu intrinzične vrijednosti imovine kroz analizu financijskih izvještaja (analiza bilance, računa dobiti i gubitka te izvještaji o novčanom toku), ekonomskih pokazatelja (BDP, stope nezaposlenosti, inflacija i kamatne stope) i industrijskih trendova. Cilj je identificirati podcijenjene ili precijenjene imovine koje imaju potencijal za rast ili pad.

Analiza sentimenta tržišta proučava stavove i mišljenja investitora i tržišnih sudionika putem analiza vijesti, društvenih medija i drugih izvora kako bi se razumjele emocionalne reakcije koje mogu utjecati na cijene. Analiza sentimenta tržišta pomaže u predviđanju kratkoročnih promjena u cijenama koje su vođene emocijama i ponašanjem tržišta, a ne temeljnim vrijednostima.

Često se ove metode koriste zajedno kako bi se dobio cjelovitiji uvid u tržišna kretanja. Kombiniranjem različitih pristupa, analitičari i investitori mogu donositi informirane odluke koje uzimaju u obzir različite aspekte tržišta. Ovi integrirani pristupi omogućuju sveobuhvatnu analizu financijskih vremenskih nizova, pružajući dublje razumijevanje tržišnih kretanja i pomažući u optimizaciji investicijskih strategija.

1.4. Izazovi u analizi financijskih vremenskih nizova

Analiza financijskih vremenskih nizova suočava se s brojnim izazovima koji mogu značajno utjecati na proces donošenja investicijskih odluka. Razumijevanje tih izazova ključno je za postizanje preciznih i pouzdanih rezultata.

Nedostatak podataka često predstavlja ključni problem u analizi financijskih vremenskih nizova, posebno u kontekstu manjih tržišta ili novih financijskih instrumenata. Nedostatak povijesnih podataka može otežati identificiranje trendova i procjenu rizika, što dovodi do veće nesigurnosti pri donošenju investicijske odluke.

Financijska tržišta poznata su po svojoj nepredvidljivosti, gdje iznenadni događaji i vanjski utjecaji mogu izazvati nagla i dramatična kretanja cijena. Politika, ekonomske krize, geopolitičke napetosti ili prirodne katastrofe samo su neki od faktora koji mogu destabilizirati tržište i izazvati neočekivane promjene.

Vanjski faktori, poput promjena u zakonodavstvu, tehnološkog napretka ili globalnih ekonomskih trendova, mogu imati značajan utjecaj na financijska tržišta. Ovi faktori mogu promijeniti uvjete trgovanja, regulatorni okvir ili konkurencijski pejzaž, što zahtijeva stalno praćenje i prilagodbu investicijskih strategija.

Unatoč ovim izazovima, temeljita i pažljiva analiza financijskih vremenskih nizova ključna je za razumijevanje tržišnih dinamika i donošenje uspješnih investicijskih odluka. Kombiniranjem različitih metoda analize i pažljivim praćenjem tržišnih trendova, investitori mogu bolje upravljati rizicima i identificirati prilike za profitabilno ulaganje.

2. Analiza glavnih komponenti (PCA)

Analiza glavnih komponentni (Principal Component Analysis – PCA) bavi se objašnjavanjem varijance i kovarijance skupa varijabli tako da iz tih varijabli pronađe nekoliko linearnih kombinacija takvih da one obuhvaćaju što je više moguće varijabilnosti. Ova tehnika ima široku primjenu u različitim područjima, uključujući financije, biologiju, računalstvo i mnoge druge discipline. PCA se koristi kako bi se količina podataka smanjila u svrhu lakše analize samih podataka, ali da pri tome i dalje možemo te podatke pouzdano interpretirati i iz njih donijeti zaključke. Ako skup podataka ima *p* atributa (komponenti), za opisati cijelu varijabilnost tih podataka potrebno je upravo tih *p* komponenti. Često velik broj komponenti ne pridonose uvelike varijabilnosti podataka te se zbog toga može *p* komponenti zamijeniti s manjim brojem *k* komponenti koje i dalje obuhvaćaju velik dio varijabilnosti podataka pa samim procesom smanjivanja broja komponenti nije izgubljena neka značajna količina informacija, a potencijalno je uvelike olakšana analiza. Tih *k* komponenti se tada nazivaju glavne komponente (principal components). Analizom glavnih komponenti se također može otkriti i nove međuovisnosti u podatcima.

2.1. Matematička osnova PCA

Gledano algebarski, glavne komponente su određene linearne kombinacije slučajnih varijabli, dok su geometrijski one zapravo osi novog koordinatnog sustava dobivene rotacijom originalnog koordinatnog sustava. Te nove osi predstavljaju smjerove najveće varijabilnosti. Glavne komponente su neovisne linearne kombinacije koje su međusobno ortogonalne odnosno između njih nema korelacije te se one bilježe redoslijedom od najveće varijance do najmanje varijance.

Izračun glavnih komponenti temelji se na svojstvenim vrijednostima i svojstvenim vektorima matrice kovarijance. Elementi matrice kovarijance se računaju po formuli:

$$Cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \tag{1}$$

gdje je n broj podataka, X_i vrijednost *i-tog* podatka prve varijable, $\bar{X}$ srednja vrijednost prve varijable, Y_i vrijednost *i-tog* podatka druge varijable, $\bar{Y}$ srednja vrijednost druge varijable

Za svojstveni vektor **e** i pripadajuću svojstvenu vrijednost λ neke matrice kovarijance Σ vrijedi:

$$\Sigma\mathbf{e} = \lambda\mathbf{e} \tag{2}$$

$$(\Sigma - \lambda I)\mathbf{e} = 0 \tag{3}$$

Jednadžba (3) ima netrivijalno rješenje kada vrijedi:

$$|\Sigma - \lambda I| = 0 \tag{4}$$

Jednadžba (4) govori da je potrebno izračunati determinantu s nepoznatom vrijednosti λ nakon čega je potrebno tu determinantu izjednačiti s nulom i riješiti dobivenu jednadžbu čime se dobiju vrijednosti λ odnosno svojstvene vrijednosti.

S obzirom na to da je sami redoslijed glavnih komponenti važan, a svaki vektor bi se jednostavno mogao pomnožiti s bilo kojom konstantom kako bi uvećao vrijednost, bitno je da su svi vektori korišteni u računu zapravo vektori jedinične duljine kako bi rezultati bili konzistentni. Stoga mora vrijediti:

$$\mathbf{e}'\mathbf{e} = 1 \tag{5}$$

Nakon što se izračunaju svojstvene vrijednosti potrebno je izračunati pripadajuće svojstvene vektore koji se dobivaju tako da u jednadžbu (3) uvrstimo svaku svojstvenu vrijednost λ_i i riješimo sustav uz (5).

Kada su poznate svojstvene vrijednosti i pripadajući svojstveni vektori, mogu se definirati glavne komponente.

Neka je Σ matrica kovarijance slučajnog vektora $\mathbf{X} = [X_1, X_2, ..., X_p]'$, s parovima svojstvenih vrijednosti i svojstvenih vektora: $(\lambda_1, \mathbf{e_1}),(\lambda_2, \mathbf{e_2}), ...,(\lambda_p, \mathbf{e_p})$, tako da vrijedi $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$. Onda je *i-ta* glavna komponenta:

$$Y_i = \mathbf{e_{i1}}X_1 + \mathbf{e_{i2}}X_2 + ... + \mathbf{e_{ip}}X_p = \mathbf{e_i}'\mathbf{X} \tag{6}$$

Uz to vrijedi da je varijanca glavne komponente zapravo iznosom jednaka svojstvenoj vrijednosti te komponente:

$$\mathrm{Var}(Y_i) = \mathbf{e_i}'\Sigma\mathbf{e_i} = \lambda_i \text{ , } i = 1, 2, ..., p \tag{7}$$

Budući da se promatraju simetrične matrice kovarijance, također vrijedi i svojstvo da su svojstveni vektori međusobno ortogonalni odnosno kovarijanca glavnih komponenti jednaka je nuli:

$$\mathrm{Cov}(Y_i \text{ , } Y_j) = \mathbf{e_i}'\Sigma\mathbf{e_j} = 0, \forall i \neq j. \tag{8}$$

Pogreška na koju treba pripaziti je korištenje korelacijske matrice umjesto matrice kovarijance pri izračunu svojstvenih vrijednosti i svojstvenih vektora jer unatoč tome što nekada rezultat može biti jednak, često se dogodi da nije pa može doći do pogrešne interpretacije. Postupak koji je nekada potrebno provesti je tzv. standardizacija varijabli. To se radi tako da se od varijable oduzme srednja vrijednost te se zatim taj iznos podijeli korijenom varijance odnosno sa standardnom devijacijom:

$$Z_p = \frac{X_p - \mu_p}{\sqrt{\sigma_{pp}}} \tag{9}$$

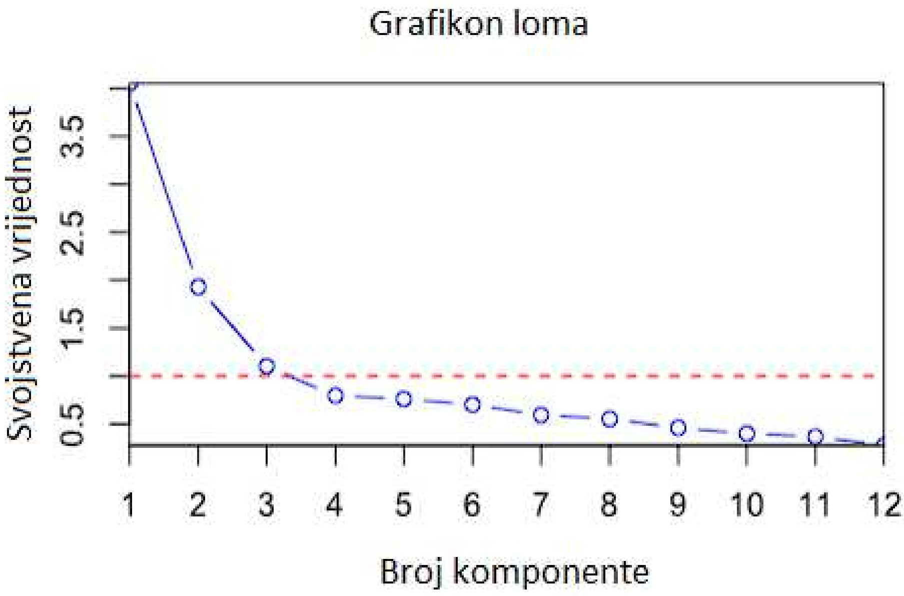
To je potrebno napraviti u slučaju da su mjerenja varijabli iznosima različitog reda veličine kako bi se izbjegla pogrešna interpretacija težina pojedinih varijabli, a standardizacijom se postiže da sve varijable budu na istoj skali što je važno za usporedbu njihovih doprinosa varijabilnosti. Činjenica koju je vrijedno spomenuti je da su korelacijska matrica nastala od početnih podataka i matrica kovarijanci nastala od standardiziranih podataka jednake.

2.2. Identifikacija glavnih komponenti i smanjenje dimenzionalnosti

Proces identifikacije glavnih komponenti u PCA započinje transformacijom originalnih varijabli u novi set varijabli koje se nazivaju glavne komponente. Te komponente su linearne kombinacije originalnih varijabli i konstruirane su tako da maksimiziraju varijancu podataka. Matematički postupak za izračun tih komponenti objašnjen je u prethodnom potpoglavlju. Nakon što su dobivene svojstvene vrijednosti i svojstveni vektori, svojstvene vrijednosti se sortiraju u silaznom redoslijedu, a odgovarajući svojstveni vektori tvore glavne komponente. Prva glavna komponenta predstavlja smjer najveće varijance u podacima, druga glavna komponenta predstavlja smjer najveće varijance preostale nakon prve komponente, i tako dalje. Svaka glavna komponenta je ortogonalna (neovisna) na prethodne komponente.

PCA smanjuje dimenzionalnost podataka projicirajući ih na prostor manje dimenzije, zadržavajući pritom što veći postotak varijabilnosti. Ovaj postupak je koristan jer smanjenje dimenzionalnosti olakšava interpretaciju podataka budući da se značajke svode na nekoliko ključnih komponenti pri čemu se uklanjaju redundantne ili neinformativne varijable, rezultirajući potencijalno kvalitetnijim modelom. Osim toga, rad s manjim brojem varijabli smanjuje računalne zahtjeve za obradu i analizu podataka. Smanjenjem dimenzionalnosti, PCA omogućava istraživačima i analitičarima da se fokusiraju na najvažnije aspekte podataka, poboljšavajući tako donošenje odluka i razumijevanje strukture podataka.

Svaka analiza ima drukčije podatke i ciljeve tj. svaka analiza ima svoje specifičnosti zbog kojih je odluku o tome koje glavne komponente odbaciti, a koje zadržati, teško univerzalno definirati. U svakom slučaju, podatak koji se gleda pri donošenju te odluke je omjer pojedine svojstvene vrijednosti i sume svih svojstvenih vrijednosti. Tim omjerom saznajemo postotak varijance koji ta svojstvena vrijednost tj. pripadajuća glavna komponenta pokriva. Jedan od najčešćih načina odluke odabira broja glavnih komponenti je definiranje udjela varijance podataka koji mora biti objašnjen unutar odabranih glavnih komponenti. Ovisno o analizi, najčešće se postavlja uvjet da je potrebno zadržati između 70% i 90% ukupne varijance. Drugi kriterij koji se često koristi je definiranje minimalnog udjela varijance koji svaka glavna komponenta mora objašnjavati. Primjerice, kriterij može biti zadan tako da se odbacuje svaka glavna komponenta koja objašnjava manje od 5% ukupne varijance, dok se ostale zadržavaju. Treća metoda je korištenje tzv. grafikona loma, poznatijeg kao scree plot koji je prikazan na slici Sl. 2.1. To je graf koji na x osi ima redne brojeve glavnih komponenti, a na y osi svojstvene vrijednosti glavnih komponenti, računajući na to da su silazno poredane po vrijednosti.



Sl. 2.1 Prikaz grafikona loma (scree plot-a) [7]

U grafu tražimo točku nakon koje su sve vrijednosti do kraja relativno slične i male što se može vidjeti naglim pregibom u grafu. Ta točka i sve nakon nje predstavljaju glavne komponente manje važnosti te se smatra da se njih može izbaciti, dok točke prije odabrane točke na pregibu predstavljaju glavne komponente koje se koriste dalje.

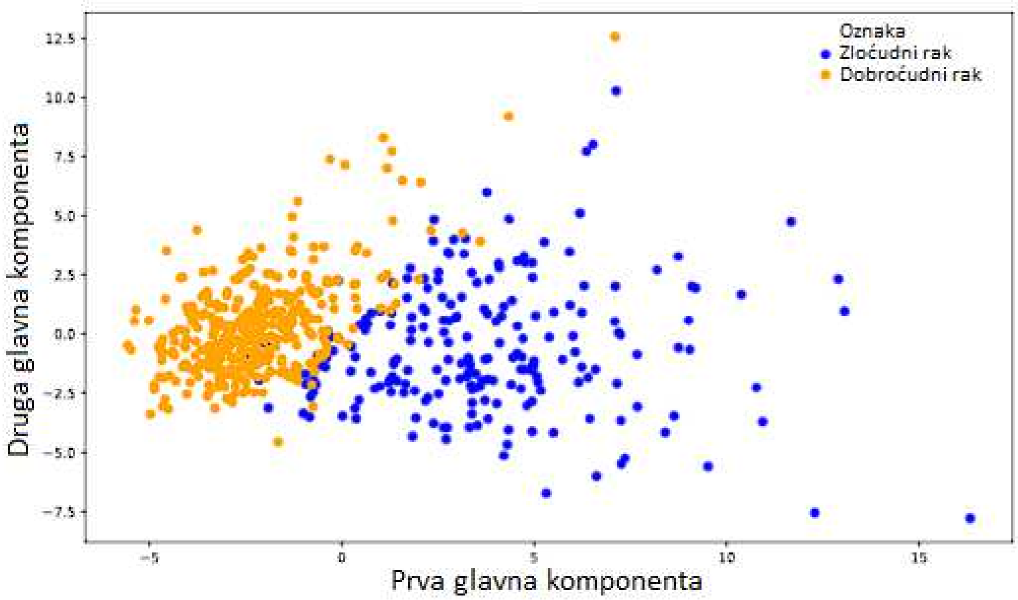
Generalno, u analizi glavnih komponenti se koriste podatci s najvećim svojstvenim vrijednostima, dok se podatke s vrlo malim svojstvenim vrijednostima odbacuje što može rezultirati gubitkom određenih međuodnosa u podatcima. Neobično male svojstvene vrijednosti zadnje glavne komponente mogu ukazivati na linearnu ovisnost u skupu podataka koja ako ostane neprimijećena može u kasnijoj analizi i računici stvoriti probleme.

Očito je da zadatak odabira broja glavnih komponenti može biti dosta kompleksan budući da je potrebno balansirati samu kompleksnost analize i njenu pouzdanost s obzirom na izgubljene informacije.

2.3. Grafički prikazi glavnih komponenti

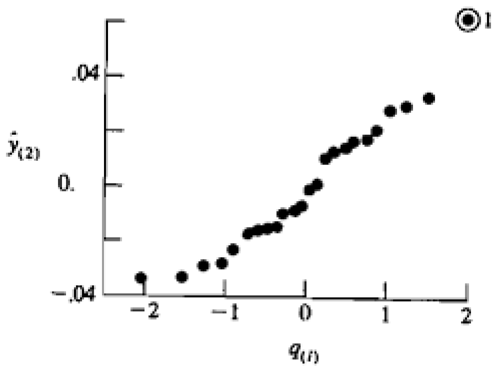
Grafička interpretacija rezultata PCA je ključna za intuitivno razumijevanje strukture podataka. Vizualni prikazi omogućuju lakše prepoznavanje obrazaca, odnosa i odstupanja u višedimenzionalnim podacima, što je posebno korisno u složenim financijskim analizama.

Raspršeni grafovi, poznatiji kao scatter plot, čiji prikaz se može vidjeti na slici Sl. 2.2, su jedan od načina grafičkog prikazivanja glavnih komponenti. Oni omogućuju vizualizaciju odnosa između dviju glavnih komponenti. Raspršeni grafovi mogu otkriti prirodne skupine ili klastere unutar podataka, kao i izuzetke (outliere) koji se značajno razlikuju od ostalih podataka. Osim toga, raspršeni grafovi omogućuju lako prepoznavanje obrazaca, poput linearnih ili nelinearnih odnosa između komponenti. Također pomažu u identifikaciji područja visoke i niske koncentracije podataka.



Sl. 2.2 Primjena grupiranja u raspršenom grafu u području biologije [8]

Ako je za analizu bitno da su podatci normalno distribuirani, to se može provjeriti s Q-Q (Quantile-Quantile) grafom. Oni se koriste za evaluaciju distribucije podataka i usporedbu s normalnom distribucijom. Q-Q grafovi uspoređuju kvantile distribuiranih podataka s kvantilima normalne distribucije. Ako podaci slijede normalnu distribuciju, točke na Q-Q grafu će ležati približno na dijagonali. Q-Q grafovi jasno pokazuju odstupanja od normalne distribucije, što može ukazivati na prisutnost izuzetka ili druge nepravilnosti u podacima. Ova informacija je korisna za daljnju analizu i prilagodbu modela. U tim grafovima se traže točke koje se izdvajaju i odskaču od ostatka grafa čime signaliziraju neko odstupanje u podatcima koje bi bilo dobro provjeriti i utvrditi radi li se o grešci u podatcima ili o stvarnom problemu. Primjer Q-Q grafa prikazan je na slici Sl. 2.3.



Sl. 2.3 Primjer izuzetka (outlier-a) u Q-Q grafu [1]

Grafički prikazi olakšavaju interpretaciju složenih višedimenzionalnih podataka. Vizualizacija omogućuje brzo prepoznavanje važnih trendova, uzoraka i anomalija u podatcima, što može biti od presudne važnosti za donošenje investicijskih odluka. Grafički prikazi čine rezultate analize dostupnijima široj publici, uključujući investitore, menadžere i druge dionike koji možda nemaju duboko tehničko znanje.

2.4. Ortogonalni portfelji

Stvaranje ortogonalnih portfelja predstavlja jednu od ključnih strategija za diverzifikaciju investicijskih portfelja. Diverzifikacija je temeljni princip u upravljanju investicijskim rizicima, a ortogonalni portfelji, koji se sastoje od nekoreliranih komponenti, pružaju dodatnu razinu zaštite protiv sistemskih rizika. Ortogonalni portfelji smanjuju izloženost investitora specifičnim rizicima povezanima s pojedinačnim imovinama ili sektorima. U volatilnim tržišnim uvjetima, ortogonalni portfelji pružaju veću stabilnost jer su osmišljeni da reagiraju na različite tržišne uvjete na nekoreliran način. To znači da dok jedan dio portfelja može gubiti na vrijednosti, drugi dio može zadržati ili čak povećati svoju vrijednost, umanjujući ukupni rizik.

Glavne komponente su osmišljene tako da obuhvaćaju maksimalnu varijabilnost podataka unutar svake komponente, čime se osigurava da svaki dodani podatak unosi što je moguće više korisnih informacija. Osim toga, glavne komponente su međusobno ortogonalne, što znači da su njihove kovarijance jednake nuli. To osigurava da promjene u jednoj komponenti ne utječu na druge komponente. Smanjivanjem broja dimenzija bez značajnog gubitka informacija, PCA omogućuje lakše i učinkovitije upravljanje portfeljem. Manji broj varijabli olakšava analizu i donošenje odluka.

Primjenom PCA, investitori mogu stvoriti portfelje koji su otporniji na tržišne fluktuacije, pružajući bolju zaštitu kapitala i veću potencijalnu dugoročnu stabilnost. PCA ne samo da olakšava razumijevanje složenih odnosa među različitim financijskim instrumentima, već i omogućava stvaranje optimiziranih portfelja koji maksimalno iskorištavaju prednosti diverzifikacije.

3. Analiza i rezultati

3.1. Podatci

U analizi koja slijedi promatrano je 8 različitih imovina. To su:

- China ETF
  - burzovni fond koji prati performanse kineskog tržišta dionica
- Europe ETF
  - burzovni fond koji prati performanse europskog tržišta dionica
- Japan ETF
  - burzovni fond koji prati performanse japanskog tržišta dionica
- US ETF
  - burzovni fond koji prati performanse američkog tržišta dionica
- Nafta
  - cijene sirove nafte, ključni pokazatelj u energetskom sektoru
- Kukuruz
  - cijene kukuruza, važnog poljoprivrednog proizvoda
- Zlato
  - cijene zlata koje se često koriste kao pokazatelj ekonomske stabilnosti
- US Bond
  - prinosi američkih državnih obveznica, važni za analizu kamatnih stopa

Podatci su dobiveni kao cijene svake od imovina za svaki dan kada se tom imovinom trgovalo u periodu od početka 2015. godine do kraja 2022. godine. Problem koji nastaje kod tih podataka je taj da se pojavljuju dani kada se neke imovine trguju, a neke ne. Na takvim podatcima se analiza ne može provesti te ih je potrebno preurediti. Podatke sam preuredio na dva različita načina.

Prva inačica podataka bila je napravljena tako da ako se u nekom danu dogodilo da postoji barem jedna imovina kojom se trgovalo, onda su se cijene za sve imovine kojima se nije trgovalo taj dan popunile tako da se prekopirala prva prethodna dostupna cijena u podatcima. Tako su se podatci popunili na način da nigdje ne nedostaje neka informacija, a dani kada se nekom imovinom nije trgovalo su se zapravo pretvorili u podatak da je taj dan povrat za tu imovinu iznosio 0%. Na tako uređenim podatcima se može provesti analiza, ali problem koji nastaje kod tih podataka je taj da se prosječan broj dana u kojima se trguje nezanemarivo poveća.

Generalno, u godini ima otprilike 252 trgovinska dana. To se temelji na prosječnom broju radnih dana, oduzimanju vikenda i državnih praznika kada su tržišta zatvorena. Taj broj može varirati ovisno o specifičnim praznicima i kalendarskim godinama, ali 252 dana je standardno korišten broj u financijskim analizama. U prvoj inačici podataka, na godišnjoj razini, broj dana u kojima se trguje je 310. Ta činjenica bi potencijalno mogla utjecati na rezultate te sam zbog toga napravio i drugu inačicu podataka.

Druga inačica podataka je napravljena tako da ako se u nekom danu dogodilo da za četiri ili više imovine nedostaje podatak o cijeni, onda se taj dan briše iz podataka za analizu. Preuređivanjem početnih podataka na drugi način dobije se otprilike 260 trgovinskih dana u godini.

Osim broja trgovinskih dana u godini, oba načina preuređivanja podataka mogu potencijalno utjecati na određene parametre koji će se promatrati tijekom analize. Stoga sam analizu proveo na obje inačice podataka. Dobiveni rezultati i analiza pokazali su se iznimno sličnima pa će iz tog razloga u nastavku biti prikazani samo analiza i rezultati dobiveni s pomoću druge inačice podataka.

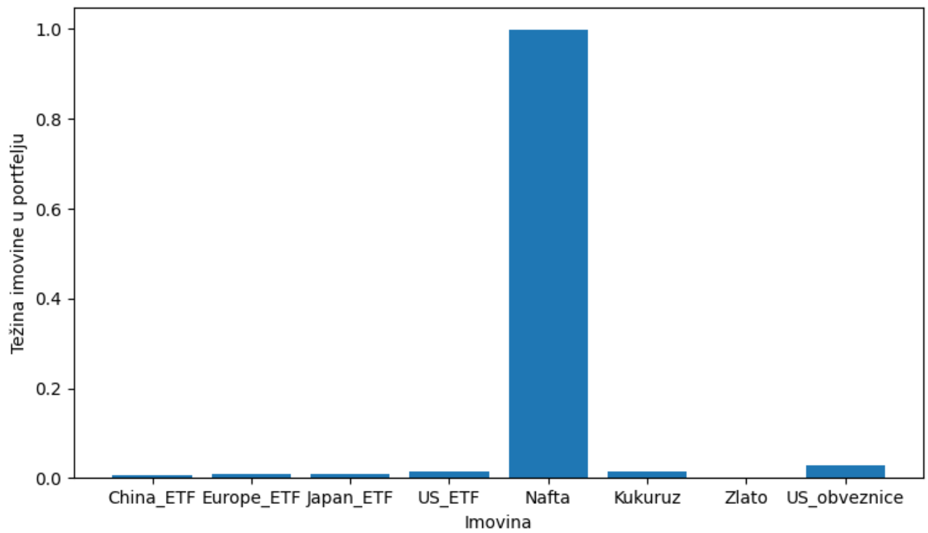
Nakon što su dobiveni podatci preuređeni, za daljnju analizu bilo je potrebno pretvoriti svakodnevne cijene svake imovine u postotne dnevne povrate svake od 8 imovina. Formula s pomoću koje se to radi je:

$$postotni\ dnevni\ povrat = \frac{današnja\ cijena - jučerašnja\ cijena}{jučerašnja\ cijena} * 100 \tag{10}$$

To je potrebno napraviti zato što su cijene različitih imovina izražene u različitim redovima veličine, npr. cijene prinosa američkih državnih obveznica su izražene u iznosima od nekoliko dolara dok su istovremeno cijene japanskog burzovnog fonda izražene u iznosima od nekoliko desetaka tisuća dolara. Takve razlike u podatcima mogu značajno utjecati na rezultate. Za razliku od cijena, svi su postotni dnevni povrati istog reda veličine. Osim toga, usporedbe i rezultati koji će se proučavati se baziraju na povratima imovina, a ne na njihovim cijenama. Dakle, sami postupak analize glavnih komponenti opisan u poglavlju

2.1. temeljit će se na izračunu svojstvenih vrijednosti i svojstvenih vektora s pomoću matrice kovarijance dobivene iz podataka o dnevnim postotnim povratima svih imovina, izračunate s pomoću formule (1).

Graf prikazan na slici Sl. 3.1 predstavlja prvi svojstveni vektor dobiven analizom glavnih komponenti u 2020. godini. Ove godine se dogodila zanimljiva pojava. Cijena nafte je 20.4.2020. dosegla -37$. S obzirom na globalni prekid ekonomske aktivnosti zbog pandemije COVID-19, potražnja za naftom dramatično je pala, dok proizvodnja nafte nije značajno smanjena. To je rezultiralo prekomjernom količinom nafte na tržištu. Budući da se višak nafte nije mogao pohraniti, posebno na fizičkim lokacijama kao što su rezerve nafte, terminali i skladišta, trgovci i investitori nisu imali mjesta za pohranu viška nafte. Svi ovi čimbenici zajedno su doveli do toga da su trgovci bili spremni platiti drugima kako bi preuzeli njihovu naftu, što je rezultiralo negativnom cijenom nafte. Ovaj događaj je očito izazvao vrlo volatilan pokret u tržištu koji je analiza glavnih komponenti prepoznala i samim time dala velik značaj nafti. Budući da je ovo anomalija u tržištu koja uvelike utječe na ovakvu analizu i stvara probleme u rezultatima, ovu godinu sam izbacio iz daljnjih analiza.



Sl. 3.1 Graf prve glavne komponente (portfelja) za 2020. godinu

3.2. Analiza portfelja dobivenog s pomoću PCA

Na podatcima opisanim u prošlom potpoglavlju provedena je analiza glavnih komponenti. Rezultat provođenja PCA metode na tim podatcima je osam svojstvenih vektora budući da se podatci sastoje od osam imovina. Kako bi se pokazala neka osnovna svojstva portfelja dobivenog s pomoću PCA, poput varijance povrata portfelja, očekivanih povrata portfelja te sposobnosti predviđanja, analiza će se provesti na vektoru prve glavne komponente odnosno na vektoru s najvećom svojstvenom vrijednošću kako bi obradili najveću moguću varijabilnost unutar podataka. Svojstveni vektor dobiven s pomoću PCA metode se zapravo interpretira kao investicijski portfelj tako da težine unutar vektora predstavljaju udio kapitala koji je potrebno investirati u svaku od imovina. Takvi portfelji izračunati su na godišnjoj razini tj. za svaku godinu od 2015. do 2022., osim 2020. godine, dobiven je jedan investicijski portfelj.

Glavna mjera koju ćemo promatrati u ovoj analizi je povrat portfelja, njegova varijanca i očekivanja. Povrati portfelja se računaju tako da se postotni dnevni povrat svake imovine u nekom danu pomnoži s težinama imovina koje su dobivene u vektoru prve glavne komponente. Drugim riječima, povrat portfelja određenog dana je skalarni produkt vektora koji sadrži postotni dnevni povrat tog dana za svih 8 imovina te vektora prve glavne komponente tj. vektora koji predstavlja investicijski portfelj sa težinama svake od imovina. Takvi povrati se računaju za svaki dan u godini. Nakon što se dobiju povrati portfelja, može se izračunati njihova varijanca:

$$Var(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{11}$$

pri čemu je n broj dana, X_i svaka pojedinačna vrijednost povrata portfelja, a $\bar{X}$ srednja vrijednost svih povrata portfelja. Očekivanje povrata portfelja može se izračunati kao srednja vrijednost svih povrata portfelja za zadanu godinu s pomoću formule:

$$E(X) = \frac{\sum_{i=1}^{n} X_i}{n} \tag{12}$$

pri čemu je n broj dana, X_i svaka pojedinačna vrijednost povrata portfelja.

3.2.1. Usporedba varijanci povrata trenutnih i predviđenih portfelja

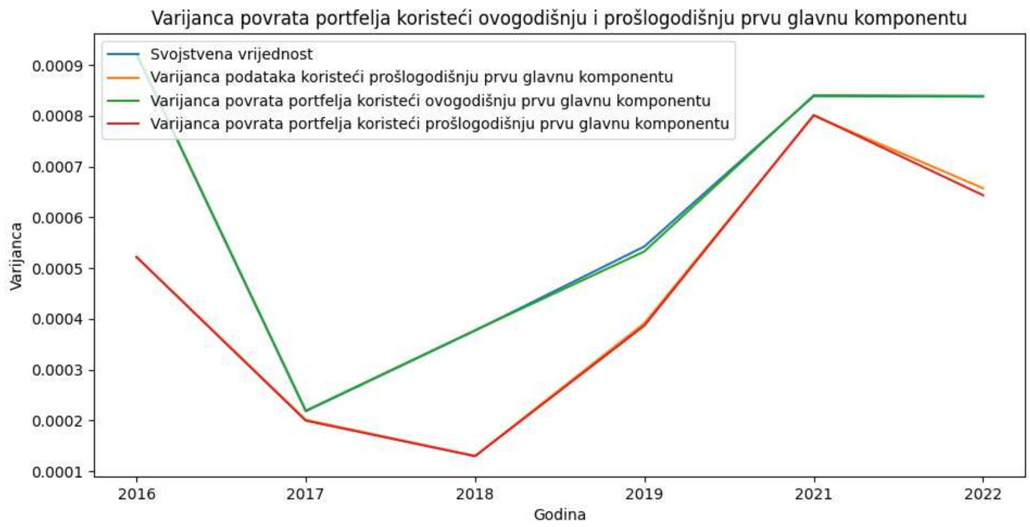
Prva usporedba koju je potrebno napraviti je usporedba svojstvene vrijednosti prve glavne komponente s varijancom povrata portfelja dobivenom s pomoću te iste prve glavne komponente. Osim toga, bitan odnos koji se mora provjeriti je varijanca povrata portfelja dobivena s pomoću prve glavne komponente izračunate za trenutnu godinu koja se promatra, s varijancom povrata portfelja dobivenom s pomoću prve glavne komponente izračunate za godinu prije. Drugim riječima, usporedit ćemo varijancu povrata portfelja dobivenog na temelju podataka trenutne godine s varijancom povrata tzv. predviđenog portfelja dobivenog na temelju podataka prethodne godine. Podatci potrebni za provesti ovu analizu dobiveni su na sljedeći način:

- Svojstvena vrijednost (plava linija na grafu prikazanom na slici Sl. 3.2) se dobiva tijekom provođenja same analize glavnih komponenti.
- Varijanca povrata portfelja dobivena s pomoću prve glavne komponente trenutne godine koja se promatra (zelena linija na grafu prikazanom na slici Sl. 3.2) računa se s pomoću formule (11) pri čemu se u izračunu povrata portfelja koriste postotni dnevni povrati svih imovina trenutne godine, a kao vektor investicijskog portfelja koristi se prva glavna komponenta trenutne godine.
- Varijanca podataka dobivena koristeći tzv. predviđeni portfelj odnosno prvu glavnu komponentu dobivenu na podatcima prethodne godine (narančasta linija na grafu prikazanom na slici Sl. 3.2) izračunava se s pomoću tzv. Rayleighov-og kvocijenta:

  $$R(v, A) = v^T A v \tag{13}$$

  pri čemu je v vektor prve glavne komponente dobivene na podatcima prethodne godine, a A je matrica kovarijance izračunata s pomoću formule (1) na temelju dnevnih postotnih povrata svih imovina trenutne godine.
- Varijanca povrata portfelja dobivena koristeći tzv. predviđeni portfelj odnosno prvu glavnu komponentu dobivenu na podatcima prethodne godine (crvena linija na grafu prikazanom na slici Sl. 3.2) računa se s pomoću formule (11) pri čemu se u izračunu povrata portfelja koriste postotni dnevni povrati svih imovina trenutne godine, a kao vektor investicijskog portfelja koristi se prva glavna komponenta dobivena na podatcima prethodne godine.

Prethodno opisane varijable prikazane su na grafu koji se nalazi na slici Sl. 3.2.



Sl. 3.2 Usporedba varijanci povrata trenutnih i predviđenih portfelja

Vidljivo je kako su zapravo svojstvena vrijednost dobivena analizom glavnih komponenti te varijanca povrata portfelja izračunata s pomoću prve glavne komponente trenutne godine jednake (plava i zelena linija). Time je potvrđena valjanost izraza (7) i pokazana je ispravnost provedene metode. Isti rezultat vidljiv je i pri primjeni prvih glavnih komponenti od prethodnih godina (crvena i narančasta linija). Drugi odnos koji je bitno istaknuti u ovom grafu je to da je varijanca povrata portfelja dobivena s pomoću prošlogodišnje prve glavne komponente manja od one dobivene s pomoću prve glavne komponente trenutne godine. Taj rezultat je i očekivan ako je poznata teorija obrađena u poglavlju 2. Naime, ako za određeni skup podataka izračunamo prvu glavnu komponentu, ona će pokazivati u smjeru najveće varijabilnosti podataka odnosno pripadajuća svojstvena vrijednost će biti najveća moguća. Dakle, ako se pokuša na taj isti skup podataka primijeniti neku drugu glavnu komponentu, poput one dobivene na skupu podataka godine prije, varijanca povrata će biti manja jer takva glavna komponenta nije najbolje prilagođena zadanim podatcima pa tako neće niti obračunati najveću moguću varijabilnost tih podataka.

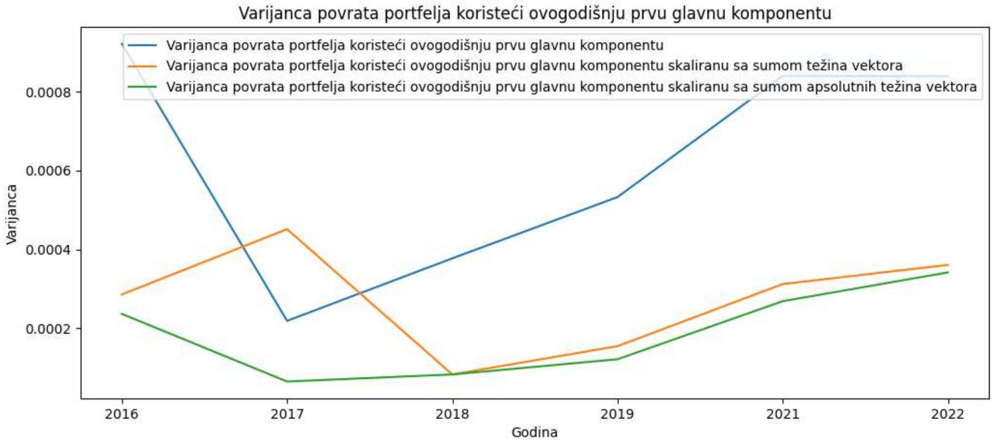
3.2.2. Usporedba skaliranja portfelja

Pri implementaciji analize glavnih komponenti za izgradnju portfelja, postoji nekoliko inačica što se tiče skaliranja samih vrijednosti težina unutar portfelja. Prva inačica je da se ne koristi nikakvo skaliranje, a interpretacija tog pristupa je korištenje tzv. poluge pri investiranju jer suma vrijednosti težina portfelja u tom slučaju može prijeći preko jedan.

Druge dvije varijante su skaliranje sa sumom težina svojstvenog vektora ili sa sumom apsolutnih vrijednosti težina svojstvenog vektora. Ova dva pristupa se koriste ako želimo približiti sumu vrijednosti težina portfelja jedinici pri čemu se skaliranjem sa sumom apsolutnih vrijednosti dobiva suma težina portfelja koja iznosi točno jedan i predstavlja investiranje isključivo s kapitalom s kojim raspolažemo dok skaliranjem sa sumom vrijednosti težina možemo dobiti i dalje sumu težina portfelja koja prelazi jedan budući da se u tom slučaju na kratku prodaju ili tzv. short-selling gleda kao na posudbu sredstava. Podatci potrebni za provesti ovu analizu dobiveni su na sljedeći način:

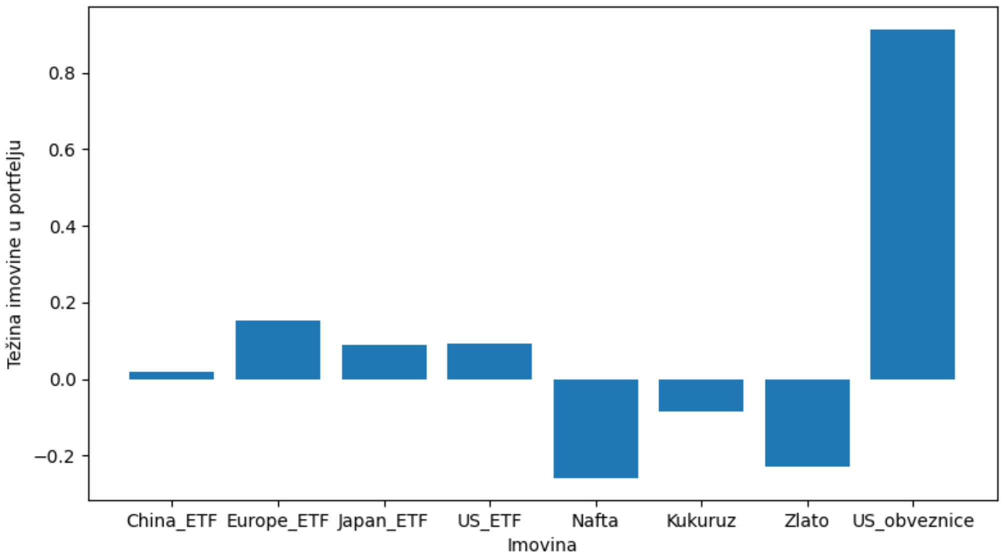
- Varijanca povrata portfelja dobivena s pomoću prve glavne komponente trenutne godine koja se promatra (plava linija na grafu prikazanom na slici Sl. 3.3) računa se s pomoću formule (11) pri čemu se u izračunu povrata portfelja koriste postotni dnevni povrati svih imovina trenutne godine, a kao vektor investicijskog portfelja koristi se prva glavna komponenta trenutne godine.
- Varijanca povrata portfelja dobivena s pomoću prve glavne komponente trenutne godine koja se promatra skalirane sa sumom težina prve glavne komponente (narančasta linija na grafu prikazanom na slici Sl. 3.3) računa se s pomoću formule (11) pri čemu se u izračunu povrata portfelja koriste postotni dnevni povrati svih imovina trenutne godine, a kao vektor investicijskog portfelja koristi se prva glavna komponenta trenutne godine čije su težine skalirane tako da je svaka podijeljena s iznosom sume svih težina.
- Varijanca povrata portfelja dobivena s pomoću prve glavne komponente trenutne godine koja se promatra skalirane sa sumom apsolutnih vrijednosti težina prve glavne komponente (zelena linija na grafu prikazanom na slici Sl. 3.3) računa se s pomoću formule (11) pri čemu se u izračunu povrata portfelja koriste postotni dnevni povrati svih imovina trenutne godine, a kao vektor investicijskog portfelja koristi se prva glavna komponenta trenutne godine čije su težine skalirane tako da je svaka podijeljena s iznosom sume apsolutnih vrijednosti svih težina.

Prethodno opisane varijable prikazane su na grafu koji se nalazi na slici Sl. 3.3.



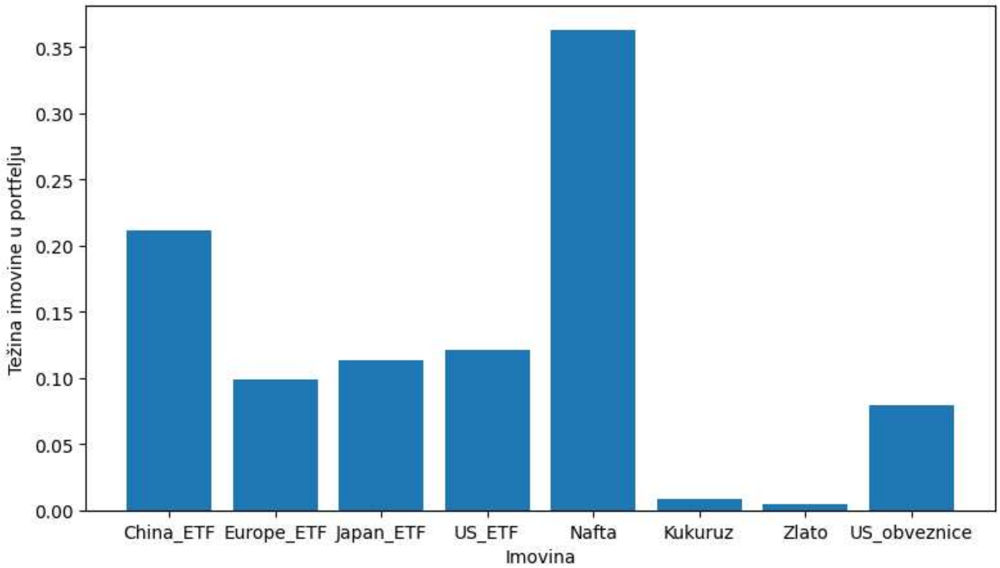
Sl. 3.3 Usporedba različitih skaliranja težina portfelja

Na grafu koji se nalazi na slici Sl. 3.3 je vidljivo kako je varijanca povrata portfelja korištenjem ne skaliranog vektora najveća, dok je varijanca povrata portfelja najmanja korištenjem vektora skaliranog sumom apsolutnih vrijednosti težina te je u tom slučaju portfelj najstabilniji odnosno taj način skaliranja ga najviše ograničava. Ističe se 2017. godina gdje je varijanca povrata portfelja korištenjem vektora skaliranog sa sumom težina najveća što nije očekivan rezultat. To se objašnjava povećanom kratkom prodajom ili tzv. short-sellingom te godine u odnosu na sve ostale, što se može vidjeti na grafu prikazanom na slici Sl. 3.4, pri čemu negativne težine zapravo predstavljaju kratku prodaju.



Sl. 3.4 Graf prve glavne komponente (portfelja) za 2017. godinu

Osim toga, na grafu prikazanom na slici Sl. 3.3 može se vidjeti kako se 2018. godine narančasta i zelena linija preklapaju. Računajući na to da razlika između skaliranja sa sumom vrijednosti težina i sumom apsolutnih vrijednosti težina nastaje pri pojavi kratkih prodaja, može se pretpostaviti da to preklapanje označava da 2018. godine nema kratkih prodaja. Ta pretpostavka je potvrđena grafom prikazanim na slici Sl. 3.5.



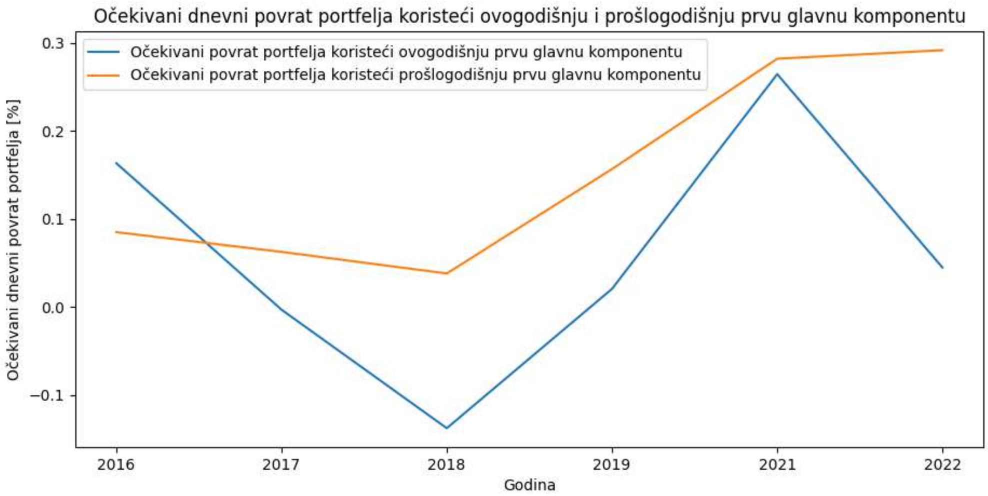
Sl. 3.5 Graf prve glavne komponente (portfelja) za 2018. godinu

3.2.3. Usporedba očekivanih povrata trenutnih i predviđenih portfelja

Osim prethodno provedenih analiza, isplati se pogledati i kako se ponašaju očekivani povrati portfelja. Usporedba očekivanih povrata je ključna za procjenu učinkovitosti predviđenog portfelja i njegove sposobnosti da generira stabilne prihode. Ova usporedba omogućuje nam da procijenimo koliko točno predviđeni portfelj može predvidjeti tržišne trendove i prilagoditi portfelj za optimalan povrat u budućem razdoblju. Podatci potrebni za provesti ovu analizu dobiveni su na sljedeći način:

- Očekivani dnevni povrati portfelja dobiveni s pomoću prve glavne komponente trenutne godine (plava linija na grafu prikazanom na slici Sl. 3.6) računa se s pomoću formule (12) pri čemu se u izračunu povrata portfelja koriste postotni dnevni povrati svih imovina trenutne godine, a kao vektor investicijskog portfelja koristi se prva glavna komponenta trenutne godine.

- Očekivani dnevni povrati portfelja dobiveni s pomoću prve glavne komponente prethodne godine (narančasta linija na grafu prikazanom na slici Sl. 3.6) računa se s pomoću formule (12) pri čemu se u izračunu povrata portfelja koriste postotni dnevni povrati svih imovina trenutne godine, a kao vektor investicijskog portfelja koristi se prva glavna komponenta dobivena na temelju podataka prethodne godine.



Sl. 3.6 Usporedba očekivanih dnevnih povrata trenutnih i predviđenih portfelja

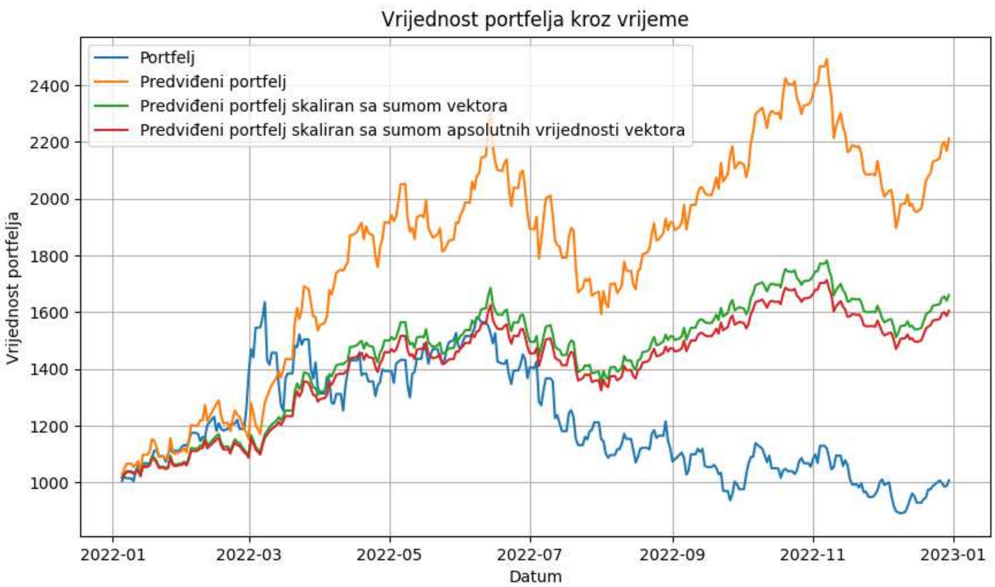
Graf koji se nalazi na slici Sl. 3.6 prikazuje vrijednosti očekivanih dnevnih povrata pri korištenju ovogodišnjeg i prošlogodišnjeg portfelja. Iz grafičkog prikaza jasno je da su očekivani povrati uglavnom veći kada se koristi prošlogodišnji portfelj. Ovaj rezultat može ukazivati na stabilnost promjena tržišnih uvjeta između godina, što sugerira da povijesni podaci mogu imati značajnu sposobnost predviđanja za buduće tržišne performanse. Međutim, ovo također naglašava potrebu za analizom više od jedne glavne komponente. Korištenje samo jedne komponente može pojednostaviti složene tržišne uvjete i zanemariti važnu međusobnu kompenzaciju između različitih faktora koji utječu na povrate. Detaljnija analiza višestrukih glavnih komponenti može pružiti bogatiji uvid u dinamiku tržišta, omogućavajući preciznije predviđanje i bolje upravljanje portfeljem. Stoga, integriranje dodatnih komponenti u analizu moglo bi dodatno povećati pouzdanost i stabilnost investicijskih strategija.

3.2.4. Simulacija investiranja

Za kraj ovog dijela analize, prikazat će se simulacija kretanja vrijednosti portfelja izgrađenog na temelju prve glavne komponente. Na grafu koji se nalazi na slici Sl. 3.7 je prikazana simulacija investiranja za 2022. godinu jer se te godine najjasnije vide odnosi između različitih portfelja. Početne vrijednosti portfelja iznosit će 1000€, a portfelji koji će se proučavat će biti:

- Portfelj dobiven s pomoću prve glavne komponente izračunate na temelju podataka 2022. godine (plava linija na grafu prikazanom na slici Sl. 3.7)
- Portfelj dobiven s pomoću prve glavne komponente izračunate na temelju podataka 2021. godine (narančasta linija na grafu prikazanom na slici Sl. 3.7)
- Portfelj dobiven s pomoću prve glavne komponente izračunate na temelju podataka 2021. godine čije su težine skalirane sa sumom svih vrijednosti težina (zelena linija na grafu prikazanom na slici Sl. 3.7)
- Portfelj dobiven s pomoću prve glavne komponente izračunate na temelju podataka 2021. godine čije su težine skalirane sa sumom apsolutnih vrijednosti svih težina (crvena linija na grafu prikazanom na slici Sl. 3.7)

Povrati portfelja izračunati su postupcima već opisanim u prethodnim potpoglavljima, a sama simulacija vrijednosti investicije izračunata je kao kumulativni umnožak povrata portfelja za svaki dan.



Sl. 3.7 Simulacija investiranja

Ono što treba istaknuti na grafu prikazanom na slici Sl. 3.7 je odnos između rezultata ne skaliranih i skaliranih predviđenih portfelja (narančasta, zelena i crvena linija). Vidljivo je da se sva tri portfelja kreću identičnim smjerom, ali s različitim amplitudama. To potvrđuje teoriju svojstvenih vektora, gdje usmjerenje vektora ostaje prema maksimalnoj varijabilnosti, dok skaliranje utječe samo na duljinu vektora. Na ovom grafu to se manifestira kao različite amplitude vrijednosti portfelja, dok kretanje vrijednosti portfelja ostaje isto. Osim toga, na grafu prikazanom na slici Sl. 3.7 se također potvrđuje odnos prikazan na grafu koji se nalazi na slici Sl. 3.6. Na grafu koji se nalazi na slici Sl. 3.6 vidljivo je kako je očekivani dnevni povrat predviđenog portfelja veći u 2022. godini dok je isto tako vidljivo na grafu prikazanom na slici Sl. 3.7 da je krajnja vrijednost predviđenog portfelja veća od vrijednosti portfelja dobivenog na temelju podataka iz 2022. godine.

3.3. Izrada ortogonalnih portfelja i njihova svojstva

Ortogonalnost portfelja odnosi se na nekoreliranost njegovih komponenti, što je ključ za minimizaciju rizika u investicijskom portfelju. Razvijanje ideje ortogonalnosti počinje razumijevanjem faktora koji utječu na nekoreliranost, kao što su različite tržišne sile i specifične karakteristike imovine. Korištenjem analize glavnih komponenti (PCA), moguće je identificirati komponente koje su ortogonalne, tj. nekorelirane, čime se omogućuje stvaranje portfelja koji su otporniji na promjene tržišnih uvjeta. PCA kao svoj rezultat daje međusobno ortogonalne vektore, no problem kod tih vektora je taj da su oni ortogonalni u periodu čije podatke smo koristili za njihovo računanje, a ono što svakog investitora zanima je budućnost, a ne sadašnjost. Cilj analiza poput ove je na temelju podataka koji su nam u danom trenutku dostupni provjeriti njihovu korisnost za buduće investicijske odluke. Stoga, ono što je zapravo bitno provjeriti je koliko kvalitetno se ortogonalnost zadržava među portfeljima dobivenim s pomoću PCA metode prilikom njihove primjene na buduće trenutke. Kako bi to provjerili, potrebno je uvesti mjeru kojom se može lako provjeriti ponašanje svojstva ortogonalnosti u različitim uvjetima. Kao što je već rečeno, korelacija je odličan pokazatelj ortogonalnosti. Točnije, što je iznos korelacije bliži nuli, ortogonalnost je bolja.

3.3.1. Postupak i definicija mjere ortogonalnosti

Ono što je potrebno napraviti je provesti analizu glavnih komponenti na nekom zadanom periodu kako bi se dobili ortogonalni vektori tj. portfelji. Nakon toga je potrebno te portfelje primijeniti na podatke koji su nastali nakon perioda korištenog za izračun samih portfelja. To se postiže tako da postotne dnevne povrate svih imovina tijekom budućeg perioda pomnožimo sa svakim od dobivenih ortogonalnih portfelja kako bi dobili dnevne povrate svakog od portfelja. Tako su se ortogonalni vektori primijenili na budući period u kojem je sada moguće promatrati kako se ponaša ortogonalnost na novom, neviđenom periodu. Ortogonalnost se onda zapravo mjeri među dnevnim povratima svaka dva portfelja.

U ovom konkretnom slučaju, rezultat prethodno navedenog postupka bit će 8 skupova dnevnih povrata portfelja, pri čemu se između svaka 2 skupa može izračunati njihova međusobna korelacija formulom:

$$Cor(X,Y) = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})}} \tag{14}$$

gdje je n broj dana u budućem periodu koji se promatra, X_i vrijednost povrata prvog portfelja u *i-tom* danu, $\bar{X}$ srednja vrijednost svih povrata prvog portfelja, Y_i vrijednost povrata drugog portfelja u *i-tom* danu, $\bar{Y}$ srednja vrijednost svih povrata drugog portfelja. Kada se izračunaju korelacije između svaka dva skupa povrata portfelja, može se dobiti konačna mjera ortogonalnosti, a to je suma apsolutnih vrijednosti korelacije svaka dva skupa povrata portfelja. Sumiraju se apsolutne vrijednosti zato što korelacija može biti i pozitivna i negativna, a za mjeru ortogonalnosti je zapravo bitna „udaljenost" iznosa korelacije od nule. Što je ta suma manja, to je svojstvo ortogonalnosti bolje zadržano primjenom portfelja na buduće podatke.

3.3.2. Uvođenje parametara o kojima ovisi korelacija

Ponašanje koje je vrijedno provjeriti je utjecaj dužine vremenskog okvira koji se koristi za izračun ortogonalnih portfelja na kvalitetu zadržavanja ortogonalnosti. Osim toga zanimljivo je provjeriti i kvalitetu zadržavanja ortogonalnosti u odnosu na dužinu vremenskog okvira budućih podataka na koje se primjenjuju ortogonalni portfelji.

Najprirodnija prva usporedba će biti usporedba kada se oba vremenska okvira postave na godinu dana i na mjesec dana. Dakle, za svaki uzastopni par godina provesti će se postupak opisan u prethodnom potpoglavlju pri čemu će prva godina iz para služiti za izračun ortogonalnih portfelja, a druga godina iz para služiti kao podatci iz budućnosti na kojima se testira kvaliteta zadržavanja ortogonalnosti. Na primjer, na podatcima iz 2015. godine će se izračunati ortogonalni portfelji te će se nakon toga ti portfelji primijeniti na podatke iz

2016. godine. Pri tome, podatci su postotni dnevni povrati svih imovina. Rezultat provedenog postupka za neki par godina bit će prethodno definirana mjera ortogonalnosti za navedeni par godina. Kada se izračunaju mjere ortogonalnosti za svaki par uzastopnih godina, može se izračunati konačna prosječna mjera ortogonalnosti. Za korištene podatke u ovoj analizi, kada se vremenski okvir koji se koristi za izračun ortogonalnih portfelja postavi na godinu dana i kada se vremenski okvir budućih podataka na kojima se primjenjuju dobiveni ortogonalni portfelji postavi na godinu dana, srednja vrijednost odstupanja korelacije od nule iznosi **0.14**. **(Slučaj 1)**

Ako se oba okvira postave na mjesec dana tj. ako se prethodni postupak provede na svaka dva uzastopna mjeseca, srednja vrijednost odstupanja korelacije od nule iznosit će **0.26**. **(Slučaj 2)**

Na temelju ta dva rezultata može se naslutiti kako postavljanje okvira na duže vremenske periode bolje zadržava svojstvo ortogonalnosti no to će se detaljnije provjeriti u kasnijem dijelu analize.

Sljedeći parametar koji bi potencijalno mogao utjecati na kvalitetu zadržavanja ortogonalnosti je izračun ortogonalnih portfelja na dnevnoj bazi. U ovom slučaju vremenski okvir koji se koristi za izračun ortogonalnih portfelja će se mijenjati na dnevnoj bazi. Njegova dužina će biti konstantna i nepromijenjena tijekom provođenja postupka, ali podatci unutar njega će se mijenjati svakodnevno. Podatci unutar vremenskog okvira budućih podataka na kojima se primjenjuju ortogonalni portfelji u ovom slučaju koristit će se jedan po jedan, a važna stavka će biti ukupna duljina tog vremenskog okvira. Postupak je sličan prethodnom, ali ima određene promjene. U ovoj inačici postupka, ortogonalni portfelji će se izračunati na temelju podataka unutar vremenskog okvira koji se koristi za izračun ortogonalnih portfelja, ali ti portfelji će se primijeniti samo na prvi idući dan koji se ne nalazi unutar podataka korištenih u izračunu ortogonalnih portfelja. Točnije, dobiveni ortogonalni portfelji će se primijeniti na jedan dan iz podataka unutar drugog vremenskog okvira. Dakle, dobiti će se dnevni povrati portfelja za jedan dan te je te vrijednosti potrebno zabilježiti. Nakon toga, iz prvog vremenskog okvira će se ukloniti prvi podatak, a na njegov kraj će se dodati podatak koji predstavlja idući dan za koji se upravo izračunao dnevni povrat portfelja. Pomicanje podataka u prvom vremenskom okviru te dnevni izračun povrata portfelja će se ponavljati dok se ne zabilježi dnevnih povrata portfelja onoliko koliko je zadana dužina drugog vremenskog okvira. Kada se zabilježi potreban broj dnevnih povrata portfelja, provodi se izračun korelacije svaka dva skupa dnevnih povrata portfelja, kao i izračun mjere ortogonalnosti tj. sume apsolutnih vrijednosti svih korelacija. Nakon što je dobivena mjera ortogonalnosti, nju se zabilježi, nakon čega se ovaj postupak provodi ispočetka sve dok se ne iskoriste svi dostupni podatci. Kada se prođe kroz sve podatke, dobije se više „mjerenja“ mjere ortogonalnosti tako da se konačni rezultat izražava kao srednja vrijednost svih dobivenih mjera ortogonalnosti.

Ako se oba vremenska okvira postave na godinu dana, srednja vrijednost odstupanja korelacije od nule u ovoj inačici postupka iznosit će **0.07**. **(Slučaj 3)**

Ako se oba vremenska okvira postave na mjesec dana, srednja vrijednost odstupanja korelacije od nule u ovoj inačici postupka iznosit će **0.20**. **(Slučaj 4)**

Kada ove rezultate usporedimo s rezultatima dobivenima bez izračuna ortogonalnih portfelja na dnevnoj bazi, vidljivo je kako je srednja vrijednost odstupanja korelacije od nule manja kada se radi izračun ortogonalnih portfelja na dnevnoj bazi za obje inačice dužina vremenskih okvira. Uz to, slučaj kada su vremenski okviri postavljeni na godinu dana kao i prije pokazuje bolju mjeru ortogonalnosti.

Sada kada su uvedeni određeni parametri po kojima se može mjeriti kvaliteta zadržavanja ortogonalnosti među portfeljima, bilo bi korisno te vrijednosti usporediti s korelacijama podataka na koje se ortogonalni portfelji primjenjuju. Cilj te usporedbe bio bi provjeriti pomaže li uopće računanje portfelja PCA metodom u smanjivanju korelacije među podatcima. Kako bi se došlo do te usporedbe, potrebno je za svaki od dosadašnjih četiri slučaja s pomoću formule (14) izračunati korelaciju ulaznih podataka na koje se primjenjuju ortogonalni vektori izračunati na zadanom periodu prije trenutno promatranog. Radi lakše usporedbe ovih rezultata te lakšeg prikaza dosadašnjih zaključaka, dobivene vrijednosti zapisat će se u tablicu (Tablica 3.1).

Tablica 3.1 Vrijednosti korelacije dobivene sa PCA portfeljima te korelacije ulaznih podataka u različitim slučajevima

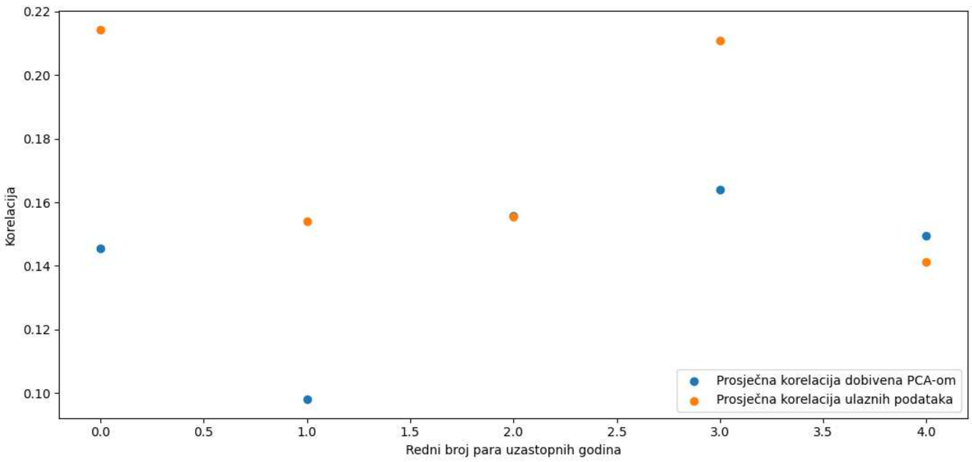
	Srednja vrijednost korelacije dobivena s PCA portfeljima	Srednja vrijednost korelacije ulaznih podataka
Slučaj 1	0.14	0.17
Slučaj 2	0.26	0.26

Slučaj 3	0.07	0.24
Slučaj 4	0.20	0.25

Na temelju vrijednosti u tablici (Tablica 3.1) vidljivo je kako je u sva četiri slučaja iznos korelacije manji ili jednak pri korištenju PCA portfelja u odnosu na korelaciju ulaznih podataka što znači da PCA metoda doista pomaže u snižavanju korelacije.

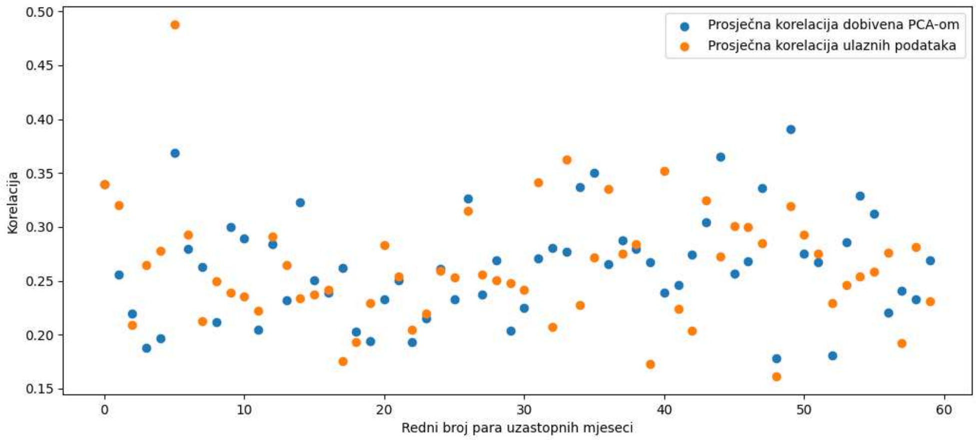
Iako su srednje vrijednosti unutar tablice (Tablica 3.1) dovoljne za donošenje zaključaka, radi lakše vizualizacije i provjere bitno je napraviti i grafičke prikaze pojedinih vrijednosti.

Graf koji se nalazi na slici Sl. 3.8 prikazuje za svaki par uzastopnih godina mjeru ortogonalnosti tj. iznos sume apsolutnih vrijednosti korelacija svaka dva skupa povrata portfelja. Drugim riječima prikazan je slučaj 1 gdje je vremenski okvir koji se koristi za izračun ortogonalnih portfelja postavljen na godinu dana i vremenski okvir budućih podataka na kojima se primjenjuju dobiveni ortogonalni portfelji postavljen na godinu dana. U istom grafu su za usporedbu postavljene i vrijednosti koje predstavljaju korelacije ulaznih podataka na kojima se primjenjuju izračunati ortogonalni portfelji.



Sl. 3.8 Usporedba prosječne korelacije dobivene PCA-om i prosječne korelacije ulaznih podataka s vremenskim okvirom od godinu dana

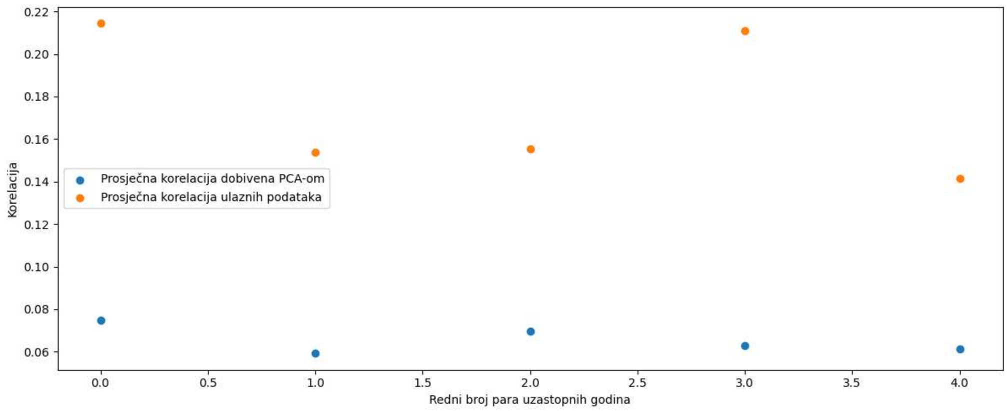
Graf koji se nalazi na slici Sl. 3.9 prikazuje za svaki par uzastopnih mjeseci mjeru ortogonalnosti tj. iznos sume apsolutnih vrijednosti korelacija svaka dva skupa povrata portfelja. Drugim riječima prikazan je slučaj 2 gdje je vremenski okvir koji se koristi za izračun ortogonalnih portfelja postavljen na mjesec dana i vremenski okvir budućih podataka na kojima se primjenjuju dobiveni ortogonalni portfelji postavljen na mjesec dana. U istom grafu su za usporedbu postavljene i vrijednosti koje predstavljaju korelacije ulaznih podataka na kojima se primjenjuju izračunati ortogonalni portfelji.



Sl. 3.9 Usporedba prosječne korelacije dobivene PCA-om i prosječne korelacije ulaznih podataka s vremenskim okvirom od mjesec dana

Usporedbom grafova koji se nalaze na slikama Sl. 3.8 i Sl. 3.9 vidljivo je kako su i narančaste i plave točke manjeg iznosa na prvom grafu. Dakle, potvrđuje se zaključak da se svojstvo ortogonalnosti bolje zadržava na dužim vremenskim okvirima.

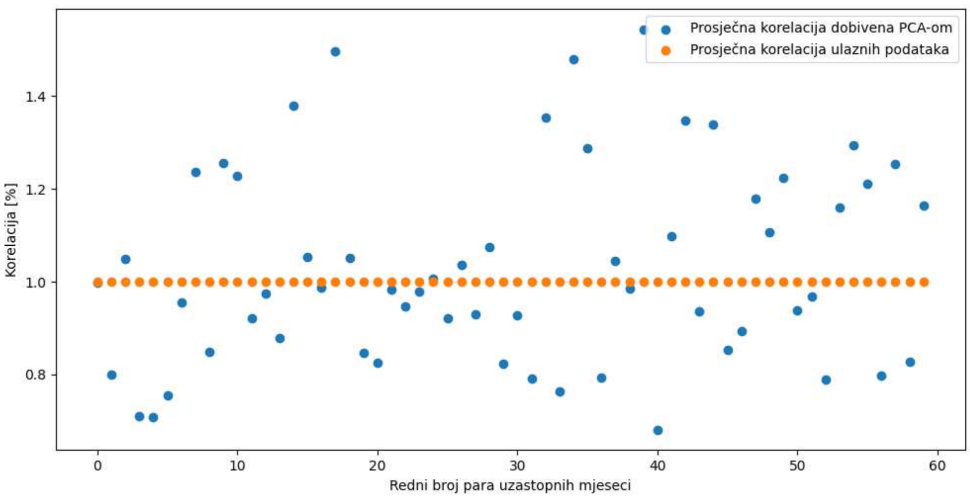
Graf koji se nalazi na slici Sl. 3.10 prikazuje mjeru ortogonalnosti tj. iznos sume apsolutnih vrijednosti korelacija svaka dva skupa povrata portfelja kada se ortogonalni portfelji računaju na dnevnoj bazi. Drugim riječima prikazan je slučaj 3 gdje je vremenski okvir koji se koristi za izračun ortogonalnih portfelja postavljen na godinu dana i vremenski okvir budućih podataka na kojima se primjenjuju dobiveni ortogonalni portfelji postavljen na godinu dana. U istom grafu su za usporedbu postavljene i vrijednosti koje predstavljaju korelacije ulaznih podataka tijekom kojih se provodi dnevno preračunavanje ortogonalnih portfelja.



Sl. 3.10 Usporedba prosječne korelacije dobivene PCA-om i prosječne korelacije ulaznih podataka s vremenskim okvirom od godinu dana uz dnevno preračunavanje

Usporedbom grafova koji se nalaze na slikama Sl. 3.8 i Sl. 3.10 vidljivo je kako su plave točke na drugom grafu manjeg iznosa. Dakle, potvrđuje se zaključak da se svojstvo ortogonalnosti bolje zadržava pri korištenju dnevnih preračunavanja.

Osim tri prethodna grafa, istaknuo bih graf koji se nalazi na slici Sl. 3.11 koji prikazuje iste podatke kao i graf prikazan na slici Sl. 3.9 uz prilagodbu da se vrijednosti mjere ortogonalnosti prikazuju kao postotci u usporedbi s korelacijom ulaznih podataka.



Sl. 3.11 Usporedba prosječne korelacije dobivene PCA-om i prosječne korelacije ulaznih podataka s vremenskim okvirom od mjesec dana u obliku postotaka

Na grafu prikazanom na slici Sl. 3.11 se ističe velika količina plavih točaka koje se nalaze iznad narančastih točaka što se ne pojavljuje ni u jednom drugom slučaju. Taj odnos točaka zapravo predstavlja slučajeve kada korelacija pri korištenju PCA portfelja u odnosu na korelaciju ulaznih podataka bude veća odnosno kada PCA ne uspije smanjiti korelaciju podataka. To znači da ako je u interesu ostvarivanje smanjenja korelacije tj. zadržavanja ortogonalnosti, nije dobro koristiti vremenske okvire od mjesec dana, barem ne bez dnevnog preračunavanja.

Ostale inačice prethodna četiri grafa nisu prikazane zbog preglednosti dokumenta te budući da se na temelju njih ne donose nikakvi drukčiji zaključci.

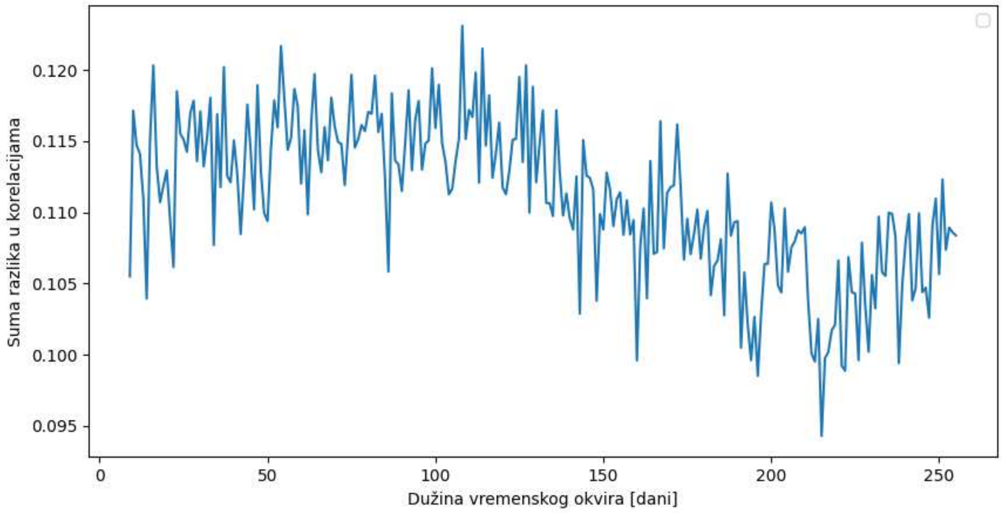
Dakle, dosadašnji rezultati pokazuju da mjera kvalitete zadržavanja ortogonalnosti ovisi o dužini vremenskih okvira koji se koriste, kao i o tome radi li se izračun ortogonalnih portfelja na dnevnoj bazi ili ne. Također je pokazano kako PCA pomaže u smanjenju iznosa korelacije. Te ovisnosti i rezultati su pokazani samo na četiri specifična slučaja. Stoga će se u idućem potpoglavlju detaljnije ispitati utjecaj dužine vremenskih okvira na kvalitetu smanjenja iznosa korelacije.

3.3.3. Ispitivanje utjecaja dužine vremenskih okvira na korelaciju

Budući da je dnevno preračunavanje u oba slučaja rezultiralo kvalitetnijim vrijednostima, u daljnjoj analizi ću koristiti inačicu postupka u kojoj se računaju ortogonalni portfelji na dnevnoj bazi, kao što je opisano za slučajeve 3 i 4 iz prethodnog potpoglavlja.

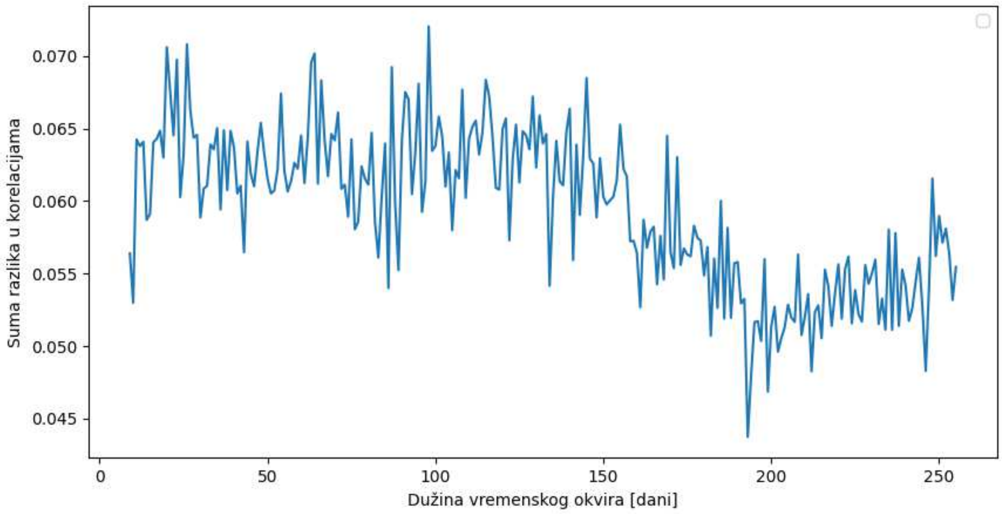
Ako se fiksira drugi vremenski okvir tj. vremenski okvir budućih podataka na kojima se primjenjuju dobiveni ortogonalni portfelji, može se proučiti utjecaj prvog vremenskog okvira, tj. vremenskog okvira koji se koristi za izračun ortogonalnih portfelja, na smanjenje iznosa korelacije. Drugi vremenski okvir ćemo najprije fiksirati na dužinu od godinu dana, a nakon toga na dužinu od mjesec dana. Prvi vremenski okvir poprimat će sve dužine između 9 i 256. Dužina prvog vremenskog okvira mora biti barem 9 budući da PCA na ulaznim podatcima rezultira s osam portfelja. Gornja granica je postavljena na 256 jer je to dužina najkraće godine u podatcima. Mjera koja će se gledati u ovoj analizi je suma razlika srednje vrijednosti korelacije povrata portfelja dobivenih s PCA portfeljima i srednje vrijednosti korelacije ulaznih podataka na kojima se primjenjuju izračunati ortogonalni PCA portfelji. Tu mjeru će se još i podijeliti s brojem godina odnosno mjeseci tijekom kojih se račun provodio kako bi se na kraju moglo odrediti i kvalitetniju dužinu drugog vremenskog okvira za smanjenje iznosa korelacije.

Na grafu prikazanom na slici Sl. 3.12 je vidljivo kako rezultati imaju puno šuma, ali taj šum je unutar jako uskog raspona vrijednosti tako da su te fluktuacije razumne. Ono što se može primijetiti je da za dužine vremenskog okvira od 9 do otprilike 135 dana, suma razlika u korelacijama poprima otprilike jednaku vrijednost uz manja odskakanja. Nakon dužine vremenskog okvira od 135 dana, vrijednosti sume razlika u korelacijama počinju padati i poprimaju minimalnu vrijednost kada je dužina vremenskog okvira otprilike 210 dana, nakon čega vrijednosti sume razlika u korelacijama polako počinju ponovno rasti.



Sl. 3.12 Mjera razlike u korelacijama ovisno o dužini vremenskog okvira u periodu od godinu dana

Na grafu prikazanom na slici Sl. 3.13 vrijede sva ista zapažanja kao i za prethodni graf osim što se minimum vrijednosti sume razlika u korelacijama javlja nešto ranije, kada je dužina vremenskog okvira otprilike 190 dana.



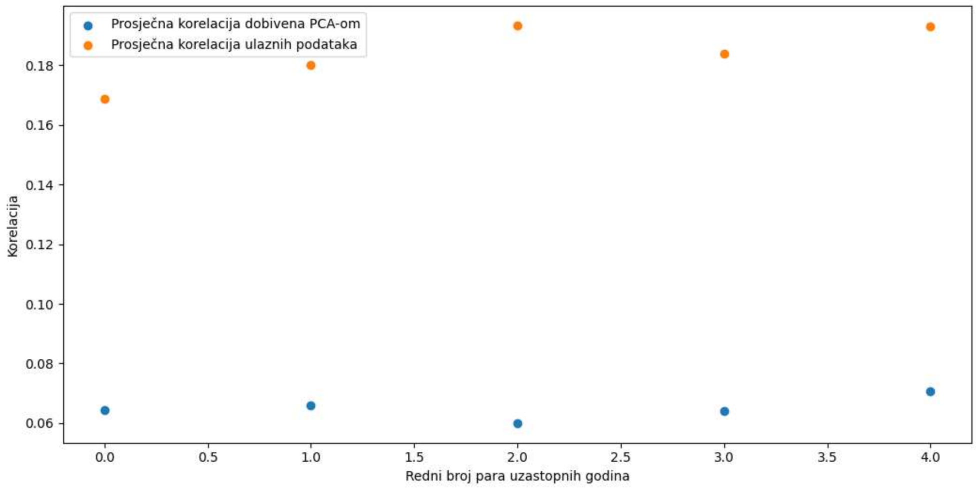
Sl. 3.13 Mjera razlike u korelacijama ovisno o dužini vremenskog okvira u periodu od mjesec dana

Ono što se može zaključiti iz ove analize je da korištenje dužine prvog okvira do 140 dana rezultira s približno jednakom kvalitetom smanjenja iznosa korelacije i za te dužine može se aproksimirati da je ta kvaliteta maksimalna. Zanimljiva pojava u ovoj analizi je ta da se minimum kvalitete smanjenja korelacije pojavljuje za dužine prvog vremenskog okvira od otprilike 200 dana. Dužina vremenskog okvira od 200 dana je približno 75% trgovinskih dana u godini. Dakle rezultat ove analize je da je kvaliteta smanjenja iznosa korelacije najmanja kada se za izračun ortogonalnih portfelja koriste podatci tri kvartala unatrag. Osim toga, ako usporedimo vrijednosti na y-osi između grafova prikazanih na slikama Sl. 3.12 i Sl. 3.13 očito je da su vrijednosti na prvom grafu skoro duplo većeg iznosa nego na drugom grafu. Na temelju toga se zaključuje da je smanjenje iznosa korelacije odnosno svojstvo održavanja ortogonalnosti kvalitetnije kada se dužina drugog vremenskog okvira, tj. vremenskog okvira budućih podataka na kojima se primjenjuju dobiveni ortogonalni portfelji, postavi na godinu dana.

3.3.4. Provjera stacionarnosti tržišnih uvjeta

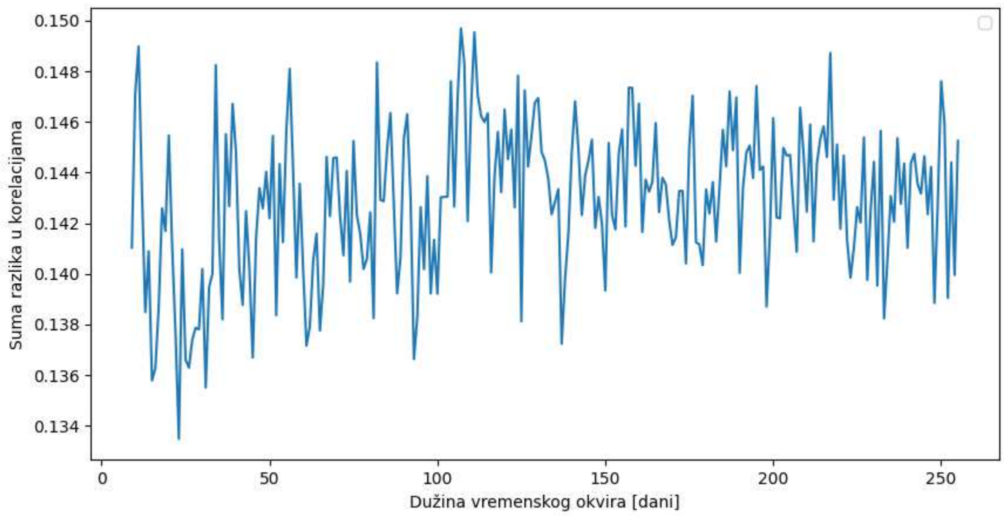
Stacionarnost tržišnih uvjeta može biti ključna za učinkovitu analizu financijskih vremenskih nizova i modeliranje portfelja. Stacionarnost podrazumijeva da statistička svojstva vremenskog niza, kao što su srednja vrijednost i varijanca, ostaju konstantne kroz vrijeme. Stacionarnost će se provjeriti tako da se na temelju ulaznih podataka izračunaju parametri srednje vrijednosti (μ) i standardne devijacije (σ) nakon čega se koristi multivarijatna normalna razdioba kako bi se generirali novi dnevni postotni povrati svake imovine koji će imati normalnu razdiobu. Ove nove podatke podvrgnut će se istoj analizi kao i originalne ulazne podatke. Ono što se pokušava naći su razlike u rezultatima analize dobivenim na temelju originalnih ulaznih podataka i u rezultatima analize dobivenim na temelju novih podataka. U nastavku su istaknute primijećene razlike.

Graf koji se nalazi na slici Sl. 3.14 je ekvivalent grafu prikazanom na slici Sl. 3.8. Narančaste točke tj. korelacije ulaznih podataka u oba grafa su približno jednakih vrijednosti. Razlika između ta dva grafa nastaje kada se usporede plave točke tj. iznos sume apsolutnih vrijednosti korelacija svaka dva skupa povrata portfelja izračunatih s pomoću PCA portfelja. Očita je razlika u iznosu tih vrijednosti pri čemu su iznosi dobiveni s novim podatcima više nego duplo manji od onih dobivenih s originalnim podatcima. Ostale inačice grafova koje su istaknute u poglavlju 3.3.2. ne razlikuju se od njihovih ekvivalenata dobivenih na temelju novih podataka. Dakle razlika koja je nastala generiranjem novih podataka koji su normalno distribuirani je takva da se značajno smanjio iznos sume apsolutnih vrijednosti korelacija svaka dva skupa povrata portfelja izračunatih s pomoću PCA portfelja u vremenskom okviru od godinu dana dok se istovremeno ta vrijednost nije smanjila u vremenskom okviru od godinu dana s dnevnim preračunavanjem portfelja. U tom slučaju sam parametar dnevnog preračunavanja portfelja gubi na značaju. Ova analiza implicira da se uvođenjem dnevnog preračunavanja portfelja smanjuje značaj stacionarnosti parametara. Dakle, iako generirani podaci pokazuju niže korelacije, dnevno preračunavanje portfelja može pomoći u održavanju stabilnosti u nestacionarnim uvjetima, čime se gubi na značaju same stacionarnosti u dugoročnim analizama.



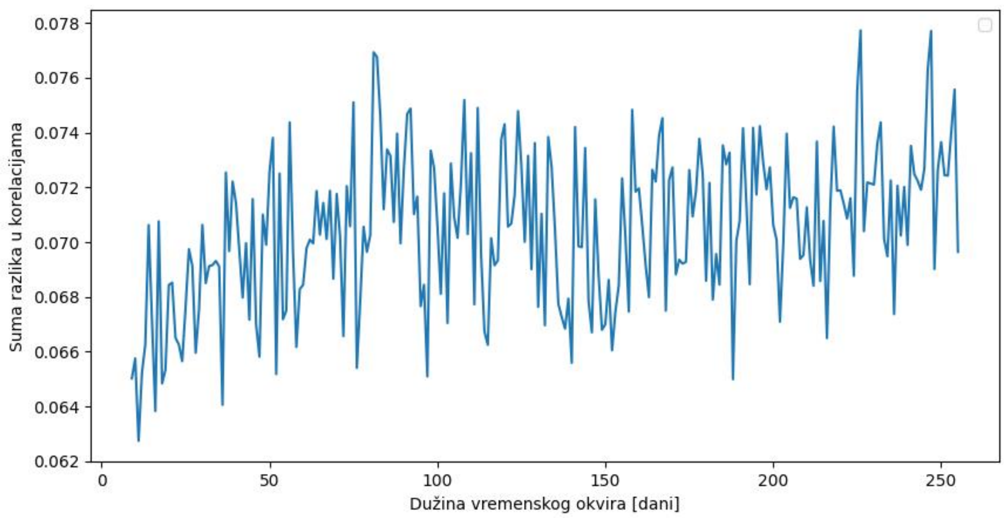
Sl. 3.14 Usporedba prosječne korelacije dobivene PCA-om i prosječne korelacije ulaznih podataka na vremenskom okviru od godinu dana i sa normalno distribuiranim podatcima

Graf koji se nalazi na slici Sl. 3.15 je ekvivalent grafu prikazanom na slici Sl. 3.12. Razlika koja je vidljiva među ova dva grafa je ta da se suma razlika u korelacijama dobivena s novim podatcima kreće cijelo vrijeme oko iste vrijednosti odnosno ne dolazi do pojave trenda pada i rasta ovisno o dužini vremenskog okvira. To sugerira da generirani podaci ne pokazuju promjene u sumama razlika korelacija koje bi bile uvjetovane duljinom promatranog vremenskog okvira, što može implicirati stacionarnost tih podataka prema sumama razlika korelacije. Osim toga, sami iznos sume razlika u korelacijama je nešto veći od iznosa dobivenog na originalnim podatcima. Ovaj rezultat može ukazivati na to da novi podatci koji su normalno distribuirani i stacionarni imaju bolju sposobnost očuvanja ortogonalnosti portfelja.



Sl. 3.15 Mjera razlike u korelacijama ovisno o dužini vremenskog okvira u periodu od godinu dana sa normalno distribuiranim podatcima

Graf koji je prikazan na slici Sl. 3.16 je ekvivalent grafu prikazanom na slici Sl. 3.13. Razlike između ova dva grafa su jednake kao i razlike između grafova prikazanih na slikama Sl. 3.15 i Sl. 3.12.



Sl. 3.16 Mjera razlike u korelacijama ovisno o dužini vremenskog okvira u periodu od mjesec dana sa normalno distribuiranim podatcima

Generiranjem novih podataka koji su normalno distribuirani i provođenjem analize koja je ekvivalentna analizi provedenoj na originalnim ulaznim podatcima uočene su neke razlike u rezultatima. To znači da originalni podatci nisu potpuno stacionarni odnosno to je pokazatelj pojava nekih specifičnih promjena unutar tržišta i dinamike financijskih tržišta.

Zaključak

Analizom provedenom na korištenom skupu podataka pokazano je kako prva glavna komponenta uistinu sadrži najveću moguću količinu varijabilnosti podataka. Osim toga, rezultati su pokazali i potencijalnu sposobnost predviđanja smjera kretanja investicija koja se očitovala u impresivnim vrijednostima očekivanih dnevnih povrata. Glavno svojstvo analize glavnih komponenti koje je potrebno proučiti je njena sposobnost stvaranja međusobno ortogonalnih portfelja sa svrhom smanjivanja sistemskog rizika pri investiranju. Rezultati analize govore da je za ostvarenje najveće ortogonalnosti portfelja pri investiranju s pomoću PCA potrebno investirati u periodu od godinu dana budući da je rezultat ortogonalnosti pri investiranju u periodu od mjesec dana skoro pa duplo gori. Ovaj rezultat i ima smisla budući da su povrati unutar mjesec dana međusobno više povezani i bliži jedan drugome što u ovom slučaju kao posljedicu ima veću vrijednost međusobne korelacije. Osim investiranja na duži period, prilagodba koja dodatno poboljšava rezultat ortogonalnosti je uvođenje dnevnog preračunavanja težina imovina unutar portfelja. Dnevnim preračunavanjem svakom danu se pridodaje posebna pažnja i težine portfelja za svaki dan će biti izračunate na temelju različitih podataka čime se smanjuje međusobna povezanost dana, a pogotovo onih dana koji su međusobno više udaljeni. Generalno, ortogonalni portfelji dobiveni s pomoću PCA metode su se pokazali kao solidan alat za redukciju iznosa korelacije među podatcima na kojima su primijenjeni, a to svojstvo se najviše ističe upravo kada se gleda period od godinu dana pri čemu se rade dnevni preračuni. Daljnjom analizom utvrđeni su detalji oko utjecaja dužine vremenskog okvira na svojstvo zadržavanja ortogonalnosti. Rezultati su pokazali kako je skoro pa svejedno koja dužina se koristi sve dok je manja od otprilike 140 dana jer su se dužine nakon te pokazale nešto lošijima, a najgori vremenski okvir za izračun portfelja iznosi tri kvartala. Usporedbom s analizom provedenom na normalno distribuiranim podatcima pojavile su se pojedine razlike u rezultatima kojima se pokazalo da dnevno preračunavanje portfelja može pomoći u održavanju stabilnosti u nestacionarnim tržišnim uvjetima te se generalno pokazalo da korišteni originalni podatci vjerojatno nisu stacionarni već pokazuju dinamične promjene tržišta. PCA se pokazao kao ključni alat za analizu i diverzifikaciju investicijskih portfelja, pružajući investitorima neophodno oruđe u upravljanju rizikom i postizanju financijske sigurnosti u svijetu sve složenijih financijskih tržišta.

Literatura

[1] Johnson, R. A., Wichern, D. W. *Applied Multivariate Statistical Analysis*. 6. izdanje. Upper Saddle River: Pearson Prentice Hall, 2007.

[2] Tsay, R. S. *Analysis of Financial Time Series*. 3. izdanje. Hoboken: Wiley, 2010.

[3] Begušić, S., Kostanjčar, Z. *Analiza multivarijatnih podataka - primjene u financijama,* Fakultet elektortehnike i računarstva, (2024.), Poveznica: https://www.fer.unizg.hr/_download/repository/Analiza_multivarijantnih_podataka_primjene_u_financijama.pdf , pristupljeno 10. lipnja 2024.

[4] TileStats, *Eigenvectors and eigenvalues - the math step-by-step,* Youtube, (veljača 2021.), Poveznica: https://www.youtube.com/watch?v=JtcNe--fsyA , pristupljeno: 8.lipnja 2024.

[5] Chavez-Bedoya, L., Rosales Marticorena, F. Orthogonal Portfolios to Assess Estimation Risk, International Review of Economics & Finance, 80,6 (travanj 2019.), str. 80-86

[6] Investopedia, Time Series, (2023, lipanj). Poveznica: https://www.investopedia.com/terms/t/timeseries.asp; pristupljeno 10. lipnja 2024.

[7] Wikipedia, Scree plot, (2023, srpanj). Poveznica: https://en.wikipedia.org/wiki/Scree_plot; pristupljeno 10. lipnja 2024.

[8] Statistics Globe, Scatterplot PCA in Python, (2023, srpanj). Poveznica: https://statisticsglobe.com/scatterplot-pca-python; pristupljeno 10. lipnja 2024.

Sažetak

Razvoj ortogonalnih portfelja pomoću metode glavnih komponenti

Ovaj rad istražuje koncept diverzifikacije kao ključne metode za smanjenje sistemskog rizika u investiranju, s posebnim naglaskom na stvaranju ortogonalnih portfelja pomoću analize glavnih komponenti (PCA). Uvodni dio rada obrađuje financijske vremenske nizove, njihova svojstva i metode analize, pružajući čitatelju temeljno razumijevanje financija i analitičkih pristupa. Nakon toga, razrađuje se teorijska osnova PCA, objašnjavajući njene ključne karakteristike. Praktični dio rada uključuje analizu financijskih podataka kako bi se demonstrirala primjena PCA i njena osnovna svojstva. Posebna pažnja posvećena je proučavanju ponašanja ortogonalnih portfelja u različitim uvjetima, kao i utjecaju različitih parametara na njihovu učinkovitost. U konačnici, ispituje se stacionarnost tržišnih uvjeta u zadanom skupu podataka.

Ključne riječi: analiza glavnih komponenti, ortogonalni portfelji, financijski vremenski nizovi, upravljanje rizikom, diverzifikacija

Summary

Development of orthogonal portfolios based on the principal component analysis

This paper explores the concept of diversification as a key method for reducing systemic risk in investing, with a particular focus on the creation of orthogonal portfolios using Principal Component Analysis (PCA). The introductory part of the paper covers financial time series, their properties, and analysis methods, providing the reader with a fundamental understanding of finance and analytical approaches. Following this, the theoretical foundation of PCA is elaborated, explaining its key characteristics. The practical part of the paper involves the analysis of financial data to demonstrate the application of PCA and its fundamental properties. Special attention is given to studying the behavior of orthogonal portfolios under various conditions and the impact of different parameters on their effectiveness. Finally, the stationarity of market conditions in the given data set is examined.

Keywords: principal component analysis, orthogonal portfolios, financial time series, risk management, diversification