Modeliranje korelacija financijskih vremenskih nizova zasnovano na latentnim faktorima

Papak, Mate

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1639

MODELIRANJE KORELACIJA FINANCIJSKIH VREMENSKIH NIZOVA ZASNOVANO NA LATENTNIM FAKTORIMA

Mate Papak

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1639

MODELIRANJE KORELACIJA FINANCIJSKIH VREMENSKIH NIZOVA ZASNOVANO NA LATENTNIM FAKTORIMA

Mate Papak

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Zagreb, 4. ožujka 2024.

ZAVRŠNI ZADATAK br. 1639

Pristupnik:	**Mate Papak (0036543707)**
Studij:	Elektrotehnika i informacijska tehnologija i Računarstvo
Modul:	Računarstvo
Mentor:	doc. dr. sc. Stjepan Begušić

Zadatak: **Modeliranje korelacija financijskih vremenskih nizova zasnovano na latentnim faktorima**

Opis zadatka:

U sklopu ovog završnog rada potrebno je razmotriti problem procjene i predviđanja korelacijskih koeficijenata parova vremenskih nizova povrata financijskih imovina, te standardne procjenitelje kao šro je uzorački Pearsonov korelacijski koeficijent. Potom je potrebno razviti model financijskih vremenskih nizova zasnovan na latentnim faktorima, kao i algoritme za procjenu koeficijenata faktorskog modela iz podataka. Također je potrebno razviti i pristup zasnovan na strojnom učenju za predviđanje budućih vrijednosti koeficijenata faktorskog modela. Poseban naglasak potrebno je staviti na integraciju predviđenih matrica korelacije u matrice kovarijance financijskih vremenskih nizova, te razviti okvir za usporedbu tako sastavljenih matrica kovarijance. Razvijene metode potrebno je primijeniti na povijesnim tržišnim podatcima i provesti statističku analizu rezultata.

Rok za predaju rada: 14. lipnja 2024.

Zahvaljujem se svom mentoru doc. dr. sc. Stjepanu Begušiću na savjetima i pomoći pri pisanju rada, te obitelji i prijateljima koji su bili uz mene za vrijeme studiranja.

Sadržaj

Uvod

U današnjem dinamičnom financijskom okruženju, točna procjena i predviđanje financijskih vremenskih nizova postaju ključni izazovi u analizi tržišta. Financijski vremenski nizovi, poput cijena dionica, predstavljaju nizove podataka koji se bilježe u redoslijedu vremena i čija analiza može pružiti vrijedne uvide u kretanja tržišta. Ključna karakteristika tih nizova je njihova inherentna volatilnost i kompleksnost, što otežava precizne procjene i predviđanja. Upravo je ova problematika središnja tema ovog rada, koji se fokusira na metode za procjenu i predviđanje korelacijskih koeficijenata među vremenskim nizovima povrata dionica.

Korelacija u financijama označava stupanj do kojeg se dva ili više financijskih instrumenata međusobno kreću. Dobra procjena tih korelacija je ključna za izgradnju učinkovitih portfelja i razumijevanje međusobnih odnosa između različitih financijskih instrumenata. Međutim, standardni pristupi poput Pearsonovog korelacijskog koeficijenta, iako široko korišteni, često nisu adekvatni za visoko-dimenzionalne financijske podatke, posebice kada se radi o velikom broju instrumenata [1]. Uzorački Pearsonov koeficijent može biti neprikladan za procjenu nesingularne korelacijske matrice kod velikih portfelja, što je neophodno za konstruiranje optimalnih portfelja.

Kako bi se prevladali ti izazovi, ovaj rad istražuje naprednije metode procjene koje bolje odgovaraju specifičnostima financijskih vremenskih nizova. Temeljni pristup je korištenje jednofaktorskog modela koji pretpostavlja da povrate dionica možemo objasniti pomoću jednog zajedničkog latentnog faktora. Ovaj model omogućava konstrukciju pozitivno definitnih korelacijskih matrica, što je nužno za izgradnju optimalnih portfelja. Modeliranje povrata kroz jednofaktorski model pomaže u identificiranju skrivenih struktura u podacima, čime se poboljšava razumijevanje kako različite dionice reagiraju na promjene ekonomskih uvjeta.

Predviđanje budućih korelacija predstavlja dodatni izazov, ali je također od velike važnosti za investitore koji žele optimizirati svoje portfelje. Tradicionalne statističke tehnike često nisu dovoljne za hvatanje složenih uzoraka u povijesnim podacima. Stoga, u ovom radu će se koristiti metode strojnog učenja koje imaju sposobnost prepoznavanja složenih obrazaca i predviđanja budućih korelacija. Implementacijom algoritama strojnog učenja nastojat ćemo predvidjeti kako će se korelacije među dionicama mijenjati u budućnosti, što će omogućiti investitorima bolje planiranje i alokaciju kapitala.

Glavni ciljevi ovog rada su: prvo, istražiti problem procjene korelacijskih koeficijenata pomoću standardnih metoda kao što je Pearsonov koeficijent; drugo, razviti jednofaktorski model za financijske vremenske nizove i analizirati njegovu učinkovitost; i treće, razviti pristup zasnovan na strojnom učenju za predviđanje budućih korelacija. Na kraju, rad će se fokusirati na integraciju predviđenih korelacija u matricu kovarijance kako bi se izgradili optimalni portfelji s minimalnim rizikom.

Kroz ovu analizu i modeliranje, rad će pružiti sveobuhvatan okvir za procjenu, predviđanje i korištenje financijskih korelacija u svrhu optimizacije portfelja, s posebnim naglaskom na minimizaciju rizika. Implementacija i evaluacija razvijenih metoda provest će se na povijesnim podacima dionica, omogućujući empirijsku analizu i validaciju predloženih pristupa.

1. Modeliranje multivarijatnih financijskih vremenskih nizova

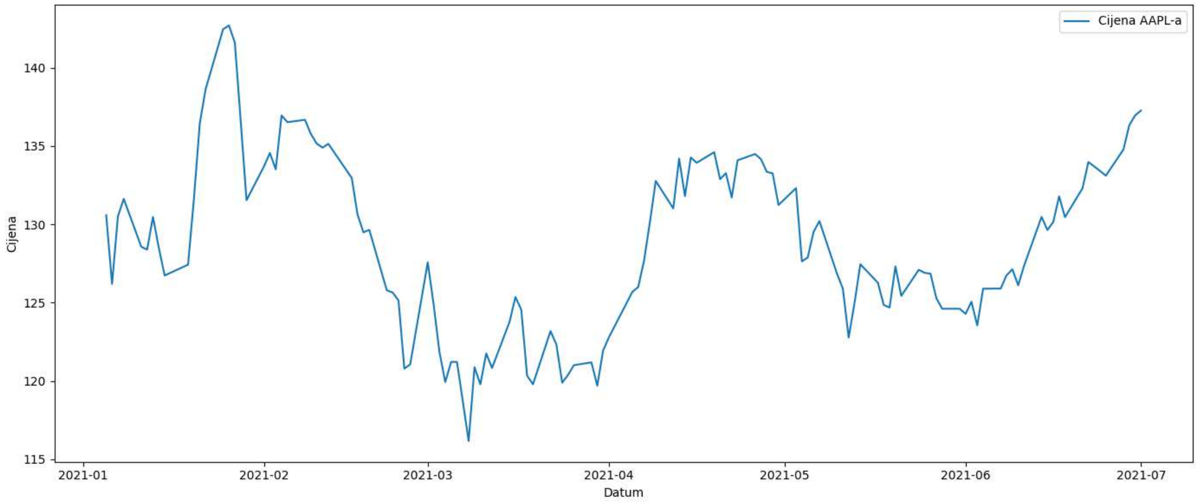
U ovom poglavlju razmotrit će se temeljni koncepti financijskih vremenskih nizova, uloga i značaj korelacije u financijama, te uvesti teorijski okvir latentnih faktora i faktorskih modela, s posebnim naglaskom na jednofaktorski model. Također će se detaljno obraditi matematička podloga ortogonalnog faktorskog modela koristeći relevantnu literaturu [2].

1.1. Financijski vremenski nizovi

Financijski vremenski nizovi predstavljaju sekvence podataka prikupljenih u pravilnim vremenskim intervalima, kao što su dnevne cijene dionica, kamatne stope ili tečajevi valuta. Ovi nizovi odražavaju povijesne performanse financijskih instrumenata i pružaju osnovu za analizu i predviđanje budućih kretanja. Ključno je razumjeti da su financijski vremenski nizovi često podložni složenim obrascima i visokim razinama volatilnosti, što čini njihovo modeliranje izazovnim zadatkom. Na primjer, cijene dionica mogu pokazivati nepredvidive skokove zbog neočekivanih tržišnih događaja, dok valute mogu fluktuirati u odgovoru na ekonomske politike ili geopolitičke tenzije.

Jedan od glavnih izazova u radu s financijskim vremenskim nizovima je osigurati da se modeli koji ih analiziraju ne koriste informacije iz budućnosti za predviđanje prošlosti. To je poznato kao problem "izlijevanja informacija", i može dovesti do nepravilnog prilagođavanja modela i pretjeranog optimizma u prognozama. Stoga je ključno pažljivo razdvojiti podatke za učenje od onih za ispitivanje kako bi se osigurala valjanost modela.

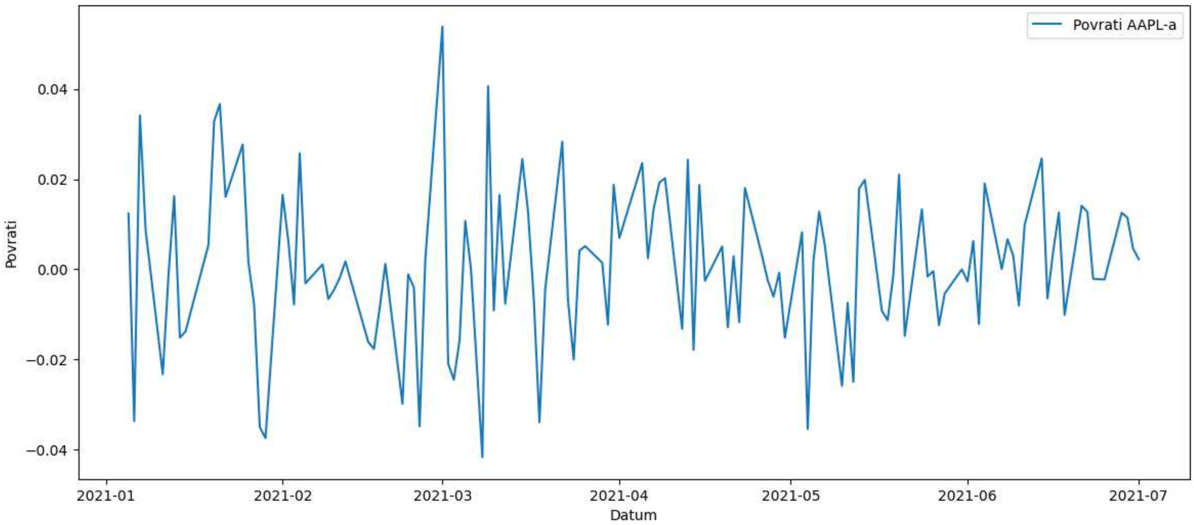
Isječak vremenskog niza koji prikazuje kretanje cijene jedne dionice Applea je prikazan slikom Slika 1.1. No, u analizi financijskih vremenskih nizova, uobičajena praksa je raditi s povratima dionica umjesto s njihovim apsolutnim cijenama. Povrati predstavljaju relativne promjene cijena dionica i obično su stacionarniji od cijena. Stacionarnost znači da statistička svojstva kao što su srednja vrijednost i varijanca ostaju približno jednaki kroz vrijeme, što olakšava modeliranje i analizu. Povrati također omogućuju usporedbu različitih dionica koje mogu imati različite apsolutne cijene. Relativne promjene cijena dionica olakšavaju analizu jer uklanjaju razlike u razmjerima.



Slika 1.1 Graf cijene dionice AAPL kroz vrijeme

Matematički, povrat dionice između dva vremenska trenutka t i t+1 računa se prema izrazu (1). gdje je P_t cijena dionice u trenutku t, a R_t povrat u periodu t. Prebacivanjem cijena sa slike Slika 1.1 u povrate, dobivamo graf na Slika 1.2.

$$R_t = \frac{P_{t+1} - P_t}{P_t} \tag{1}$$



Slika 1.2 Graf povrata dionice AAPL kroz vrijeme

1.1.1. Varijanca

Varijanca je mjera disperzije podataka oko njihove srednje vrijednosti i ključna je za razumijevanje volatilnosti u financijskim vremenskim nizovima. Varijanca daje kvantitativnu procjenu koliko su vrijednosti udaljene od prosjeka. Varijanca vektora populacije **X** je kao populacijski parametar definirana preko operatora očekivanja, prikazano jednadžbom (2).

$$\mathrm{Var}(\mathbf{X})=\mathbb{E}(\mathbf{X}^2)-\mathbb{E}(\mathbf{X})^2 \tag{2}$$

Međutim, u praksi se često radi s uzorcima, tj. podskupom cijele populacije. U takvim slučajevima, za procjenu populacijske varijance se koristi uzoračka varijanca. Uzoračka varijanca se računa pomoću formule

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2 \tag{3}$$

gdje je $\overline{X}$ srednja vrijednost uzorka, a *n* broj promatranja u uzorku.

Korištenje $n-1$ umjesto *n* u nazivniku poznato je kao Besselova korekcija. Ovo se radi kako bi se kompenzirala pristranost u procjeni varijance. Prilikom uzimanja uzorka, srednja vrijednost uzorka $\overline{X}$ je procjena populacijskog očekivanja, ali ona ne reflektira nužno cijelu populaciju, već samo njen dio. Oduzimanjem 1 od broja promatranja smanjujemo pristranost i osiguravamo da je procjena varijance konzistentnija i nepristrana, što je ključno za ispravne statističke analize.

Radi jednostavnosti kasnijih računa, poželjno je da varijance svih instrumenata budu zapisane u jednu strukturu. S obzirom na potrebe rada, koristit će se dijagonalna matrica **V** koja sadrži varijance svih instrumenata na dijagonali, ostali elementi su 0.

1.1.2. Standardna devijacija

Jedan od problema s varijancom je njena izraženost u kvadratnim jedinicama izvorne varijable ***X***. Na primjer, ako su cijene dionica izražene u dolarima, varijanca će biti izražena u kvadratnim dolarima, što je teško interpretirati. Kako bi se riješio ovaj problem, koristi se standardna devijacija iz jednadžbe (4). Standardna devijacija vraća mjerne jedinice natrag u originalne jedinice varijable ***X***, što omogućava lakšu interpretaciju širine distribucije podataka. Za uzorak, standardna devijacija se procjenjuje preko (5).

$$\sigma=\sqrt{\mathrm{Var}(\mathbf{X})} \tag{4}$$

$$s=\sqrt{\frac{1}{\text{n-1}}\sum_{\text{i=1}}^{\text{n}}(\text{X}_\text{i}\text{-}\overline{\mathbf{X}})^2} \tag{5}$$

U kontekstu financija, standardna devijacija je često korištena kao mjera volatilnosti dionica ili portfelja, omogućujući investitorima bolje razumijevanje rizika povezanog s varijabilnošću njihovih ulaganja. Dijagonalna matrica sa standardnim devijacijama na dijagonali je ekvivalentna izrazu $\mathbf{V}^{1/2}$.

1.1.3. Kovarijanca

Kovarijanca mjeri zajedničku varijaciju dviju varijabli. U financijama je ključna za razumijevanje kako se povrati različitih dionica kreću zajedno. Matematički, kovarijanca između dvije varijable $\boldsymbol{X_1}$ i $\boldsymbol{X_2}$ se definira kao očekivana vrijednost odstupanja proizvoda od njihovih srednjih vrijednosti. Za empirijske podatke, kovarijanca se procjenjuje formulom

$$\text{Cov}(\boldsymbol{X}_1,\boldsymbol{X}_2)=\frac{1}{p}\sum_{i=1}^{p}(\text{X}_{1,\text{i}}\text{-}\mu_{X_1})(X_{2,\text{i}}\text{-}\mu_{X_2}) \tag{6}$$

gdje μ_{X_1} i μ_{X_2} predstavljaju srednje vrijednosti $\boldsymbol{X_1}$ i $\boldsymbol{X_2}$. Formulom (7) se može doći do kovarijacijske matrice. Ovdje je $\mathbf{X}$ matrica dimenzija $n \times p$ (broj varijabli × broj stanja), a $\overline{\mathbf{X}}$ je vektor srednjih vrijednosti, gdje i-ti element predstavlja srednju vrijednost i-tog retka matrice $\mathbf{X}$.

$$\mathbf{\Sigma}=\frac{1}{\text{n-1}}(\mathbf{X}-\overline{\mathbf{X}})^{\text{T}}(\mathbf{X}-\overline{\mathbf{X}}) \tag{7}$$

Pozitivna kovarijanca znači da varijable imaju tendenciju kretati se u istom smjeru, dok negativna kovarijanca znači da se kreću u suprotnim smjerovima. U analizi portfelja, kovarijanca omogućuje procjenu kako kombinacija različitih dionica može smanjiti ili povećati rizik. Ako dvije dionice imaju negativnu kovarijancu, one mogu pružiti prirodnu diversifikaciju, smanjujući ukupni rizik portfelja. Ovo svojstvo se koristi u izradi strategija minimalne varijance, gdje se ciljaju kombinacije investicija koje minimiziraju ukupnu volatilnost portfelja.

1.1.4. Korelacija

Korelacija je normalizirana mjera kovarijance koja kvantificira jačinu i smjer linearnog odnosa između dviju varijabli. Korelacija je uvijek unutar raspona od -1 do 1, gdje vrijednosti blizu 1 ili -1 označavaju snažnu linearnu vezu, dok vrijednosti blizu 0 sugeriraju slab ili nikakav linearni odnos. Korelacijski koeficijent $\rho_{1,2}$ između varijabli $\boldsymbol{X_1}$ i $\boldsymbol{X_2}$ definira se pomoću (8), a korelacijska matrica **R** je jednadžbom (9). **R** je matrica korelacija između dionica, gdje svaki element R_{ij} predstavlja korelaciju između dionica i i j. $\mathbf{V}^{-1/2}$ je dijagonalna matrica recipročnih standardnih devijacija. Standardne devijacije koristimo za skaliranje kovarijanci kako bismo dobili korelacije [3][2].

$$\rho_{1,2}=\frac{\mathrm{Cov}(\boldsymbol{X}_1,\boldsymbol{X}_2)}{\sigma_2\sigma_2} \tag{8}$$

$$\mathbf{R}=\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{\Sigma}\mathbf{V}^{-\frac{1}{2}} \tag{9}$$

Korelacija, za razliku od kovarijance, nema mjernu jedinicu, što olakšava interpretaciju njezine vrijednosti bez obzira na mjerne jedinice varijabli.

U financijama, korelacija je korisna za analizu povezanosti između različitih financijskih instrumenata. Korelacija između povrata različitih dionica može pomoći u procjeni kako kombinacije tih dionica utječu na volatilnost i rizik portfelja. Na primjer, dionice koje pokazuju nisku ili negativnu korelaciju mogu ponuditi značajne prednosti u smislu diversifikacije, smanjujući rizik portfelja bez značajnog utjecaja na očekivani povrat.

1.2. Portfelj minimalne varijance

1.2.1. Definicija portfelja

Portfelj je zbirka financijskih instrumenata poput dionica, obveznica, derivativa i drugih vrsta imovine koje investitor posjeduje. Glavni cilj izgradnje portfelja je optimizirati odnos između rizika i povrata. Glavni cilj izgradnje portfelja je optimizacija odnosa između rizika i povrata. Na primjer, diverzifikacijom ulaganja investitori mogu smanjiti ukupni rizik bez proporcionalnog smanjenja povrata. Portfelji se mogu sastojati od različitih klasa imovina i industrijskih sektora kako bi se postigla diverzifikacija.

Težine pojedinih instrumenata u portfelju označavaju udio investiranog kapitala u svaki instrument pojedinih instrumenata u portfelju označavaju udio investiranog kapitala u svaki instrument. Matematički, težine portfelja w_i su definirane

$$\sum_{i=1}^{N} w_i = 1 \tag{10}$$

gdje je w_i težina i-tog instrumenta u portfelju, a N ukupan broj instrumenata. Težine određuju koliko kapitala se alocira u svaki pojedini instrument.

1.2.2. Duga i kratka pozicija

Zauzimanje duge pozicije (*long position*) podrazumijeva kupnju financijskog instrumenta s očekivanjem da će njegova cijena rasti, čime se ostvaruje dobit prilikom buduće prodaje po višoj cijeni. Ova strategija je uobičajena i jednostavna za većinu investitora.

Nasuprot tome, kratka pozicija (*short position*) uključuje posuđivanje financijskog instrumenta i njegovu prodaju na tržištu s očekivanjem pada cijene. Kada cijena padne, investitor ponovno kupuje instrument po nižoj cijeni i vraća ga posuđivaču, ostvarujući profit od razlike u cijenama. Kratke pozicije mogu biti riskantne jer cijena instrumenta može porasti neograničeno, potencijalno stvarajući neograničene gubitke.

1.2.3. Kratki povijesni pregled

Koncept portfelja minimalne varijance dolazi iz teorije portfelja Harryja Markowitza iz 1952. godine. Markowitzeva teorija, poznata kao *Modern Portfolio Theory (MPT)*, predstavlja temelj za razumijevanje odnosa između rizika i povrata u izgradnji portfelja. Portfelj minimalne varijance je strategija koja ima za cilj minimizirati varijancu povrata portfelja, čineći ga privlačnim za investitore koji žele smanjiti rizik bez značajnog smanjenja očekivanog povrata.

Koristeći *MPT*, investitori mogu konstruirati portfelje koji maksimiziraju povrat za zadani nivo rizika ili minimiziraju rizik za zadani nivo povrata. Portfelj minimalne varijance optimizira diverzifikaciju, smanjujući izloženost volatilnosti pojedinačnih instrumenata i osiguravajući stabilniji povrat.

1.2.4. Težine portfelja

U izračunu težina za globalni portfelj minimalne varijance (*GMV*), cilja se na minimiziranje ukupne varijance portfelja uz dopuštenje kratkih pozicija [4], koje su prikazane negativnim težinama. Težine **w** koje minimiziraju varijancu portfelja, uz uvjet (10), određene su rješavanjem problema minimizacije

$$\min_{\mathbf{w}} \mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w} \tag{11}$$

gdje je **Σ** kovarijacijska matrica povrata instrumenata u portfelju. Ovaj problem se može riješiti korištenjem Lagrangeove metode multiplikatora ili metoda za optimizaciju kvadratnih funkcija. Optimalne težine **w** za portfelj minimalne varijance mogu se izraziti kao

$$\mathbf{w}=\frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{1}} \tag{12}$$

gdje je **1** vektor jedinica duljine *N*. Ova formula omogućuje izračunavanje težina za portfelj koji minimizira varijancu povrata, uzimajući u obzir kovarijancu između različitih instrumenata. Pozitivne težine ukazuju na alokacije dugih pozicija, a negativne kratkih.

1.2.5. Ograničavanje težina

Portfelj minimalne varijance s ograničenim težinama *(constrained minimum variance portfolio)* je varijanta gdje su težine ograničene na interval [0,1], čime se isključuju kratke pozicije. Ovo ograničenje znači da investitor može samo kupovati instrumente, bez mogućnosti kratke prodaje. Time se dodatno smanjuje rizik povezan s volatilnošću instrumenata, jer kratke pozicije mogu potencijalno dovesti do većih gubitaka.

U ovoj konfiguraciji, težine *CMV* portfelja su optimizirane tako da minimiziraju varijancu, ali sve težine su nenegativne i zbroj težina je jednak 1. Ograničavanje težina na nenegativne vrijednosti pruža sigurnost konzervativnijim investitorima i smanjuje izloženost ekstremnim gubicima koji mogu nastati zbog kratke prodaje. Ovaj pristup omogućuje izgradnju portfelja koji je stabilniji i manje podložan fluktuacijama tržišta.

U oba slučaja, cilj je smanjiti ukupnu volatilnost portfelja, čime se optimizira odnos između rizika i povrata, omogućujući investitorima da maksimiziraju svoje prinose uz minimalnu izloženost riziku.

1.3. Problem inverza kovarijacijske matrice

1.3.1. Kovarijacijska matrica

Matrica kovarijance je simetrična matrica koja prikazuje varijance i kovarijance između skupa podataka. U kontekstu portfelja dionica, u formulu (7) za **X** se koristi matrica $N \times T$ gdje se nalaze povrati N različitih dionica za svaki trenutak na prozoru duljine T. Tada je kovarijacijska matrica oblika (13).

$$\mathbf{\Sigma} = \begin{bmatrix} Cov(X_1, X_1) & \cdots & Cov(X_1, X_N) \\ \vdots & \ddots & \vdots \\ Cov(X_N, X_1) & \cdots & Cov(X_N, X_N) \end{bmatrix} \tag{13}$$

Korisno je primijetiti da vrijedi $Cov(X_i, X_i) = Var(X_i)$, te se **Σ** može promatrati kao (14). Nadalje, ako se svi elementi postave na nulu, osim dijagonalnih, dobije se matrica **V**.

$$\mathbf{\Sigma} = \begin{bmatrix} Var(X_1) & \cdots & Cov(X_1, X_N) \\ \vdots & \ddots & \vdots \\ Cov(X_N, X_1) & \cdots & Var(X_N) \end{bmatrix} \tag{14}$$

1.3.2. Inverz matrice

Inverz matrice je koncept linearne algebre koji se koristi u mnogim područjima, uključujući financije i optimizaciju portfelja. Za kvadratnu matricu **A** dimenzija $N \times N$, inverz, označen s $\mathbf{A}^{-1}$, je matrica takva da vrijedi

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \tag{15}$$

gdje je **I** jedinična matrica istih dimenzija kao i **A**. Postojanje inverza matrice je uvjetovano time da matrica **A** bude nesingularna. Ako **A** nije invertibilna, njena determinanta je nula (det(**A**)=0), što znači da je rang matrice manji od njenog broja redaka ili stupaca.

U kontekstu financija, posebno pri konstrukciji portfelja, inverz kovarijacijske matrice se koristi za izračun težina portfelja minimalne varijance. Kada radimo s velikim brojem instrumenata N, problem nastaje ako imamo relativno mali broj vremenskih točaka T. U slučaju gdje je $N \geq T$, kovarijacijska matrica može postati singularna jer nema dovoljno informacija za stabilnu procjenu svih kovarijanci i varijanci [5]. Ako su povrati različitih instrumenata visoko korelirani, može doći do linearne zavisnosti među stupcima matrice, što također vodi do singularnosti.

1.3.3. Povećanje vremenskih prozora

Jedan od načina za rješavanje problema singularnosti je korištenje većeg broja vremenskih točaka T za izračunavanje kovarijacijske matrice. Povećanjem T možemo povećati stabilnost procjene kovarijacijske matrice jer se povećava broj stupaca koji mogu biti uspješno uključeni bez gubitka ranga matrice.

Međutim, korištenje dužih vremenskih prozora dolazi s vlastitim problemima. Povijesni podaci koji obuhvaćaju dulje razdoblje mogu sadržavati zastarjele informacije koje nisu relevantne za trenutne tržišne uvjete, zbog čega se promjene u dinamici tržišta moraju uzeti u obzir. Financijski instrumenti često mijenjaju svoje obrasce ponašanja tijekom vremena, što može učiniti stare podatke manje korisnima za trenutne procjene rizika.

1.3.4. Pozitivno definitne matrice

Zbog prijašnje navedenih matričnih svojstava, posebno u uvjetima kada se procjenjuje kovarijacijska matrica s mnogo instrumenata na malom vremenskom prozoru, često nije moguće jednostavno izračunati inverz kovarijacijske matrice $\mathbf{\Sigma}$. Bez stabilnog inverza $\mathbf{\Sigma}$, standardne metode za računanje težina portfelja minimalne varijance postaju neprimjenjive ili vrlo nepouzdane.

Kako bi se riješio problem singularnosti i osiguralo da je kovarijacijska matrica $\mathbf{\Sigma}$ invertibilna, mogu se tražiti pozitivno definitne matrice. Pozitivno definitna matrica je matrica koja zadovoljava sljedeće uvjete:

1. Simetrija: Matrica mora biti simetrična, tj. $\mathbf{A}=\mathbf{A}^{\mathrm{T}}$.
2. Pozitivno definitna svojstva: Za svaki nenulti vektor x vrijedi $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}>0$.

Pozitivno definitne matrice imaju važnu osobinu da su uvijek invertibilne. To znači da njihova determinanta nikada nije nula, osiguravajući uspješan izračun inverza.

1.4. Faktorski model

1.4.1. Latentni faktori

Latentni faktori su neopažene varijable koje utječu na promatrane podatke. U kontekstu financija, latentni faktori mogu uključivati makroekonomske varijable, industrijske trendove ili druge sile koje oblikuju kretanja cijena dionica. Identificiranje i modeliranje ovih latentnih faktora omogućava analitičarima bolje razumijevanje strukture i dinamike tržišta.

Jednofaktorski modeli koriste jedan zajednički latentni faktor za objašnjavanje povrata dionica. Iako postoji mnogo mogućih faktora koji mogu utjecati na cijene dionica, stanje sveukupnog tržišta se pokazalo kao najbolji. Stoga ćemo se u ovom radu fokusirati na taj glavni faktor. Ovo se često naziva tržišni faktor, a može biti reprezentiran tržišnim indeksom ili drugim relevantnim ekonomskim indikatorom.

1.4.2. Povijesni pregled faktorskih modela

Povijest faktorskih modela seže do početka 20. stoljeća, kada su prvi put korišteni za analizu psihometrijskih podataka. U financijama, faktorski modeli postali su popularni za analizu povrata dionica, omogućujući procjenu utjecaja zajedničkih faktora na različite financijske instrumente.

1.4.3. Ortogonalni jednofaktorski model

Ortogonalni jednofaktorski model koristi se za analizu podataka s ciljem razumijevanja odnosa između više varijabli kroz utjecaj jednog zajedničkog faktora. Ovaj model pretpostavlja da je promatrani vektor varijabli **X** linearno ovisan o jednom latentnom faktoru F i specifičnim čimbenicima ili pogreškama e. Pojedinačne varijable X_i mogu se izraziti preko

$$X_i - \mu_i = \beta_i F + e_i \tag{16}$$

gdje je:

- X_i promatrana varijabla,
- μ_i sredina varijable X_i,
- β_i koeficijent varijable X_i na latentni faktor F,

- F latentni faktor,
- e_i specifični čimbenik varijable X_i.

S obzirom na količinu podataka, za rad će biti korisniji vektorski zapis koji će za N varijabli izgledati

$$\mathbf{X}\text{-}\boldsymbol{\mu}=\boldsymbol{\beta}F+\mathbf{e} \tag{17}$$

gdje je:

- $\mathbf{X}$ vektor promatranih varijabli (dimenzije $N\times 1$),
- $\boldsymbol{\mu}$ vektor sredina (dimenzije $N\times 1$),
- $\boldsymbol{\beta}$ vektor koeficijenata faktorskog modela (dimenzije $N\times 1$),
- F latentni faktor (dimenzije 1×1),
- $\mathbf{e}$ vektor specifičnih čimbenika (dimenzije $N\times 1$).

Ortogonalni jednofaktorski model zahtijeva nekoliko pretpostavki koje moraju biti zadovoljene:

1. $\mathbb{E}[F\,]=0$ (Očekivana vrijednost latentnog faktora je nula)
2. $\mathrm{Var}(F\,)=1$ (Varijanca latentnog faktora je jedan)
3. $\mathbb{E}[\mathbf{e}]=0$ (Očekivana vrijednost specifičnih čimbenika je nula)
4. $\mathrm{Cov}(\mathbf{e})=\boldsymbol{\Psi}=\mathrm{diag}(\psi_1,\psi_2,\ldots,\psi_N)$ (Specifični čimbenici su nezavisni; njihova kovarijanca je dijagonalna matrica)
5. $\mathrm{Cov}(F,\mathbf{e})=0$ (Latentni faktor i specifični čimbenik su nezavisni)

Prilikom korištenja ovog modela, pažljivo će se vršiti provjera da su ove pretpostavke zadovoljene kako bismo osigurali točnost i valjanost rezultata.

Koristeći svojstva kovarijance (18) i pretpostavke jednofaktorskog modela, moguće je izvesti formulu kovarijacijske matrice $\boldsymbol{\Sigma}$ za podatke objašnjene tim modelom.

$$\begin{aligned}\mathrm{Cov}(\mathbf{X}) &= \mathrm{Cov}(\boldsymbol{\beta}F + \mathbf{e})\\ &= \mathrm{Cov}(\boldsymbol{\beta}F) + \mathrm{Cov}(\mathbf{e}) + \mathrm{Cov}(\boldsymbol{\beta}F,\mathbf{e}) + \mathrm{Cov}(\mathbf{e},\boldsymbol{\beta}F)\end{aligned} \tag{18}$$

Zbog ortogonalnosti faktora i grešaka, $\mathrm{Cov}(\boldsymbol{\beta}F,\mathbf{e})$ i $\mathrm{Cov}(\mathbf{e},\boldsymbol{\beta}F)$ su nula, a uz pretpostavku da je $\mathrm{Var}(F)= 1$, vrijedi (19).

$$\mathrm{Cov}(\mathbf{X}) = \mathrm{Cov}(\boldsymbol{\beta}F) + \mathrm{Cov}(\mathbf{e}) \tag{19}$$

$$= \boldsymbol{\beta} \cdot \mathrm{Var}(F) \cdot \boldsymbol{\beta}^{\mathrm{T}} + \mathrm{Cov}(\mathbf{e}) = \boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}} + \mathrm{Cov}(\mathbf{e})$$

Znajući da je Cov(**e**)= $\boldsymbol{\Psi}$, konačna matrica $\boldsymbol{\Sigma}$ je

$$\boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}} + \boldsymbol{\Psi} \tag{20}$$

gdje je:

- $\boldsymbol{\Sigma}$ kovarijacijska matrica promatranih varijabli (dimenzije $N \times N$),
- $\boldsymbol{\beta}$ vektor koeficijenata faktorskog modela (dimenzije $N \times 1$),
- $\boldsymbol{\Psi}$ dijagonalna matrica specifičnih varijanci (dimenzije $N \times 1$).

1.4.4. Dokaz pozitivne definitnosti matrice Σ

Da bi matrica $\boldsymbol{\Sigma}$ bila pozitivno definitna, mora vrijediti (21) za svaki nenulti vektor $\mathbf{x} \in \mathbb{R}^N$.

$$\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{x} > 0 \tag{21}$$

Analiza sastavnih dijelova $\boldsymbol{\Sigma}$:

1. Svojstva matrice $\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}$:
   - Matrica $\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}$ je rezultat vanjskog produkta vektora $\boldsymbol{\beta}$ sa samim sobom.
   - $\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}$ je semi pozitivno definitna jer je $(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})^2 \geq 0$ za svaki $\mathbf{x}$, ali može imati rang manji od N pa može biti degenerirana (ne pozitivno definitna).
2. Svojstva dijagonalne matrice $\boldsymbol{\Psi}$:
   - Dijagonalna matrica $\boldsymbol{\Psi}$ ima pozitivne elemente $\psi_i > 0$ na dijagonali.
   - $\boldsymbol{\Psi}$ je pozitivno definitna jer za svaki nenulti vektor $\mathbf{x}$ vrijedi (22).

$$\mathbf{x}^{\mathrm{T}}\boldsymbol{\Psi}\mathbf{x} = \sum_{i=0}^{n} \psi_i x_i^2 > 0 \tag{22}$$

3. Suma $\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}$ i $\boldsymbol{\Psi}$:
   - Kombinacijom $\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}$ i $\boldsymbol{\Psi}$ dobivamo (20).
   - Za svaki nenulti vektor $\mathbf{x}$ vrijedi (23).

$$\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{x} = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\boldsymbol{\Psi}\mathbf{x} \tag{23}$$

   - Prvi dio, $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}=(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})^2$, je uvijek nenegativan (≥ 0).
   - Drugi dio, $\mathbf{x}^{\mathrm{T}}\boldsymbol{\Psi}\mathbf{x}$, je pozitivan jer $\boldsymbol{\Psi}$ ima pozitivne dijagonalne elemente.

Kombinacijom semi pozitivno definitne matrice $\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}$ i pozitivno definitne matrice $\boldsymbol{\Psi}$, uvijek će se dobiti pozitivno definitna matrica $\boldsymbol{\Sigma}$. Matrica $\boldsymbol{\Psi}$ osigurava pozitivnu vrijednost kvadratne forme $\mathbf{x}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{x}$, što garantira da je $\mathbf{\Sigma}$ pozitivno definitna. Dakle, u ortogonalnom jednofaktorskom modelu, kovarijacijska matrica $\mathbf{\Sigma} = \boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}} + \mathbf{\Psi}$ je uvijek pozitivno definitna, i to svojstvo omogućava računanje težina portfelja minimalne varijance.

2. Modeli za procjenu i predviđanje korelacijskih matrica

Trenutno poglavlje će se baviti primjenom i proširivanjem spomenutih teorijskih koncepata. Istražit će načine ekstrakcije i modifikacije podataka, te njihovu implementaciju u jednofaktorski model. Uz to će proći kroz razne procjenitelje budućih vrijednosti, koji će u teoriji omogućiti izgradnju boljih portfelja.

2.1. Procjena koeficijenata faktorskog modela

2.1.1. Tržišni indeks

Kako bi se izgradio jednofaktorski model, potrebno je identificirati zajednički latentni faktor koji najbolje opisuje tržišne uvjete. U ovom kontekstu, koristi se tržišni indeks kao latentni faktor. Tržišni indeks je mjera koja predstavlja prosječno kretanje odabranog skupa dionica i može se koristiti kao pokazatelj ukupnog tržišnog stanja. Umjesto da se oslanja na postojeće indekse, izgradit će se vlastiti, koristeći pristup portfelja jednakih težina (*equal weight portfolio*).

U takvom portfelju svaka dionica se smatra jednako važnom, tj. svaka dionica doprinosi jednako ukupnom povratu portfelja. Za razliku od portfelja ponderiranih tržišnom kapitalizacijom, gdje veće tvrtke dominiraju, u *equal weight* portfelju svaka dionica, bez obzira na njezinu veličinu, cijenu ili volumen trgovanja, ima jednak utjecaj na ukupni povrat portfelja.

Za portfelj s N dionica, a povrati ovih dionica kroz vrijeme su organizirani u matricu (24) gdje je $r_{i,t}$ povrat i-te dionice u trenutku t. Neka $\mathbf{r}$ bude matrica dimenzija $N \times T$, gdje je N broj dionica, a T broj vremenskih perioda.

$$\boldsymbol{r} = \begin{bmatrix} r_{1,1} & \cdots & r_{1,T} \\ \vdots & \ddots & \vdots \\ r_{N,1} & \cdots & r_{N,T} \end{bmatrix} \tag{24}$$

U matričnom obliku, povrat portfelja jednakih težina kroz sve vremenske trenutke $t=1,2,\ldots,T$ može se predstaviti kao vektor $\mathbf{r}_{\mathrm{eq}}$ dimenzija $1 \times T$, gdje je svaki element $\mathbf{r}_{\mathrm{eq,t}}$ prosječni povrat u trenutku t. Koristeći vektor jedinica $\mathbf{1}$ dimenzija $N \times 1$, može se zapisati formula (25).

$$\boldsymbol{r}_{\mathrm{eq}}=\frac{1}{\mathrm{N}}\mathbf{1}^{\mathrm{T}}\boldsymbol{r} \tag{25}$$

Kreiranjem ovakvog portfelja iz svih dionica koje će se koristiti u radu, stvara se kompozitni indikator koji reprezentira prosječne tržišne uvjete. Ovaj portfelj će služiti kao latentni faktor F u faktorskom modelu, i kao mjerilo za usporedbu s konstruiranim portfeljima minimalne varijance.

2.1.2. Standardizacija povrata

Za implementaciju ortogonalnog faktorskog modela, potreban je latentni faktor F koji zadovoljava uvjet $\mathbb{E}(F)=0$ i $\mathrm{Var}(F)=1$. Da bi se to postiglo, povrate tržišnog indeksa treba normalizirati. Normalizacija uključuje oduzimanje srednje vrijednosti i dijeljenje sa standardnom devijacijom povrata. Ovo osigurava da takvi standardizirani povrati imaju srednju vrijednost 0 i standardnu devijaciju jednaku 1. Formula za standardizirani povrat Z u vremenskom trenutku t je

$$Z=\frac{r_t-\mu}{\sigma} \tag{26}$$

gdje je r_t povrat portfelja ili dionice u trenutku t, μ je srednja vrijednost povrata, a σ je standardna devijacija povrata. Kada je standardna devijacija povrata nula (što implicira da su svi povrati isti), standardizirani povrati su također nula. Ovakvi povrati će se koristiti kao latentni faktor F.

U modelu koji će se koristiti, povrati dionice i kroz vrijeme će biti označeni vektorom $\mathbf{X}_i$, a bit će izraženi kao linearna kombinacija faktora $\mathbf{F}$ i specifičnih pogrešaka $\mathbf{e}$

$$\boldsymbol{X}_i-\boldsymbol{\mu}=\beta_i\mathbf{F}+\mathbf{e} \tag{27}$$

gdje je β_i koeficijent za dionicu i. Da bi model bio konzistentan, povrati pojedinačnih dionica koje želimo objasniti modelom će se također standardizirati. Standardizacija ovdje također uključuje oduzimanje srednje vrijednosti $\mu_{\boldsymbol{X}_i}$ i dijeljenje sa standardnom devijacijom $\sigma_{\boldsymbol{X}_i}$ na određenom vremenskom rasponu

$$Z_i=\frac{X_{i,t}-\mu_{\boldsymbol{X}_i}}{\sigma_{\boldsymbol{X}_i}} \tag{28}$$

gdje $Z_{X_{i,t}}$ predstavlja standardizirani povrat dionice *i* u trenutku *t*. Time je osigurano da su sve varijable u modelu usporedive i da model ne mora predviđati srednje standardizirane povrate (jer su 0), već samo odstupanja od prosječnog povrata. Time se izgled modela mijenja u (29).

$$\mathbf{Z}_i = \beta_i \mathbf{F} + \mathbf{e} \tag{29}$$

2.1.3. Metoda najmanjih kvadrata

Sljedeći fokus rada je izračunavanje parametra β, koji je bitan za povezivanje povrata pojedinačnih dionica s tržišnim indeksom. Oni koji su upoznati sa statistikom mogu primijetiti da formula modela ima sličnosti s modelom linearne regresije bez slobodnog člana. Jedan od načina procjene parametra **β** je koristeći metodu najmanjih kvadrata (*least squares linear regression*).

Linearna regresija je osnovna tehnika u statistici koja se koristi za modeliranje odnosa između zavisne varijable (odgovora) *Y* i jedne ili više nezavisnih varijabli (prediktora) *X*. U svom najjednostavnijem obliku, linearna regresija s jednom nezavisnom varijablom *X* može se izraziti kao

$$Y = \alpha + \beta X + e \tag{30}$$

gdje je:

- *Y* zavisna varijabla,
- α slobodni član (*intercept*),
- β nagib (*slope*),
- *X* nezavisna varijabla,
- *e* slučajna pogreška.

Cilj regresije je pronaći vrijednosti α i β koje minimiziraju razliku između stvarnih vrijednosti *Y* i predviđenih vrijednosti $\hat{Y}$ dobivenih modelom. Metoda najmanjih kvadrata minimizira sumu kvadrata razlika između stvarnih i predviđenih vrijednosti.

2.1.4. Faktorski model bez slobodnog člana

U kontekstu faktorskog modela ovoga rada, nemamo slobodni član α, jer se pretpostavlja da su povrati standardizirani tako da imaju srednju vrijednost nula. To znači da se takva linearna regresija može izraziti preko (31), te pritom:

- $\mathbf{X}_i$ predstavlja standardizirane povrate pojedine dionice i (dimenzije $T\times 1$),
- $\mathbf{F}$ su standardizirani povrati tržišnog indeksa (dimenzije $T\times 1$),
- β_i je koeficijent regresije ili faktorskog modela,
- $\mathbf{e}$ je komponenta specifičnih pogrešaka (dimenzije $T\times 1$).

$$\boldsymbol{X}_i=\beta_i\mathbf{F}+\boldsymbol{e} \tag{31}$$

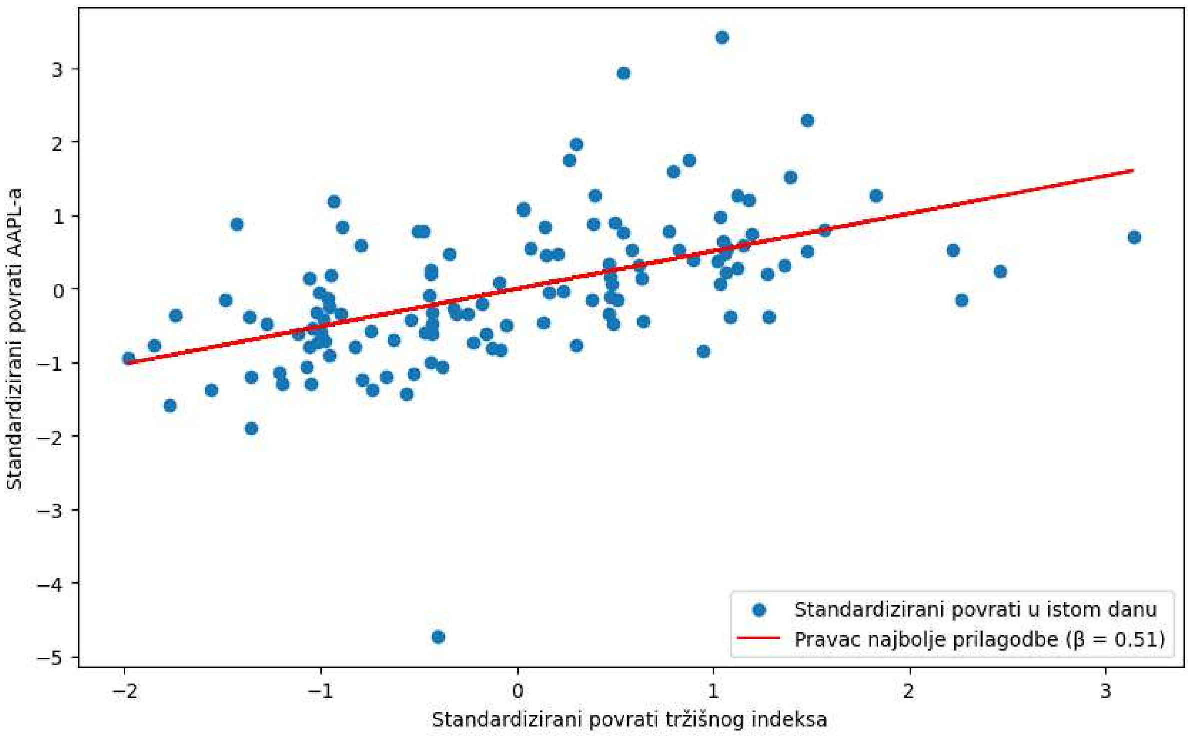
Cilj je procijeniti vrijednost β koja najbolje opisuje odnos između povrata dionice X i tržišnog indeksa F.

Procjena β_i se svodi na pronalaženje koeficijenata koji minimiziraju sumu kvadrata razlika između stvarnih vrijednosti zavisne varijable i vrijednosti predviđenih modelom. Do rješenja se može doći formulom (32). Ona dolazi iz minimiziranja sume kvadrata razlika između stvarnih povrata $\mathbf{X}_i$ i predviđenih vrijednosti $\beta_i\mathbf{F}$. Rezultat daje procjenu koliko se povrat dionice $\mathbf{X}_i$ mijenja u odnosu na promjene tržišnog indeksa $\mathbf{F}$.

$$\beta_i=\frac{\mathbf{F}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{i}}}{\mathbf{F}^{\mathrm{T}}\mathbf{F}} \tag{32}$$

β_i je mjera senzitivnosti povrata dionice na povrate tržišnog indeksa. Viša vrijednost β_i implicira jaču povezanost s tržišnim uvjetima, dok niža vrijednost β_i sugerira veću nezavisnost dionice od tržišnih fluktuacija.

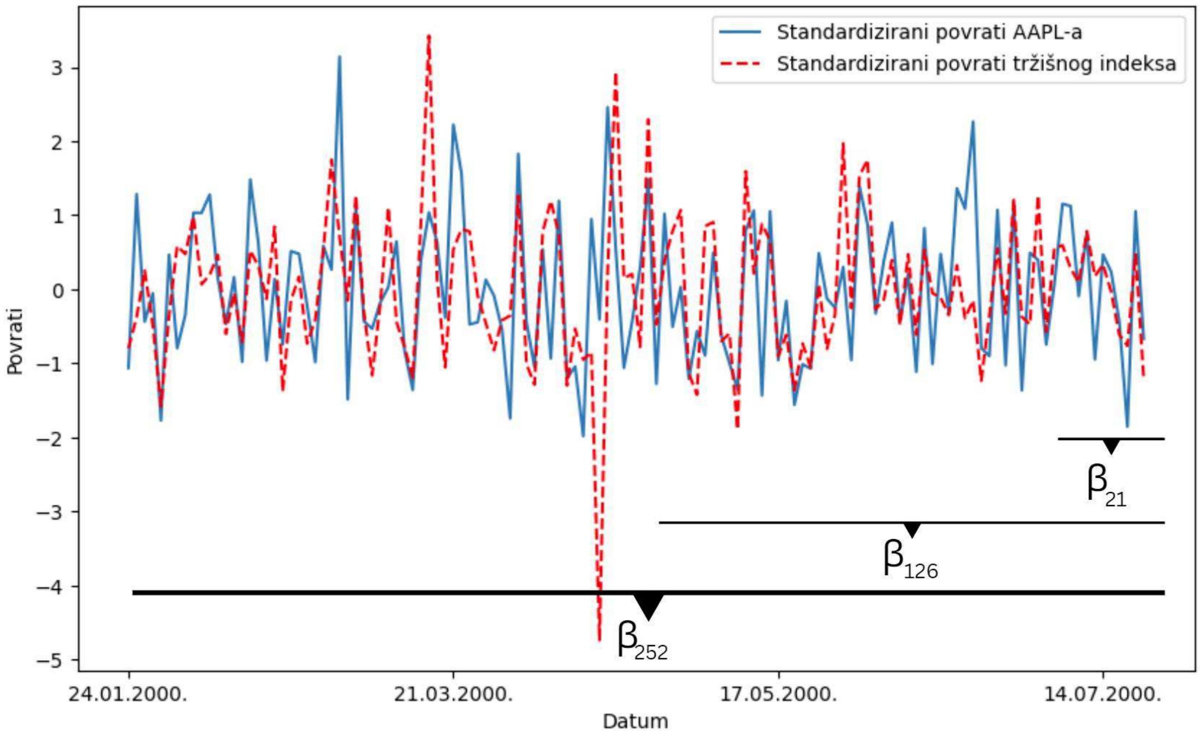
Korištenje linearne regresije za procjenu β_i je efikasna metoda jer je jednostavna za implementaciju i omogućuje korištenje standardiziranih povrata kako bi se dobila jasna slika o povezanosti između dionice i tržišta. Ovaj pristup omogućuje kvantifikaciju utjecaja zajedničkog tržišnog faktora na povrate pojedinačnih dionica kao što je vidljivo na Slika 2.1.



Slika 2.1 Pravac najbolje prilagodbe na standardizirane povrate AAPL-a i tržišnog indeksa

2.1.5. Klizeći prozori

Klizeći prozori omogućuju segmentiranje vremenskog niza povrata na manje podskupove podataka. Za svrhe rada koriste se tri različite veličine prozora: 21, 126 i 252 radna dana, što otprilike predstavlja jedan mjesec, šest mjeseci i jednu godinu trgovačkih dana. U trenutku *t*, gledaju se svi podaci unazad za duljinu prozora, te na tom rasponu se standardiziraju povrati dionica i tržišnog indeksa. Za taj prozor se izračunaju koeficijenti faktorskog modela kao ovisnost standardiziranih povrata dionica o standardiziranim povratima tržišnog indeksa (Slika 2.2). Prozori se pomiču naprijed u koracima od 21 radni dan (jedan kalendarski mjesec), te se ponavlja isti postupak.



Slika 2.2 Standardizirani povrati AAPL-a i tržišnog indeksa

2.1.6. *Oracle* procjenitelj

Osim prozora koji gledaju unazad, ima i prozor koji gleda 21 dan unaprijed. Na ovom prozoru se također računa koeficijent $\boldsymbol{\beta}$, te će se njega pokušavati predvidjeti. Ovaj pristup omogućuje usporedbu predikcija sa stvarnim budućim vrijednostima, pružajući mjerilo za procjenu točnosti modela. Takav procjenitelj će se zvati *Oracle*, u prijevodu prorok, jer zna buduće faktorske koeficijente.

2.2. Kovarijacijska matrica standardiziranih povrata

U ortogonalnom faktorskom modelu, kovarijacijska matrica povrata je definirana s (20). Međutim, standardizacijom povrata se osigurava da sve varijable imaju srednju vrijednost 0 i standardnu devijaciju 1. Stoga, kovarijanca između dvije standardizirane varijable postaje identična njihovoj korelaciji, jer je korelacijski koeficijent normalizirana kovarijanca. Odnosno, uvrštavanjem standardiziranih povrata $\mathbf{Z_1}$ i $\mathbf{Z_2}$ u formulu korelacije (8), dobiva se (33). Budući da su $\sigma_\mathbf{1}$ i $\sigma_\mathbf{2}$ jednaki 1, kovarijanca i korelacija se savršeno podudaraju.

$$\rho_{1,2} = \frac{\mathrm{Cov}(\mathbf{Z}_1, \mathbf{Z}_2)}{\sigma_1 \sigma_2} = \mathrm{Cov}(\mathbf{Z}_1, \mathbf{Z}_2) \tag{33}$$

Stoga, korištenjem ortogonalnog faktorskog modela i standardiziranih podataka, dobivaju se matrice koje predstavljaju korelacije umjesto kovarijanci. Ova svojstva su korisna jer onda automatski znamo i izgled dijagonalne matrice $\mathbf{\Psi}$. Dijagonalni elementi korelacijske matrice su uvijek 1 jer predstavljaju korelaciju varijable sa samom sobom, što je uvijek savršena korelacija. Budući da su povrati standardizirani, svaki povrat ima varijancu 1. Time se zna da je specifična greška ψ_i (koja nije objašnjena modelom) za dionicu i jednaka razlici između 1 i korelacije objašnjene modelom (element ii matrice $\mathbf{\Sigma}$). Dakle, specifična greška pokazuje koliko nedostaje do 1 da bi model potpuno objasnio povrat dionice.

2.3. Modeli za nadzirano učenje korelacijskih matrica

Predviđanje korelacijskih matrica za portfelje minimalne varijance zahtijeva korištenje raznih modela i pristupa. U ovom dijelu je objašnjena metodologija i koraci za izgradnju i evaluaciju tih modela.

2.3.1. Skup za učenje i ispitivanje

Podjela podataka na skupove za učenje i ispitivanje je ključan korak u izgradnji prediktivnih modela.

- **Skup za učenje**: traži se model koji dobro opisuje dostupne podatke. Ovo uključuje optimizaciju parametara modela kako bi se minimizirala pogreška predikcije na poznatim podacima. Cilj je naučiti model obrascima i relacijama unutar podataka koji će se moći koristiti za predviđanje budućih vrijednosti.
- **Skup za ispitivanje**: provjerava se koliko dobro ti modeli opisuju nikad dosad viđene podatke, tj. koliko dobro generaliziraju. Ispitni podaci simuliraju stvarne uvjete primjene modela, pružajući realnu procjenu njegove učinkovitosti i robusnosti. Evaluacija na skupu omogućuje identifikaciju potencijalnih problema poput prenaučenosti (engl. *overfitting*), što je pretjerana prilagodba modela podacima iz skupa za učenje.

Primjenom ovog pristupa, osigurava se da ne samo da prediktivni modeli dobro funkcioniraju na poznatim podacima već da su i sposobni pružiti točna predviđanja na novim, nepoznatim podacima.

2.3.2. Uzorački koeficijenti faktorskog modela

Uzoračke analize koriste se za procjenu koliko dobro izračunati faktorski koeficijenti na određenim prozorima opisuju buduće *Oracle* koeficijente. Ovo će omogućiti da se vidi koliko su korisni i koliko griješe rezultati izračunati jednostavnim tehnikama i bez kombiniranja više varijabli metodama strojnog učenja.

Za najjednostavniji model predviđanja budućih vrijednosti koeficijenata koristit će se 3 procjenitelja iz . Tako će se budući faktorski koeficijenti jednostavno procijeniti na temelju prethodno izračunatih koeficijenata na određenim prozorima.

$$\hat{\beta}_{21}=\beta_{21}$$
$$\hat{\beta}_{126}=\beta_{126} \quad (34)$$
$$\hat{\beta}_{126}=\beta_{126}$$

2.3.3. Linearna regresija

Linearni regresijski model koristi se za predviđanje budućih *Oracle* koeficijenata na temelju onih izračunatih na prijašnjim prozorima. Karakteristike ovog modela su:

- **Slobodni član**: Linearni regresijski model uključuje slobodni član koji omogućuje bolju prilagodbu podacima.
- **Ulazne varijable**: Ulazne varijable u ovom modelu su koeficijenti (β_{21}, β_{126}, β_{252}) izračunati na prijašnjim prozorima.

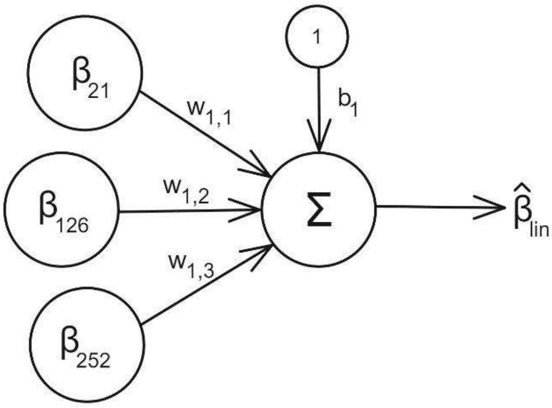
Model linearne regresije može se prikazati sljedećom jednadžbom

$$\hat{\beta}_{lin}=w_{1,1}\cdot\beta_{21}+w_{1,2}\cdot\beta_{126}+w_{1,3}\cdot\beta_{252}+b_1 \quad (35)$$

gdje su:

- $\hat{\beta}_{lin}$ predviđeni koeficijent,
- $w_{1,1}$, $w_{1,1}$, i $w_{1,1}$ težine modela,
- β_{21}, β_{126}, i β_{252} koeficijenti izračunati na prozorima od 21, 126 i 252 radna dana,
- b_1 slobodni član.

Grafički prikaz modela je vidljiv na Slika 2.3 (Σ ovdje predstavlja operaciju zbrajanja). Važno je napomenuti da u ovom modelu ne postoje nikakva ograničenja za koeficijente, pa model može dati predikcije izvan raspona [-1,1], što ne bi bila smislena vrijednost za korelaciju. Ovo zahtijeva pažljivo praćenje i validaciju rezultata kako bi se osiguralo da predikcije ostanu realistične.

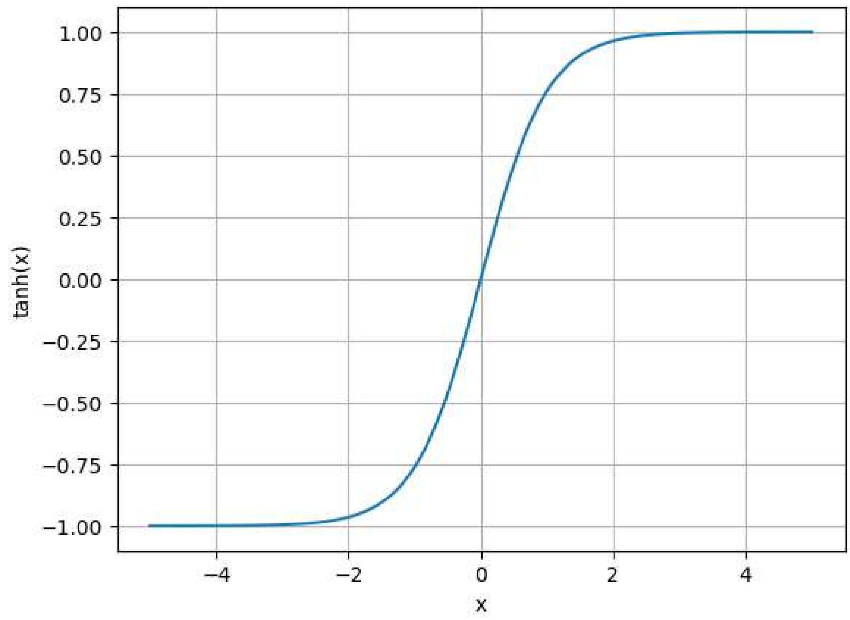


Slika 2.3 Grafički prikaz modela linearne regresije

2.3.4. Jednostavni perceptron s aktivacijskom funkcijom

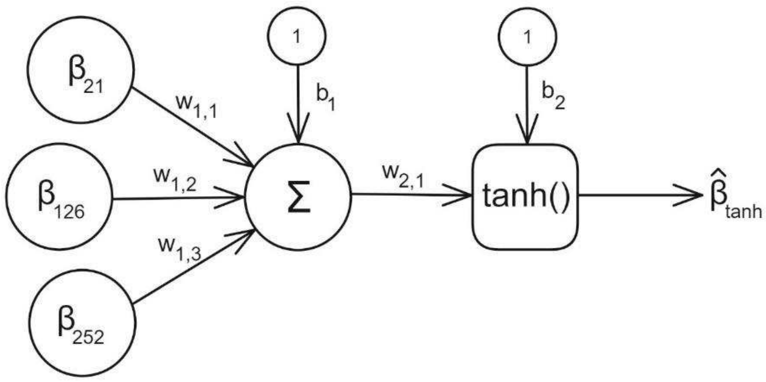
Jednostavni perceptroni se koriste za modeliranje složenih odnosa između ulaznih i izlaznih podataka, često pružajući bolje performanse u usporedbi s tradicionalnim metodama. Jedna od glavnih prednosti perceptrona je njihova sposobnost da koriste aktivacijske funkcije za rješavanje problema ograničenja koeficijenata. Ovo je posebno korisno kada radimo s financijskim podacima, gdje je važno da predikcije budu unutar razumnog raspona.

Funkcija tangens hiperbolni je idealna za rješavanje problema rada jer transformira ulazne vrijednosti u rasponu od $-\infty$ do ∞ u izlazne vrijednosti unutar raspona od -1 do 1. Grafički prikaz tanh funkcije (Slika 2.4) jasno pokazuje to svojstvo, što pomaže u održavanju stabilnosti modela.



Slika 2.4 Graf funkcije tangens hiperbolni

Jednostavni perceptron koji će se koristiti u radu nije mnogo drugačiji od linearne regresije, ali uključuje dodatni korak aktivacije koji omogućuje bolje upravljanje rasponom vrijednosti. Skica modela Slika 2.5 prikazuje kako se β koeficijenti izračunati iz klizećih prozora koriste kao ulazi, koji se zatim ponderiraju, sumiraju i transformiraju pomoću tanh funkcije. Na skici modela, β koeficijenti izračunati iz prozora od 21, 126 i 252 se množe sa svojim odgovarajućim težinama $w_{1,1}$, $w_{1,2}$ i $w_{1,3}$, te se zatim dodaje slobodni član b_1. Dobivena suma se skalira po $w_{2,1}$, nakon čega se dodaje drugi slobodni član b_2, te prolazi kroz tanh aktivacijsku funkciju kako bi se dobila konačna predikcija $\widehat{\beta}_{tanh}$.



Slika 2.5 Grafički prikaz modela jednostavnog perceptrona s aktivacijskom funkcijom tangens hiperbolni

Ovaj pristup omogućuje fleksibilnije i robusnije modeliranje financijskih podataka, osiguravajući da predikcije budu unutar očekivanog raspona i bolje odgovaraju stvarnim tržišnim uvjetima. Model prikazan u obliku jednadžbe bi bio (36).

$$\hat{\beta}_{tanh} = \tanh(w_{2,1} \cdot (w_{1,1} \cdot \beta_{21} + w_{1,2} \cdot \beta_{126} + w_{1,3} \cdot \beta_{252} + b_1) + b_2) \tag{36}$$

2.3.5. Procjena buduće standardne devijacije povrata

Malo drukčijim zapisom jednadžbe (9) se otkriva formula kovarijacijske matrice (37) preko dijagonalne matrice standardnih devijacija i korelacijske matrice. S obzirom na to da ovaj rad pokušava pronaći buduću matricu kovarijance, $\mathbf{V}^{1/2}$ također mora predstavljati buduće standardne devijacije.

$$\mathbf{\Sigma} = \mathbf{V}^{1/2} \mathbf{R} \mathbf{V}^{1/2} \tag{37}$$

Za buduću standardnu devijaciju povrata će se pretpostaviti da je jednaka uzoračkoj za prethodni mjesec. Ako povrati neke dionice imaju standardnu devijaciju jednaku nuli (što može ukazivati na nepromjenjivost cijene ili nedostatak trgovanja), njihova standardna devijacija se postavlja na vrlo malu vrijednost, konkretno 1×10^{-10} u ovom radu. To se čini da se izbjegne singularnost u kovarijacijskoj matrici, što bi moglo uzrokovati probleme pri inverziji matrice. Ova kovarijacijska matrica je pozitivno definitna jer smo pozitivnu definitnu matricu **R** pomnožili s pozitivnim vektorima standardnih devijacija. Zahvaljujući svim prijašnjim koracima, predikcija matrice kovarijance je sigurno invertibilna, te iskoristiva za računanje težina u optimalnim portfeljima.

2.4. Implementacija

Za implementaciju svih modela korišten je Python zbog njegove bogate zbirke biblioteka koje olakšavaju analizu podataka, razvoj modela i vizualizaciju rezultata. Implementacija je izvedena u Jupyter bilježnici, koja omogućuje interaktivnu obradu i lakše pokretanje dijelova koda.

Korištene knjižnice:

- **NumPy**: Osnovna knjižnica za rad s velikim nizovima i matricama te za obavljanje različitih matematičkih operacija.

- **Pandas**: Moćna biblioteka za manipulaciju i analizu podataka u obliku tablica (DataFrame objekti).
- **scikit-learn**: Popularna biblioteka za strojno učenje koja nudi jednostavne alate za izgradnju i evaluaciju prediktivnih modela.
- **SciPy**: Knjižnica koja pruža dodatne funkcionalnosti za znanstvene i tehničke računarske zadatke, uključujući napredne matematičke funkcije.
- **Keras**: API za visokorazinske neuronske mreže, koji omogućuje jednostavnu i brzu izgradnju i učenje perceptrona.
- **PyPortfolioOpt**: Biblioteka specijalizirana za optimizaciju portfelja koja koristi moderne tehnike za izradu optimalnih portfelja.

Koristeći ove alate, implementacija omogućuje učinkovitu obradu podataka, razvoj i evaluaciju modela te jasno prikazivanje rezultata analize.

3. Rezultati

3.1. Korišteni podaci

Za analizu su korišteni podaci o cijenama dionica 1161 tvrtke iz indeksa Russell 3000, koji obuhvaćaju razdoblje od 3. siječnja 2000. do 16. srpnja 2021. Russell 3000 je burzovni indeks koji prati performanse 3000 najvećih javnih kompanija sa sjedištem u SAD-u, što čini otprilike 98 % ukupnog tržišta dionica u SAD-u.

Podaci o cijenama dionica pretvoreni su u dnevne povrate. U analizu su uključeni samo radni dani, a dani kada su povrati za sve dionice bili 0 uklonjeni su iz skupa podataka.

3.1.1. Podjela na skupove za učenje i ispitivanje

Podjela podataka na skup za učenje i ispitivanje je približno u omjeru 3:1. Period učenja obuhvaća podatke od 23. siječnja 2001. do 2. siječnja 2015., dok ispitni period obuhvaća podatke od 2. siječnja 2015. do 2. srpnja 2021. godine.

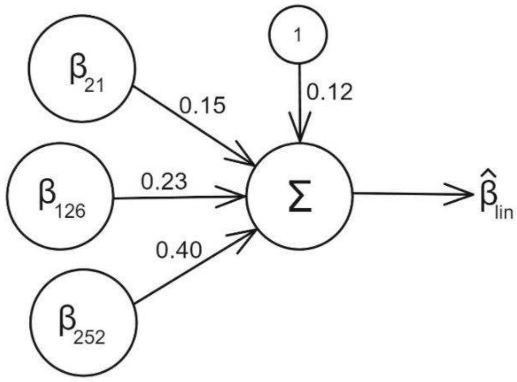
Ova podjela omogućuje da učimo modele na povijesnim podacima i potom procijenimo njihovu učinkovitost na nikad dosad viđenim podacima, što omogućuje procjenu sposobnosti modela za generalizaciju.

3.2. Prediktivni modeli

Nakon izračuna težina za linearnu regresiju i jednostavni perceptron na temelju podataka za učenje slijedi analiza kako svaki model koristi povijesne podatke o povratima dionica da bi predvidio buduće koeficijente faktorskog modela.

3.2.1. Linearna regresija

Nakon izračuna svih težina za linearnu regresiju na period za učenje, Slika 3.1 prikazuje izgled modela.



Slika 3.1 Grafički prikaz konačnog modela linearne regresije

Matematički, model za linearnu regresiju može se opisati pomoću (38). Ova jednadžba jasno pokazuje kako linearna regresija koristi povijesne bete, ponderirane njihovim pripadajućim težinama, da bi predvidjela buduće faktorske koeficijente.

$$\hat{\beta}_{lin}=0.15\cdot\beta_{21}+0.23\cdot\beta_{126}+0.40\cdot\beta_{252}+0.12 \tag{38}$$

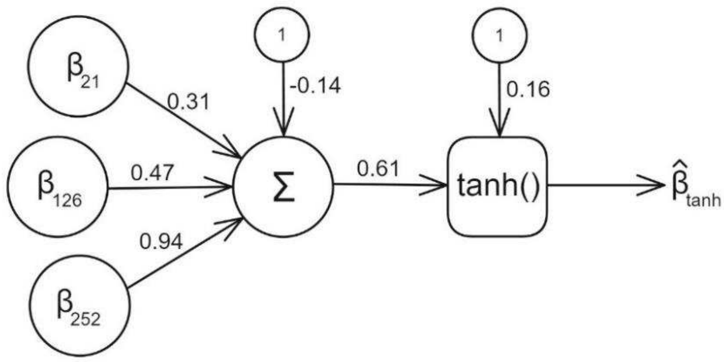
3.2.2. Jednostavni perceptron s aktivacijskom funkcijom

Izgled modela perceptrona je prikazan na Slika 3.2. Matematički, model za jednostavni perceptron može se opisati kao

$$\hat{\beta}_{tanh}=\tanh(0.61\cdot(0.31\cdot\beta_{21}+0.47\cdot\beta_{126}+0.94\cdot\beta_{252}-0.14)+0.16) \tag{39}$$

a nakon množenja, možemo izraziti funkciju jednostavnije s (40).

$$\hat{\beta}_{tanh}=\tanh(0.19\cdot\beta_{21}+0.29\cdot\beta_{126}+0.57\cdot\beta_{252}-0.10) \tag{40}$$



Slika 3.2 Grafički prikaz konačnog modela jednostavnog perceptrona s aktivacijskom funkcijom tangens hiperbolni

3.2.3. Usporedba težina

Kada se usporede težine dobivene iz linearne regresije i jednostavnog perceptrona, može se primijetiti nekoliko razmatranja:

- β_{252}: U oba modela, koeficijent izračunat na prozoru od 252 radna dana ima najveći utjecaj. U linearnom modelu, težina je 0.40, dok je u perceptronu efektivna težina 0.57.
- β_{126}: Ovaj koeficijent također ima značajan utjecaj, ali manji od koeficijenta na 252 dana. U linearnom modelu težina je 0.23, dok u perceptronu nakon kombiniranja težina iznosi 0.29.
- β_{21}: Najmanje je značajan u oba modela, s težinama 0.19 u linearnom modelu i 0.18 u jednostavnom perceptronu.

Iz ovih usporedbi se da zaključiti da oba modela daju veći naglasak na dugoročne povrate (252 radna dana), dok kratkoročni povrati (21 radni dan) imaju relativno manji utjecaj. Ovo može ukazivati na to da povijesni povrati preko duljih vremenskih razdoblja imaju veći prediktivni značaj za buduće bete u ovim modelima.

3.3. Mjere za evaluaciju predviđanja

3.3.1. Prosječna kvadratna pogreška (*mean square error*)

MSE mjeri prosječnu kvadratnu pogrešku između stvarnih i predviđenih vrijednosti. Računa se kao

$$\mathrm{MSE}=\frac{1}{n}\sum_{i=1}^{n}(\beta_i-\hat{\beta}_i)^2 \tag{41}$$

gdje je β_i stvarna vrijednost, $\hat{\beta}_i$ predviđena vrijednost, a n broj predviđanja. Niže vrijednosti *MSE* ukazuju na manju grešku modela.

3.3.2. Koeficijent determinacije (R^2)

R^2 mjeri proporciju varijance stvarnih vrijednosti koja je objašnjena predviđenim vrijednostima. Računa se sa

$$R^2=1-\frac{\sum_{i=1}^{n}(\beta_i-\hat{\beta}_i)^2}{\sum_{i=1}^{n}(\beta_i-\overline{\beta})^2} \tag{42}$$

gdje je $\overline{\beta}$ prosječna stvarna vrijednost. Vrijednost R^2 može biti između 0 i 1, ali je moguće da bude i negativna ako je model lošiji od jednostavnog modela s prosječnom vrijednošću. Koristi se kao mjera primjerenosti modela (*goodness-of-fit*) za linearne modele.

3.3.3. Srednja log-izglednost (*mean log-likelihood*)

Srednja log-izglednost procjenjuje vjerojatnost da se promatrani podaci pojave s obzirom na predviđene vrijednosti i pretpostavljenu distribuciju pogrešaka. Koristi se logaritamska izglednost naspram obične jer se računa s vrlo malim vrijednostima. Ovime se osiguravaju čitljiviji rezultati, i još bitnije, u Pythonu se neće izgubiti male vrijednosti zbog zaokruživanja na nulu.

U kontekstu modela rada, *LL* se računa kao prosjek logaritma izglednosti za sve buduće povrate. Ako imamo *n* skupova povrata $\mathbf{r}_1,...,\mathbf{r}_T$ s predviđenom kovarijacijskom matricom, log-izglednost za multivarijatnu normalnu distribuciju definira se kao

$$LL(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{r}_1,\ldots,\mathbf{r}_T)=-\frac{TN}{2}\ln(2\pi)-\frac{T}{2}\ln(\det(\boldsymbol{\Sigma}))-\frac{1}{2}\sum_{i=1}^{T}(\mathbf{r}_i-\mu)^T\boldsymbol{\Sigma}^{-1}(\mathbf{r}_i-\mu) \tag{43}$$

gdje je:

- T broj promatranih skupova povrata,
- N dimenzija vektora $\mathbf{r}_i$,
- $\boldsymbol{\mu}$ vektor srednjih vrijednosti,
- $\boldsymbol{\Sigma}$ kovarijacijska matrica,
- $\mathbf{r}_i$ promatrani vektor povrata.

U kontekstu rada koristi se korelacijska matrica **R** umjesto kovarijacijske, i standardizirani povrati $\mathbf{Z}_i$ prozora za koji predviđamo korelacije. Tim povratima je srednja vrijednost 0 , te se log-likelihood funkcija pojednostavljuje na (44).

$$LL(\mathbf{0}, \mathbf{R}|\mathbf{Z}_1,\ldots,\mathbf{Z}_T)=-\frac{TN}{2}\ln(2\pi)-\frac{T}{2}\ln(\det(\mathbf{R}))-\frac{1}{2}\sum_{i=1}^{T}\mathbf{Z}_i^T\mathbf{R}^{-1}\mathbf{Z}_i \tag{44}$$

Da bi se dobila srednja log-izglednost (45), još se uzima prosjek izračunatih vrijednosti po svim *n* predviđenim korelacijama za *n* skupova podataka.

$$LL=\frac{1}{n}\sum_{i=1}^{n} LL\left(\mathbf{0}, \mathbf{R_i} \middle| (\mathbf{Z}_1,\ldots,\mathbf{Z}_T)_i\right) \tag{45}$$

T je ujedno i duljina *Oracle* prozora (21 radni dan), a *N* broj dionica. *LL* indikator pomaže objasniti koliko su predviđajući modeli vjerodostojni u opisivanju budućih povrata, čime se može procijeniti kvaliteta modela gdje veća izglednost govori da predviđanja modela bolje odgovaraju stvarnim podacima.

3.3.4. Hipoteza o valjanosti modela

Linearna regresija i jednostavni perceptron iz rada, po svojoj vrsti minimiziraju *MSE*, pa će se isto tako provjeriti hipoteza da smanjivanje srednje kvadratne greške vodi do boljih ostalih mjera, kao što su R^2 i *LL*. Tako će se utvrditi koliko dobro modeli generaliziraju i predviđaju.

3.4. Evaluacija portfeljskih modela

Za svaki model se na temelju predviđenih kovarijacijskih matrica izgrađuje optimalni *GMV* (*global minimum variance*) i *CMV* (*constrained minimum variance*) portfelj. Kao mjera za procjenu korisnosti predikcija u izradi optimalnih portfelja će se koristiti volatilnost *GMV* i *CMV* portfelja. Procjenjivat će se standardna devijacija za svaki period između mijenjanja težina portfelja, te će se izračunati prosjek svih. Množenjem s $\sqrt{252}$ se dobije anualizirana standardna devijacija povrata portfelja, što je volatilnost portfelja u kontekstu rada.

Uz sve dosad viđene modele izračunate opisanim tehnikama, za mjeru najoptimalnijeg portfelja koristit će se *Oracle+*. Ovaj model ne samo da zna buduće koeficijente faktorskog modela, nego mu je pri izračunu matrice kovarijance poznata i buduća varijanca. Za svako mjerilo u nastavku, rezultat najboljeg realističnog modela će biti podebljan.

3.5. Rezultati na skupu za učenje (*in-sample*)

Ovi rezultati govore kako se modeli ponašaju na podacima na kojima su zapravo naučeni. Tablica 3.1 daje mjere za predviđanja svakog modela, gdje je očekivano *Oracle* najbolji po *MSE*-u i R^2-u, ali ima i najbolji *LL*. Predviđanja linearnom regresijom i perceptronom se po svim metrikama čine dosta bolji od jednostavnih uzoračkih procjenitelja. Model koji za procjenu uzima samo koeficijente procijenjene na prozoru od 21 dan je daleko najlošiji, a uz to jedini ima negativan R^2.

Tablica 3.1 Mjere za evaluaciju predviđanja na skupu za učenje

Model	**MSE**	**R²**	**LL**
$\hat{\beta}_{\text{Oracle}}$	0.0	1.0	-1360.144
$\hat{\beta}_{21}$	0.0759	-0.0659	-1481.441
$\hat{\beta}_{126}$	0.0523	0.2656	-1437.497
$\hat{\beta}_{252}$	0.0513	0.2800	-1436.480
$\hat{\beta}_{\text{lin}}$	**0.0468**	**0.3422**	**-1428.644**
$\hat{\beta}_{\text{tanh}}$	0.0469	0.3415	-1428.897

Tablica 3.2 prikazuje volatilnosti za *CMV* i *GMV* portfelje, čije su težine dobivene procijenjenim kovarijacijskim matricama različitih modela. Dominantnosti portfelja baziranih na *Oracle+* modelu ukazuju na važnost dobre procjene budućih volatilnosti, a odmah iza njega je *Oracle*. Od realističnih portfelja, najmanji rizik su pokazali ograničeni portfelji bazirani na uzoračkim koeficijentima od 6 i 12 mjeseci. Najlošiji portfelj je izveden iz uzoračkih procjena na prozoru od 1 mjesec. Za usporedbu, *equal weight* portfelj je u tom periodu imao standardnu devijaciju 18.93 %.

Tablica 3.2 Volatilnosti GMV i CMV portfelja na skupu za učenje

Model	$\boldsymbol{\sigma_{GMV}}(\%)$	$\boldsymbol{\sigma_{CMV}}(\%)$
$\hat{\Sigma}_{\text{Oracle}}$	10.60	6.53
$\hat{\Sigma}_{\text{Oracle+}}$	04.02	3.75
$\hat{\Sigma}_{21}$	11.44	12.22
$\hat{\Sigma}_{126}$	10.71	**9.91**
$\hat{\Sigma}_{252}$	10.71	9.98

$\hat{\Sigma}_{\mathrm{lin}}$	**10.62**	11.84
$\hat{\Sigma}_{\mathrm{tanh}}$	10.63	12.21

3.6. Rezultati na skupu za ispitivanje (*out-of-sample*)

Sudeći po Tablica 3.3, na nikad viđenim podacima, rang-lista modela se nije znatno promijenila. No, može se uočiti da modeli po svim metrikama lošije procjenjuju na neviđenim podacima, što je za očekivati. Usprkos tome, kao na skupu za učenje, jasna je poveznica između *MSE* i ostalih mjera. Iz dobivenih rezultata, da se zaključiti da, u kontekstu rada, modeli s manjom kvadratnom greškom imaju veće R^2 i *LL*, i obrnuto.

Tablica 3.3 Mjere za evaluaciju predviđanja na skupu za ispitivanje

Model	**MSE**	**R²**	**LL**
$\hat{\beta}_{\mathrm{Oracle}}$	0.0	1.0	-1359.623
$\hat{\beta}_{21}$	0.0858	-0.2002	-1497.707
$\hat{\beta}_{126}$	0.0590	0.1750	-1444.958
$\hat{\beta}_{252}$	0.0577	0.1930	-1443.861
$\hat{\beta}_{\mathrm{lin}}$	**0.0533**	**0.2546**	**-1436.933**
$\hat{\beta}_{\mathrm{tanh}}$	0.0540	0.2445	-1438.139

Unatoč boljim procjenama složenijih modela, priloženi rezultati iz Tablica 3.4 ne podupiru hipotezu da to vodi ka boljim portfeljima. Iako su im greške manje od jednostavnih uzoračkih procjenitelja, rezultati ukazuju da je recentnost koeficijenata procjenjenih na manjim prozorima bitnija karakteristika za izgradnju portfelja minimalne varijance od smanjivanja kvadratne greške u predviđanju. Za isti period je portfelj s jednakim težinama imao standardnu devijaciju 16.81 %. Također, suprotno očekivanjima *CMV* je pokazao manju volatilnost od *GMV*, no nekad može biti isplativije ograničiti težine portfelja [6].

Tablica 3.4 Volatilnosti GMV i CMV portfelja na skupu za ispitivanje

Model	$\boldsymbol{\sigma_{GMV}}(\%)$	$\boldsymbol{\sigma_{CMV}}(\%)$
$\hat{\Sigma}_{\text{Oracle}}$	9.48	8.95
$\hat{\Sigma}_{\text{Oracle+}}$	7.05	6.45
$\hat{\Sigma}_{21}$	**11.65**	11.45
$\hat{\Sigma}_{126}$	12.00	**10.97**
$\hat{\Sigma}_{252}$	12.11	11.40
$\hat{\Sigma}_{\text{lin}}$	12.27	11.29
$\hat{\Sigma}_{\text{tanh}}$	12.25	11.29

Zaključak

Za multivarijatne podatke, korištenje uzoračke kovarijacijske matrice, procijenjene standardnim statističkim metodama, daje neiskoristive rezultate za izgradnju optimalnih portfelja. U ovom radu je prikazano moguće rješenje u obliku procjene korelacijske matrice iz koeficijenata jednofaktorskog modela, koristeći standardizirane povrate portfelja jednakih težina kao jedini faktor. Pokušajima predikcije budućih koeficijenata, pokazalo se da kombinacija uzoračkih koeficijenata raznim metodama, po svim mjerilima, daje bolja predviđanja od samih uzoračkih. Isto se ne može reći za performanse portfelja koje nisu iskazale znatne rezultate za kompleksnije procjenitelje, štoviše rezultati su im lošiji od običnih uzoračkih. Moglo bi se raspravljati da uzorački koeficijenti na manjim prozorima daju recentnije i relevantnije podatke, što je za izgradnju optimalnih portfelja potencijalno bitnije od minimiziranja kvadratne greške u predviđanju.

Moguća poboljšanja problema rada se mogu podijeliti na 3 glavne grane. Prva mogućnost je korištenje više faktora, i/ili da drukčiji faktor predstavlja stanje tržišta. U teoriji bi takav faktorski model mogao bolje objasniti povrate, time i koeficijente/korelaciju. Drugi pristup optimizaciji je korištenje naprednijih metoda za predviđanje budućih koeficijenata iz prijašnjih. Tu također spada uporaba više ulaznih parametara, i uporaba parametara koji nisu samo uzorački koeficijenti. I posljednji način poboljšanja je složenijim predviđanjem buduće varijance, što se očitovalo iz razlike u volatilnosti između portfelja konstruiranih *Oracle* i *Oracle+* modelima.

Literatura

[1] Lam, C. *High-dimensional covariance matrix estimation*, WIREs Computational Statistics, 12,2 (2020), e1485.

[2] Johnson, R. A., Wichern, D. W. *Applied multivariate statistical analysis*. 6. izdanje. London: Pearson, 2014.

[3] Steiger, J. H., *Matrix Algebra of Sample Statistics*, Vanderbilt University, (2010). Poveznica: https://www.statpower.net/Content/312/Lecture%20Slides/MatrixStat.pdf; pristupljeno 14. ožujka 2024.

[4] Clarke, R., De Silva, H., Thorley, S. *Minimum-Variance Portfolio Composition, The Journal of Portfolio Management*, 37,2 (2011), str. 31-45.

[5] Fan, J., Fan, Y., Lv, J. *High dimensional covariance matrix estimation using a factor model*, Journal of Econometrics, 147,1 (2008), str. 186-197.

[6] Jagannathan, R., Ma, T. *Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps*, The Journal of Finance, 58,4 (2003), str. 1651-1683.

Sažetak

Modeliranje korelacija financijskih vremenskih nizova zasnovano na latentnim faktorima

Ovaj završni rad istražuje problem procjene i predviđanja korelacijskih koeficijenata parova vremenskih nizova povrata financijskih imovina, s naglaskom na primjenu latentnih faktorskih modela. Ključna komponenta rada je razvoj algoritama za procjenu koeficijenata faktorskog modela iz povijesnih podataka, te primjena metoda strojnog učenja za predviđanje budućih vrijednosti tih koeficijenata. Poseban naglasak stavljen je na integraciju predviđenih matrica korelacije u matrice kovarijance financijskih vremenskih nizova, te na evaluaciju performansi različitih modela u kontekstu konstrukcije portfelja minimalne varijance. Rezultati pokazuju da kombiniranje koeficijenata faktorskog modela smanjuje predikcijske greške, ali ne poboljšava performanse portfelja minimalne varijance.

Ključne riječi: financijski vremenski nizovi, korelacija, latentni faktori, faktorski modeli, portfelj minimalne varijance, strojno učenje, Pearsonov koeficijent, klizeći prozori, koeficijenti faktorskog modela, linearna regresija

Summary

Modelling financial time series correlation based on latent factors

This thesis investigates the problem of estimating and predicting the correlation coefficients of pairs of time series returns of financial assets, with a focus on the application of latent factor models. A key component of the work is the development of algorithms for estimating factor model coefficients from historical data, and the application of machine learning methods to predict future values of these coefficients. Special emphasis is placed on integrating the predicted correlation matrices into covariance matrices of financial time series, and on evaluating the performance of different models in the context of constructing minimum variance portfolios. The results show that combining factor model coefficients reduces prediction errors, but does not improve the performance of minimum variance portfolios.

Key words: financial time series, correlation, latent factors, factor models, minimum variance portfolio, machine learning, Pearson's coefficient, rolling windows, factor model coefficients, linear regression