

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1099

**Analiza vremenskih nizova
zasnovana na kompleksnim
mrežama**

Lovre Mrčela

Zagreb, srpanj 2017.

Zahvaljujem svojoj obitelji, svim dragim prijateljima i profesorima na Fakultetu elektrotehnike i računarstva, na višegodišnjoj potpori i pružanju novih znanja i vještina bez kojih ovaj rad ne bi mogao nastati.

SADRŽAJ

1. Uvod	1
2. Osnovni financijski pojmovi	2
2.1. Pretpostavke osnovnog modela trgovanja	4
2.2. Teorem o nearbitraži	4
3. Metode i koncepti	7
3.1. Statistička arbitraža	7
3.1.1. Razlika normaliziranih log-cijena	8
3.1.2. Normalizirana razlika log-cijena	11
3.1.3. Mjera kointegracije slučajnih procesa	12
3.1.4. Nedostaci statističke arbitraže	14
3.2. Relacija preferencije i funkcija korisnosti	16
3.3. Graf toka preferencija	17
3.4. Metoda potencijala	20
3.4.1. Računanje potencijala čvorova	20
3.4.2. Konzistentna rekonstrukcija grafa	22
3.4.3. Primjer korištenja metode potencijala	22
3.5. Koeficijent obrtaja	24
4. Praktično ostvarenje algoritma trgovanja	26
4.1. Konstrukcija grafa toka preferencija	27
4.2. Signal trgovanja	29
4.2.1. Optimiranje praga trgovanja	33
4.2.2. Optimiranje veličine vremenskog prozora	33
4.3. Konstrukcija portfelja	34
4.4. Testiranje	36

5. Rezultati	40
5.1. S&P 203	40
5.2. S&P 500	43
5.3. CROBEX	45
6. Zaključak	47
Literatura	48
A. Računanje momenata vremenskog niza na pomičnom prozoru	50

1. Uvod

Klasične metode statističke arbitraže uzimaju u obzir parove vrijednosnica čije cijene se ponašaju slično tijekom određenog razdoblja. Sličnost se mjeri kointegracijom, korelacijom, ili nekom drugom mjerom, s ciljem pronalaska trenutka kada te cijene izlaze izvan statistički utvrđeno visoko pouzdanog intervala. Takve prilike mogu se iskoristiti predviđanjem da će se cijene u idućem trenutku ponovno vratiti unutar intervala visoke razine pouzdanosti te se u skladu s tim predviđanjem može provesti trgovanje.

U ovom radu predložena je nova metoda temeljena na predviđanjima koja su dobivena metodom statističke arbitraže, pri čemu se statističke mjere koriste za opisivanje relacije preferencije među parovima vrijednosnica. Razlikuju se dvije vrste preferencije: preferencija jedne vrijednosnice nad drugom i individualna mjera preferencije za svaku vrijednosnicu. Iz dobivene relacije preferencije konstruira se graf toka preferencija iz kojeg se može proučiti međusobna interakcija svih vrijednosnica koje su na raspolaganju. Konačno, za svaku vrijednosnicu izračunava se individualna mjera preferencije, na temelju čega se donosi odluka o sastavljanju portfelja.

Ideja ove metode je ostvariti generalizaciju klasičnih metoda statističke arbitraže koja će biti robusnija i ostvarivati bolje rezultate kada na raspolaganju ima velik broj vrijednosnica tako što će izvlačiti zajedničku interakciju iz svih vrijednosnica.

Organizacija poglavlja je sljedeća: u poglavlju 2 dan je uvod u osnovne pojmove iz svijeta financija koji su bitni za razumijevanje teme rada. U poglavlju 3 dana je teorijska podloga temeljena na matematici koja je također bitna za razumijevanje rada, a opisani su i novi koncepti koji su razvijeni u radu. Praktično ostvarenje zadatka i zapažanja opisani su u poglavlju 4. Sustavni prikaz rezultata dan je u poglavlju 5. Naposljetku, sam zaključak rada dan je u poglavlju 6.

2. Osnovni financijski pojmovi

Ovdje su navedeni osnovni financijski pojmovi koji su korišteni u radu, te su dane njihove definicije.

Vrijednosnica. Vrijednosnica, ili vrijednosni papir (engl. *security*) u širem smislu je oblik imovine kojim se može trgovati. Postupci u ovom radu prvenstveno su namijenjeni trgovanju dionicama, iako bi se trgovanje moglo generalizirati i na druge vrijednosnice, kao što su obveznice, opcije, ili *futuresi*.

Portfelj. Pod pojmom portfelja misli se na količine vrijednosnica koje investitor posjeduje, ili je dužan kupiti ako se radi o negativnoj količini. Formalno, ako je na tržištu dostupno N različitih vrijednosnica, portfelj se može predstaviti vektorom $\alpha = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_N]$, gdje je α_i količina i -te vrijednosnice.

Likvidnost. Tržište je likvidno za neku vrijednosnicu ako je na njemu moguće trgovati gotovo svakom količinom te vrijednosnice u bilo kojem trenutku; odnosno, na njemu uvijek postoji dovoljna ponuda i potražnja za određenim vrijednosnicama. Likvidnost je neizbježna pretpostavka svih algoritama optimizacije portfelja kod kojih se odvija visokofrekventno trgovanje, kao što je slučaj kod algoritma statističke arbitraže.

Divizibilnost. Divizibilnost je pretpostavka da količine vrijednosnica prisutnih u portfelju nisu nužno cijeli brojevi. U matematičkim modelima trgovanja često se pretpostavlja divizibilnost radi fleksibilnosti u izračunima, odnosno dopušta se $\alpha_i \in \mathbb{R}$ ili $\alpha_i \in \mathbb{Q}$. Na primjer, novac je divizibilan financijski instrument, dok dionice nisu; međutim, u praktične svrhe smatra se da je vrijednosnica divizibilna ako je njezina raspoloživa količina na tržištu puno veća od njezine cijene.

Pozicija. Pozicija označava stanje određene vrijednosnice u portfelju. Pozicija može biti duga (engl. *long*), kratka (engl. *short*) i neutralna (engl. *neutral*).

U dugoj poziciji, količina i -te vrijednosnice u portfelju α_i je pozitivna, odnosno investitor posjeduje određenu količinu kupljene vrijednosnice. Suprotno tome, u kratkoj poziciji, količina vrijednosnice α_i u portfelju je negativna, što znači da je investitor prethodno prodao određenu količinu vrijednosnice koju nije posjedovao, i time se obavezao na kasniju kupnju iste količine. Na nekim tržištima nije dopuštena kratka pozicija u vrijednosnicama. Neutralna pozicija $\alpha_i = 0$ označava da se u portfelju ne nalazi ta vrijednosnica. Kada se govori o zatvaranju pozicije, radi se o prelasku u neutralnu poziciju.

Solventnost. Solventnost označava da je vrijednost portfelja $V(t) \geq 0$ u svakom trenutku t . Portfelj koji zadovoljava svojstvo solventnosti je ostvariv portfelj; u suprotnom je neostvariv portfelj.

Indeks. Indeks je ukupna mjera rasta nekog tržišta. Računa se kao težinski zbroj trenutnih vrijednosti odabranih vrijednosnica, a težine pritom mogu biti određene na nekoliko načina, a najčešće uniformno — sve težine su jednake; prema cijeni — svaka vrijednosnica doprinosi indeksu proporcionalno svojoj trenutnoj vrijednosti; ili prema kapitalu — uz trenutnu vrijednost se uzima u obzir i količina vrijednosnica koja je u opticaju. Primjer značajnijeg indeksa na hrvatskom tržištu dionica je CROBEX, a na američkom tržištu dionica je Standard & Poor's 500 (S&P 500).

Logaritamska cijena. Logaritamska cijena dobiva se jednostavnim logaritmiranjem obične cijene. Koristi se pri računanju logaritamskog povrata zbog svojstva prebacivanja kvocijenta u razliku: $\log\left(\frac{y}{x}\right) = \log(y) - \log(x)$.

Povrat. Povrat $r(t_1, t_2)$ je udio promjene cijene između dva vremenska trenutka t_1 i t_2 :

$$r(t_1, t_2) = \frac{V(t_2) - V(t_1)}{V(t_1)}.$$

No, u praksi se češće koristi logaritamski povrat $R(t_1, t_2)$ radi lakšeg računanja, definiran kao razlika logaritamskih cijena:

$$R(t_1, t_2) = \log(V(t_2)) - \log(V(t_1)).$$

Logaritamski povrat odgovara godišnjoj kontinuiranoj kamatnoj stopi uz koju bi se vrijednost iz $V(t_1)$ promijenila u $V(t_2)$ nakon godine dana:

$$V(t_1) \cdot \exp(R(t_1, t_2)) = V(t_2).$$

Vrijedi da je $r(t_1, t_2) \approx R(t_1, t_2)$ kada je $V(t_1) \approx V(t_2)$, što je čest slučaj kada se promatraju dnevne cijene:

$$r(t_1, t_2) = \frac{V(t_2)}{V(t_1)} - 1 \approx \log\left(1 + \frac{V(t_2)}{V(t_1)} - 1\right) = \log\left(\frac{V(t_2)}{V(t_1)}\right) = R(t_1, t_2),$$

uz korištenje procjene $x \approx \log(1 + x)$ kada je $x \approx 0$.

2.1. Pretpostavke osnovnog modela trgovanja

S ciljem pojednostavljivanja i boljeg opisivanja trgovanja u stvarnom svijetu, osmišljeni su različiti modeli trgovanja. Model trgovanja je matematička idealizacija koja obuhvaća vrstu imovine, kretanje cijena, pravila trgovanja i ostalo. Kod osnovnog modela trgovanja pretpostavlja se sljedeće (prema [1]):

1. Portfelj se sastoji od rizične i nerizične imovine: cijena rizične imovine u svakom trenutku je diskretna slučajna varijabla s barem dvije moguće različite vrijednosti, a cijena nerizične imovine je iznos koji je poznat u svakom trenutku.
2. Vrijednosti rizične i nerizične imovine su uvijek pozitivni realni brojevi.
3. Trgovanje je likvidno, divizibilno, i dopuštena je kratka pozicija. Ovom pretpostavkom dopušta se da udjeli imovine u portfelju budu elementi iz skupa realnih brojeva.
4. Portfelj mora biti solventan, odnosno njegova vrijednost mora biti u svakom trenutku pozitivna; u suprotnom bi to značilo da je investitor bankrotirao.
5. Cijene rizične imovine mogu u svakom trenutku poprimiti samo konačno mnogo različitih vrijednosti.

2.2. Teorem o nearbitraži

Uz navedene pretpostavke osnovnog modela trgovanja, jedan od osnovnih teorema u trgovanju je teorem o nearbitraži koji tvrdi da nije moguće ostvariti pozitivan profit bez preuzimanja rizika. Drugim riječima, ako je vrijednost portfelja u trenutku 0, $V(0) = 0$, onda je vjerojatnost $\mathbf{P}(V(t) > 0) = 0$ u svakom trenutku $t > 0$.

Teorem o nearbitraži također se može iskazati na malo složeniji način (prema [2]), koji omogućuje bolje shvaćanje kada je arbitraža moguća, i kako se situacije u kojima nearbitražni uvjeti ne vrijede mogu iskoristiti. Neka je na tržištu raspoloživo N vrijednosnica označenih brojevima od 1 do N , te je za svaku od njih moguće M različitih ishoda (kretanja cijene) u sljedećem razdoblju. Neka je s $r_i[j]$ označen iznos koji se dobije od jediničnog ulaganja u i -tu vrijednosnicu ako je ostvaren j -ti ishod. Portfelj $\alpha = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_N]$ tako u slučaju ostvarivanja j -tog ishoda donosi ukupni dobitak $R[j] = \sum_{i=1}^N \alpha_i r_i[j]$.

Neka slučajna varijabla X opisuje razdiobu mogućih M ishoda:

$$X \sim \begin{pmatrix} 1 & 2 & \cdots & M \\ p_1 & p_2 & \cdots & p_M \end{pmatrix}.$$

U nearbitražnim uvjetima tada mora vrijediti da je očekivanje dobitka od i -te vrijednosnice:

$$\mathbf{E}[r_i[X]] = \sum_{j=1}^M p_j r_i[j] = 0, \quad (2.1)$$

za sve $i \in [1, 2, \dots, N]$. No, ponekad se za proizvoljno postavljene dobitke $r_i[j]$ ne može odrediti vjerojatnosni vektor $\mathbf{p} = [p_1 \ p_2 \ \cdots \ p_M]$ koji bi ispunjavao (2.1). Teorem o nearbitraži tvrdi da je moguće samo jedno od sljedećeg:

- ili postoji vjerojatnosni vektor \mathbf{p} koji ispunjava (2.1),
- ili je moguće konstruirati portfelj α tako da uvijek vrijedi:

$$R[j] = \sum_{i=1}^N \alpha_i r_i[j] > 0, \quad (2.2)$$

za sve $j \in [1, 2, \dots, M]$.

Teorem o nearbitraži iskazan na ovaj način proizlazi iz rješavanja problema linearnog programiranja (linearne optimizacije). Problemi linearnog programiranja su optimizacijski problemi kod kojih je potrebno maksimizirati linearnu funkciju cilja uz zadana ograničenja nejednakosti na argument funkcije. Osnovna formulacija optimizacijskog problema jest primarni problem, koji se može preinačiti u dualni problem u kojem je cilj minimizirati dualnu funkciju cilja, uz dualna ograničenja jednakosti na argument dualne funkcije. Za primarnu i dualnu formulaciju problema vrijedi načelo dualnosti koje tvrdi da, ako je bilo koji od ta dva problema rješiv, onda je maksimum primarnog jednak minimumu dualnog problema; no ako jedan od dvaju problema nije rješiv, onda za onaj drugi ne postoji jedinstveno rješenje.

U teoremu o nearbitraži primarni problem predstavlja situaciju u kojoj je moguća arbitraža. Potrebno je maksimizirati ukupne dobitke $R[j]$, za sve $j \in [1, 2, \dots, M]$. Neka je donja granica tih dobitaka označena s ρ . Ograničenja nejednakosti u primarnom problemu su $\sum_{i=1}^N \alpha_i r_i[j] \geq 0$, koja se razlikuju od (2.2) u tome što dopuštaju jednakost. U njegovom pripadnom dualnom problemu, ciljna funkcija koju treba minimizirati je konstantna i iznosi 0. Ograničenja jednakosti u dualnom problemu odgovaraju izrazu (2.1). Posljedica ovog odnosa dualnosti između izraza (2.1) i (2.2) jest sljedeća: ako postoji vjerojatnosni vektor \mathbf{p} koji ispunjava (2.1), tada je dualni problem rješiv i njegov minimum koji iznosi 0 je ispunjen, a prema načelu dualnosti to je ujedno i maksimum primarnog problema, $\rho = 0$. No, ograničenja nejednakosti primarnog problema mogu biti zadovoljena samo ako su $\alpha_1 = \alpha_2 = \dots = \alpha_N = 0$. Takav portfelj ne donosi nikakav profit, što potvrđuje da je ispunjeno svojstvo nearbitraže. S druge strane, nepostojanje vjerojatnosnog vektora \mathbf{p} koji ispunjava (2.1) znači da 0 nije ostvariv minimum dualnog problema. Tada rješenje $\rho > 0$ primarnog problema nije jedinstveno i moguće je sastaviti portfelj $\boldsymbol{\alpha}$ koji će zadovoljiti (2.2).

3. Metode i koncepti

Ovdje su opisane matematičke metode i koncepti korišteni u radu, kao i definicije korištenih osnovnih pojmova.

3.1. Statistička arbitraža

Statistička arbitraža (engl. *statistical arbitrage*) je algoritam trgovanja zasnovan na statistici. Pojam arbitraža u kontekstu financija označava okolnosti pod kojima se može konstruirati portfelj koji ostvaruje pozitivan profit s vjerojatnošću 1, a da pritom njegova početna vrijednost iznosi 0:

$$\begin{aligned} V(0) &= 0 \\ \mathbf{P}(V(t) > 0) &= 1, \end{aligned} \tag{3.1}$$

gdje $V(t)$ označava vrijednost portfelja u trenutku t .

Trivijalan primjer arbitraže je situacija u kojoj se ista vrijednosnica nudi po različitoj cijeni na dva različita tržišta. U tom slučaju arbitražu je moguće ostvariti kupovanjem vrijednosnice po nižoj cijeni na jednom tržištu i prodavanjem po višoj na drugom.

Cilj statističke arbitraže je eksploatirati statističke zakonitosti koje vrijede među parovima vrijednosnica kako bi se ostvario profit. Valja napomenuti da, iako nosi naziv arbitraža, statistička arbitraža zapravo ne zadovoljava nužno (3.1) jer korištene statističke mjere imaju interval pouzdanosti manji od 100%. Tako se može dogoditi da bude $V(t) < 0$ za neke t . Ipak, uz pretpostavku dovoljne likvidnosti tržišta, kako $t \rightarrow +\infty$, tako vjerojatnost $\mathbf{P}(V(t) > 0) \rightarrow 1$. U praksi, potrebno je osigurati dovoljno velik početni iznos $V(0)$ kako bi se izbjegla mogućnost *defaulta*.

Algoritam statističke arbitraže u svakom vremenskom trenutku najčešće čine sljedeći koraci:

1. određivanje neke mjere sličnosti među parovima vrijednosnica nad proteklom periodom zadane duljine,
2. uzimanje u obzir samo onih parova kod kojih mjera sličnosti ne prelazi zadani prag, odnosno odbacivanje preostalih parova;
3. za parove čije je ponašanje međusobno dovoljno slično utvrđuje se nalaze li se u sadašnjem trenutku njihove cijene izvan očekivanog intervala,
4. u skladu s utvrđenim provodi se trgovanje dvjema vrijednosnicama u parovima tako da se u jednoj od njih zauzme kratka, a u drugoj duga pozicija, sve dok se njihove cijene ne vrate unutar očekivanog intervala.

Najčešće korištene mjere sličnosti su: razlika normaliziranih log-cijena, normalizirana razlika log-cijena, i kointegracija slučajnih procesa. U nastavku su ukratko opisane navedene mjere.

3.1.1. Razlika normaliziranih log-cijena

Neka je $P_X^{(t)}$ označena log-cijena vrijednosnice X u trenutku t , i neka se koristi period od T vremenskih koraka. Neka je ukupno M vremenskih koraka, tj. $t \in [0, 1, \dots, M-1]$. Za svaku vrijednosnicu X u trenucima $t \geq T$ izračunaju se normalizirane log-cijene $\bar{P}_X^{(t)}$:

$$\bar{P}_X^{(t)} = \frac{P_X^{(t)} - \mathbf{E} [P_X^{(\tau)} \mid t - T \leq \tau < t]}{\sqrt{\mathbf{Var} [P_X^{(\tau)} \mid t - T \leq \tau < t]}}. \quad (3.2)$$

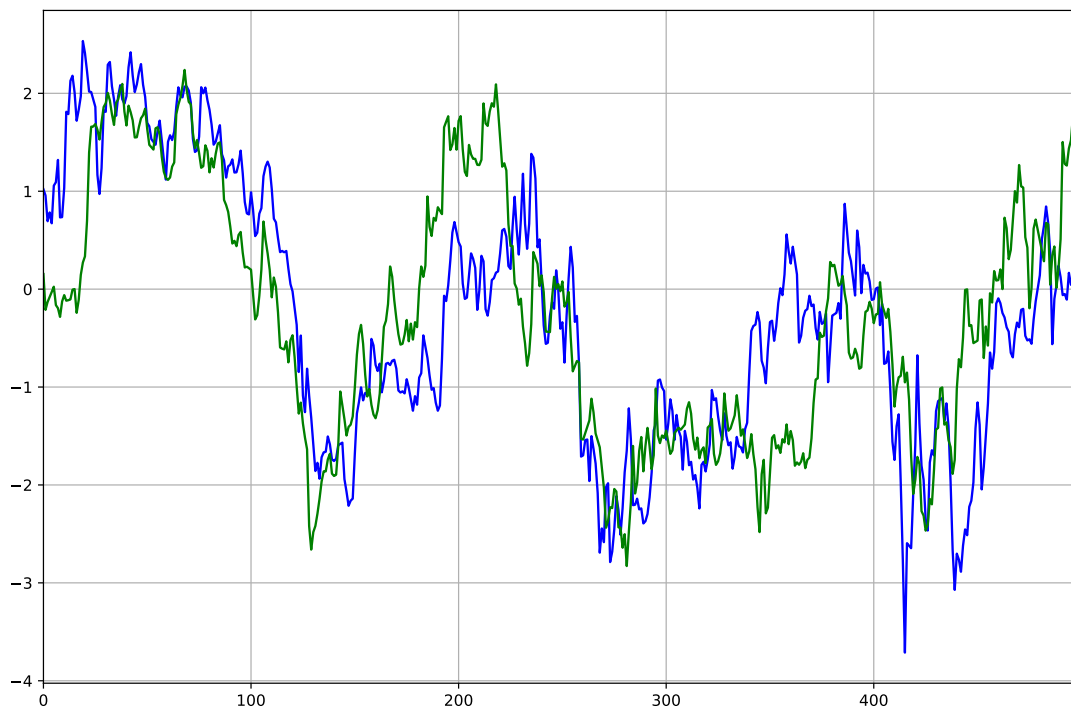
Oznake $\mathbf{E} [P_X^{(\tau)} \mid t_1 \leq \tau < t_2]$ i $\mathbf{Var} [P_X^{(\tau)} \mid t_1 \leq \tau < t_2]$ predstavljaju očekivanje i varijancu cijene vrijednosnice X nad intervalom $[t_1, t_2]$. Nadalje, za svaki par vrijednosnica X i Y izračuna se razlika njihovih normaliziranih log-cijena $D_{X,Y}^{(t)}$ u trenutku t :

$$D_{X,Y}^{(t)} = \bar{P}_X^{(t)} - \bar{P}_Y^{(t)} \quad (3.3)$$

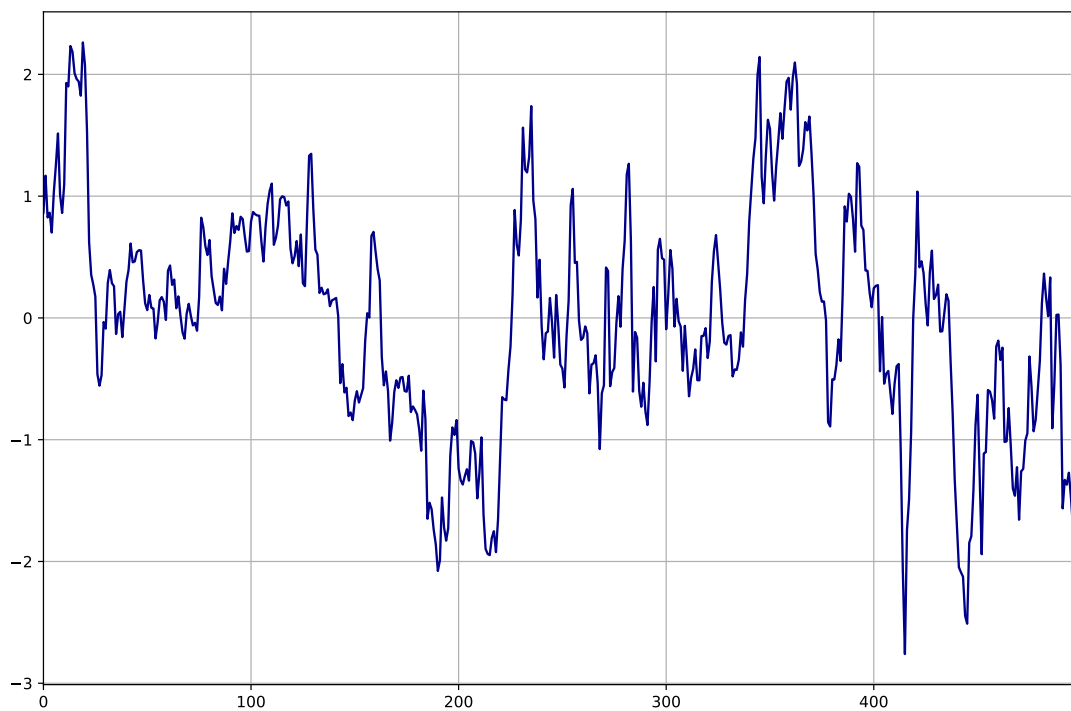
Razlika normaliziranih log-cijena koristi se kao mjera sličnosti u algoritmu statističke arbitraže. Kod ove mjere normalizacijom se postiže ravnopravna usporedba dviju vrijednosnica čije su volatiliteti različiti, što znači da će se prilikom usporedbe dviju dionica, čiji su intenziteti promjene cijene različiti, nakon normalizacije oni dovesti na istu razinu. Ilustrativni primjer kretanja cijena vrijednosnica prikazan je na slici 3.1, normalizirane cijene na slici 3.2, i razlika normaliziranih cijena na slici 3.3.



Slika 3.1: Prikaz kretanja cijena dviju vrijednosnica, X i Y , u razdoblju od 500 dana. Plavom bojom prikazana je cijena $P_X^{(t)}$, a zelenom cijena $P_Y^{(t)}$. Svjetlijim nijansama tih boja prikazano je punom linijom očekivanje, i crtkanom linijom udaljenost za jednu devijaciju od očekivanja; očekivanje i devijacija izračunati su u svakom trenutku na proteklom periodu T duljine 120 dana.



Slika 3.2: Plavo: normalizirana cijena $\bar{P}_X^{(t)}$, zeleno: normalizirana cijena $\bar{P}_Y^{(t)}$.



Slika 3.3: Razlika normaliziranih cijena, $D_{X,Y}^{(t)}$.

3.1.2. Normalizirana razlika log-cijena

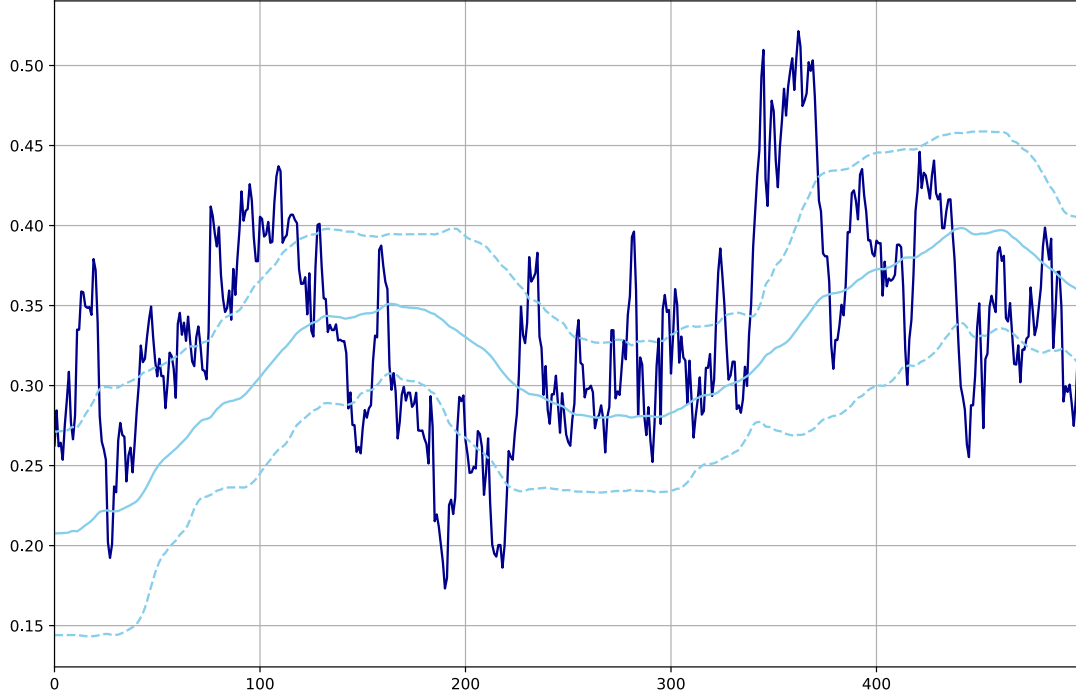
Za razliku od prethodne mjere sličnosti, u ovom slučaju razlika se računa nad nenormaliziranim log-cijenama za sve t , označeno s $d_{X,Y}^{(t)}$:

$$d_{X,Y}^{(t)} = P_X^{(t)} - P_Y^{(t)}, \quad (3.4)$$

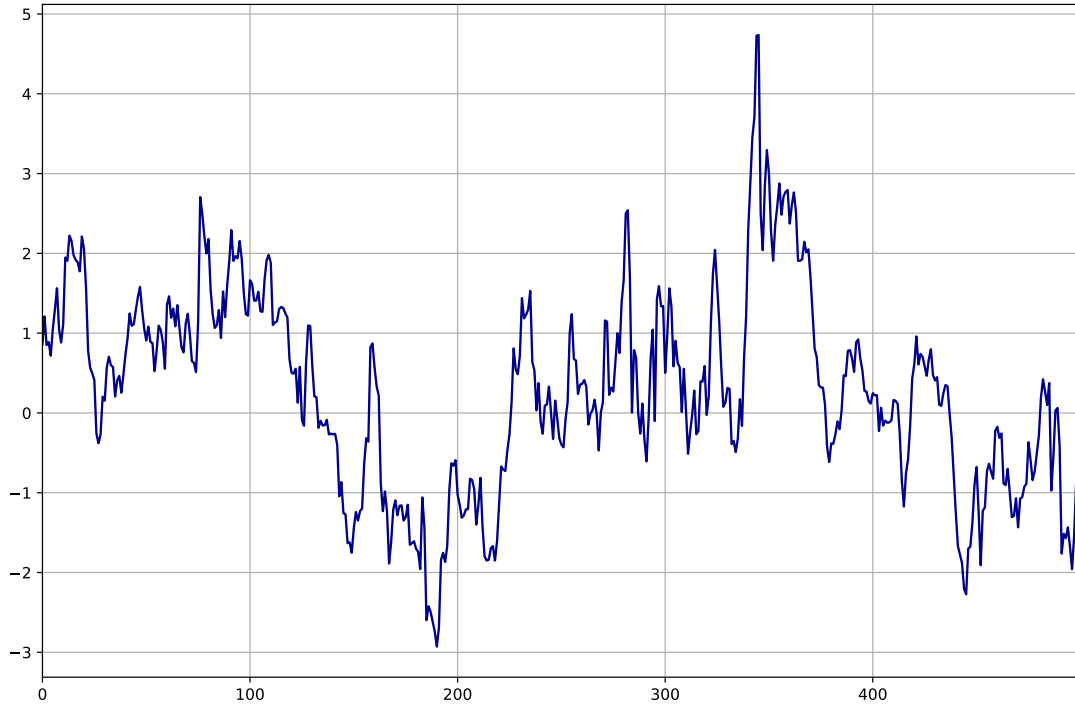
a zatim se iz razlika $d_{X,Y}^{(t)}$ računaju normalizirane razlike $\bar{d}_{X,Y}^{(t)}$ za $t \geq T$:

$$\bar{d}_{X,Y}^{(t)} = \frac{d_{X,Y}^{(t)} - \mathbf{E} \left[d_{X,Y}^{(\tau)} \mid t - T \leq \tau < t \right]}{\sqrt{\mathbf{Var} \left[d_{X,Y}^{(\tau)} \mid t - T \leq \tau < t \right]}}. \quad (3.5)$$

Ovako dobivena normalizirana razlika log-cijena koristi se kao mjera sličnosti. U odnosu na prethodnu, ova mjera je više osjetljiva na promjene log-cijena neovisno o volatilitetima samih vrijednosnica, što će doći do izražaja prilikom usporedbe dviju vrijednosnica čije se cijene mijenjaju različitom brzinom i intenzitetom. Primjer razlike nenormaliziranih log-cijena istih dionica X i Y , kao i u prethodnom slučaju, prikazan je na slici 3.4, te normalizirane razlike na slici 3.5.



Slika 3.4: Tamnijom bojom prikazana je razlika nenormaliziranih cijena $d_{X,Y}^{(t)}$, a svjetlijom punom linijom očekivanje i crtkanom linijom udaljenost od očekivanja za jednu devijaciju.



Slika 3.5: Normalizirana razlika cijena, $\bar{d}_{X,Y}^{(t)}$.

3.1.3. Mjera kointegracije slučajnih procesa

Za razliku od prethodne dvije mjere sličnosti gdje se cijene dionica tretiraju kao vremenski nizovi, kod ove mjere one se tretiraju kao realizacije diskretnog slučajnog procesa. Za neki slučajni proces zanimljivo je proučiti svojstvo stacionarnosti — slučajni proces $U[t]$ je stacionaran ako i samo ako vrijedi:

$$\mathbf{E}[U[t]] = \text{const.} \quad (3.6)$$

$$\mathbf{R}_{UU}[t, \tau] = \mathbf{E}[U[t] \cdot U[t + \tau]] = \mathbf{R}_{UU}[\tau]. \quad (3.7)$$

Za stacionarne slučajne procese vrijedi da njihova statistička svojstva ne ovise o vremenu: jednačba (3.6) tvrdi da je za stacionaran proces $U[t]$ očekivanje $\mathbf{E}[U[t]]$ neovisno o vremenu t , a jednačba (3.7) tvrdi da njegova autokorelacija $\mathbf{R}_{UU}[t, \tau]$ ovisi samo o vremenskom pomaku τ , ali ne i o t . Ovakva definicija stacionarnosti naziva se još i “stacionarnost u širem smislu”.

Red integracije slučajnog procesa, označen s $I(d)$, gdje je d red integracije, opisuje njegovu stacionarnost. Stacionarni slučajni procesi su $I(0)$. Za općenit proces $V[t]$ koji je $I(1)$ vrijedi da je proces $\Delta V[t]$ dobiven njegovim diferenciranjem

$$\Delta V[t] = (1 - L)V[t] = V[t] - V[t - 1] \quad (3.8)$$

stacionaran proces, a za općenit proces $W[t]$ koji je $I(d)$ vrijedi da je proces $\Delta^d W[t]$ dobiven njegovim uzastopnim diferenciranjem d puta

$$\begin{aligned}\Delta^d W[t] &= (1 - L)^d W[t] = \left(\sum_{n=0}^d \binom{d}{n} (-1)^n L^n \right) W[t] \\ &= \sum_{n=0}^d \binom{d}{n} (-1)^n W[t - n]\end{aligned}\quad (3.9)$$

stacionaran proces. U (3.8) i (3.9) korišten je operator kašnjenja L koji djeluje na sljedeći način: $L\{U[t]\} = U[t - 1]$.

Trivijalno je provjeriti da linearna kombinacija dvaju slučajnih procesa koji su $I(0)$ jest također $I(0)$, kao i da linearna kombinacija dvaju slučajnih procesa od kojih je jedan $I(0)$, a drugi $I(1)$ jest $I(1)$. Malo je teže odrediti kojeg je reda integracije linearna kombinacija dvaju slučajnih procesa koji su $I(1)$: to može biti $I(1)$, ali u posebnim slučajevima može biti i $I(0)$. Kointegracija dvaju slučajnih procesa $X[t]$ i $Y[t]$ koji su $I(1)$ označava da je njihova linearna kombinacija $U[t]$ stacionaran proces, odnosno $I(0)$:

$$X[t] + \beta Y[t] = U[t], \quad (3.10)$$

gdje je β konstanta koja se još naziva i kointegracijski koeficijent. Stacionarni slučajni proces $U[t]$ pritom se može prikazati kao:

$$U[t] = \mu + \varepsilon[t], \quad (3.11)$$

gdje je $\mu = \mathbf{E}[U[t]]$, očekivanje, a $\varepsilon[t]$ rezidualna vrijednost za koju vrijedi $\mathbf{E}[\varepsilon[t]] = 0$.

Pretpostavka kod ove mjere sličnosti jest da su dva nestacionarna slučajna procesa $P_X[t]$ i $P_Y[t]$ koji predstavljaju kretanje log-cijena vrijednosnica X i Y kointegrirana. Postoji nekoliko metoda za provjeru kointegracije, najčešće korištene su Engle-Granger metoda, Johansen metoda i Phillips-Ouliaris metoda. Najjednostavnija od navedenih je Engle-Granger metoda, koja je opisana u nastavku.

Ideja metode Engle-Granger je u prvom koraku estimirati stacionaran proces $U[t]$ koji je linearna kombinacija cijena $P_X[t]$ i $P_Y[t]$, zatim u drugom koraku provjeriti je li estimirani proces zaista stacionaran [3]. Postupkom najmanjih kvadrata najprije se određuju procjene parametara μ i β , označene s $\hat{\mu}$ i $\hat{\beta}$:

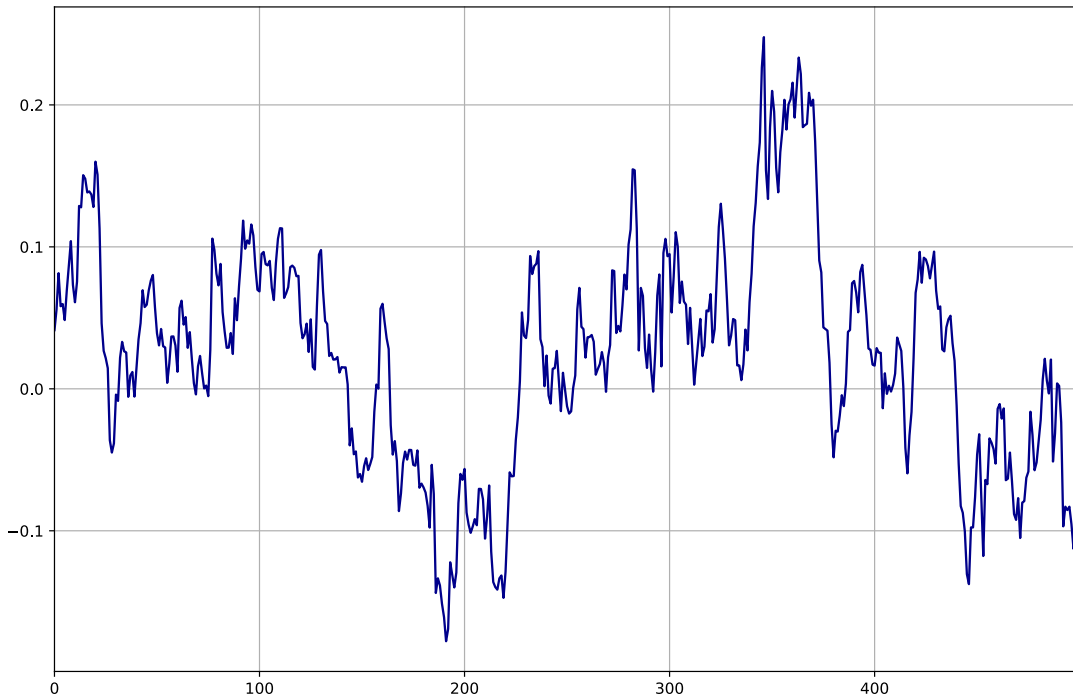
$$\hat{\mu}, \hat{\beta} = \arg \min_{\mu, \beta} \left\{ \sum_{\tau=t-T}^{t-1} |P_X[\tau] + \beta P_Y[\tau] - \mu|^2 \right\}, \quad (3.12)$$

a zatim se izraze procijenjene rezidualne vrijednosti $\hat{\varepsilon}[t]$:

$$\hat{\varepsilon}[t] = -\hat{\mu} + P_X[t] + \hat{\beta}P_Y[t]. \quad (3.13)$$

U drugom koraku provjerava se je li $\hat{\varepsilon}[t]$ zaista stacionaran proces: za dobivenu procjenu statističkim metodama računaju se t -vrijednosti kako bi se utvrdila pouzdanost kointegracije. t -vrijednost može se dobiti Dickey-Fullerovim testom [4].

Procijenjene rezidualne vrijednosti $\hat{\varepsilon}[t]$ koriste se kao mjera sličnosti. Primjer dobivenih rezidualnih vrijednosti za iste vrijednosnice X i Y prikazan je na slici 3.6. U odnosu na prethodne dvije mjere, ova mjera je rezultat dublje statističke analize, zbog čega je i računski zahtjevnija, što je ujedno i prednost i nedostatak. Kada se radi o velikom broju parova vrijednosnica, mjera kointegracije može se pokazati nepraktičnom za računanje, stoga je bolje koristiti neku od dvije prethodno navedene mjere sličnosti. U ovom radu korištena je normalizirana razlika log-cijena.



Slika 3.6: Procijenjene rezidualne vrijednosti $\hat{\varepsilon}[t]$ u slučaju vrijednosnica X i Y .

3.1.4. Nedostaci statističke arbitraže

Jedan od nedostataka statističke arbitraže jest relativno niska točnost u predviđanju kretanja cijene, ako se odluke o trgovanju promatraju kao klasifikacijski

problem predviđanja rasta ili pada cijene vrijednosnice. Točnost Acc računa se kao:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN}, \quad (3.14)$$

gdje su TP (engl. *true positive*), FP (engl. *false positive*), FN (engl. *false negative*), i TN (engl. *true negative*) odgovarajući brojevi pogođenih ili promašenih predviđanja pada ili rasta cijene kako je opisano na slici 3.7. Može se pretpostaviti

		ostvareno	
		raste	pada
predviđeno	raste	TP	FP
	pada	FN	TN

Slika 3.7: Moguće kombinacije predviđenih i ostvarenih kretanja cijene.

da je broj trgovanja u kojima cijena stvarno raste i pada podjednak. Pri testiranju algoritma statističke arbitraže nad testnim skupovima podataka točnost rijetko kada prelazi 60%, a najčešće je manja od 50%. Jedan od razloga slabe točnosti predviđanja je činjenica da je samo kretanje cijena slučajan proces čije je ponašanje teško predvidjeti. Korištene metode u algoritmu pokušavaju na temelju statističkih zakonitosti utvrditi vjerojatnosti kretanja cijena. Također, uzrok može biti i prividna korelacija (engl. *spurious correlation*) u kojoj kretanje cijena dviju vrijednosnica izgleda povezano, no zapravo je uzrok tome neki nepoznati vanjski faktor. Prividna korelacija na tržištima opisana je u [5]. Na primjer, ako se dogodi da na tržištu istovremeno raste potražnja za sladoledom i rashladnim uređajima, moglo bi se zaključiti kako postoji povezanost među njima, ali je vjerojatnije povećanje potražnje uzrokovao faktor koji nije bio uzet u obzir, primjerice visoka temperatura. Tada predviđeno ponašanje može biti poprilično različito od ostvarenog, budući da vanjski faktor može prestati utjecati u bilo kojem trenutku, a time se i gubi prividna korelacija. Isto tako, statističke metode ne mogu razlikovati prirodne devijacije u ponašanju cijena koje su karakteristične za slučajne procese, od onih prisilnih za koje stvarno postoji fundamentalan uzrok, kao što je npr. značajna promjena u načinu poslovanja neke firme, promjena vlasništva i tome slično. No, unatoč tome što algoritam griješi u više od pola slučajeva, ukupni profit je ipak pozitivan, jer pogreške donose znatno manje gubitke u odnosu na zaradu ostvarenu pogocima.

Drugi nedostatak tiče se načina na koji se raspoložive vrijednosnice uzimaju u obzir. Svaki par vrijednosnica gleda se zasebno i pritom je isključena dublja analiza međusobnih interakcija svih vrijednosnica zajedno. Trgovanje se također odvija uvijek u parovima, zauzimanjem kratke pozicije u jednoj i duge u drugoj, što zahtijeva da je kratka pozicija dopuštena na tržištu, a to ne mora biti slučaj. Također, pokazalo se da ovaj način promatranja parova zasebno rezultira jako visokim koeficijentom obrtaja¹, odnosno portfelj dobiven isključivo metodom statističke arbitraže je vrlo promjenljiv. Češće mijenjanje vrijednosnica u portfelju uzrokuje i veće troškove trgovanja, što u konačnici smanjuje i ukupnu vrijednost portfelja. U ovom radu istražena je metoda koja omogućuje trgovanje kupovanjem i prodajom svih vrijednosnica neovisno o parovima, koristeći pritom metode statističke arbitraže za opisivanje relacije preferencije među vrijednosnicama. Interakcija među vrijednosnicama opisuje se grafom toka preferencija, iz čega se naposljetku izvlači poredak prema individualnoj preferenciji svih vrijednosnica.

3.2. Relacija preferencije i funkcija korisnosti

Neka je Ω skup općenitih dobara. Relacija preferencije, označena simbolom \succ i definirana nad $\Omega \times \Omega$, je strogi slabi uređaj koji odgovara načinu na koji ljudi preferiraju jednu vrijednosnicu u odnosu na drugu. Između dvaju dobara $a, b \in \Omega$ relacija može, ali i ne mora postojati. Primjerice, a može biti više preferirano u odnosu na b , ili b u odnosu na a , ali moguća je situacija gdje su oba dobra podjednako preferirana. U tom slučaju radi se o indiferentnosti između a i b , što se označava kao $a \sim b$.

Ova relacija specifična je po tome što je:

- *irefleksivna*: $\forall x \in \Omega: \neg(x \succ x)$ — ni za jedno dobro ne vrijedi da je više preferirano od samog sebe,
- *asimetrična*: $\forall x, y \in \Omega: x \succ y \Rightarrow \neg(y \succ x)$ — ako je x više preferirano od y , onda y nije više preferirano od x ,
- *tranzitivna*: $\forall x, y, z \in \Omega: x \succ y \wedge y \succ z \Rightarrow x \succ z$ — ako je x više preferirano od y , i y više preferirano od z , tada je i x više preferirano od z ,
- *tranzitivna po indiferentnosti*: $\forall x, y, z \in \Omega: x \sim y \wedge y \sim z \Rightarrow x \sim z$ — ako je x podjednako preferirano kao i y , i y podjednako preferirano kao i z , tada je i x podjednako preferirano kao i z .

¹Koeficijent obrtaja opisan je u odjeljku 3.5.

Relacija preferencije koristi se u mikroekonomici u svrhu vrednovanja međusobnog odnosa različitih dobara [6]. Ovakva vrsta relacije prirodno opisuje odnose među različitim vrijednosnicama, npr. dionica A u nekom trenutku može biti više preferirana od dionice B . Razlog tome je što je za čovjeka lakše ocijeniti odnos (više, manje, jednako preferirano) između svakog para vrijednosnica, nego pridijeliti svakoj vrijednosnici individualnu mjeru preferencije, pogotovo ako se radi o velikom broju vrijednosnica. No ipak, u svrhu konstruiranja portfelja korisnije je posjedovati individualnu mjeru preferencije za svaku vrijednosnicu. Stoga je poželjno pronaći način da se iz relacije preferencije dobiju individualne mjere preferencije.

Funkcija korisnosti (engl. *utility function*) $U: \Omega \rightarrow \mathbb{R}$ je preslikavanje iz skupa dobara u skup realnih brojeva, na način da poredak preslikanih realnih brojeva odgovara poretку dobara prema individualnoj preferenciji, tj. vrijedi:

$$\forall x, y \in \Omega: x \succ y \Leftrightarrow U(x) > U(y).$$

Posljedično, za indiferentnost vrijedi:

$$\forall x, y \in \Omega: x \sim y \Leftrightarrow U(x) = U(y).$$

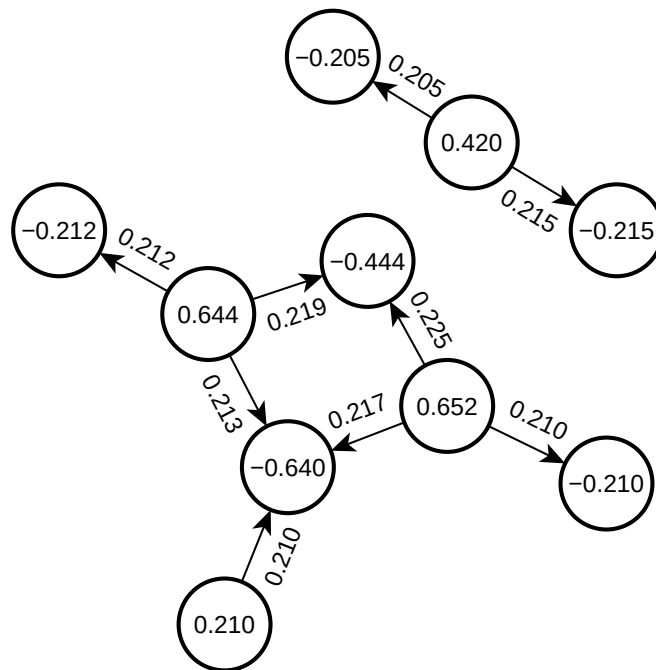
Jedno takvo preslikavanje ostvaruje se korištenjem metode potencijala koja je opisana u poglavlju 3.4. Ekvivalencija metode potencijala i funkcije korisnosti pokazana je u radu [7]. Osim što određuje poredak dobara prema individualnoj preferenciji, funkcija korisnosti također kvantificira preferenciju time što unosi intenzitet preferencije jedne vrijednosnice u odnosu na drugu; stoga funkcija korisnosti nosi više informacija od same relacije.

3.3. Graf toka preferencija

Graf toka preferencija je pomoćna struktura koja služi pri prevodenju toka preferencija među čvorovima u individualne mjere preferencije svakog čvora. U odnosu na relaciju preferencije koja je implicirana njime, graf sadrži još i intenzitete preferencija jednih čvorova nad drugima. Ako se zanemare težine bridova u grafu i ostave samo smjerovi, iz njega se direktno može pročitati relacija preferencije koju on opisuje. Težine u grafu daju dodatnu informaciju o individualnim mjerama preferencije čvorova koje odgovaraju toj relaciji preferencije, za razliku od situacije kada težina nema, što je ekvivalentno tome da su sve preferencije jednakog intenziteta. U odnosu na individualne mjere preferencije

svakog čvora zasebno, graf toka preferencija ne određuje poredak čvorova prema preferenciji.

Graf toka preferencija je težinski usmjereni graf, bez višestrukih bridova i petlji (bridova koji počinju i završavaju u istom grafu). Njegovi čvorovi predstavljaju vrijednosnice, usmjereni bridovi preferenciju jedne vrijednosnice nad drugom, a težine bridova odgovaraju intenzitetu preferencije. Ako između dva čvora nedostaje brid, smatra se da su pripadne vrijednosnice podjednako preferirane (indiferentnost u odlučivanju kod relacije preferencije). Graf kao cjelina opisuje tok preferencija među vrijednosnicama. Primjer grafa toka preferencija prikazan je na slici 3.8.



Slika 3.8: Primjer grafa toka preferencija. Na bridovima su prikazani intenziteti preferencija jedne vrijednosnice u odnosu na drugu, a u čvorove su upisane individualne mjere preferencija za pripadni čvor izračunate kao razlika zbroja težina bridova koji izlaze iz tog čvora i zbroja težina bridova koji ulaze u taj čvor. Opravdanje za takav postupak dano je u odjeljku 3.4. Ovaj graf je konzistentan s relacijom preferencije koju predstavlja.

Konstrukcija grafa u ovom radu temelji se na metodi statističke arbitraže. Brid koji ide od čvora i do čvora j s težinom $w_{i,j}$ je prisutan u grafu ako i samo ako su se dvije pripadne vrijednosnice i i j ponašale slično tijekom proteklog perioda, ali su se prema određenim statističkim mjerama trenutačno razdvojile.

Težina $w_{i,j}$ opisuje intenzitet ovog razdvajanja i služi kao mjera intenziteta preferencije jedne vrijednosnice u odnosu na drugu. Podrazumijeva se da je $w_{i,j} = -w_{j,i}$, odnosno negativan intenzitet preferencije označava da je jedna vrijednosnica toliko manje preferirana u odnosu na drugu. Detaljan opis postupka dobivanja korištenih statističkih mjera dan je u odjeljku 4.1.

Veze u ovom grafu opisuju relaciju preferencije među vrijednosnicama opisanima grafom, tako da brid koji ide iz čvora i u čvor j pokazuje da je vrijednosnica i više preferirana od vrijednosnice j . Prethodno spomenuta svojstva irefleksivnosti i asimetričnosti uvijek vrijede i u grafu toka preferencija: kako graf ne sadrži petlje, tj. nijedna vrijednosnica nije u relaciji sama sa sobom, ispunjeno je svojstvo irefleksivnosti; a kako nema višestrukih bridova među čvorovima vrijedi i svojstvo asimetričnosti. Ipak, problemi se pojavljuju kod svojstava tranzitivnosti, i tranzitivnosti po indiferentnosti, koja ne moraju uvijek vrijediti za proizvoljno konstruiran graf. Ako graf toka preferencija zadovoljava sva 4 svojstva, tada je on konzistentan s relacijom preferencije koju predstavlja. Poželjno bi bilo da relacija preferencije koju graf nameće zadovoljava sva četiri prethodno navedena svojstva; međutim konstruiranje grafa koji posjeduje takva svojstva je nepraktično, pogotovo kada se radi o većem broju vrijednosnica.

S druge strane, graf toka preferencija je intrinzično konzistentan ako vrijedi, za sve parove čvorova i i j : kada postoji više puteva u grafu koji od čvora i do čvora j , tada je zbroj težina duž tih puteva jednak. Na primjer, ako se u grafu nalaze tri čvora, i , j , i k , te je i više preferirano u odnosu na j s intenzitetom 1, a j više preferirano u odnosu na k također s intenzitetom 1, tada mora i biti više preferirano u odnosu na k s intenzitetom 2; u suprotnom graf nije intrinzično konzistentan. Intrinzično konzistentan graf također istovremeno zadovoljava i svojstvo tranzitivnosti relacije preferencije (obrat ne mora vrijediti).

Umjesto da cilj bude konstrukcija grafa koji će biti konzistentan sa svojstvima relacije preferencije, definirana je mjera konzistentnosti koja opisuje koliko je neki graf sličan sa svojom najbližom konzistentnom rekonstrukcijom. Mjera konzistentnosti koristi se kao dodatan parametar pri konstruiranju portfelja. Opis mjere konzistentnosti i način dobivanja najbliže konzistentne rekonstrukcije grafa dani su u odjeljku 3.4.2.

3.4. Metoda potencijala

Iz prethodno dobivenog grafa moguće je utvrditi u kojem paru vrijednosnica je tok preferencije najveći. Ipak, još uvijek nije moguće izravno utvrditi koja konkretna vrijednosnica je najviše ili najmanje poželjna, ili odrediti individualne mjere preferencije za svaku vrijednosnicu. Kako bi to bilo moguće korištena je metoda potencijala opisana u radu [8].

3.4.1. Računanje potencijala čvorova

Neka je za promatrani graf \mathcal{G} ukupno N čvorova, te najviše $E = \binom{N}{2}$ bridova, ako se radi o potpunom grafu. Ako \mathcal{G} nije potpun graf, on se može dopuniti tako da se dodaju bridovi između nepovezanih čvorova, čija je težina jednaka 0, a sam smjer nije bitan. U nastavku se podrazumijeva da je \mathcal{G} potpun graf.

Neka je \mathbf{B} matrica incidencije grafa \mathcal{G} dimenzija $[E \times N]$. Matrica incidencije je matrica za koju vrijedi:

1. broj redaka matrice jednak je broju bridova E , i broj stupaca matrice jednak je broju čvorova N ;
2. za svaki od E bridova grafa postoji odgovarajući redak u matrici koji ga opisuje, na sljedeći način: ako brid ide od čvora i do čvora j , tada taj redak sadrži redom -1 i 1 u stupcima koji odgovaraju čvorovima i i j ;
3. svi ostali elementi matrice jednaki su 0.

Neka je \mathbf{f} vektor dimenzija $[E \times 1]$ koji sadrži težine bridova, tj. tokove preferencije, i neka je ϕ vektor dimenzija $[N \times 1]$ koji sadrži potencijale čvorova, koje je potrebno pronaći. Redoslijed pripadnih čvorova i bridova moraju odgovarati redoslijedu čvorova i bridova u matrici incidencija \mathbf{B} .

U idealnom slučaju, kada je \mathcal{G} konzistentan, \mathbf{B} , ϕ , i \mathbf{f} zadovoljavaju jednadžbu:

$$\mathbf{B}\phi = \mathbf{f}. \quad (3.15)$$

Jednadžba (3.15) tvrdi da razlike potencijala dvaju čvorova odgovaraju toku preferencije, odnosno težini brida koji povezuje te čvorove (do na predznak). Ovo je zadovoljivo samo za konzistentne grafove, što često i nije slučaj u ovom zadatku. U slučaju kada je graf \mathcal{G} nekonzistentan, od interesa je pronaći rješenje ϕ^* koje ima minimalnu kvadratnu pogrešku. Tako originalni problem postaje:

$$\phi^* = \arg \min_{\phi} \left\{ \|\mathbf{B}\phi - \mathbf{f}\|^2 \right\} \Leftrightarrow \frac{\partial \|\mathbf{B}\phi^* - \mathbf{f}\|^2}{\partial \phi^*} = \mathbf{0}. \quad (3.16)$$

Matričnim deriviranjem (3.16) dobiva se sljedeća jednađžba:

$$\begin{aligned} 2\mathbf{B}^\top [\mathbf{B}\boldsymbol{\phi}^* - \mathbf{f}] &= \mathbf{0} \\ \mathbf{B}^\top \mathbf{B}\boldsymbol{\phi}^* &= \mathbf{B}^\top \mathbf{f}. \end{aligned} \quad (3.17)$$

Jednađžba (3.17) određuje $\boldsymbol{\phi}^*$ do na konstantu, tj. rješenje ima jedan stupanj slobode. Stoga se dodaje sljedeće ograničenje:

$$\mathbf{j}^\top \boldsymbol{\phi}^* = 0, \quad (3.18)$$

gdje je \mathbf{j} vektor jedinica istih dimenzija kao i $\boldsymbol{\phi}^*$. Jednađžba (3.18) osigurava jedinstveno rješenje u kojem je ukupan zbroj potencijala svih čvorova jednak nuli.

Združivanjem prethodnih dviju jednađžbi u jednu, tako da se (3.18) doda svakom retku u (3.17), dobiva se:

$$\begin{aligned} \mathbf{B}^\top \mathbf{B}\boldsymbol{\phi}^* + \mathbf{J}\boldsymbol{\phi}^* &= \mathbf{B}^\top \mathbf{f} \\ [\mathbf{B}^\top \mathbf{B} + \mathbf{J}] \boldsymbol{\phi}^* &= \mathbf{B}^\top \mathbf{f}, \end{aligned} \quad (3.19)$$

gdje je \mathbf{J} matrica jedinica istih dimenzija kao i $\mathbf{B}^\top \mathbf{B}$. Konačno, rješavanjem (3.19) po $\boldsymbol{\phi}^*$ dobiva se:

$$\boldsymbol{\phi}^* = [\mathbf{B}^\top \mathbf{B} + \mathbf{J}]^{-1} \mathbf{B}^\top \mathbf{f}. \quad (3.20)$$

Dodatno, izraz (3.20) sadrži u sebi Laplaceovu matricu $\mathbf{L} = \mathbf{B}^\top \mathbf{B}$. Za Laplaceovu matricu vrijedi:

- $(\mathbf{L})_{ii}$ je jednak broju susjeda čvora i ,
- ostali elementi jednaki su -1 .

Kako je graf \mathcal{G} potpun, odnosno broj susjeda svakog čvora jednak je $N - 1$, to znači da su elementi na dijagonali njegove Laplaceove matrice jednaki $N - 1$, dok su ostali elementi jednaki -1 . Tako se izraz $[\mathbf{B}^\top \mathbf{B} + \mathbf{J}]^{-1}$ može svesti na $[N\mathbf{I}]^{-1} = \frac{1}{N}\mathbf{I}$; i stoga se (3.20) može pojednostavniti u:

$$\boldsymbol{\phi}^* = \frac{1}{N} \mathbf{B}^\top \mathbf{f}, \quad (3.21)$$

tako da se dobije računalno optimalan izraz. Jednađžba (3.21) tvrdi da se rješenje u smislu najmanjih kvadrata $\boldsymbol{\phi}^*$ jednađžbe (3.15) dobije tako da se za svaki čvor pribroje težine svih bridova koji izlaze iz njega, i oduzmu težine svih bridova koji ulaze u njega.

3.4.2. Konzistentna rekonstrukcija grafa

Sada se na temelju dobivenog potencijala čvorova ϕ^* može izračunati tok preferencije f^* koji je konzistentan s potencijalom, tako da se u (3.15) supstituira ϕ^* umjesto ϕ :

$$f^* = B\phi^*. \quad (3.22)$$

Rekonstruirani graf toka preferencije f^* u odnosu na originalni f može sadržavati neke nove, pa čak i izgubiti neke stare bridove. Dodatno, B , ϕ^* , i f opisuju konzistentan graf \mathcal{G}^* koji je najbliža konzistentna rekonstrukcija grafa \mathcal{G} . Intrinzična konzistentnost u rekonstruiranom grafu je posljedica djelovanja matrice incidencije B : ona je linearni operator preslikavanja u čijoj slici su konzistentni tokovi preferencija.

Mjera konzistentnosti κ definira se kao:

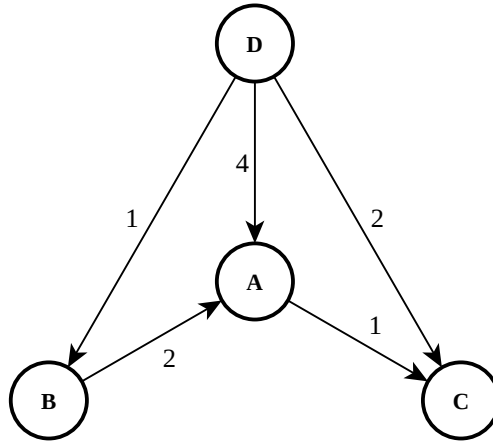
$$\kappa = \frac{\|f^*\|}{\|f\|}. \quad (3.23)$$

Jednadžba (3.23) predstavlja kosinus kuta između f i f^* u prostoru određenom vektor stupcima matrice incidencije B . Mjera konzistentnosti κ opisuje koliko je graf \mathcal{G} sličan grafu \mathcal{G}^* , i poprima vrijednosti u rasponu od 0 do 1, gdje 0 znači potpunu nekonzistentnost (u praktičnoj primjeni se ne događa), a 1 potpunu konzistentnost.

3.4.3. Primjer korištenja metode potencijala

Na sljedećem primjeru demonstrirana je metoda potencijala čvorova za određivanje individualnih preferencija čvorova u grafu. Na raspolaganju su četiri vrijednosnice, A , B , C i D . Metodom statističke arbitraže dobivene su preferencije među vrijednosnicama s intenzitetima prikazanim na grafu na slici 3.9.

Taj graf nije konzistentan s relacijom preferencije koju predstavlja: prema njemu B nije izravno usporediva s C , ali kako je $B \succ A$ i $A \succ C$, po tranzitivnosti bi trebalo biti i $B \succ C$, dok je prema grafu $B \sim C$. Također, postoje dva puta od D prema A : to su $D \rightarrow A$ i $D \rightarrow B \rightarrow A$; duž prvog puta je zbroj težina jednak 4, a duž drugog jednak je 3, pa graf nije intrinzično konzistentan. Isto je i s putevima od D do C : duž puta $D \rightarrow C$ zbroj težina jednak je 2, a duž puta $D \rightarrow B \rightarrow C$ zbroj težina jednak je 5, što je čak i veće nepodudaranje nego u prethodnom slučaju.



Slika 3.9: Primjer nekonzistentnog grafa prije izračunavanja potencijala čvorova. U čvorove su upisane oznake vrijednosnica, a na bridove su upisane težine koje opisuju tok preferencije.

Matrica incidencije \mathbf{B} i vektor toka preferencije \mathbf{f} grafa sa slike 3.9 su sljedeći:

$$\mathbf{B} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} -2 \\ 1 \\ -4 \\ 0 \\ -1 \\ -2 \end{bmatrix}$$

Matrica incidencije \mathbf{B} konstruirana je tako da redci odgovaraju leksikografskom poretку bridova: \overrightarrow{AB} , \overrightarrow{AC} , \overrightarrow{AD} , \overrightarrow{BC} , \overrightarrow{BD} , i \overrightarrow{CD} ; a stupci prirodnom poretку čvorova: A , B , C i D . Istim redoslijedom i težine bridova su upisane u vektor \mathbf{f} . Neke težine su negativne jer se nametnuti smjer brida u matrici incidencije ne podudara sa stvarnim smjerom brida u grafu.

Računanjem prema metodi potencijala $\phi^* = \frac{1}{N} \mathbf{B}^\top \mathbf{f}$ i $\mathbf{f}^* = \mathbf{B} \phi^*$, dobivaju se:

$$\phi^* = \begin{bmatrix} -1.25 \\ 0.75 \\ -0.25 \\ 1.75 \end{bmatrix}, \quad \mathbf{f}^* = \begin{bmatrix} -1.5 \\ -0.5 \\ -1 \\ 1 \\ -1.5 \\ -2.5 \end{bmatrix}$$

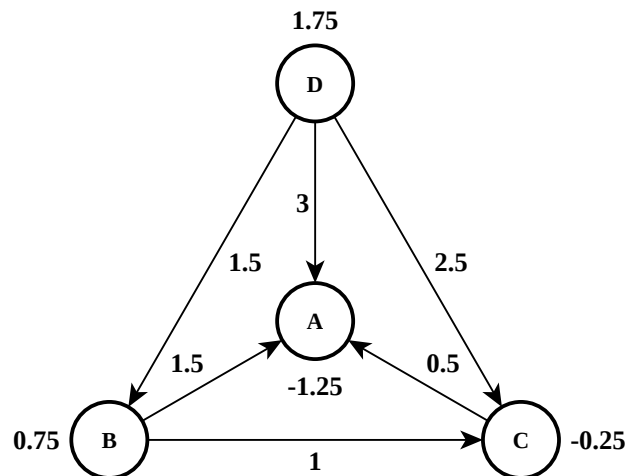
Prema dobivenim potencijalima čvorova ϕ^* može se zaključiti kako je najviše preferiran čvor D , zatim B , zatim C , i naposljetku A . Rekonstrukcija grafa

prikazana je na slici 3.10. Težine u rekonstruiranom vektoru toka preferencija \mathbf{f}^* su promijenjene u odnosu na originalni \mathbf{f} , i to je promjena veća gdje je brid uzrokovao veću nekonzistentiju u originalnom grafu — tako je najveća promjena u težini brida \overrightarrow{AD} , s -4 na -1 . Također, pojavio se i novi brid \overrightarrow{BC} koji prije nije postojao u originalnom grafu. Smjer brida \overrightarrow{AC} je preokrenut. Sada postoje tri puta od čvora D do A , i duž sva tri puta zbroj težina bridova iznosi 3. Iz toga se vidi da je rekonstruirani graf doista intrinzično konzistentan.

Mjera konzistentnosti grafa na slici 3.9 računa se po formuli (3.23):

$$\kappa = \frac{\|\mathbf{f}^*\|}{\|\mathbf{f}\|} = \frac{\sqrt{13}}{\sqrt{26}} \approx 0.70711.$$

Dobivena mjera konzistentnosti manja je od 1 što potvrđuje da početni graf nije bio intrinzično konzistentan. Što je mjera konzistentnosti niža, to je odluka koju početni graf implicira manje pouzdana.



Slika 3.10: Primjer najbliže konzistentne rekonstrukcije grafa nakon izračunavanja potencijala čvorova.

3.5. Koeficijent obrtaja

Koeficijent obrtaja (engl. *turnover ratio*) je mjera koja određuje koliki udio vrijednosnica u portfelju je promijenjen između dva vremenska koraka. Ova mjera može biti u rasponu od 0 do 2, gdje 0 označava da su vrijednosnice u portfelju ostale iste između dva vremenska koraka, dok 2 označava da među vrijednosnicama uključenima u portfelj u prethodnom i sadašnjem koraku nijedna

nije ostala ista. Ova mjera je važna između ostalog jer opisuje troškove trgovanja: u svakom vremenskom koraku troškovi trgovanja su direktno proporcionalni količini vrijednosnica koje su prodane ili kupljene u tom koraku.

Koeficijent obrtaja $\eta^{(t)}$ u vremenskom koraku t u odnosu na $t - 1$ računa se na sljedeći način. Neka je na raspolaganju N vrijednosnica. Vektor $\boldsymbol{\alpha}^{(t)} = [\alpha_1^{(t)} \ \alpha_2^{(t)} \ \dots \ \alpha_N^{(t)}]$ opisuje portfelj u koraku t ; $\alpha_i^{(t)}$ predstavlja udio i -te vrijednosnice, te vrijedi $\alpha_i \geq 0$, $\sum_{i=1}^N \alpha_i = 1$. Tada se koeficijent obrtaja $\eta^{(t)}$ računa na način:

$$\eta^{(t)} = \sum_{i=1}^N \left| \alpha_i^{(t)} - \alpha_i^{(t-1)} \right|. \quad (3.24)$$

Iz (3.24) može se vidjeti da se minimum $\eta^{(t)} = 0$ doista postiže ako i samo ako je ispunjeno $\boldsymbol{\alpha}^{(t)} = \boldsymbol{\alpha}^{(t-1)}$, a maksimum $\eta^{(t)} = 2$ ako i samo ako je ispunjeno $\alpha_i^{(t)} \neq 0 \Rightarrow \alpha_i^{(t-1)} = 0$, odnosno $\alpha_i^{(t-1)} \neq 0 \Rightarrow \alpha_i^{(t)} = 0$, jer tada vrijedi:

$$\eta^{(t)} = \sum_{i=1}^N \left| \alpha_i^{(t)} - \alpha_i^{(t-1)} \right| = \sum_{i=1}^N \left| \alpha_i^{(t)} \right| + \sum_{i=1}^N \left| \alpha_i^{(t-1)} \right| = 1 + 1 = 2.$$

Neka troškovi trgovanja, bilo kupovanja bilo prodavanja vrijednosnice, iznose p dijelova vrijednosti po kojoj se vrijednosnica kupuje, odnosno prodaje. Ako se računa s logaritmima cijena umjesto s običnim cijenama, troškovi trgovanja mogu se obračunati tako da se u svakom trenutku od ostvarenog profita oduzme $p \cdot \eta^{(t)}$.

4. Praktično ostvarenje algoritma trgovanja

Neka u zadanom skupu podataka ima ukupno N vrijednosnica označenih brojevima od 1 do N , kroz D vremenskih koraka. Ulazni podaci za algoritam čine vremenski nizovi povijesnih cijena vrijednosnica koje su na raspolaganju. Pretpostavka je da su vrijednosnice na tržištu likvidne, odnosno da se njima može trgovati u svakom vremenskom koraku u proizvoljnim količinama. Neka je s $a_i^{(t)}$ označena cijena vrijednosnice i u vremenskom koraku t , za $i \in [1, 2, \dots, N]$, i $t \in [0, 1, \dots, D - 1]$.

Logaritamska cijena $b_i^{(t)}$ vrijednosnice i , i razlika logaritamskih cijena $c_{i,j}^{(t)}$ para vrijednosnica i i j dobivaju se na sljedeći način, $\forall i, j \in [1, 2, \dots, N]$:

$$b_i^{(t)} = \log(a_i^{(t)}) \quad (4.1)$$

$$c_{i,j}^{(t)} = b_i^{(t)} - b_j^{(t)}. \quad (4.2)$$

Očekivanje $m_{i,j}^{(t)}$, i standardna devijacija $d_{i,j}^{(t)}$ razlika logaritamskih cijena $c_{i,j}^{(t)}$ tijekom proteklog perioda duljine T procjenjuju se sljedećim izrazima:

$$m_{i,j}^{(t)} = \frac{1}{T} \sum_{\tau=t-T}^{t-1} c_{i,j}^{(\tau)} \quad (4.3)$$

$$d_{i,j}^{(t)} = \sqrt{\frac{1}{T-1} \sum_{\tau=t-T}^{t-1} (c_{i,j}^{(\tau)} - m_{i,j}^{(t)})^2}. \quad (4.4)$$

Procjena srednje vrijednosti u (4.3) je nepristrana, što znači da je očekivanje od $m_{i,j}^{(t)}$ doista jednako srednjoj vrijednosti toga uzorka. Procjena standardne devijacije u (4.4) nije nepristrana, jer je dobivena korjenovanjem nepristrane procjene varijance. Kako je korjenovanje nelinearna operacija, ono unosi pristranost u procjenu te je ona nešto manja od veličine koju procjenjuje. Međutim, pristranost je manja što je broj uzoraka veći i postaje zanemariva za $T \geq 10$. Valja primijetiti kako je u izrazima sumiranja u (4.3) i (4.4) vremenski korak t

namjerno izostavljen jer se radi nad proteklom periodom trajanja T , zbog čega sumiranje ide do $t - 1$. Izračuni (4.1) – (4.4) su temeljni za daljnji tijek algoritma.

Također valja primijetiti da bi izračunavanje očekivanja i standardnih devijacija zasebno za svaki vremenski korak bilo računski jako neefikasno, pogotovo kada se radi o velikom skupu podataka. Stoga je preporučljivo koristiti algoritam za računanje s pomičnim prozorima, koji je opisan u dodatku A. Što se tiče redukcije memorijske potrošnje, $c_{i,j}^{(t)}$, $m_{i,j}^{(t)}$, i $d_{i,j}^{(t)}$ su idejno trodimenzionalni tenzori, dimenzija $[(D - T) \times N \times N]$. Međutim, sva tri tenzora su simetrična ili antisimetrična, odnosno vrijedi $c_{i,j}^{(t)} = -c_{j,i}^{(t)}$, $m_{i,j}^{(t)} = -m_{j,i}^{(t)}$, i $d_{i,j}^{(t)} = d_{j,i}^{(t)}$. Stoga je polovina informacije sadržane u njima redundantna. Imajući to u vidu, ti tenzori se mogu gusto upakirati u matrice dimenzija $[(D - T) \times N \cdot (N - 1)/2]$, koristeći sljedeće preslikavanje — za kodiranje para (i, j) u kod k koristi se izraz:

$$k = N \cdot (i - 1) + j - 1 - i \cdot (i + 1)/2, \quad (4.5)$$

a za dekodiranje (i, j) iz k izrazi:

$$i = \left\lfloor N + 1/2 - \sqrt{(N + 1/2)^2 - 2(N + k)} \right\rfloor, \quad (4.6)$$

$$j = k + i \cdot (i + 1)/2 - N \cdot (i - 1) + 1. \quad (4.7)$$

Na slici 4.1 prikazano je kodiranje u slučaju $N = 5$ vrijednosnica.

i/j	1	2	3	4	5
1	.	0	1	2	3
2	.	.	4	5	6
3	.	.	.	7	8
4	9
5

k	0	1	2	3	4	5	6	7	8	9
i	1	1	1	1	2	2	2	3	3	4
j	2	3	4	5	3	4	5	4	5	5

Slika 4.1: Primjer predložene sheme kodiranja i dekodiranja za $N = 5$. Točka (.) označava da se određena kombinacija ne koristi.

4.1. Konstrukcija grafa toka preferencija

Korištenjem prethodno dobivenih $c_{i,j}^{(t)}$, $m_{i,j}^{(t)}$, i $d_{i,j}^{(t)}$ sada je moguće kreirati graf toka preferencija među vrijednosnicama u svakom vremenskom koraku t .

Promatrajući fiksni vremenski korak t , izdvajaju se svi parovi vrijednosnica (i, j) za koje vrijedi:

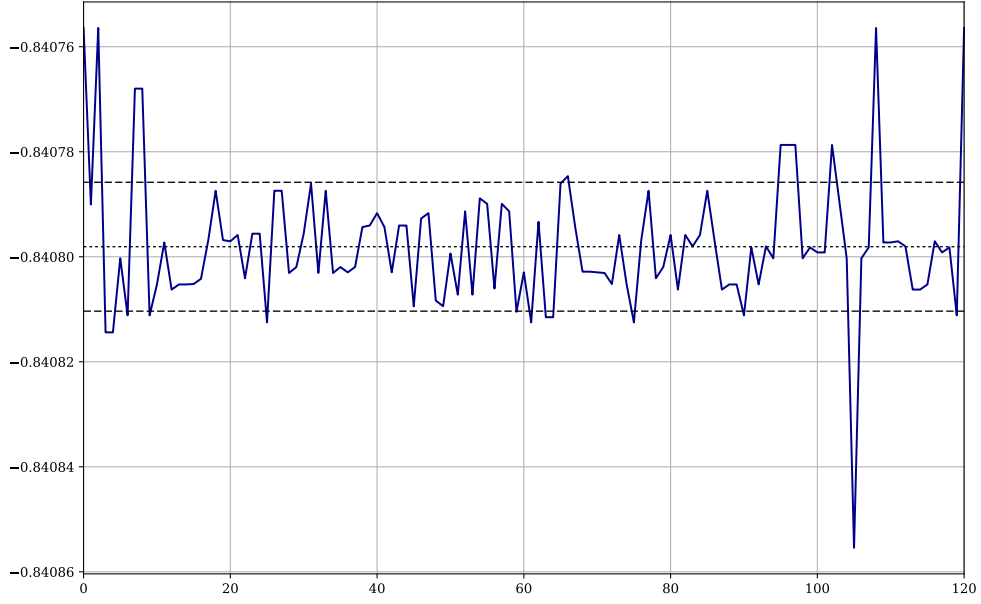
$$\left| c_{i,j}^{(t)} - m_{i,j}^{(t)} \right| > \beta \cdot d_{i,j}^{(t)}, \quad (4.8)$$

tj. trenutna apsolutna vrijednost razlike log-cijena para vrijednosnica (i, j) je barem za β standardnih devijacija udaljena od očekivanja, izmjerenih u proteklom periodu. Veličina β je parametar algoritma koji se zove prag trgovanja. Nakon toga, za svaki par (i, j) koji zadovoljava (4.8) dodaju se u graf čvorovi i i j , te brid koji ide od i do j s težinom $w_{i,j}^{(t)}$. Težina $w_{i,j}^{(t)}$ dobiva se kao:

$$w_{i,j}^{(t)} = \frac{c_{i,j}^{(t)} - m_{i,j}^{(t)}}{d_{i,j}^{(t)}}, \quad (4.9)$$

što odgovara udaljenosti razlike log-cijena od očekivanja, mjerene u standardnim devijacijama, i po apsolutnoj vrijednosti je veća od β . Negativan predznak težine $w_{i,j}^{(t)}$ označava da je stvarni smjer preferencije obrnut od pretpostavljenog, odnosno brid koji ide od j do i s pozitivnom težinom $w_{j,i}^{(t)} = -w_{i,j}^{(t)}$ ekvivalentan je prethodno pretpostavljenom bridu. Ilustracija ove mjere prikazana je na slici 4.2.

Tako je moguće konstruirati graf toka preferencija za bilo koji vremenski korak $t \in [T, T + 1, \dots, D - 1]$. U nekim vremenskim koracima moguće je da dobiveni graf bude prazan, ako nijedan par vrijednosnica (i, j) ne zadovoljava (4.8). Postavljanjem nižih vrijednosti za β dobivaju se gušći grafovi i manje često se događa da je graf prazan, dok se postavljanjem $\beta = 0$ uvijek dobivaju potpuni grafovi.



Slika 4.2: Razlika log-cijena u paru vrijednosnica (i, j) tijekom perioda od $T + 1$ vremenskih koraka, uz $T = 120$. Točka-crta linija predstavlja očekivanje, a područje između dviju crtkanih linija predstavlja β standardnih devijacija udaljenosti od očekivanja; obje mjere su izračunate nad prvih T vremenskih koraka. Tijekom vremenskog koraka $T + 1$, razlika log-cijena prelazi β standardnih devijacija iznad očekivanja proteklog perioda duljine T . To znači da će vrijednosnice i i j biti dodane u graf toka preferencija u vremenskom koraku $T + 1$. Težina $w_{i,j}^{(T+1)}$ opisuje devijaciju od očekivanja.

4.2. Signal trgovanja

Signal trgovanja $\Gamma_{X,Y}^{(t)}$ između vrijednosnica X i Y u trenutku t je funkcija koja određuje težine bridova kako je opisano u prethodnom odjeljku. Dobiva se na isti način kao i težine bridova:

$$\Gamma_{X,Y}^{(t)} = \frac{c_{X,Y}^{(t)} - m_{X,Y}^{(t)}}{d_{X,Y}^{(t)}}, \quad (4.10)$$

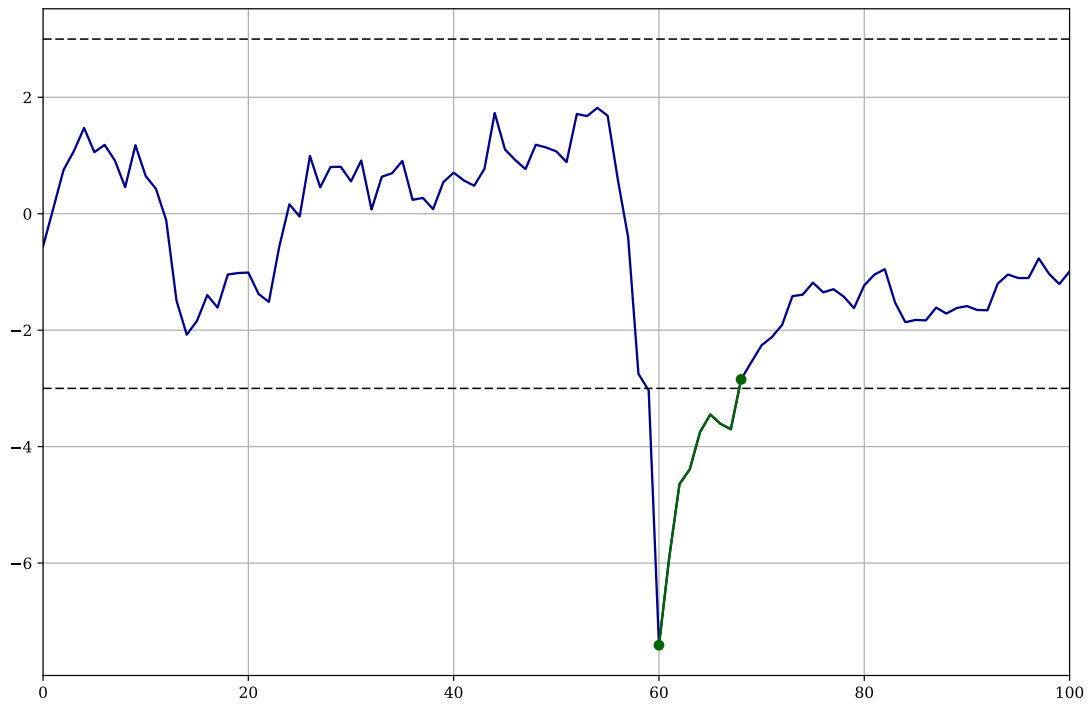
s bitnom razlikom da težine bridova postoje samo ako je prethodno zadovoljen prag β , a signal trgovanja je vremenski kontinuirana funkcija podložna daljnjoj analizi. Signal trgovanja $\Gamma_{X,Y}^{(t)}$, kao što mu i ime sugerira, opisuje željenu odluku o trgovanju vrijednosnicama X i Y u trenutku t na način opisan tablicom 4.1. U područjima I i III signal je aktivan, odnosno potiče trgovanje, dok je u području II neaktivan, jer ne dopušta trgovanje.

Na slici 4.4 prikazano je kretanje logaritamskih cijena vrijednosnica koje se slično ponašaju, te je vidljiv trenutak u kojem slično ponašanje prestaje, a to

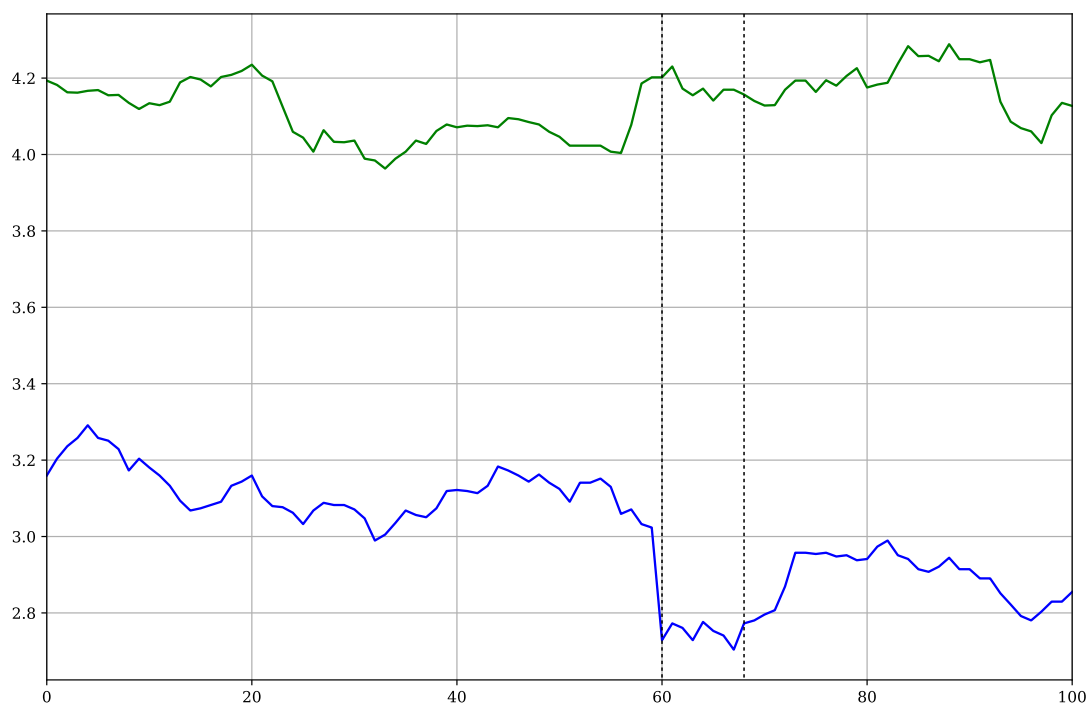
Tablica 4.1: Opis akcija kod signala trgovanja.

područje	uvjet	odluka
I	$-\infty < \Gamma_{X,Y}^{(t)} < -\beta$	duga pozicija u X , kratka u Y
II	$-\beta \leq \Gamma_{X,Y}^{(t)} \leq \beta$	nema akcije
III	$\beta < \Gamma_{X,Y}^{(t)} < \infty$	kratka pozicija u X , duga u Y

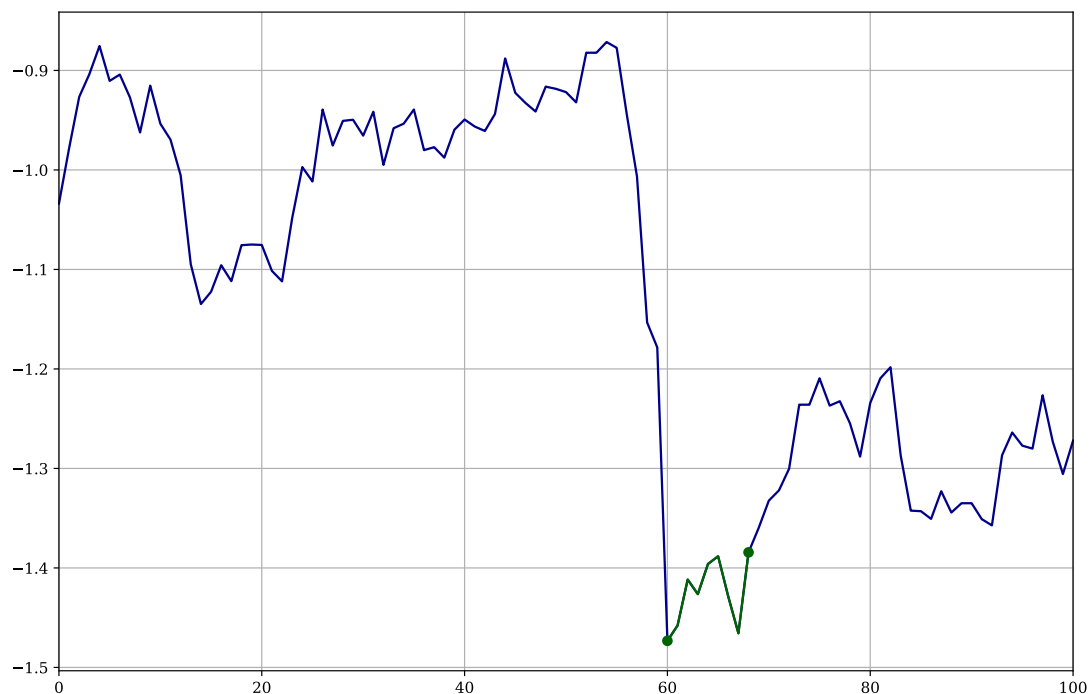
je trenutak kada se signal trgovanja mora aktivirati. Na slici 4.3 prikazan je pripadni signal trgovanja tih dviju dionica tijekom istog razdoblja s označenim aktivnim područjem. Naposljetku, na slici 4.5 prikazana je razlika logaritamskih cijena tijekom istog razdoblja kako bi se lakše mogao pratiti profit ostvaren trgovanjem u skladu sa signalom trgovanja.



Slika 4.3: Signal trgovanja $\Gamma_{X,Y}^{(t)}$ za vrijednosnice X i Y , dobiven uz korištenje $T = 120$. Prag β je postavljen na 3. U trenutku $t_1 = 60$ (označeno prvom zelenom točkom) signal trgovanja prelazi zadani prag i ostaje u aktivnom području (čiji je razpon označen zelenom bojom) do trenutka $t_2 = 69$ (označeno drugom zelenom točkom). Kako je signal u aktivnom području manji od $-\beta$, zauzima se kratka pozicija u vrijednosnici X i duga u Y , a po izlasku iz aktivnog područja te pozicije se zatvaraju.



Slika 4.4: Kretanje logaritamskih cijena vrijednosnica X (plavo) i Y (zeleno). Između crtkanih linija nalazi se područje u kojem je signal trgovanja aktivan. Uočljiva je sličnost u kretanju cijena prije i poslije aktivnog područja signala trgovanja.



Slika 4.5: Razlika logaritamskih cijena vrijednosnica X i Y . Početak i kraj aktivnog područja označen je zelenim točkama. Ukupan profit jednak je razlici iznosa na kraju i iznosa na početku aktivnog područja i iznosi 0.889 u ovom primjeru.

Signal trgovanja ovisi o odabranoj veličini vremenskom prozora T , a odluka koju donosi dodatno ovisi o odabranom iznosu praga trgovanja β . Unutar iduća dva pododjeljka opisane su metode za biranje pogodnih vrijednosti za parametre T i β .

4.2.1. Optimiranje praga trgovanja

U slučaju kada se trgovanje radi nad 10 ili više vrijednosnica, značajna poboljšanja mogu se postići ako se selekcija parova radi na sljedeći način. Za svaki par vrijednosnica X i Y s pripadnim cijenama $P_X^{(t)}$ i $P_Y^{(t)}$ u trenutku t , zahtijeva se da varijanca razlike cijena u proteklom periodu duljine T bude manja od zbroja varijanci cijena pojedinačno:

$$\begin{aligned} \mathbf{Var} \left[\left(P_X^{(\tau)} - P_Y^{(\tau)} \right) \mid t - T < \tau \leq t \right] &< \mathbf{Var} \left[P_X^{(\tau)} \mid t - T < \tau \leq t \right] \\ &+ \mathbf{Var} \left[P_Y^{(\tau)} \mid t - T < \tau \leq t \right]. \end{aligned} \quad (4.11)$$

Uvjet (4.11) ekvivalentan je uvjetu $\mathbf{Cov} \left[P_X^{(\tau)}, P_Y^{(\tau)} \mid t - T < \tau \leq t \right] > 0$, jer je:

$$\begin{aligned} \mathbf{Var} \left[\left(P_X^{(\tau)} - P_Y^{(\tau)} \right) \mid t - T < \tau \leq t \right] &= \mathbf{Var} \left[P_X^{(\tau)} \mid t - T < \tau \leq t \right] \\ &+ \mathbf{Var} \left[P_Y^{(\tau)} \mid t - T < \tau \leq t \right] \\ &- 2 \mathbf{Cov} \left[P_X^{(\tau)}, P_Y^{(\tau)} \mid t - T < \tau \leq t \right], \end{aligned}$$

ali je (4.11) računalno optimalniji budući da su navedene varijance već izračunate.

4.2.2. Optimiranje veličine vremenskog prozora

Jedan način optimiranja veličine vremenskog prozora T može se postići jednostavnom pretragom prostora vrijednosti za T testiranjem nad povijesnim podacima (npr. prethodnih 1000 dana) kako bi se pronašla vrijednost T koja daje najbolje kriterije (ukupan profit, ili Sharpeov omjer). Smislene vrijednosti parametra T koje vrijedi ispitati uglavnom su veće od 30 i manje od 250 dana. No, kao i u prethodnom slučaju, bolji rezultati mogu se postići optimiranjem veličine prozora za svaki par vrijednosnica zasebno. Ispitivanje svih mogućih veličina prozora za svaki par nije izvedivo u praktičnom vremenu, stoga se mora posegnuti za nekom bržom metodom optimiranja.

Testiranjem nad povijesnim podacima, prilikom trgovanja s 10 ili više vrijednosnica, empirijski je utvrđen način za poboljšanje dobivenih rezultata. Promatrajući signal trgovanja $\Gamma_{X,Y}$ za vrijednosnice X i Y određuje se vrijednost

parametra T koja maksimizira varijancu signala trgovanja. U tablici 4.2 prikazani su rezultati prije i poslije optimizacije praga trgovanja i veličine vremenskog prozora.

Tablica 4.2: Rezultati prije i poslije optimizacije parametara T i β .

	prije optimizacije	poslije optimizacije
bez troškova trgovanja:		
ukupni profit	0.32509	0.48489
Sharpeov omjer	2.28115	3.36565
uz uključene troškove trgovanja od 0.10%:		
ukupni profit	0.17093	0.30617
Sharpeov omjer	1.19602	2.12594
prosječni koeficijent obrtaja	0.77078	0.89355
prosječna konzistentnost	0.94716	0.67343
prosječna preciznost	0.51500	0.60500

4.3. Konstrukcija portfelja

Individualna mjera preferencije za svaku vrijednosnicu dobivena je metodom potencijala, na način kako je ranije opisano u 3.4.1. Vrijednosnice s trenutno najvećom preferencijom preporučeno je kupiti, dok je one s najmanjom preferencijom preporučeno prodati, odnosno zauzeti kratku poziciju, ako je to moguće.

Neka je $\phi^{(t)} = [\phi_1^{(t)} \ \phi_2^{(t)} \ \dots \ \phi_N^{(t)}]$ vektor individualnih preferencija vrijednosnica u vremenskom koraku t , te $\phi_i^{(t)}$ predstavlja individualnu preferenciju vrijednosnice i u vremenskom koraku t . Pri odabiru vrijednosnica za portfelj, u obzir se također uzima i mjera konzistentnosti κ . Uključivanje mjere konzistentnosti pri odabiru omogućuje zaštitu od fundamentalnog rizika (propadanja neke dionice) putem diverzifikacije portfelja. Interpretacija mjere konzistentnosti je sljedeća: niska mjera konzistentnosti sugerira da procjena u sadašnjem vremenskom koraku nije pouzdana, stoga je bolje diverzificirati portfelj uzimanjem većeg broja vrijednosnica, dok visoka vrijednost mjere konzistentnosti sugerira da je sigurno trgovati s manjim brojem vrijednosnica.

Granica mjere preferencije za vrijednosnice koje će biti uključene u portfelj je proporcionalna mjeri konzistentnosti κ . Ovisno o prirodi skupa vrijednosnica koji je na raspolaganju, mjera konzistentnosti može se preinačiti tako da algoritam bude više ili manje sklon diverzifikaciji portfelja, na sljedeći način:

$$\kappa' = a + (1 - a) \cdot \kappa^b, \quad (4.12)$$

gdje je $a \in [0, 1], b \in \mathbb{R}^+$. Za vrijednosti $a = 0, b = 1$, κ' postaje jednak κ . Parametar a regulira utjecaj mjere konzistentnosti na određivanje portfelja: za $a = 0$, diverzifikacija u potpunosti ovisi o mjeri konzistentnosti κ , dok za $a = 1$ mjera konzistentnosti ne dolazi do izražaja te se u portfelju zadržava samo vrijednosnica s najvećom individualnom preferencijom. S druge strane, parametar b regulira diverzifikaciju portfelja: kada je $0 < b < 1$, algoritam je manje privržen diverzifikaciji čak i kada je konzistentnost mala, a kada je $b > 1$, algoritam je više privržen diverzifikaciji čak i kada je mjera konzistentnosti velika.

Duga pozicija zauzima se u vrijednosnicama i za koje u vremenskom koraku t vrijedi:

$$\phi_i^{(t)} \geq \kappa' \cdot \Phi_{max}^{(t)},$$

gdje je $\Phi_{max}^{(t)} = \max_j \{|\phi_j^{(t)}|\}$, maksimum po apsolutnoj vrijednosti. Isto tako, kratka pozicija zauzima se u onim vrijednosnicama i za koje u vremenskom koraku t vrijedi:

$$\phi_i^{(t)} \leq -\kappa' \cdot \Phi_{max}^{(t)}.$$

Ponekad je poželjno koristiti alternativnu strategiju za odabir vrijednosnica koja u svakom trenutku koristi fiksni broj vrijednosnica u portfelju. Neka je željeni broj vrijednosnica označen sa $q > 0$. Prag $\Phi_q^{(t)}$ određuje se kao q -ta po apsolutnoj vrijednosti najveća mjera preferencije u vektoru $\phi^{(t)}$:

$$\Phi_q^{(t)} = \begin{cases} \Phi_{max}^{(t)}, & q = 1, \\ \text{abs max} \left\{ \phi^{(t)} \setminus \left\{ \Phi_i^{(t)} \mid 1 \leq i < q \right\} \right\}, & q > 1, \end{cases}$$

gdje je $\text{abs max} \{ \cdot \}$ maksimum po apsolutnoj vrijednosti. Duga pozicija zauzima se u vrijednosnicama i za koje u vremenskom koraku t vrijedi:

$$\phi_i^{(t)} \geq \Phi_q^{(t)},$$

i kratka pozicija u vrijednosnicama i za koje u vremenskom koraku t vrijedi:

$$\phi_i^{(t)} \leq -\Phi_q^{(t)}.$$

4.4. Testiranje

Simulacija algoritma trgovanja implementirana je u programskom jeziku *Python 3* uz korištenje standardnog paketa znanstvenih računalnih alata za *Python*, *SciPy* [9]. Ključne komponente tog paketa uključuju:

NumPy Standardna biblioteka za računanje, koja je optimirana za efikasno računanje nad matricama i vektorima.

pandas Nadogradnja nad *NumPy* biblioteku koja donosi naprednije podatkovne strukture, te mogućnost čitanja i pisanja u raznovrsne datotečne formate.

SciPy Kolekcija efikasnih algoritama iz linearne algebre, statistike, obrade signala, optimizacije i ostaloga, koji koriste *NumPy*.

Matplotlib Biblioteka koja pruža podršku za prikazivanje podataka u obliku raznih vrsta grafova.

Uz navedene komponente korišten je i *Jupyter* poslužitelj koji pruža mogućnost rada u zasebnim bilježnicama u kojima je moguće interaktivno pokretati napisani kod i pregledavati dobivene rezultate. Svi korišteni alati su otvorenog koda. U svrhu testiranja rada algoritma implementirane su dolje navedene pomoćne funkcije.

Računanje srednje vrijednosti i varijance na pomičnom prozoru

```
def rolling_mean_variance(series, T, weight=1.0,
                        return_stds=False):
    """
    Calculates mean and variance of time series on rolling windows.
    :param series: 2D matrix of multiple time series packed in
    columns.
    :param T: size of rolling window.
    :param weight: (optional) for exponential mean and average.
    :param return_stds: return standard deviation instead of
    variance.
    :return: a tuple consisting of mean and variance.
    """
    L, N = series.shape
    if L < T:
        return np.empty((0, N)), np.empty((0, N))
```



```

init_ws = np.tile(np.array([[weight ** i
                             for i in range(T - 2, -1, -1)]]
                             ).transpose(), [1, N])
mean = np.zeros((L - T + 1, N))
var = np.zeros((L - T + 1, N))
s0 = ((np.sum(init_ws[:, 0]) + weight ** (T - 1))
      * np.ones((1, N)))
s1 = np.sum(series[:T - 1, :] * init_ws, axis=0,
             keepdims=True)
s2 = np.sum(series[:T - 1, :] ** 2 * init_ws, axis=0,
             keepdims=True)
s1_prev = np.zeros((1, N))
s2_prev = np.zeros((1, N))
for i in range(L - T + 1):
    s1 *= weight
    s2 *= weight
    s1 += series[i + T - 1, :] - s1_prev * weight ** T
    s2 += series[i + T - 1, :] ** 2 - s2_prev * weight ** T
    s1_prev = series[i, :]
    s2_prev = series[i, :] ** 2
    mean[i, :] = s1 / s0
    var[i, :] = (s0 * s2 - s1 * s1) / (s0 * (s0 - 1))
return (mean, var) if not return_stds else (mean, np.sqrt(var))

```

Računanje koeficijenata obrtaja

```

def turnover_ratio(series):
    """
    Calculates turnover ratio.
    :param series: list of sets of assets
    :return: list of turnover ratios
    """
    turnovers = []
    S1 = set()
    for S2 in (set(s) for s in series):
        lS1 = len(S1)
        lS2 = len(S2)
        turnover = 0
        if lS1:
            turnover += len(S1 - S2) / lS1
        if lS2:
            turnover += len(S2 - S1) / lS2
        if lS1 and lS2:

```

```

        turnover += len(S1 & S2) * abs(1 / lS1 - 1 / lS2)
    turnovers += [turnover]
    S1 = S2
    return np.array(turnovers)

```

Algoritam statističke arbitraže

```

def stat_arb(log_prices, T, d='filtered',
             return_trading_signal=False, v=0.0,
             trading_signal=None):
    """
    Statistical arbitrage algorithm.
    :param log_prices: assets log price series
    :param T: windows size
    :param d: selection threshold, real number or 'filtered' if using
    selection based on change of variance
    :param return_trading_signal: (optional) include return trading
    signal in returned values
    :param v: (optional) portion of most volatile assets that should
    be discarded
    :param trading_signal: supply precalculated trading signal
    :return: triple containing list of timestamps, list of coded
    pairs and list of weights
    """
    N = log_prices.shape[1]
    log_pairs = calculate_pairwise_diffs(log_prices)
    _, std_prices = rolling_mean_variance(log_prices, T,
                                         return_stds=True)
    mean_pairs, std_pairs = rolling_mean_variance(log_pairs, T,
                                                  return_stds=True)

    if trading_signal is None:
        trading_signal = ((log_pairs[T:] - mean_pairs[:-1])
                          / std_pairs[:-1])

    cond = None
    if d == 'filtered':
        trading_threshold = np.array([np.sqrt(std_prices[:-1, i] ** 2
                                              + std_prices[:-1, j] ** 2)
                                     for i in range(N)
                                     for j in range(i + 1, N)]
                                   ).transpose()
        cond = std_pairs[:-1] <= trading_threshold
    elif isinstance(d, numbers.Number):

```

```

    cond = np.abs(trading_signal) >= d
else:
    raise ValueError('d_must_be_"filtered"_or_number.')
if v != 0.0:
    most_volatile = np.argsort(std_prices, axis=1)[: , -int(v * N):]
    make_pairs = lambda row: [encode_pair(i + 1, j + 1, N)
                               for ii, i in enumerate(row)
                               for j in row[ii + 1:]]
    most_volatile_pairs = np.apply_along_axis(make_pairs, 1,
                                              most_volatile)

    if most_volatile_pairs.shape[1] != 0:
        cond[:, most_volatile_pairs] = False

cond[std_pairs[:,-1] == 0] = False
ts, codes = np.where(cond)
weights = trading_signal[ts, codes]
ts += T
return_values = (ts, codes, weights)
if return_trading_signal:
    return_values += (trading_signal,)
return return_values

```

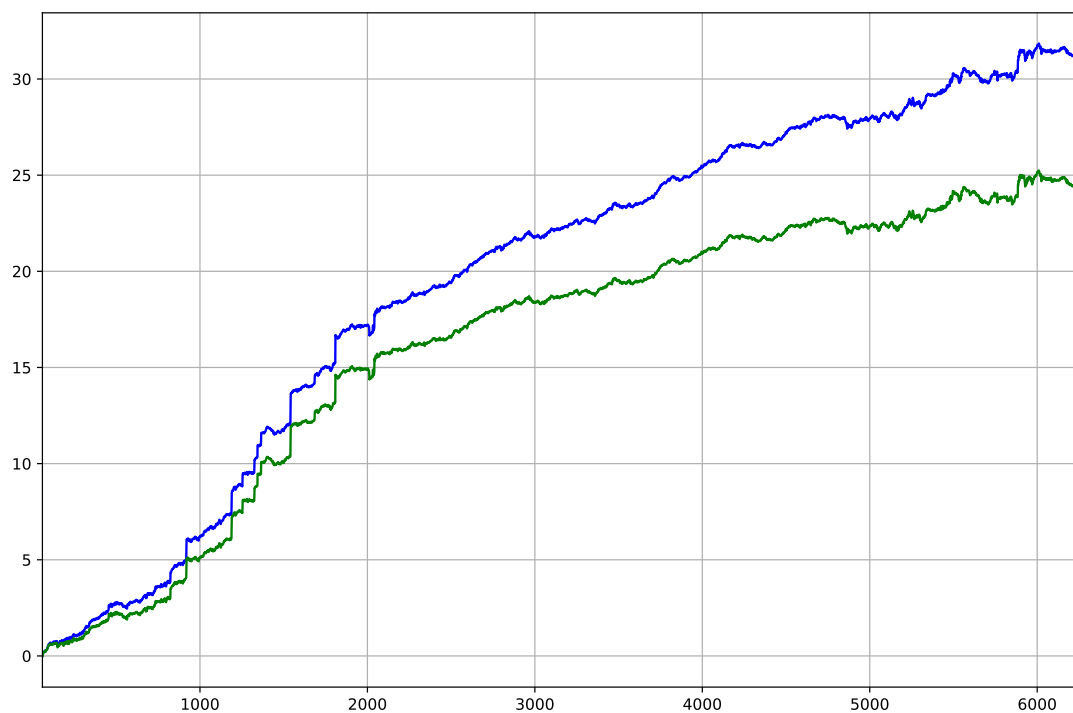
5. Rezultati

Ovdje su prikazani rezultati testiranja razvijenog algoritma trgovanja vrijednosnicama na američkom (S&P 500 index), i hrvatskom tržištu (CROBEX).

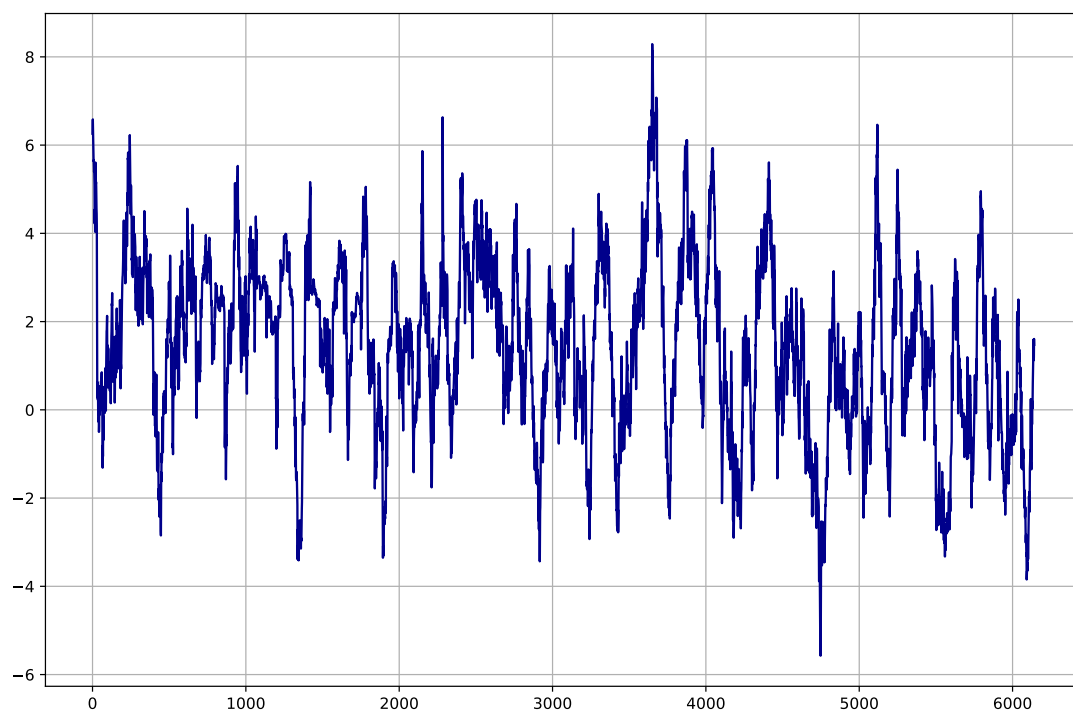
5.1. S&P 203

Ovaj skup podataka sadrži odabrane 203 od 500 dionica koje su bile kontinuirano uključene u S&P 500 indeks u razdoblju od 1. siječnja 1980. do 31. prosinca 2003. Razdoblje uključuje 6261 dan trgovanja, odnosno vremenskih koraka. Ukupno 20503 parova dionica je ispitano metodom statističke arbitraže u svakom vremenskom koraku. Simulirani su troškovi trgovanja u iznosu 0.10% za svako obavljeno trgovanje. Sažeti prikaz rezultat dan je u tablici 5.1.

Godišnji Sharpeov omjer S definiran je kao omjer srednjeg godišnjeg povrata μ_r i srednje godišnje volatilnosti σ_r . Analizirana su zasebno trgovanja koja rezultiraju pozitivnim i negativnim profitom, njihova razlika i omjer. Mjera točnosti definirana je kao omjer broja trgovanja koja rezultiraju pozitivnim profitom u ukupnom broju trgovanja. Mjera obrtaja (engl. *turnover ratio*) opisuje promjenu portfelja između dva vremenska koraka. Ova mjera može biti u rasponu od 0 do 2, gdje vrijednost 0 znači da nema promjene, vrijednost 1 odgovara prelasku iz ili u prazan portfelj, a vrijednost 2 znači da su vrijednosnice između dva vremenska koraka potpuno različite. Na kraju je izračunat i ukupan profit uz uključene troškove trgovanja.



Slika 5.1: Rezultati simulacija trgovanja nad podskupom dionica S&P 500 indeksa. Plavom bojom je prikazana krivulja kumulativnog profita bez uključenih troškova trgovanja, a zelenom uz uključene troškove trgovanja od 0.10% po transakciji.



Slika 5.2: Sharpeov omjer izražen na godišnjoj razini, na prozorima veličine $T = 60$, za simulaciju trgovanja nad podskupom dionica S&P 500 indeksa.

Tablica 5.1: Rezultati testiranja nad S&P 203 skupu, uz $T = 60$, $\beta = 0$.

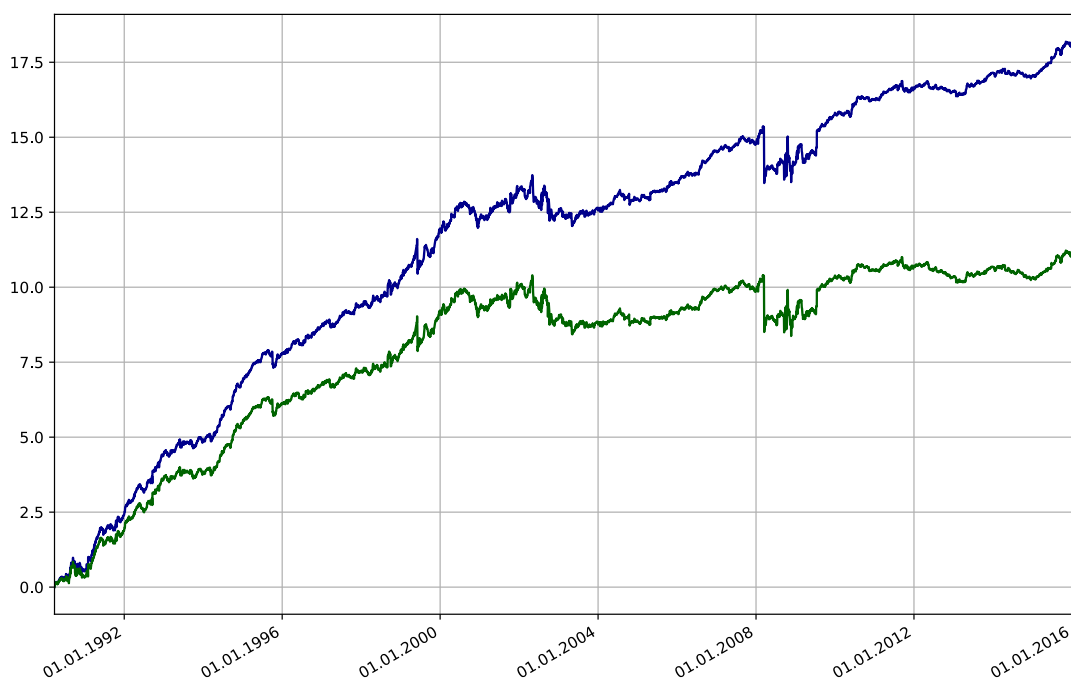
Parametar:	0.0				0.5		1.0	
	0.5	1.0	2.0	0.5	1.0	2.0	1.0	/
a								
b								
Prosječni povrat (godišnji)	0.95339	0.88967	0.84463	0.98336	0.95663	0.89704	1.00223	
Volatilnost (godišnja)	0.77042	0.76595	0.74077	0.77905	0.77054	0.76660	0.78363	
Sharpeov omjer (godišnji)	1.23750	1.16152	1.14020	1.26225	1.24150	1.17015	1.27896	
Profit:								
samo pozitivan	89.27624	88.89440	88.32548	89.58775	89.29840	89.04414	89.55020	
samo negativan	-58.37779	-59.03715	-58.41396	-58.24852	-58.32385	-59.02220	-58.05316	
ukupan	30.89846	29.85725	29.91152	31.33923	30.97456	30.02195	31.49704	
omjer pozitivnog i negativnog	1.52928	1.50574	1.51206	1.53803	1.53108	1.50866	1.54256	
Prosječna točnost	0.36485	0.39276	0.43413	0.34902	0.36458	0.39145	0.33241	
Prosječni koeficijent obrtaja	0.59976	0.64224	0.73597	0.57585	0.59947	0.64089	0.55112	
Stvarni profit, uz troškove trgovanja od 0.10%	23.46019	21.89215	20.78402	24.19757	23.53996	22.07361	24.66204	

5.2. S&P 500

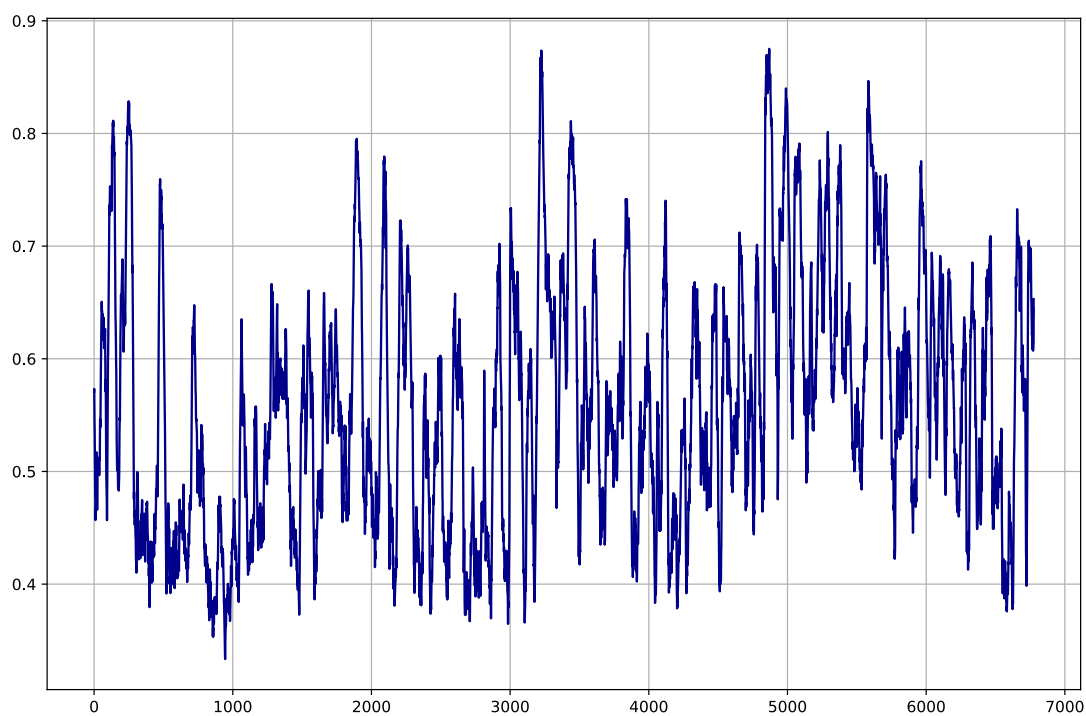
Rezultati testiranja nad punim S&P 500 indeksom. Za razliku od prethodne situacije gdje su sve dionice ostale “žive” tokom cijelog razdoblja, u ovom skupu postoje dionice koje propadaju, odnosno njihova vrijednost padne na nulu. Pokazalo se da algoritam ne može predvidjeti propadanje dionice, između ostalog jer ne raspolaže mjerodavnim podacima iz kojih bi se takvo što moglo zaključiti. Stoga trpi gubitke u određenim trenucima. Kao nadomjesna mjera mogućnosti propadanja dionice koristi se volatilnost cijena. Prilikom testiranja izbačeno je 20% najvolatilnijih dionica. Rezultati su prikazani u tablici 5.2 i na slikama 5.3 i 5.4.

Tablica 5.2: Rezultati testiranja nad S&P 500 skupom.

Rezultati:	
bez troškova trgovanja:	
ukupni profit	18.09261
Sharpeov omjer	0.87507
uz uključene troškove trgovanja od 0.10%:	
ukupni profit	11.03961
Sharpeov omjer	0.53382
prosječni koeficijent obrtaja	1.04088
prosječna konzistentnost	0.55817
prosječna preciznost	0.52642



Slika 5.3: Rezultati simulacija trgovanja nad punim skupom dionica S&P 500 indeksa. Plavom bojom je prikazana krivulja kumulativnog profita bez uključenih troškova trgovanja, a zelenom uz uključene troškove trgovanja od 0.10% po transakciji.



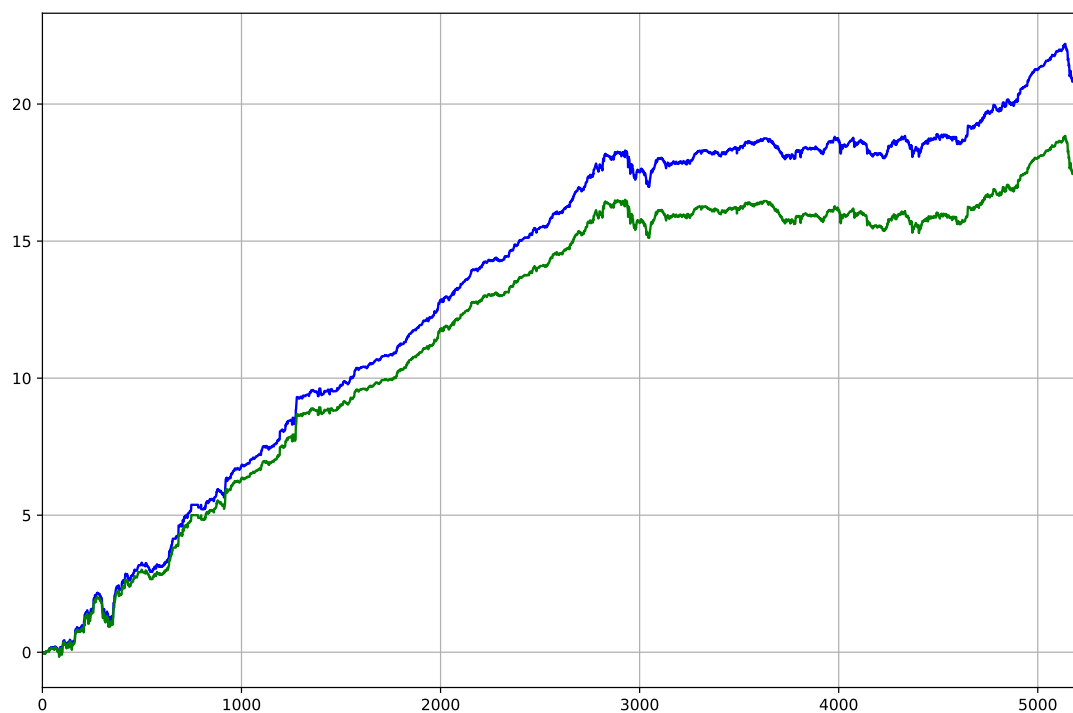
Slika 5.4: Prikaz dobivenih mjera konzistentnosti za simulaciju trgovanja nad punim skupom dionica S&P 500 indeksa.

5.3. CROBEX

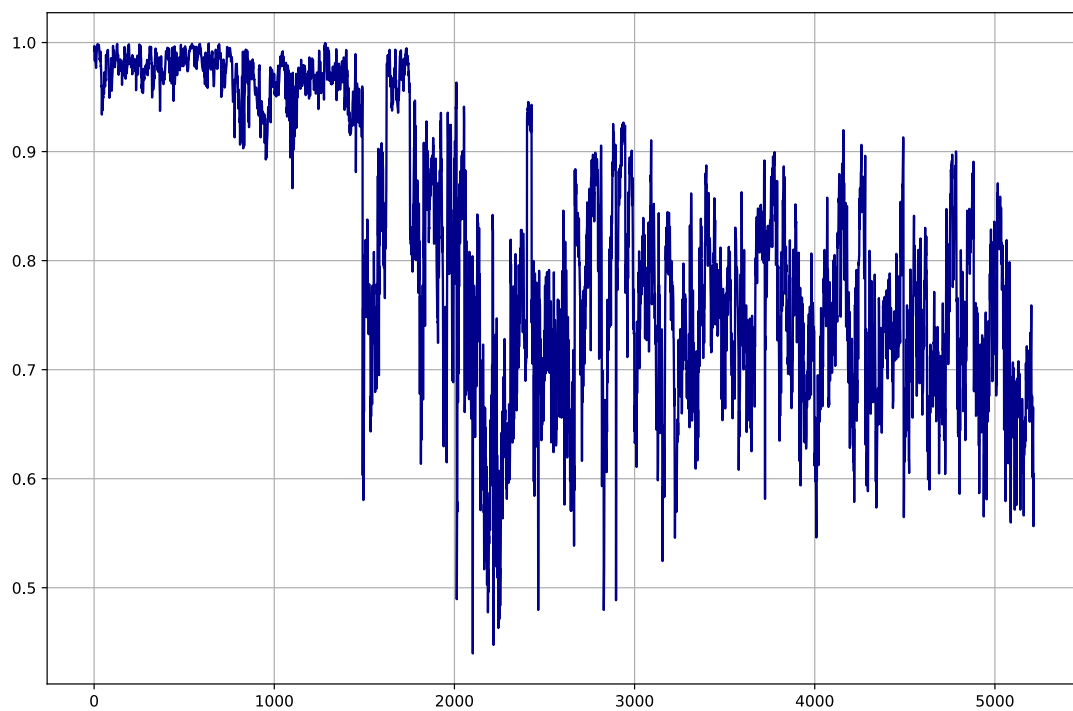
Algoritam je testiran i nad skupom dionica na hrvatskom tržištu koje su uključene u CROBEX indeks. Razlika u odnosu na prethodna dva testiranja je što u ovom slučaju nije bila dopuštena kratka pozicija. Rezultati su prikazani u tablici 5.3 i na slikama 5.5 i 5.6

Tablica 5.3: Rezultati testiranja nad CROBEX skupom.

Rezultati:	
bez troškova trgovanja:	
ukupni profit	24.32031
Sharpeov omjer	2.16176
uz uključene troškove trgovanja od 0.10%:	
ukupni profit	20.77131
Sharpeov omjer	1.84749
prosječni koeficijent obrtaja	0.68041
prosječna konzistentnost	0.81656
prosječna preciznost	0.43558



Slika 5.5: Rezultati simulacija trgovanja nad punim skupom dionica CROBEX indeksa. Plavom bojom je prikazana krivulja kumulativnog profita bez uključenih troškova trgovanja, a zelenom uz uključene troškove trgovanja od 0.10% po transakciji.



Slika 5.6: Prikaz dobivenih mjera konzistentnosti za simulaciju trgovanja nad punim skupom dionica CROBEX indeksa.

6. Zaključak

Razvijeni algoritam trgovanja temelji se na metodi statističke arbitraže koja radi nad parovima vrijednosnica, i traži takve parove kod kojih se pojavljuju statistički značajna odstupanja, predviđajući pritom da su ona privremena i da će se vratiti u normalno stanje u kratkom periodu. Glavni nedostatak metode statističke arbitraže leži u činjenici da ne dolazi do izražaja ukupna interakcija svih raspoloživih vrijednosnica, već se primjećuju samo statistička odstupanja na razini parova. To za posljedicu ima jako promjenljiv portfelj i uzrokuje značajne troškove trgovanja, što čini algoritam statističke arbitraže neostvarivim na dugi rok. U ovom radu ostvareno je poboljšanje uvođenjem kompleksnih mreža koje omogućuju modeliranje cjelokupne interakcije raspoloživih vrijednosnica, čime je smanjena promjenljivost portfelja. Kao posljedica toga, niži su i troškovi trgovanja te se pokazalo u simulacijama nad povijesnim podacima da je novi algoritam trgovanja ostvariv i donosi dobre rezultate.

U načelu, algoritam radi bolje kada je dostupan veći broj vrijednosnica, i samim time kvadratno veći broj parova vrijednosnica. Prilagodba na nekonzistentnost dobivenih preferencija postignuta je u vidu diverzifikacije portfelja proporcionalno nekonzistentnosti. Sama mjera nekonzistentnosti služi kao informacija o smislenosti odluka o trgovanju dobivenih algoritmom, te je poželjno da ona bude što veća. Unatoč tome, do sada nisu poznate metode za izravnu optimizaciju mjere konzistentnosti, ali su empirijski utvrđene metode za optimizaciju parametara algoritma koje neizravno poboljšavaju performanse algoritma u određenim uvjetima. Mogući prostor za napredak ovog algoritma je u korištenju metoda dubokog učenja kojima bi se iz velikih količina podataka o kretanju cijena vrijednosnica mogli naučiti parametri algoritma trgovanja.

LITERATURA

- [1] Capinski, Marek i Zastawniak, Tomasz: *Mathematics for Finance: An Introduction to Financial Engineering*. Springer, 2003, ISBN 9780387142227,0-387-14222-3,1852333308.
- [2] Ross, S.M.: *An Elementary Introduction to Mathematical Finance*. Cambridge University Press, 2011, ISBN 9781139498036. <https://books.google.hr/books?id=roYOpM96m5sC>.
- [3] Engle, R. F. i Granger, C. W. J.: *Co-Integration and Error Correction: Representation, Estimation, and Testing*. Econometrica, 1987. <https://www.jstor.org/stable/pdf/1913236.pdf>.
- [4] Dickey, David A. i Fuller, Wayne A.: *Distribution of the Estimators for Autoregressive Time Series With a Unit Root*. Journal of the American Statistical Association 74, no. 366, 1979. <https://www.jstor.org/stable/pdf/2286348.pdf>.
- [5] Banerjee, Anindya: *Co-integration, Error Correction, and the Econometric Analysis of Non-stationary Data*. Oxford, 2003, ISBN 0-19-828810-7.
- [6] Sabolić, Dubravko: *Uvod u mikroekonomiku*. Bilješke s predavanja, 2013.
- [7] Čaklović, Lavoslav i Šego, Vedran: *Potential Method Applied on Exact Data*. Proceedings of 9th International Conference on Operational Research KOI2002, Trogir, 2002. <https://pdfs.semanticscholar.org/75a3/364feaf512ccb994dfe3f0775f25a31c487b.pdf>.
- [8] Čaklović, Lavoslav: *Decision Making by Potential Method*. 2003. <https://pdfs.semanticscholar.org/3f2a/12f13a36ef89c475ab51ad364086ec5b415b.pdf>.

- [9] Jones, Eric i Oliphant, Travis i Peterson Pearu i ostali: *SciPy: Open source scientific tools for Python*, 2001. <http://www.scipy.org/>.
- [10] Finch, Tony: *Incremental calculation of weighted mean and variance*. 2009. <http://people.ds.cam.ac.uk/fanf2/hermes/doc/antiforgery/stats.pdf>.

Dodatak A

Računanje momenata vremenskog niza na pomičnom prozoru

Pri izračunavanju momenata (očekivanje, varijanca) vremenskih nizova nad prozorima konačne duljine, pokazalo se neefikasnim izračunavati svaki prozor zasebno u svakom koraku, jer se većina uzoraka između dva susjedna prozora ponavlja. Neka vremenski niz traje ukupno D vremenskih koraka, i neka je veličina vremenskog prozora T . Momenti se računaju u ukupno $D - T + 1$ vremenskih koraka, svaki put preko prozora veličine T , tako da je vremenska složenost nerekurzivne metode računanja momenata $\mathcal{O}((D - T + 1) \cdot T)$, odnosno približno $\mathcal{O}(DT)$, kako je $D \gg T$. Stoga se isplati potražiti rekurzivne izraze za očekivanje i varijancu kako bi se izračun mogao obaviti na učinkovitiji način. Jedan izvod rekurzivnim izraza dan je u radu [10].

Neka je $a[t]$ vremenski niz, za $t \in [0, 1, \dots, D - 1]$. Nenormirani momenti n -tog reda S_n niza $a[t]$ nad vremenskim prozorom duljine T definirani su na sljedeći način:

$$S_n^{(t)} = \sum_{\tau=t-T+1}^t a^n[\tau]. \quad (\text{A.1})$$

Za nulti moment uvijek vrijedi $S_0^{(t)} = T$ neovisno o t . Rekurzivni izraz za ažuriranje nenormiranog momenta je očigledan:

$$\begin{aligned}
S_n^{(t+1)} &= \sum_{\tau=t+1-T+1}^{t+1} a^n[\tau] \\
&= -a^n[t-T+1] + \sum_{\tau=t-T+1}^t a^n[\tau] + a^n[t+1] \\
&= -a^n[t-T+1] + S_n^{(t)} + a^n[t+1].
\end{aligned} \tag{A.2}$$

Preko nenormiranih momenata mogu se izraziti očekivanje i varijanca. Očekivanje $\mathbf{E}[a[\tau] | t-T < \tau \leq t]$ je normirani moment prvog reda, i može se izraziti kao:

$$\mathbf{E}[a[\tau] | t-T < \tau \leq t] = \frac{S_1^{(t)}}{S_0^{(t)}}, \tag{A.3}$$

što je bilo relativno jednostavno za izračunati i bez korištenja nenormiranih momenata. No, kod računanja varijance situacija je nešto složenija. Varijanca je centrirani i normirani moment drugog reda. Prema definiciji, varijanca $\mathbf{Var}[a[\tau] | t-T < \tau \leq t]$ može se računati još i kao:

$$\begin{aligned}
\mathbf{Var}[a[\tau] | t-T < \tau \leq t] &= \mathbf{E}[(a[\tau] - \mathbf{E}[a[\tau]])^2 | t-T < \tau \leq t] \\
&= \mathbf{E}[a^2[\tau] - 2a[\tau]\mathbf{E}[a[\tau]] + \mathbf{E}^2[a[\tau]] | t-T < \tau \leq t] \\
&= \mathbf{E}[a^2[\tau] | t-T < \tau \leq t] \\
&\quad - 2\mathbf{E}^2[a[\tau] | t-T < \tau \leq t] + \mathbf{E}^2[a[\tau] | t-T < \tau \leq t] \\
&= \mathbf{E}[a^2[\tau] | t-T < \tau \leq t] - \mathbf{E}^2[a[\tau] | t-T < \tau \leq t]
\end{aligned} \tag{A.4}$$

Izraz $\mathbf{E}[a^2[\tau] | t-T < \tau \leq t]$ u (A.4) može se preko nenormiranih momenata izraziti kao $S_2^{(t)}/S_0^{(t)}$, a izraz $\mathbf{E}^2[a[\tau] | t-T < \tau \leq t]$ kao $(S_1^{(t)})^2/(S_0^{(t)})^2$. Konačno, varijanca se može izraziti preko nenormiranih momenata kao:

$$\mathbf{Var}[a[\tau] | t-T < \tau \leq t] = \frac{S_2^{(t)}}{S_0^{(t)}} - \frac{(S_1^{(t)})^2}{(S_0^{(t)})^2} = \frac{S_0^{(t)}S_2^{(t)} - (S_1^{(t)})^2}{(S_0^{(t)})^2}. \tag{A.5}$$

Na ovaj način moguće je izračunati očekivanje i varijancu nad pomičnim prozorom na mnogo učinkovitiji način nego nerekurzivnim postupkom. U početku je potrebno izračunati momente $S_0^{(T-1)}$, $S_1^{(T-1)}$ i $S_2^{(T-1)}$ prema (A.1), te iz njih $\mathbf{E}[a[\tau] | 0 \leq \tau < T]$ i $\mathbf{Var}[a[\tau] | 0 \leq \tau < T]$ prema (A.3) i (A.5); u složenosti $\mathcal{O}(T)$. Nadalje, za svaki sljedeći pomak vremenskog prozora rekurzivno se ažuriraju vrijednosti nenormiranih momenata koristeći (A.2), te se putem njih izračunavaju očekivanje i varijanca ponovno koristeći (A.3) i (A.5), u složenosti $\mathcal{O}(T-D)$. Time je ukupna vremenska složenost algoritma pala s $\mathcal{O}(DT)$ na $\mathcal{O}(D)$.

Analiza vremenskih nizova zasnovana na kompleksnim mrežama

Sažetak

U ovom radu istražene su metode za analizu financijskih vremenskih nizova, temeljene na algoritmu statističke arbitraže. Rezultati istraživanja zaokruženi su u samostalan algoritam za simulaciju trgovanja vrijednosnicama temeljenu na povijesnim podacima, i trgovanje vrijednosnicama nad podacima koji dolaze u stvarnom vremenu. Ulazne podatke za algoritam čini kretanje cijena nekog broja vrijednosnica u nekom razdoblju. Na temelju njih, konstruira se kompleksna mreža koja opisuje odnos vrijednosnica u svakom vremenskom koraku. Iz dobivene mreže procjenjuje se poredak vrijednosnica po preferenciji i računa se pouzdanost same procjene poretka. Konačno, na temelju poretka i pouzdanosti njegove procjene konstruira se portfelj koji čine vrijednosnice s najvećom preferencijom. Algoritam je simuliran nad stvarnim povijesnim podacima i polučio je uspješne rezultate. Iako su istraženi specijalno financijski vremenski nizovi, korišteni matematički i statistički postupci koji su korišteni lako mogu poslužiti i u druge svrhe.

Ključne riječi: vremenski nizovi, financije, statistička arbitraža, relacija preferencije, graf toka preferencija, metoda potencijala.

Complex network based time series analysis

Abstract

In this thesis, methods for financial time series analysis using statistical arbitrage algorithm are researched. Results of the research are summed up into standalone algorithm for simulation of assets trading based on historical data, and also for real world assets trading on live data. Input data for the algorithm are prices of some number of assets during some time span. Next, a complex network is constructed using the input data that describes relations among the assets at the each time step. Then, the ranking based on preference of assets is derived from the network and also the confidence of derived ranking is obtained. Finally, portfolio is constructed from the most preferred assets, by using ranking and its confidence. Simulating the algorithm on the real historical data has yielded significant results. Although the research in this thesis has been done on financial time series, mathematical and statistical methods used can easily prove useful in other applications.

Keywords: time series, finances, statistical arbitrage, preference relation, preference flow graph, potential method.