

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1099

**Analiza vremenskih nizova  
zasnovana na kompleksnim  
mrežama**

Lovre Mrčela

Zagreb, lipanj 2017.

*Zahvaljujem.*

# SADRŽAJ

<b>1. Uvod</b>	<b>1</b>
<b>2. Metode i koncepti</b>	<b>2</b>
2.1. Statistička arbitraža . . . . .	2
2.1.1. Razlika normaliziranih log-cijena . . . . .	3
2.1.2. Normalizirana razlika log-cijena . . . . .	6
2.1.3. Mjera kointegracije slučajnih procesa . . . . .	7
2.1.4. Nedostaci statističke arbitraže . . . . .	9
2.2. Relacija preferencije i funkcija korisnosti . . . . .	10
2.3. Graf toka preferencija . . . . .	11
2.4. Metoda potencijala . . . . .	14
2.4.1. Računanje potencijala čvorova . . . . .	14
2.4.2. Konzistentna rekonstrukcija grafa . . . . .	16
2.4.3. Primjer korištenja metode potencijala . . . . .	16
2.5. Koeficijent obrtaja . . . . .	18
<b>3. Algoritam</b>	<b>20</b>
3.1. Konstrukcija grafa toka preferencija . . . . .	21
3.2. Signal trgovanja . . . . .	23
3.3. Konstrukcija portfelja . . . . .	26
3.4. Minimizacija koeficijenta obrtaja . . . . .	27
<b>4. Rezultati</b>	<b>28</b>
<b>5. Zaključak</b>	<b>31</b>
<b>Literatura</b>	<b>32</b>
<b>A. Računanje momenata vremenskog niza na pomičnom prozoru</b>	<b>33</b>

# 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 visoko pouzdanog intervala, 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 ima velik broj vrijednosnica na raspolaganju tako što izvlači zajedničku interakciju iz svih vrijednosnica.

## 2. Metode i koncepti

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

### 2.1. 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 pritome njegova početna vrijednost iznosi 0:

$$\begin{aligned} V^{(0)} &= 0 \\ P(V^{(t)} > 0) &= 1, \end{aligned} \tag{2.1}$$

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, i tada je arbitražu moguće ostvariti kupovanjem vrijednosnice po nižoj cijeni na jednom tržištu i prodavanjem po višoj na drugom.

U statističkoj arbitraži cilj 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 (2.1), jer korištene statističke mjere imaju određeni interval pouzdanosti koji je 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  $P(V^{(t)} > 0) \rightarrow 1$ . 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 parovima vrijednosnica tako da se u jednoj od njih zauzme kratka, a u drugoj duga pozicija, te se u idućem trenutku izlazi iz tih pozicija, osim ako se i u idućem trenutku utvrdi da treba ostati u istim pozicijama.

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

### 2.1.1. Razlika normaliziranih log-cijena

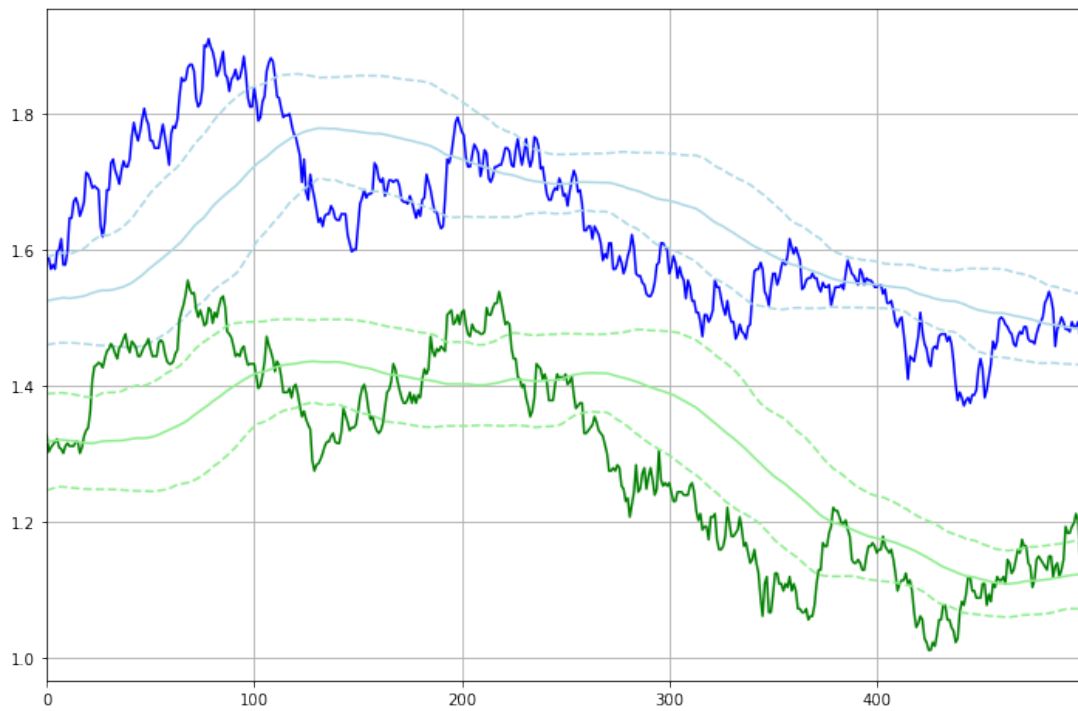
Neka je sa  $P_X^{(t)}$  označena log-cijena vrijednosnice  $X$  u trenutku  $t$ , i neka se koristi period od  $T$  vremenskih koraka. Neka je ukupno  $M$  vremenskih koraka, tj.  $t \in [0, 1, \dots, M-1]$ . Za svaku vrijednosnicu  $X$  u trenucima  $t \geq T$  izračunaju se normalizirane log-cijene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

$$\bar{P}_X^{(t)} = \frac{P_X^{(t)} - \mathbf{E} [P_X^{(\tau)} \mid t - T \leq \tau < t]}{\sqrt{\mathbf{Var} [P_X^{(\tau)} \mid t - T \leq \tau < t]}}. \quad (2.2)$$

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  $t$ :

$$D_{X,Y}^{(t)} = \bar{P}_X^{(t)} - \bar{P}_Y^{(t)} \quad (2.3)$$

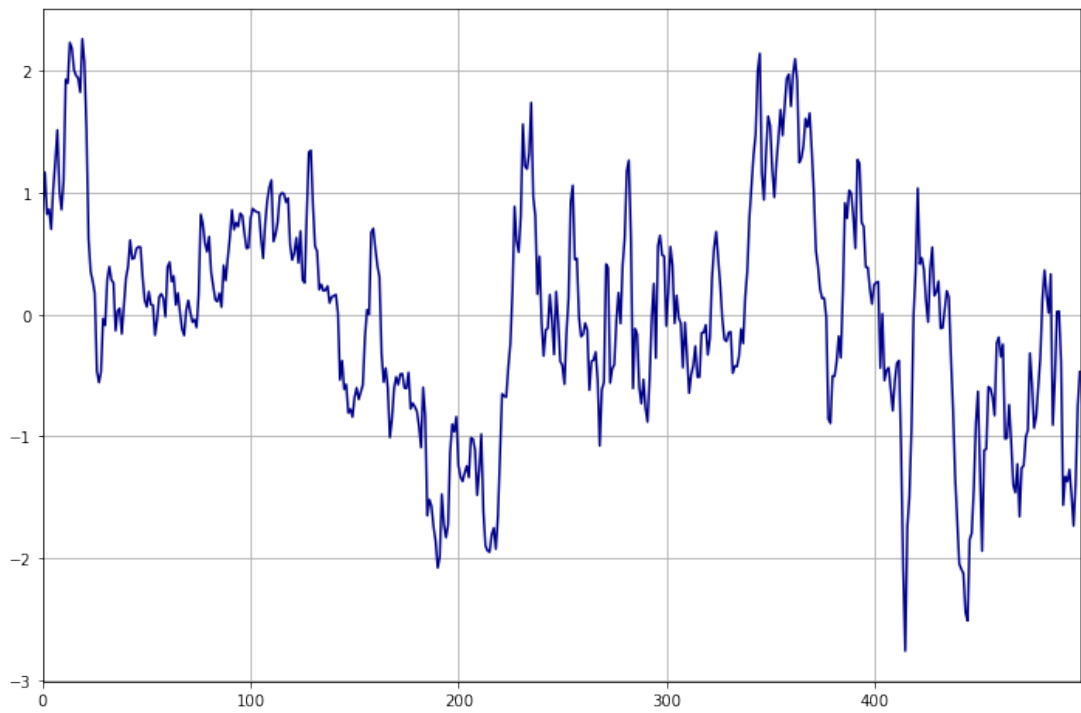
Razlika normaliziranih log-cijena se koristi kao mjera sličnosti u algoritmu statističke arbitraže. Kod ove mjere normalizacijom se postiže ravnopravna usporedba dviju vrijednosnica različitih volatilnosti, što znači da će se prilikom usporedbe dviju dionica čiji su intenziteti promjene cijene različiti nakon normalizacije oni dovesti na istu razinu. Ilustrativni primjer kretanja cijena vrijednosnica prikazan je na slici 2.1, normalizirane cijene na slici 2.2, i razlika normaliziranih cijena na slici 2.3.



**Slika 2.1:** 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 2.2:** Plavo: normalizirana cijena  $\bar{P}_X^{(t)}$ , zeleno: normalizirana cijena  $\bar{P}_Y^{(t)}$ .



**Slika 2.3:** Razlika normaliziranih cijena,  $D_{X,Y}^{(t)}$ .



### 2.1.2. Normalizirana razlika log-cijena

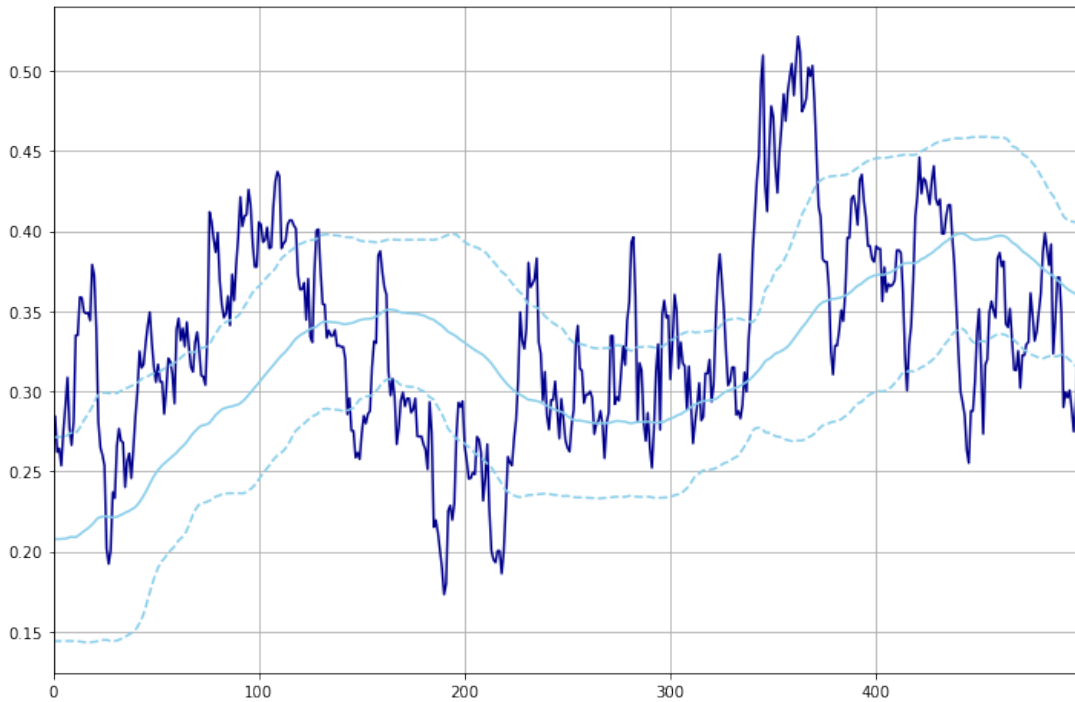
Za razliku od prethodne mjere sličnosti, u ovom slučaju se razlika računa nad nenormaliziranim log-cijenama za sve  $t$ , označeno sa  $d_{X,Y}^{(t)}$ :

$$d_{X,Y}^{(t)} = P_X^{(t)} - P_Y^{(t)}, \quad (2.4)$$

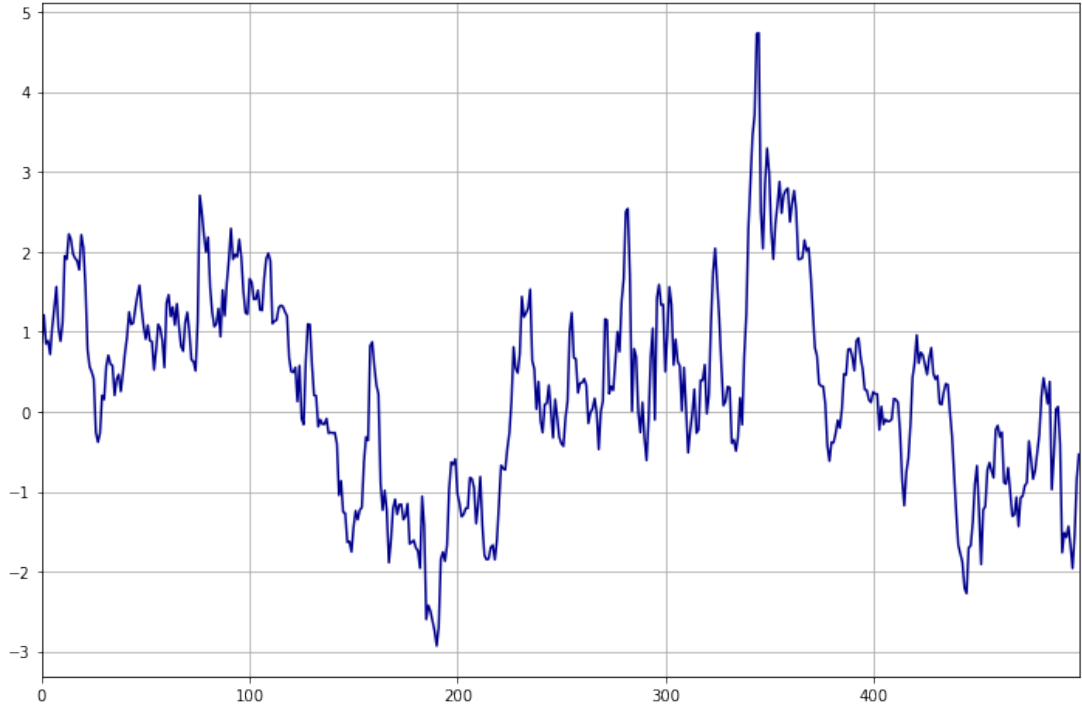
a zatim se iz razlika  $d_{X,Y}^{(t)}$  računaju normalizirane razlike  $\bar{d}_{X,Y}^{(t)}$  za  $t \geq T$ :

$$\bar{d}_{X,Y}^{(t)} = \frac{d_{X,Y}^{(t)} - \mathbf{E} \left[ d_{X,Y}^{(\tau)} \mid t - T \leq \tau < t \right]}{\sqrt{\mathbf{Var} \left[ d_{X,Y}^{(\tau)} \mid t - T \leq \tau < t \right]}}. \quad (2.5)$$

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 2.4, te normalizirane razlike na slici 2.5.



**Slika 2.4:** Tamnijom bojom prikazana je razlika nenormaliziranih cijena  $d_{X,Y}^{(t)}$ , a svijetlijom punom linijom očekivanje i crtkanom linijom udaljenost od očekivanja za jednu devijaciju.



**Slika 2.5:** Normalizirana razlika cijena,  $\bar{d}_{X,Y}^{(t)}$ .

### 2.1.3. 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 slučajnog procesa. Za neki slučajni proces zanimljivo je proučiti svojstvo stacionarnosti — slučajni proces  $U[t]$  je stacionaran ako i samo ako vrijedi:

$$\mathbf{E}[U[t]] = \text{const.} \quad (2.6)$$

$$\mathbf{R}_{UU}[t, \tau] = \mathbf{E}[U[t] \cdot U[t + \tau]] = \mathbf{R}_{UU}[\tau]. \quad (2.7)$$

Za stacionarne slučajne procese vrijedi da njihova statistička svojstva ne ovise o vremenu: jednačba (2.6) tvrdi da je za stacionaran proces  $U[t]$  očekivanje slučajnog procesa  $\mathbf{E}[U[t]]$  neovisna o vremenu  $t$ , a jednačba (2.7) tvrdi da njegova autokorelacija  $\mathbf{R}_{UU}[t, \tau]$  ovisi samo o vremenskom pomaku  $\tau$ , ali ne i o  $t$ . Ovakva definicija stacionarnosti naziva se još i “stacionarnost u širem smislu”.

Red integracije slučajnog procesa, označen sa  $I(d)$ , 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

$$\Delta V[t] = (1 - L)V[t] = V[t] - V[t - 1] \quad (2.8)$$

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

$$\begin{aligned}\Delta^d W[t] &= (1 - L)^d W[t] = \left( \sum_{n=0}^d \binom{d}{n} (-1)^n L^n \right) W[t] \\ &= \sum_{n=0}^d \binom{d}{n} (-1)^n W[t - n]\end{aligned}\quad (2.9)$$

stacionaran proces. U (2.8) i (2.9) korišten je operator kašnjenja  $L$  sa značenjem  $LU[t] = U[t - 1]$ .

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  $I(0)$ :

$$X[t] + \beta Y[t] = U[t], \quad (2.10)$$

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

$$U[t] = \mu + \varepsilon[t], \quad (2.11)$$

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[1] 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. Postupkom najmanjih kvadrata najprije se određuju procjene parametara  $\mu$  i  $\beta$ , označene sa  $\hat{\mu}$  i  $\hat{\beta}$ :

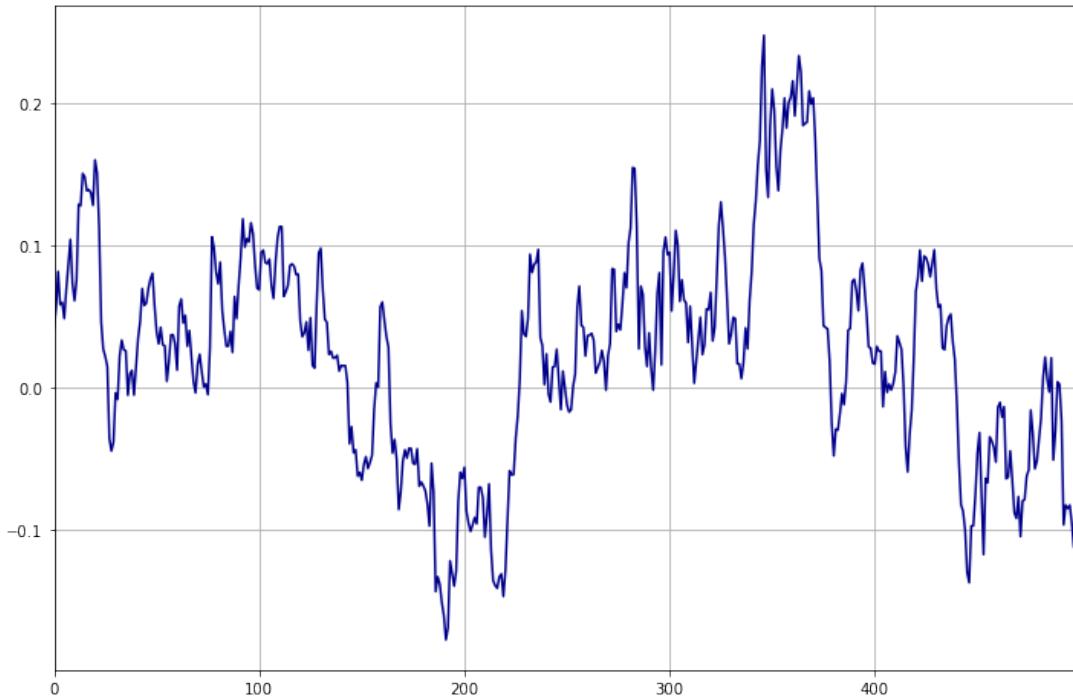
$$\hat{\mu}, \hat{\beta} = \arg \min_{\mu, \beta} \left\{ \sum_{\tau=t-T}^{t-1} |P_X[\tau] + \beta P_Y[\tau] - \mu|^2 \right\}, \quad (2.12)$$

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

$$\hat{\varepsilon}[t] = -\hat{\mu} + P_X[t] + \hat{\beta}P_Y[t]. \quad (2.13)$$

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[2].

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 2.6. U odnosu na prethodne dvije mjere, ova mjera je rezultat dublje statističke analize, zbog čega je i računski zahtjevnija, što je ujedno prednost i nedostatak. Kada se radi o velikom broju parova vrijednosnica, kointegracije se može 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 2.6:** Procijenjene rezidualne vrijednosti  $\hat{\varepsilon}[t]$  u slučaju vrijednosnica  $X$  i  $Y$ .

#### 2.1.4. Nedostaci statističke arbitraže

Jedan od nedostataka statističke arbitraže jest generalno slaba preciznost izražena kao udio broja trgovanja koji rezultira pozitivnim profitom u ukupnom broju trgovanja. Uzrok tome su statističke mjere na kojima se algoritam temelji,

koje ne mogu dati apsolutno pouzdanu odluku samim time što je kretanje cijena slučajan proces. Također, problem može biti i tzv. varljiva korelacija u kojoj je kretanje cijena dviju vrijednosnica naizgled povezano, no zapravo je uzrok tome neki nepoznati vanjski faktor; tada predviđeno ponašanje može biti poprilično različito od stvarnog. Isto tako, statističke metode ne mogu razlikovati one devijacije u ponašanju cijena koje su svojstvene koreliranim slučajnim procesima od onih za koje stvarno postoji fundamentalan uzrok, kao što je npr. značajna promjena u načinu poslovanja neke firme, promjena vlasništva i tome slično. Pri testiranju algoritma statističke arbitraže nad testnim skupovima podataka preciznost rijetko kada prelazi 60%, a većinom je manja od 50%. 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 pogotcima.

Drugi nedostatak tiče se načina tretiranja vrijednosnica koje su na raspolaganju i koje se mogu uključiti u portfelj. Svaki par vrijednosnica gleda se zasebno i pritom je isključena dublja analiza međusobnih interakcija svih vrijednosnica zajedno. Trgovanje se također odvija uvijek u parovima, odnosno jednaki broj vrijednosnica se kupuje i prodaje u svakom trenutku. U ovom radu istražena je metoda koja omogućuje trgovanje kupovanjem i prodajom svih vrijednosnica nezavisno, koristeći pritom metode statističke arbitraže kao opis relacije preferencije među vrijednosnicama. Interakcija među vrijednosnicama opisuje se grafom toka preferencija, iz čega se naposljetku izvlači poredak prema individualnoj preferenciji svih vrijednosnica.

## 2.2. Relacija preferencije i funkcija korisnosti

Neka je  $\Omega$  skup općenitih dobara. Relacija preferencije, označena sa  $\succ$  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*:  $\forall x \in \Omega: \neg(x \succ x)$  — za nijedno 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  $y$ , te  $y$  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  $y$ , te  $y$  podjednako preferirano kao i  $z$ , tada je i  $x$  podjednako preferirano kao i  $z$ .

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.

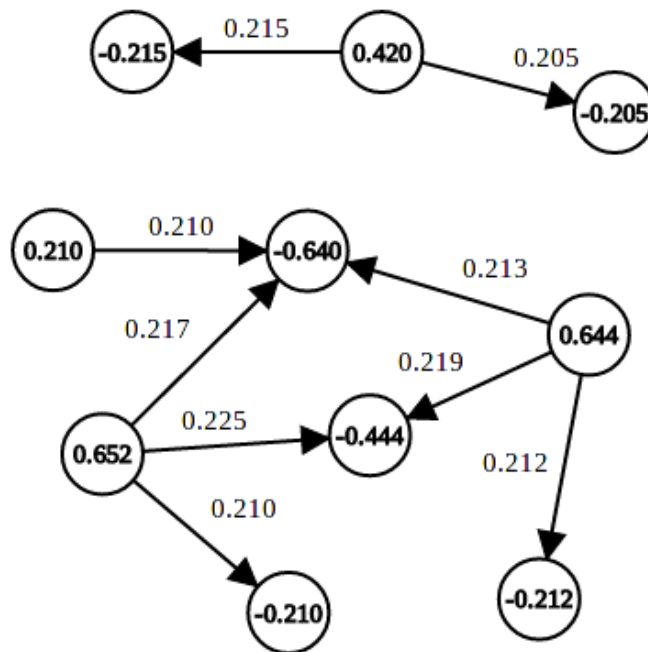
Funkcija korisnosti (engl. *utility function*)  $U: \Omega \rightarrow \mathbb{R}$  je preslikavanje iz skupa dobara u skup realnih brojeva, tako da poredak preslikanih realnih brojeva odgovara poretku dobara prema individualnoj preferenciji, tj. vrijedi  $\forall x, y \in \Omega: x \succ y \Leftrightarrow U(x) > U(y)$ . 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 2.4. Uz to što određuje poredak dobara prema individualnoj preferenciji, funkcija korisnosti također unosi i mjeru intenziteta preferencije prema nekoj vrijednosnici, što znači da nosi više informacija od same relacije.

## 2.3. 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 2.7.



**Slika 2.7:** Primjer grafa toka preferencija. Na bridovima su prikazani intenziteti preferencija jedne vrijednosnice u odnosu na drugu, a u čvorove su upisane izračunate individualne mjere preferencija za pripadni čvor 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 2.4. 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 prema određenim statističkim mjerama trenutačno razdvojile. Težina  $w_{i,j}$  opisuje magnitudu 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 3.1.

Veze u ovom grafu opisuju relaciju preferencije među vrijednosnicama koje su opisane 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; no 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, i  $j$  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. Graf koji je intrinzično konzistentan također istovremeno zadovoljava i svojstvo tranzitivnosti relacije preferencije (dok 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 se koristi kao dodatan parametar pri konstruiranju portfelja. Opis mjere konzistentnosti, te način dobivanja najbliže konzistentne rekonstrukcije grafa dan je u odjeljku 2.4.2.



## 2.4. 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[3] koja je opisana u nastavku.

### 2.4.1. 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 čvorova koji nisu povezani, 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  $-1$  i  $1$  u stupcima koji odgovaraju čvorovima  $i$  i  $j$ ;
3. svi ostali elementi matrice su jednaki 0.

Neka je  $\mathbf{f}$  vektor dimenzija  $[E \times 1]$  koji sadrži težine bridova, tj. tokove preferencije, i neka je  $\phi$  vektor dimenzija  $[N \times 1]$  koji sadrži potencijale čvorova, koji se žele 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,  $\mathbf{B}$ ,  $\phi$ , i  $\mathbf{f}$  zadovoljavaju jednadžbu:

$$\mathbf{B}\phi = \mathbf{f}. \quad (2.14)$$

Jednadžba (2.14) 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  $\phi^*$  koje ima minimalnu kvadratnu pogrešku. Tako originalni problem postaje:

$$\phi^* = \arg \min_{\phi} \{ \|\mathbf{B}\phi - \mathbf{f}\|^2 \} \Leftrightarrow \frac{\partial \|\mathbf{B}\phi^* - \mathbf{f}\|^2}{\partial \phi^*} = \mathbf{0}. \quad (2.15)$$

Matričnim deriviranjem (2.15) dobiva se sljedeća jednađžba:

$$\begin{aligned} 2\mathbf{B}^\top [\mathbf{B}\phi^* - \mathbf{f}] &= \mathbf{0} \\ \mathbf{B}^\top \mathbf{B}\phi^* &= \mathbf{B}^\top \mathbf{f}. \end{aligned} \quad (2.16)$$

Jednađžba (2.16) određuje  $\phi^*$  do na konstantu, tj. rješenje ima jedan stupanj slobode. Stoga se dodaje sljedeće ograničenje:

$$\mathbf{j}^\top \phi^* = 0 \quad (2.17)$$

gdje je  $\mathbf{j}$  vektor jedinica istih dimenzija kao i  $\phi^*$ . Jednađžba (2.17) osigurava jedinstveno rješenje u kojem je ukupan zbroj potencijala svih čvorova jednak nuli.

Združivanjem prethodnih dviju jednađžbi u jednu, tako da se (2.17) doda svakom retku u (2.16), dobiva se:

$$\begin{aligned} \mathbf{B}^\top \mathbf{B}\phi^* + \mathbf{J}\phi^* &= \mathbf{B}^\top \mathbf{f} \\ [\mathbf{B}^\top \mathbf{B} + \mathbf{J}]\phi^* &= \mathbf{B}^\top \mathbf{f}, \end{aligned} \quad (2.18)$$

gdje je  $\mathbf{J}$  matrica jedinica istih dimenzija kao i  $\mathbf{B}^\top \mathbf{B}$ . Konačno, rješavanjem (2.18) po  $\phi^*$  dobiva se:

$$\phi^* = [\mathbf{B}^\top \mathbf{B} + \mathbf{J}]^{-1} \mathbf{B}^\top \mathbf{f}. \quad (2.19)$$

Dodatno, izraz (2.19) 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  $N - 1$ , Laplaceova matrica grafa  $\mathcal{G}$  ima  $N - 1$  na dijagonali, a  $-1$  drugdje. Tako se izraz  $[\mathbf{B}^\top \mathbf{B} + \mathbf{J}]^{-1}$  može pojednostavniti u  $[N\mathbf{I}]^{-1} = \frac{1}{N}\mathbf{I}$ ; i stoga se (2.19) može svesti na:

$$\phi^* = \frac{1}{N} \mathbf{B}^\top \mathbf{f}, \quad (2.20)$$

tako da se dobije računalno optimalan izraz. Jednađžba (2.20) tvrdi da se rješenje u smislu najmanjih kvadrata  $\phi^*$  jednađžbe (2.14) dobije tako da se za svaki čvor pribroje težine svih bridova koji izlaze iz njega, i oduzmu težine svih bridova koji ulaze u njega.

### 2.4.2. Konzistentna rekonstrukcija grafa

Sada se na temelju dobivenog potencijala čvorova  $\phi^*$  može izračunati tok preferencije  $f^*$  koji je konzistentan s potencijalom, tako da se u (2.14) supstituira  $\phi^*$  umjesto  $\phi$ :

$$f^* = B\phi^*. \quad (2.21)$$

Rekonstruirani graf toka preferencije  $f^*$  u odnosu na originalni  $f$  može sadržavati neke nove, pa čak i izgubiti neke stare bridove. Dodatno,  $B$ ,  $\phi^*$ , 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 koji potencijale čvorova preslikava u odgovarajući tok preferencija.

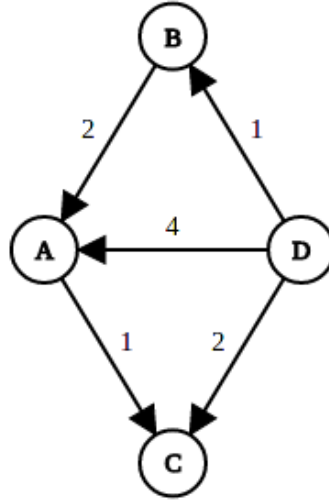
Mjera konzistentnosti  $\kappa$  definira se kao:

$$\kappa = \frac{\|f^*\|}{\|f\|}. \quad (2.22)$$

Jednadžba (2.22) predstavlja kosinus kuta između  $f$  i  $f^*$  u prostoru određenom vektor stupcima matrice incidencije  $B$ . Mjera konzistentnosti  $\kappa$  opisuje koliko je sličan graf  $\mathcal{G}$  s grafom  $\mathcal{G}^*$ , te je u rasponu od 0 do 1, gdje 0 znači potpunu nekonzistentnost (u praksi nedostižno), a 1 potpunu konzistentnost.

### 2.4.3. 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 2.8. Taj graf nije konzistentan s relacijom preferencije koju predstavlja: prema njemu  $B$  nije izravno usporediva sa  $C$ , ali kako je  $B \succ A$  i  $A \succ C$ , po tranzitivnosti bi trebalo biti i  $B \succ C$ , što u grafu nije slučaj. Također, postoje dva puta od  $D$  prema  $A$ : to su  $D \rightarrow A$  i  $D \rightarrow B \rightarrow A$ , te je duž prvog puta zbroj težina jednak 4, a duž drugog je jednak 3, pa graf nije niti intrinzično konzistentan. Isto je i s putevima od  $D$  do  $C$ : duž puta  $D \rightarrow C$  zbroj težina je jednak 2, a duž puta  $D \rightarrow B \rightarrow C$  zbroj težina je jednak 5, što je čak i veće nepodudaranje nego u prethodnom slučaju.



**Slika 2.8:** Primjer nekonzistentnog grafa prije izračunavanja potencijala čvorova. U čvorove su upisani nazivi vrijednosnica, a na bridove su upisane težine koje opisuju tok preferencije.

Matrica incidencije  $\mathbf{B}$  i vektor toka preferencije  $\mathbf{f}$  grafa sa slike 2.8 su sljedeći:

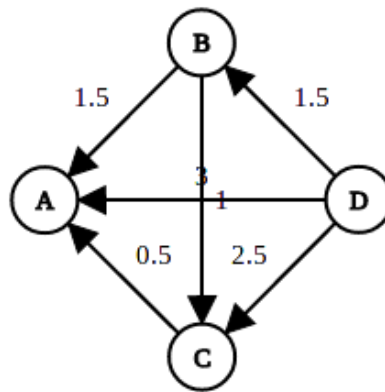
$$\mathbf{B} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} -2 \\ 1 \\ -4 \\ 0 \\ -1 \\ -2 \end{bmatrix}$$

Matrica incidencije  $\mathbf{B}$  konstruirana je tako da redci odgovaraju redom bridovima  $\overrightarrow{AB}$ ,  $\overrightarrow{AC}$ ,  $\overrightarrow{AD}$ ,  $\overrightarrow{BC}$ ,  $\overrightarrow{BD}$ , i  $\overrightarrow{CD}$ , a stupci redom čvorovima  $A$ ,  $B$ ,  $C$  i  $D$ . Istim redoslijedom su i težine bridova upisane u vektor  $\mathbf{f}$ . Neke težine su negativne jer se smjer brida u matrici incidencije ne podudara sa 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 dobivenom potencijalu čvorova  $\phi^*$  može se zaključiti kako je najviše preferiran čvor  $D$ , zatim  $B$ , zatim  $C$ , i naposljetku  $A$ . Rekonstrukcija grafa

prikazana je na slici 2.9. 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 nekonzistenciju u originalnom grafu — tako je najveća promjena u težini brida  $\overrightarrow{AD}$ , sa  $-4$  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 možemo zaključiti kako je rekonstruirani graf doista intrinzično konzistentan.



**Slika 2.9:** Primjer najbliže konzistentne rekonstrukcije grafa nakon izračunavanja potencijala čvorova.

## 2.5. 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  $t$ ;  $\alpha_i^{(t)}$  je 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:

$$\eta^{(t)} = \sum_{i=1}^N \left| \alpha_i^{(t)} - \alpha_i^{(t-1)} \right|. \quad (2.23)$$

Iz (2.23) 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  $p \cdot \eta^{(t)}$ .

### 3. Algoritam

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. Neka je sa  $a_i^{(t)}$  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,  $\forall i, j \in [1, 2, \dots, N]$ :

$$b_i^{(t)} = \log(a_i^{(t)}) \quad (3.1)$$

$$c_{i,j}^{(t)} = b_i^{(t)} - b_j^{(t)}. \quad (3.2)$$

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:

$$m_{i,j}^{(t)} = \frac{1}{T} \sum_{\tau=t-T}^{t-1} c_{i,j}^{(\tau)} \quad (3.3)$$

$$d_{i,j}^{(t)} = \sqrt{\frac{1}{T-1} \sum_{\tau=t-T}^{t-1} (c_{i,j}^{(\tau)} - m_{i,j}^{(t)})^2}. \quad (3.4)$$

Procjena srednje vrijednosti u (3.3) je nepristrana, što znači da je očekivanje od  $m_{i,j}^{(t)}$  doista jednako srednjoj vrijednosti toga uzorka. Procjena standardne devijacije u (3.4) nije nepristrana, jer je dobivena korjenovanjem nepristrane procjene varijance. Kako je korjenovanje nelinearna operacija, ono unosi pristranost u procjenu. Valja primijetiti kako je u izrazima sumiranja u (3.3) i (3.4) vremenski korak  $t$  namjerno izostavljen jer se radi nad proteklom periodom trajanja  $T$ , zbog čega sumiranje ide do  $t - 1$ . Izračuni (3.1) – (3.4) 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

kad se radi o velikom skupu podataka. Stoga je preporučljivo koristiti algoritam za računanje s pomičnim prozorima, koji je opisan u dodatku. Što se tiče redukcije memorijske potrošnje,  $c_{i,j}^{(t)}$ ,  $m_{i,j}^{(t)}$ , i  $d_{i,j}^{(t)}$  su idejno tenzori dimenzija  $[(D - T) \times N \times N]$ . Međutim, sva tri tenzora su simetrična ili antisimetrična, u smislu da 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:

$$k = N \cdot (i - 1) + j - 1 - i \cdot (i + 1)/2, \quad (3.5)$$

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

$$i = \left\lfloor N + 1/2 - \sqrt{(N + 1/2)^2 - 2(N + k)} \right\rfloor, \quad (3.6)$$

$$j = k + i \cdot (i + 1)/2 - N \cdot (i - 1) + 1. \quad (3.7)$$

Na slici 3.1 prikazano je kodiranje u slučaju  $N = 5$  vrijednosnica.

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

**Slika 3.1:** Primjer predložene sheme kodiranja i dekodiranja za  $N = 5$ . Točka  $(\cdot)$  označava da se određena kombinacija ne koristi.

### 3.1. 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:

$$\left| c_{i,j}^{(t)} - m_{i,j}^{(t)} \right| > \alpha \cdot d_{i,j}^{(t)}, \quad (3.8)$$

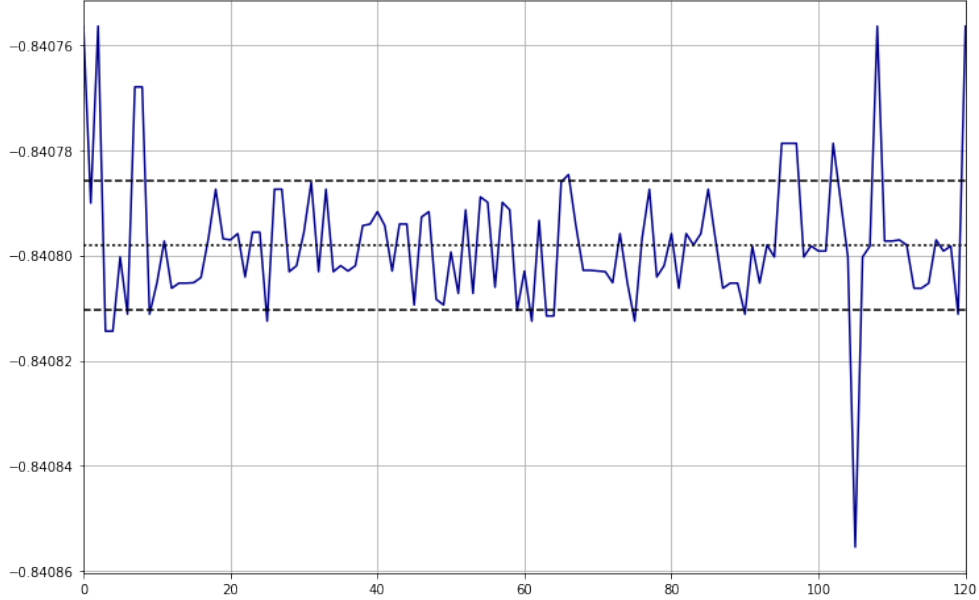
tj. trenutna apsolutna vrijednost razlike log-cijena para vrijednosnica  $(i, j)$  je barem za  $\alpha$  standardnih devijacija udaljena od očekivanja, izmjerenih u proteklom



periodu. Nakon toga, za svaki par  $(i, j)$  koji zadovoljava (3.8) 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:

$$w_{i,j}^{(t)} = \frac{c_{i,j}^{(t)} - m_{i,j}^{(t)}}{d_{i,j}^{(t)}}, \quad (3.9)$$

što odgovara udaljenosti razlike log-cijena od očekivanja, mjerene u standardnim devijacijama, i po apsolutnoj vrijednosti je veća od  $\alpha$ . 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 prvotno pretpostavljenom bridu. Ilustracija ove mjere prikazana je na slici 3.2.



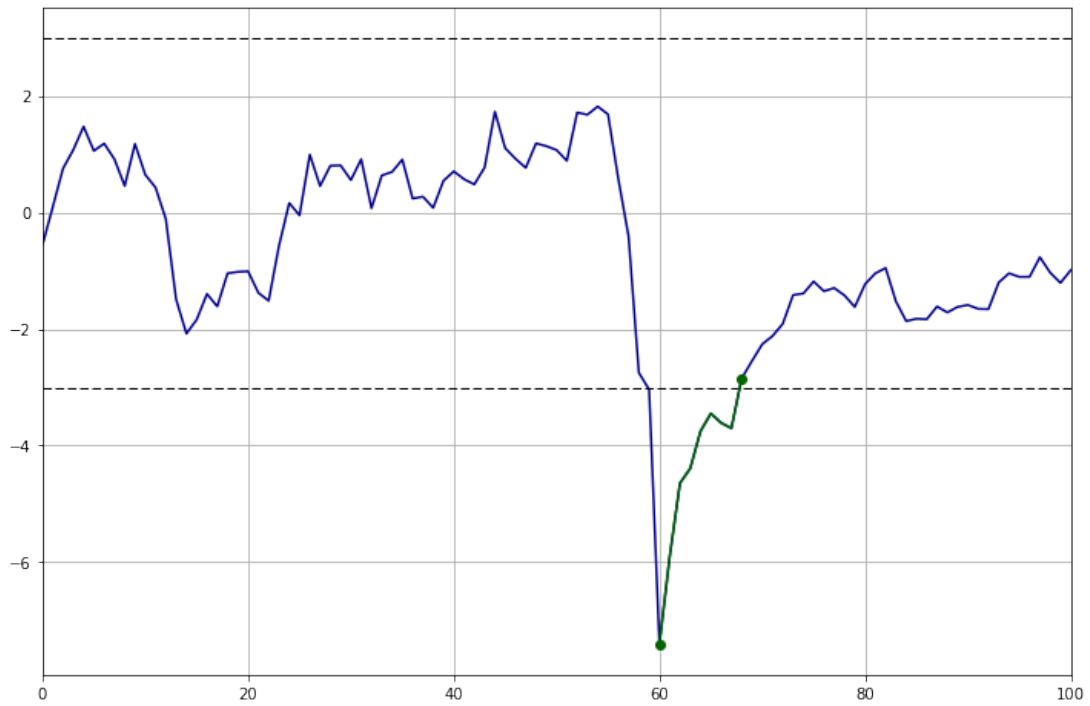
**Slika 3.2:** Razlika log-cijena u paru vrijednosnica  $(i, j)$  tijekom perioda od  $T + 1$  vremenskih koraka, uz  $T = 120$ . Točka-crta linija predstavlja očekivanje, a područje između dviju crtanih linija predstavlja  $\alpha$  standardnih devijacija udaljenosti od očekivanja; obje mjere su izračunate nad prvih  $T$  vremenskih koraka. Tijekom vremenskog koraka  $T + 1$ , razlika log-cijena prelazi preko  $\alpha$  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.

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 (3.8). Postavljanjem nižih vrijednosti za  $\alpha$  dobivaju se gušći grafovi i manje često

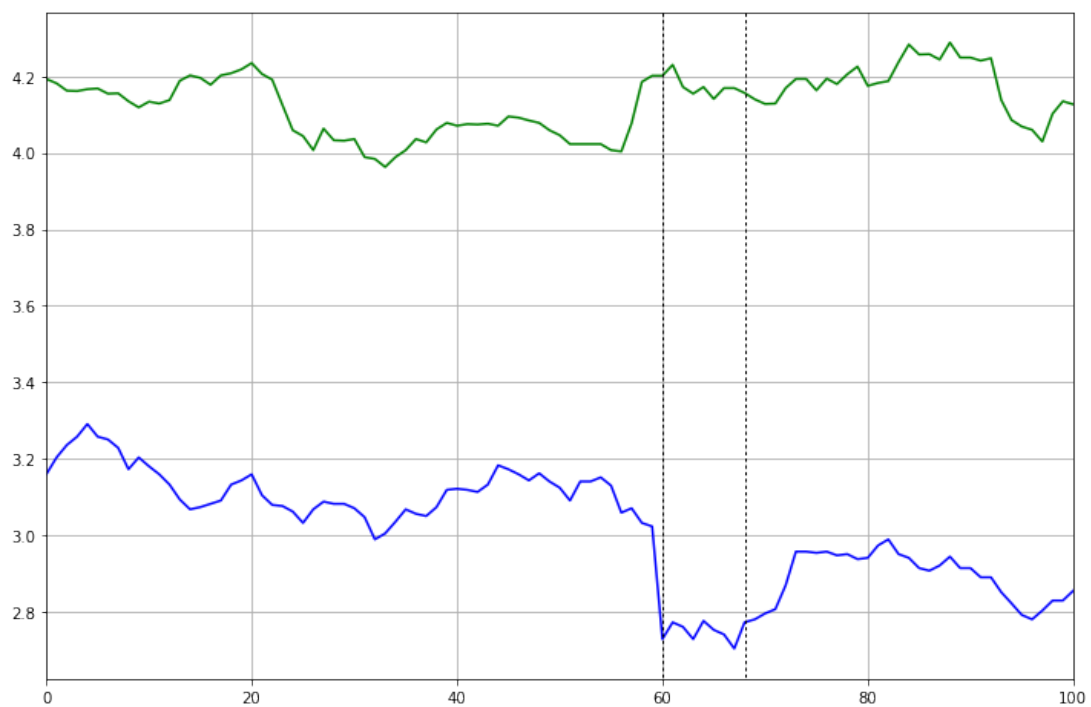
se događa da je graf prazan, dok se postavljanjem  $\alpha = 0$  uvijek dobivaju potpuni grafovi.

## 3.2. Signal trgovanja

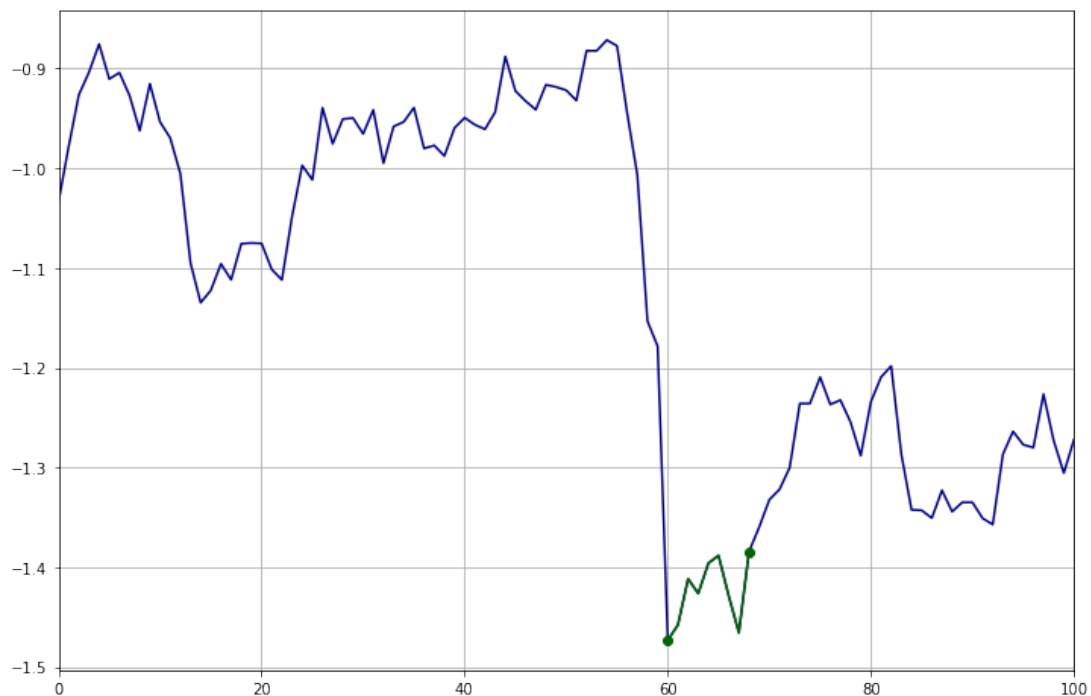
Signal trgovanja je funkcija određena



**Slika 3.3:** Signal trgovanja  $\Gamma_{X,Y}(t)$  za vrijednosnice  $X$  i  $Y$ , dobiven korištenjem  $T = 120$ . Prag  $\alpha$  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 razspan označen zelenom bojom) do trenutka  $t_2 = 69$  (označeno drugom zelenom točkom). Kako je signal u aktivnom području negativan, zauzima se kratka pozicija u vrijednosnici  $X$  i duga u  $Y$ , a po izlasku iz aktivnog područja te pozicije se zatvaraju.



**Slika 3.4:** Kretanje logaritamskih cijena vrijednosnica  $X$  (plavo) i  $Y$  (zeleno). Crtkanim linijama označeno je područje u kojem je signal trgovanja aktivan. Uočljiva je sličnost u ponašanju kretanja cijena prije i poslije aktivnog područja signala trgovanja.



**Slika 3.5:** 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 0.0889 u ovom primjeru.

### 3.3. Konstrukcija portfelja

Individualna mjera preferencije za svaku vrijednosnicu dobivena je metodom potencijala, na način kako je ranije opisano u 2.4.1. Vrijednosnice s trenutačno 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 preferenci vrijednosnica u vremenskom koraku  $t$ , te  $\phi_i^{(t)}$  individualna preferencija vrijednosnice  $i$  u vremenskom koraku  $t$ . Pri odabiru vrijednosnica za portfelj, u obzir se također uzima i mjera konzistentnosti  $\kappa$ . 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  $\kappa$ . 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:

$$\kappa' = a + (1 - a)\kappa^b, \quad (3.10)$$

gdje je  $a \in [0, 1], b \in \mathbb{R}^+$ . Za vrijednosti  $a = 0, b = 1$ ,  $\kappa'$  postaje jednak  $\kappa$ . Parametar  $a$  regulira utjecaj mjere konzistentnosti na određivanje portfelja: za  $a = 0$ , diverzifikacija u potpunosti ovisi o mjeri konzistentnosti  $\kappa$ , 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.

Za odabir vrijednosnica koje će ući u portfelj u vremenskom koraku  $t$ , pronalaze se sve vrijednosnice  $i$  za koje u trenutku  $t$  vrijedi:

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

gdje je  $\Phi_{max} = \max_j \{|\phi_j|\}$ . Isto tako, za prodaju se biraju vrijednosnice  $i$ , za koje u trenutku  $t$  vrijedi:

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

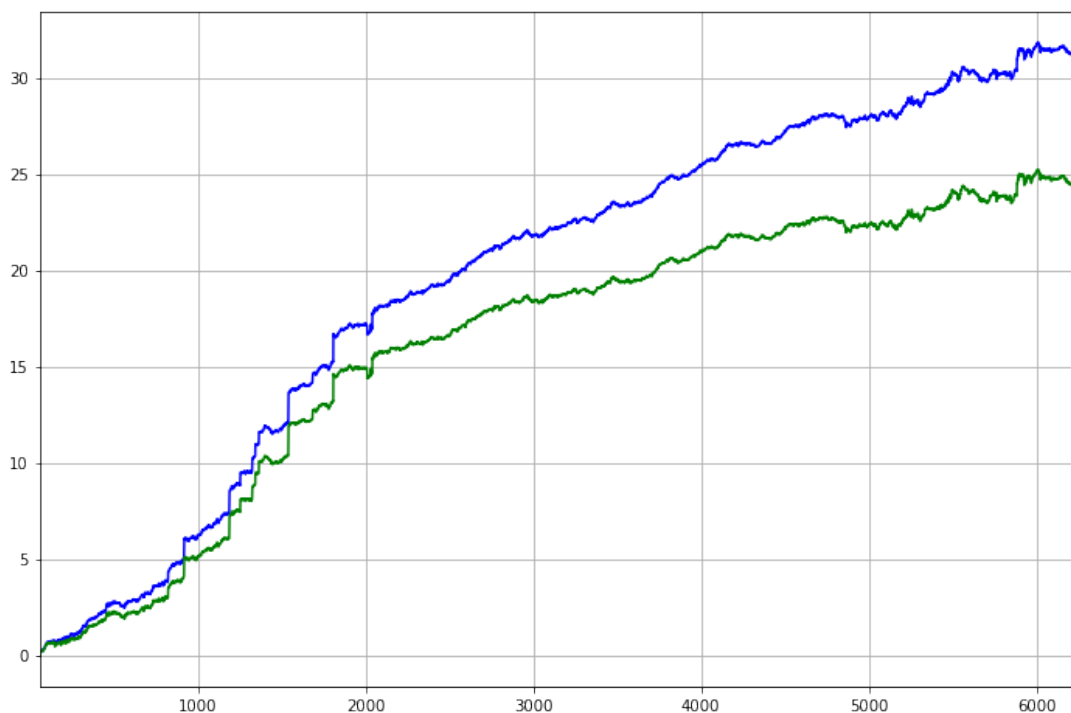
### 3.4. Minimizacija koeficijenta obrtaja

Minimizacija: energija signala.

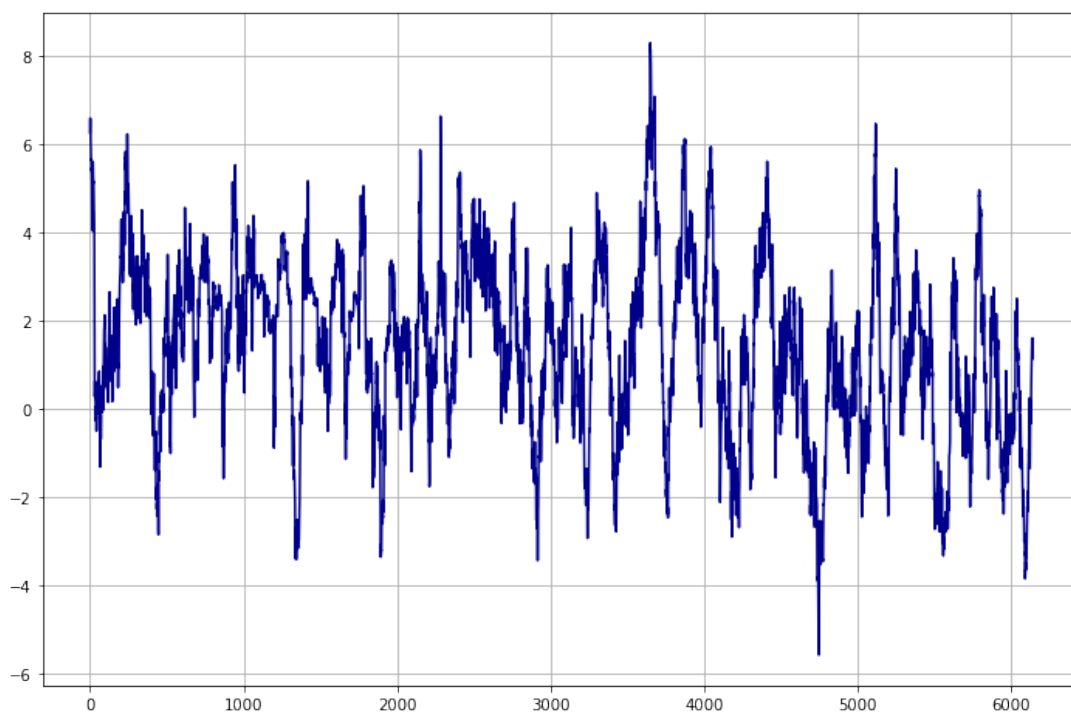
## 4. Rezultati

Skup podataka nad kojim je algoritam testiran je podskup koji sadrži 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 4.1.

Godišnji Sharpeov omjer  $S$  definiran je kao omjer srednjeg godišnjeg povrata  $\mu_r$  i srednje godišnje volatilnosti  $\sigma_r$ . Analizirana su zasebno trgovanja koja rezultiraju pozitivnim i negativnim profitom, njihova razlika i omjer. Mjera preciznosti 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. Konačno, izračunat je ukupan profit uz uključene troškove trgovanja.



**Slika 4.1:** Rezultati simulacija trgovanja nad podskupom dionica S&P 500 indeksa. Plavom bojom je prikazana krivulja kumulativnog profita bez uključenih troškova trgovanja, a zelenom uz troškove trgovanja od 0.10%.



**Slika 4.2:** Sharpeov omjer izražen na godišnjoj razini, na prozorima veličine  $T = 60$ , za simulaciju trgovanja nad podskupom dionica S&P 500 indeksa.



**Tablica 4.1:** Rezultati za  $T = 60, \alpha = 0$ .

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 preciznost	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. Zaključak

Algoritam radi nad parovima vrijednosnica, i traži takve parove kod kojih se pojavljuju statistički značajna odstupanja. U načelu, algoritam radi bolje kada je dostupno veći broj vrijednosnica, i samim time kvadratno veći broj parova vrijednosnica. Prilagodba na nekonzistentnost dobivenih preferencija postignuta je u vidu diverzifikacije portfelja proporcionalno nekonzistentnosti.

# LITERATURA

- [1] Robert F. Engle and C. W. J. Granger. Co-integration and error correction: Representation, estimation, and testing. 1987. URL [www.jstor.org/stable/pdf/1913236.pdf](http://www.jstor.org/stable/pdf/1913236.pdf).
- [2] David A. Dickey and Wayne A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. 1979. URL <https://www.jstor.org/stable/pdf/2286348.pdf>.
- [3] Lavoslav Čaklović. Decision making by potential method. 1991. URL <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.111.350>.

# 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, te 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.

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:

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

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:

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

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

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

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:

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

Na ovaj način moguće je izračunati očekivanje i varijancu nad pomičnim prozorom na mnogo učinkovitiji način nego nerekurzivnim postupkom. Inicijalno 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 rekursivno 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 uživo. Ulazne podatke za algoritam čini kretanje cijena nekog broja vrijednosnica u nekom vremenskom 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, matematički i statistički postupci koji su korišteni mogu potencijalno 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 also prove useful in other applications.

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