

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ KATEDRA

Laiko eilučių empiriniai tyrimai

Empirical analysis of time series

Magistro baigiamasis darbas

Atliko: Vytautas Valiukonis (parašas)

Darbo vadovas: doc. Gintautas Bareikis (parašas)

Recenzentas: doc. Karolis Petrauskas (parašas)

Vilnius – 2017

Santrauka

Šiame darbe analizuojami akcijų indeksų laiko eilučių indikatoriai, kurie gali būti panaudojami šių eilučių prognozavimui. Yra atliekama hursto koeficiento, slenkančių vidurkių, juostų ir kanalų analizė.

Darbo metu sukurta lengvai plečiama platforma galinti patenkinti visus veiksmus reikalingus pilnai analizei. Ši platforma panaudota finansinių laiko eilučių duomenims surinkti ir jų indikatorių analizei atlikti.

Hursto parametro analizė naudojama laiko eilučių generavimui. Atlikus slenkančio vidurkio analizę buvo pastebėta, kad laiko eilutės ir slenkančio vidurkio susikirtimo taškai gali būti naudojami tolimesnei tendencijai nustatyti. Atlikta kanalų analizė leido nustatyti laiko eilutės elgesį kai kanalas yra pralaužiamas.

Raktiniai žodžiai: laiko eilutės, techninė analizė, indikatoriai

Summary

Indicators of stock price time series which can be used for forecasting are analyzed here. Analyzed indicators include hurst coefficient, moving averages, bands and channels.

An easily expandable platform capable of performing all required actions for analysis is created. This platform is used for collecting financial time series data and analyzing its indicators.

Analysis of hurst coefficient is used for generating time series. According to moving average analysis, points where time series value intersects moving average value can be used for forecasting time series direction. Channel analysis gives insight on time series behavior at channel breakouts.

Keywords: time series, technical analysis, indicators

TURINYS

ĮVADAS	4
1. LITERATŪROS APŽVALGA	7
1.1. Laiko eilutės	7
1.2. Prognozavimas ir simuliacijos	8
1.3. Techninė analizė	8
1.3.1. Slenkantis vidurkis	9
1.3.1.1. Paprastas slenkantis vidurkis	9
1.3.1.2. Tiesiškai svorinis slenkantis vidurkis	10
1.3.1.3. Eksponentiškai išlygintas slenkantis vidurkis	11
1.3.1.4. Kiti slenkantys vidurkiai	13
1.3.2. Juostos ir kanalai	13
1.3.2.1. Bollingerio juosta	13
1.3.2.2. Kitos juostos ir kanalai	14
1.3.3. Paprasti grafikų dėsniniai	15
1.4. Fraktaliniai parametrai	15
1.5. Indikatorių gyvenimo ciklas	16
2. PLATFORMA	18
2.1. Reikalavimai platformai	18
2.2. Alternatyvios platformos	19
2.2.1. Komercinės platformos	19
2.2.2. Statistiniai paketai	19
2.2.3. Specifinė implementacija	20
2.3. Naudotos technologijos	20
2.4. Platformos struktūra	20
3. DUOMENYS	23
4. BANDYMAI	24
4.1. Hursto parametras	24
4.2. Slenkantis vidurkis	27
4.3. Kanalai	31
4.3.1. Koreliacijos koeficientas	32
4.3.2. Paprastas kanalas	36
4.4. Modeliavimas	44
REZULTATAI IR IŠVADOS	46
LITERATŪRA	47

Įvadas

Kiek bus vertos Microsoft akcijos metų gale? Koks bus dolerio kursas euro atžvilgiu po savaitės? Šie ir panašūs klausimai rūpi daugeliui žmonių, bet atsakyti į juos nėra paprasta, o dažnai ir neįmanoma. Siekiant atlikti ateities procesų prognozę galime pasikliauti aklu spėjimu arba remtis paplitusiais stereotipais. Norint gauti argumentuotą atsakymą, tenka naudoti moksliniais metodais sukurtus įrankius.

Atsakant į aukščiau pateiktus klausimus, nepakanka tiesiog stebėti duomenų srautą ir remiantis šiais stebėjimas, naudojant, pavyzdžiui, analogiją, daryti apibendrinančias išvadas. Atliekdami ateities duomenų prognozę turime taikyti sudėtingus modelius, pagrįstus duomenų statistine analize, o po to gautus rezultatus prognozuoti su tam tikra tikimybe [CH02]. Atliekant duomenų analizę paprastai išskiriamos esminės jų charakteristikos, dar vadinamos rodikliais, kuriais remiantis galima numatyti tiriamų duomenų tendencijas. Rodiklių išskyrimo procesas, gana sudėtingas darbas, kadangi tenka apdoroti didelį duomenų kiekį, reikia remtis specifinėmis metodikomis, naudotis statistine analize. Atliekdami šį darbą arba modeliuodami ateities procesus turime remtis kompiuterinėmis programomis, kurias galima arba parašyti remiantis savo patirtimi, arba remtis jau esamomis, jas kūrybiškai taikant.

Nesunku suprasti, kad esminis faktorius, prognozuojant ateities procesus, yra sukaupta istorinė informacija, kuri paprastai yra gana gausi. Turint daugiau duomenų, galima atlikti detalesnį jų sąryšių tyrimą. Modeliuojant ateities procesus turime remtis istoriniais ir realiu laiku gaunamais duomenimis. Duomenų reikšmių seka laiko intervale paprastai vadinama *laiko eilute*. Šiame darbe nagrinėjamos akcijų ir valiutų rinkų laiko eilutės. Primename, kad laiko eilute suprantame taškų porų $(t_i; d_i)$, $i = 1, 2, \dots$ aibę, čia d_i duomenų reikšmė laiko t_i momentu. Šią porų aibę reprezentuodami plokštumoje gauname laiko eilutės grafiką. Fiksavę tam tikrą reiškinį laiko intervale, išskyrę skaitinę reiškinio reikšmę ir ją pažymėję plokštumoje, kai argumentas yra laikas, gausime laiko eilutės grafiką.

Laiko eilutės gali būti tolydžiosios arba diskrečios. Tolydžias laiko eilute galima analizuoti naudojant teorinius argumentus, jas interpretuojant kaip tam tikrų funkcijų grafikus. Šiame darbe laiko eilučių analizei pasitelksime specialią platformą, kurią galima taikyti diskrečioms laiko eilutėms. Pastebėsime, kad bet kokią tolydžią laiko eilutę galima konvertuoti į diskrečią norimu tikslumu. Šiame darbe bus nagrinėjamos tik pastarojo tipo duomenų eilutės, kadangi bet koki tolydų dydį suskirsčius į intervalus ir paėmus nustatytą (mažiausią, didžiausią, vidutinę) reikšmę laiko eilutę diskretizuotume.

Analizuojant laiko eilutę naudojamos įvairios priemonės: skaičiuojamos rodiklių (charakteristikų) reikšmės, nustatomi fraktaliniai dėsningumai, atliekamos duomenų transformacijos, išskiriamos anomalijos. Informacija gauta atlikus eilučių analizę gali būti pavaizduota grafikuose, kurie paprastai būna labiau informatyvūs negu pradinių duomenų laiko eilutė. Remiantis šiais rezultatais gali būti daromos pirminės išvados ir apibendrinimai. Šiame darbe bus atliekama ir gilesnė laiko eilučių analizė, kuri bus grindžiama statistiniais ir fraktaliniais metodais. Laiko eilučių analizė taikant įvairius statistinius ir grafinius metodus yra vadinama technine laiko eilučių analize arba tiesiog *technine analize* [KD10]. Metodai, kuriais remiantis yra atliekami laiko eilučių

skaitiniai vertinimai, yra vadinami *indikatoriais*. Indikatoriai gali būti pakankamai paprasti, tokie kaip slenkantis vidurkis, bet bus skaičiuojami ir daug sudėtingesni indikatoriai ir osciliatoriai (angl.: oscillator), kuriuos nustatant tenka naudoti statistinius bei fraktalinius metodus.

Atlikus laiko eilučių analizę galima kurti naujus indikatorius, kurie būtų taikomi modeliuojant eilutes. Be to, būtų galima patvirtinti arba paneigti jau esamų indikatorių efektyvumą pasirinktų duomenų analizei. Pastebėsime, kad iš turimų duomenų galima išgauti paprastus statistinius rodiklius tokius kaip vidurkis ir standartinis nuokrypis. Šie rodikliai nėra tiesiogiai naudingi prognozavimui ir patys nėra indikatoriai, bet yra naudojami daugelyje kitų skaičiavimų ir indikatorių išvedimui.

Paprastai laiko eilutė turi vieną reikšmę vienu laiko momentu, bet tuo pačiu laiko momentu gali būti ir daugiau nei viena susijusi skaitinė reikšmė. Tokiu atveju n reikšmių turinti laiko eilutė suprantama kaip $n + 1$ ilgio vektorių $(t_i; d_i^1; d_i^2; \dots; d_i^n), i = 1, 2, \dots$ aibė, čia d_i^k yra k -oji reikšmė laiko momentu t_i . Akcijų kainos, indekso ar fondo veiklos duomenys kartu su dabartine verte visuomet gali būti susieti su tūriu (prekiautų vienetų kiekis). Kadangi duomenų taškas laiko eilutėje gali turėti daugiau nei vieną dimensiją, atsiranda daug galimybių jų analizei.

Moksliniuose darbuose yra aprašyta labai daug įvairių indikatorių, kuriuos naudojant yra atliekama techninė analizė, tačiau ne visi indikatoriai naudojami techninėje analizėje yra pilnai ištestuojami naudojantis statistiniais metodais. Rodikliai, siejami su vienomis laiko eilutėmis nebūtinai bus tinkami kitoms. Kyla natūralus klausimas, jei pasirinkti rodikliai tinka pasirinktai laiko eilutei, tai kokioms dar eilutėms jie yra efektyvūs? Ar galima nurodyti laiko eilučių klasę, kuriai tinka vienas ar kitas nustatytas rodiklis? Kitaip tariant ar galima charakterizuoti laiko eilutes pagal pasirinktus rodiklius? Šiame darbe bus bandoma ištirti įvairius laiko eilučių rodiklius, bei atsakyti į aukščiau suformuluotus klausimus. Ypatingas dėmesys šiame darbe bus skiriamas laiko eilučių prognozavimo problemai spręsti. Naudosime fraktalinės analizės [Cro95] bei statistinius metodus tiriant pasirinktas laiko eilutes ir siekiant atskirti neesminius duomenis nuo laiko eilutę charakterizuojančių duomenų.

Prognozuojant laiko eilutes bus remiamasi empiriškai išnagrinėtų (analizė bus atliekama remiantis sukurta platforma) indikatorių statistinėmis bei fraktalinėmis savybėmis. Pasirinkus norimus parametrus ir sumodeliavus didelį kiekį laiko eilučių galima atlikti jų analizę lyginant jas su laiko eilutėmis gautomis iš realių duomenų. Šio tyrimo metu bus siekiama sukurti kriterijus, kurių pagrindu būtų galima atsakyti ar sukurtas simuliacinis modelis atitinka tiriamus duomenis. Jei generuojamų duomenų laiko eilutė turės panašias savybes ir išlaikys tas pačias tendencijas su pakankamai didele tikimybe, tuomet būtų galima teigti, jog indikatorių rinkinys pagal kurį buvo generuojamos laiko eilutės, gali būti naudojamas jų prognozavimui.

Darbo tikslas:

Atlikti laiko eilučių techninės analizės metodų bei indikatorių statistinę ir fraktalinę analizę, pritaikyti gautus rezultatus laiko eilučių prognozavimui.

Darbo uždaviniai:

1. Atlikti literatūros apžvalgą, pasirinkti dažniausiai naudojamus ir naujausius indikatorius bei techninės analizės metodus.

2. Sukurti platformą, kuria pasinaudojus būtų galima surinkti ir apdoroti labai didelius kiekius istorinių duomenų iš nurodytų šaltinių.
3. Atlikti laiko eilučių (akcijų indeksų) indikatorių analizę bei vizualiai atvaizduoti nagrinėtas laiko eilutes (naudojantis sukurta platforma).
4. Kurti naujus indikatorius bei grafinius metodus, kurie parodytų laiko eilutės specifines charakteristikas arba būtų taikomi eilučių prognozavimui.
5. Prognozuoti laiko eilutes pagal pasirinktų indikatorių bei techninės analizės metodų rinkinį realiomis sąlygomis.

Laukiami rezultatai:

1. Sukurta platforma, gebanti analizuoti, generuoti ir prognozuoti laiko eilutes.
2. Atlikta laiko eilučių indikatorių analizė, duomenys buvo naudojami iš akcijų indeksų laiko eilučių. Jos metu bus pasirinktas metodų ir indikatorių rinkinys, kuris bus taikomas laiko eilučių prognozavimui.
3. Prognozavimas atliekamas naudojant daugiamatį tiesinį modelį akcijų indeksų kainoms.



2 pav. Japoniškų žvakių laiko eilutės grafiko pavyzdys

1.2. Prognozavimas ir simuliacijos

Prognozė yra teiginys apie ateitį: jis gali būti pagrįstas arba neturėti jokio logiško paaiškinimo, išsipildyti arba ne pasirinktu laiko momentu arba vidutiniškai per pasirinktą laiko intervalą, tikslus arba ne, pagrįstas modeliu arba neformalus [CH02]. Prognozės paprastai remiasi ekspertų išvadomis, laiko eilučių modeliais arba ekonometrinėmis lygčių sistemomis. Šiame darbe nebus atsižvelgiama į jokus ekspertines išvadas ar ekonometrinės lygčių sistemas. Turimi duomenys turės tik savo skaitines reikšmes ir jokio kito konteksto, neskaitant to, jog jie yra finansinės kilmės.

Simuliacijos yra bandymas sukurti vaizdą, panašų į realybę. Skirtingai nei prognozavimo metu, nėra siekiama nusakyti, kas bus ateityje, bet yra bandoma sukurti „alternatyvią dabartį“. Simuliacijos metu sudaromos laiko eilutės yra naudingos nustatant parametrus, su kuriais galima gauti panašias laiko eilutes. Sudarytos panašios laiko eilutės yra naudingos atliekant prognozavimą. Jeigu simuliuojamas modelis atitinka realybę, galima simuliuoti daugybę kartų ir taip patikrinti, kaip gerai veikia prognozavimas.

1.3. Techninė analizė

Techninė analizė yra finansinė analizė, kurioje yra aptinkami dėsniniai rinkos duomenyse siekiant atrasti tendencijas ir daryti prognozes. Techninė analizė remiasi laisvosios rinkos apibrėžimu, jog vertybinių popierių kaina yra nustatoma paklausos ir pasiūlos. Be to, kaina yra linkusi judėti pagal tendenciją, o tos tendencijos gali būti matomos grafikuose [KD10].

Naudojant pažangias skaičiavimo sistemas yra įmanoma efektyviai pritaikyti ir kitokius me-

todus, tokius kaip neuroniniai tinklai [ET05], paieška pagal šablonų taikymą (angl.: template-matching) [LMP⁺02] ar SVM (angl.: support vector machines) [CGB⁺14]. Šiame darbe yra apsiribojama laiko eilučių grafikų dėsningumo metodais.

Techninėje analizėje yra naudojami indikatoriai – išvestiniai duomenys, apskaičiuojami pagal pasirinktas formules, kuriais remiantis yra atliekami laiko eilučių skaitiniai vertinimai. Jų sudėtingumas varijuoja nuo paprastų, tokių kaip vidurkio skaičiavimas, iki daug sudėtingesnių. Čia apžvelgsime dažniausiai naudojamus indikatorius, kurie plačiai taikomi analizuojant laiko eilutes.

1.3.1. Slenkantis vidurkis

Dažniausiai naudojamas ir plačiai taikomas indikatorius yra slenkantis vidurkis. Jo tikslas yra sumažinti smulkius svyravimus, kurie iškreipia bendrą vaizdą. Pritaikius slenkančio vidurkio indikatorių laiko eilutėje, yra labai gerai matoma esama tendencija. Vis dėlto, naudojant šį indikatorių iš laiko eilutės duomenų aibės yra sudaroma viena indikatoriaus reikšmė, dėl to išsišokusios reikšmės yra slopinamos. Be to, slenkantis vidurkis dažniausiai atsilieka nuo dabartinės tendencijos ir apie jos pasikeitimą indikuoja tik praėjus tam tikram laiko tarpui.

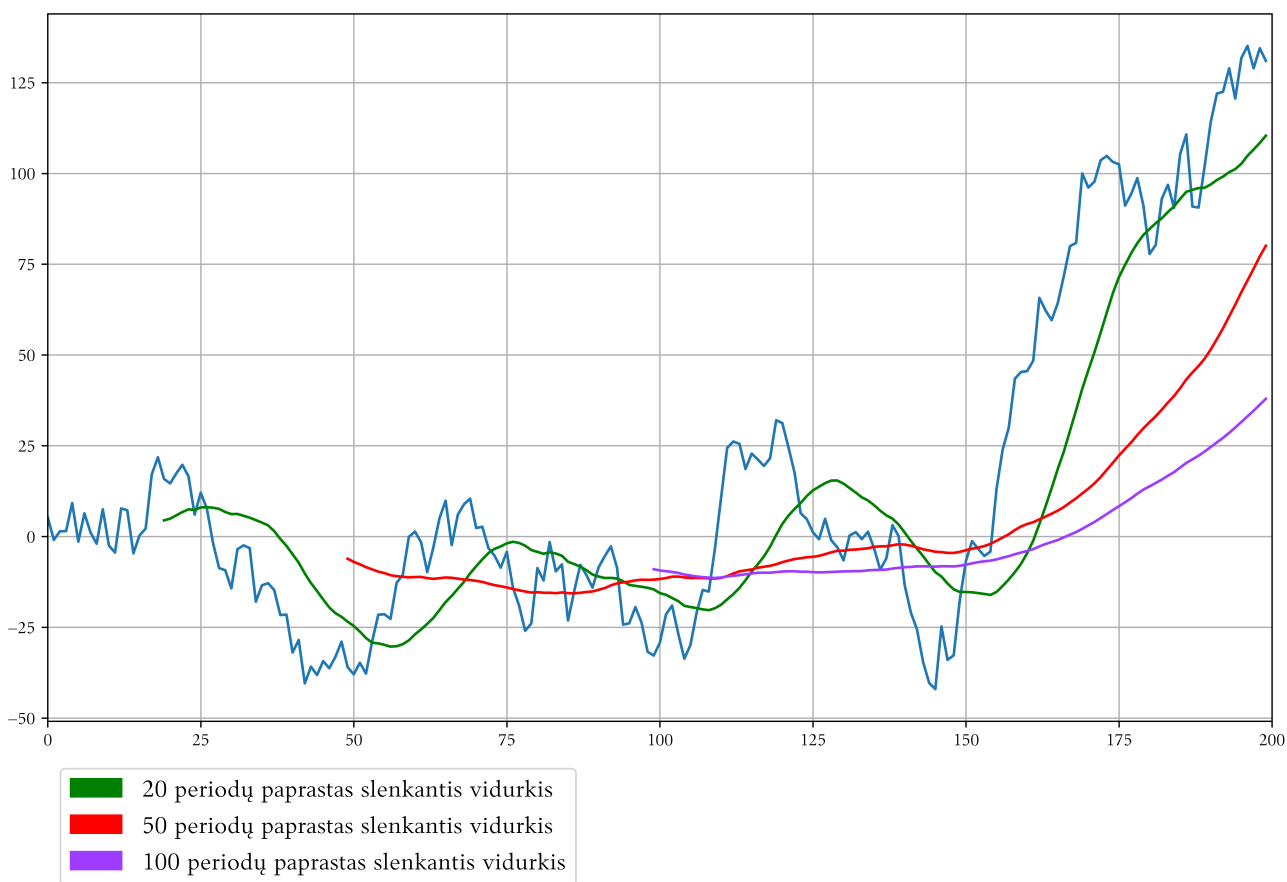
Paprastai praktikoje yra iškart taikomi keli slenkantys vidurkiai su skirtingais parametrais – taip yra pastebimos skirtingo dydžio tendencijos, trumpesnės ir ilgesnės, priklausomai nuo pasirinktų parametrų. Skirtingais parametrais skaičiuojamų slenkančių vidurkių susikirtimo taškai taip pat turi specialią reikšmę. Paprastai jie prognozuoja apie būsimą tendencijos pasikeitimą.

1.3.1.1. Paprastasis slenkantis vidurkis

Ši slenkančio vidurkio rūšis yra pati paprasčiausia ir intuityviausia. Norint apskaičiuoti slenkantį vidurkį, reikia paimti paskutines n laikotarpių skaitines reikšmes ir suskaičiuoti jų aritmetinį vidurkį.

$$SMA_n = \frac{\sum_{i=1}^n x_i}{n}$$

SMA – paprastas slenkantis vidurkis (angl.: simple moving average). Kiekvienam pasirinktam laiko vienetui apskaičiavę slenkantį vidurkį, gautume slenkančio vidurkio kreivę su parametru n .



3 pav. Skirtingų periodų skaičiaus paprastų slenkantių vidurkių pavyzdžiai

3 pav. parodyta, kaip atrodo skirtingi slenkantys vidurkiai. Iš čia nesunku pastebėti, kaip esant mažesniai periodų skaičiui slenkancio vidurkio linija greitai reaguoja į laiko eilutės pasikeitimus. Didėjant periodų skaičiui ši linija mažiau banguoja ir atspindi ilgalaikę laiko eilutės tendenciją.

1.3.1.2. Tiesiškai svorinis slenkantis vidurkis

Paprastas slenkantis vidurkis visus n paskutinių laikotarpių vertina vienodai. Tai neatspindi realios situacijos, kai naujausia informacija turi didžiausią svorį. Sudarant svorinį slenkantį vidurkį yra atsižvelgiama į šią aplinkybę.

Tiesiškai svorinis slenkantis vidurkis (angl.: *LWMA* - lineary weighted moving average) naudoja tiesiškai didėjančius svorinius koeficientus. Koeficientai yra parenkami nuo 1 iki n . Kiekviena iš n laikotarpių reikšmių yra dauginama iš atitinkamo koeficiento, pradedant seniausia reikšme, kuri yra dauginama iš mažiausio koeficiento 1 ir baigiant naujausia reikšme, kuri yra

dauginama iš koeficiento n . Visų šių sandaugų suma yra padalijama iš koeficientų sumos.

$$LWMA_n = \frac{\sum_{i=1}^n x_i i}{\sum_{i=1}^n i}$$

Nesunku įsitikinti, jog tiesiškai svorinis slenkantis vidurkis būna stabilus, kai duomenys su didžiausiais priskirtais koeficientais yra pastovūs, o duomenys su mažiausiais koeficientais smarkiai svyruoja. Tai reiškia, jog šis indikatorius gerai neatspindi duomenų, kuriems yra priskiriami mažiausi koeficientai. Skaičiuojant naują šio slenkančio vidurkio vertę (jei laiko eilutėje yra m duomenų vienetų ir n laikotarpių, slenkantis vidurkis skaičiuojamas nuo i iki $i + n$ laikotarpio, kur $0 < i \leq m - n - 2$, nauja slenkančio vidurkio vertė skaičiuojama nuo $i + 1$ iki $i + n + 1$ laikotarpio), galima gauti didelį vertės šuolį, jei i laikotarpio duomenys smarkiai skiriasi nuo likusių. Šis efektas yra vadinamas „atkritimo efektu“ ir yra labiausiai kritikuotinas slenkančio vidurkio aspektas [Kau13].

1.3.1.3. Eksponentiškai išlygintas slenkantis vidurkis

Atkritimo efekto problema gali būti išspręsta naudojant eksponentiškai išlygintą slenkantį vidurkį (angl.: exponentially smoothed moving average). Skirtingai nei tiesiškai svoriniame slenkančiame vidurkyje, čia yra naudojami geometriškai progresyvūs svoriniai koeficientai

$$a^0, a^1, a^2, \dots, a^n$$

, tačiau čia koeficientas a atitinka $0 < a < 1$ nelygybę. Tokiu atveju didžiausias svorinis koeficientas yra pirmasis ir jis priskiriamas naujausiai laiko eilutės duomenų reikšmei. Tuomet eksponentiškai išlygintas svorinis vidurkis būtų:

$$E_n = \frac{\sum_{i=1}^n x_i a^{n-i}}{\sum_{i=1}^n a^i}$$

Čia E_n - n laikotarpių eksponentiškai išlygintas slenkantis vidurkis, x_i - laiko eilutės reikšmė i -uoju laikotarpiu, a - svorinis koeficientas.

Kadangi tokį vidurkį yra sudėtinga greitai apskaičiuoti, dažniau yra naudojama kita jo forma:

$$E_n = E_{n-1} + a(x_n - E_{n-1}), \text{ kur } E_1 = x_1$$

Tokiu atveju svoriniai koeficientai prilygsta begalinei mažėjančiai geometrinei progresijai.

Galima paskaičiuoti, kokia turi būti koeficiento a reikšmė, kad pati naujausia laiko eilutės reikšmė, naudojama šio slenkančio vidurkio skaičiavimui, būtų tokia pati, kaip n laikotarpių

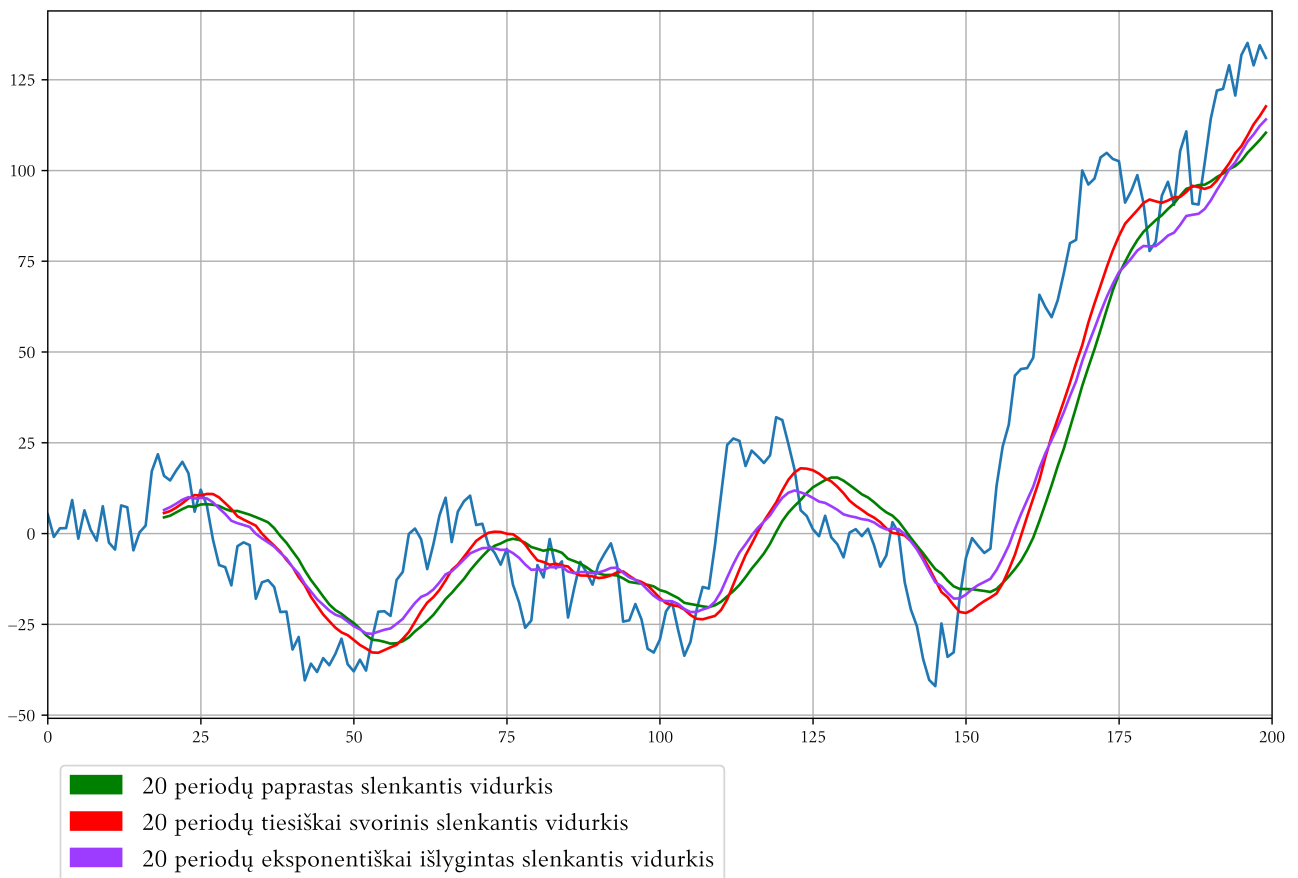
tiesiškai svoriniame slenkančiame vidurkyje. Pasinaudoję tiesiškai svorinio slenkančio vidurkio skaičiavimo metodu matome, jog naujausias duomenų taškas turi n kartų didesnę svorio koeficientą už paskutinį. Apskaičiuojame, kokią dalį visos sistemos svorio jis sudaro ir prilyginame ją eksponentiškai išlyginto slenkančio vidurkio koeficientui [Hut84]:

$$\frac{n}{\sum_{i=1}^n i} = \frac{n}{\frac{n+1}{2} \cdot n} = \frac{2}{n+1} = a$$

Čia x_i yra i -oji duomenų reikšmė slenkančiame vidurkyje. Šią svorio reikšmę panaudosime kaip eksponentiškai išlyginto slenkančio vidurkio svorinį koeficientą. Tuomet jo forma bus:

$$E_n = E_{n-1} + \frac{2}{n+1}(x_n - E_{n-1})$$

4 pav. parodyta, kaip skiriasi skirtingi slenkančių vidurkių skaičiavimo metodai. Iš čia labai lengva matyti, kad visi išvardinti metodai yra pakankamai panašūs. Didžiausi skirtumai yra pastebimi svarbiausiose vietose, kur laiko eilutė keičia savo kryptį. Tiesiškai svorinis vidurkis yra artimiausias laiko eilutei, bet jame yra pastebimi didesni išsišokimai, kurie gali trukdyti automatizuotai analizei.



4 pav. Skirtingų skaičiavimo metodų slenkančių vidurkių pavyzdžiai

1.3.1.4. Kiti slenkantys vidurkiai

Yra ir daugiau naudojamų slenkančio vidurkio skaičiavimo variacijų, tokių kaip geometrinė ar varijuojama, tačiau jos šio darbo kontekste yra neaktualios ir mažiau paplitusios nei darbe pateikiami indikatoriai.

1.3.2. Juostos ir kanalai

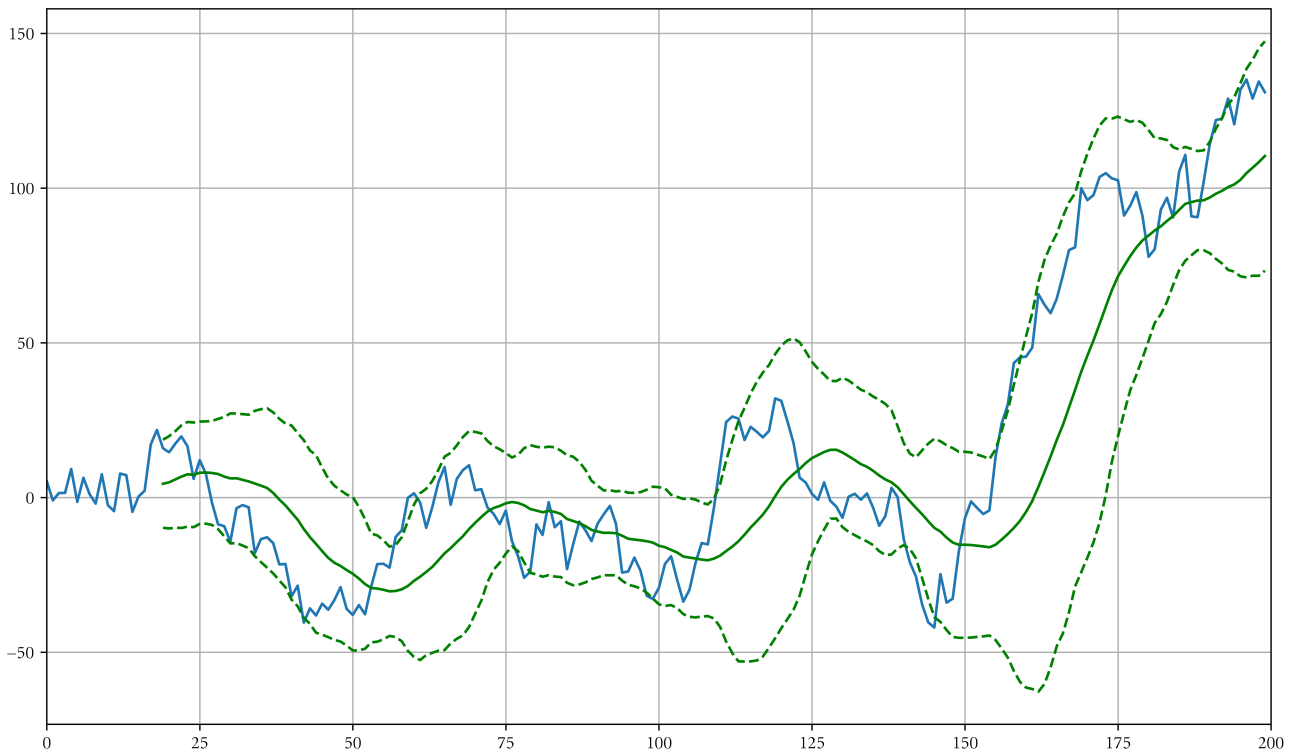
Slenkantis vidurkis parodo, kokia kryptimi juda laiko eilutė, tačiau jame neatsispindi informacija apie svyravimų amplitudes. Siekiant jas pamatyti, naudojamos juostos ir kanalai, kurie susideda iš minimalios ir maksimalios slenkančios ribos. Skirtingai nei slenkantys rodikliai, kurie turi vieną skaitinę reikšmę vienu laiko momentu, juostos ir kanalai turi 2 reikšmes. Turint juostą, galima tikrinti, ar laiko eilutė yra jos viduje ar išorėje. Didžioji dalis visų reikšmių patenka į juostos vidų. Jei laiko eilutė peržengia juostos ribą, tai signalizuoja apie galimus pokyčius tendencijoje.

1.3.2.1. Bollingerio juosta

Vienas populiariausių ir seniausių juostų ir kanalų tipo rodiklių yra Bollingerio juosta. Ji sudaroma iš slenkančio vidurkio ir standartinio nuokrypio.

$$BB(N, K) = MA_N \pm K\sigma$$

Čia MA_N yra N periodų slenkantis vidurkis, K - skaitinė reikšmė, nurodanti juostos plotį, σ - standartinis nuokrypis. Dažniausiai praktikoje naudojami parametrai yra šie: paprastas arba eksponentiškai išlygintas slenkantis vidurkis, N ir K - atitinkamai pasirenkamos 20 ir 2 vertės, o periodų ilgis yra viena diena. 5 pav. parodyta Bollingerio juosta su šiomis standartinėmis parametru vertėmis. Bollingeris manė, kad tokios parametru vertės yra optimalios daugeliui situacijų, todėl šie parametrai naudojami ir dabar.



5 pav. Bollingerio juostos su $N = 20$ ir $K = 2$ pavyzdys

Jei laiko eilutė būtų Brauno judesys, tuomet pagal standartinio nuokrypio apibrėžimą į Bollingerio juostos vidų pakliūtų 95% visų verčių.

1.3.2.2. Kitos juostos ir kanalai

Yra ir kitokio tipo juostų, tačiau jos yra daug rečiau naudojamos. Anksčiau buvo plačiai naudojamas Keltnerio kanalas [Kel60], tačiau jo nuokrypis yra sukonstruojamas iš keleto vidurkių nesiremiant standartiniu nuokrypiu. Toks skaičiavimo būdas buvo paprastesnis neturint kompiuterių, tačiau dabar nėra prasmės naudoti jį vietoje standartinių nuokrypių.

Donchian kanalas – kitas dažnai taikomas rodiklis. Jis yra sudaromas iš didžiausios ir mažiausios laiko eilutės vertės per pasirinktą periodą. Taip pat, kaip kitose juostose, laiko eilutė, peržengusi nustatytą ribą, signalizuoja apie galimus pokyčius tendencijoje. Vis dėlto, kitaip nei kituose panašiuose rodikliuose, išėjimas iš kanalo ribų yra praktiškai garantuotas. Taip atsitinka, nes kanalas naudoja didžiausią ir mažiausią reikšmes per paskutinius N periodų. Jei laiko eilutės reikšmės niekada neišeis iš kanalo ribų, tai viršutinė ir apatinė riba pastoviai trauksis. Dėl šios savybės Donchian kanalas yra nepatikimas, nes automatizuotai ieškant esminių taškų gaunamas labai didelis triukšmo kiekis.

Dar vienas naudojamas metodas yra vizualaus tipo kanalo piešimas. Tendencijos sekimui yra nupiešiamos lygiagrečios linijos palei maksimalias ir minimalias ribas, į kurias patenka laiko eilutės reikšmės. Kadangi šis metodas yra silpnai apibrėžtas ir neturi konkrečių skaitinių kriterijų, jis taip pat yra netinkamas automatizuotiems veiksams.

1.3.3. Paprasti grafikų dėsningumai

Grafikų dėsningumai (angl.: chart patterns) yra indikatoriai, kurie yra vizualiai aptinkami grafikuose [KD10]. Jie buvo ypač naudojami iki kompiuterių paplitimo. Nėra pateiktų vienareikšmiškų statistinių įrodymų visiems naudojamiems grafikų dėsningumams, egzistuoja tik įvairūs neatkartojami tyrimai [Bul03].

Priklausomai nuo grafiko rūšies skiriasi ir pritaikomos dėsningumų rūšys. Japoniškų žvakių grafikuose vienas laiko periodas turi 4 informacijos vienetus, o linijinis grafikas tik 1. Iš čia matoma, jog gali egzistuoti daugiau skirtingų dėsningumų rūšių, kurios yra pritaikomos tik japoniškų žvakių tipo grafikams. Kai kurie tyrimai rodo šių dėsningumų pagrindžiamumą [LCH15], tačiau šiame darbe apsiribojama tik linijiniais grafikais bei jų dėsningumais.

Kai kuriems grafiniams dėsningumams atliktos statistinės analizės. Pavyzdžiui, „vėliavos“ dėsningumas turi patikrintą ir patvirtintą patikimumo analizę [LFH⁺08]. Norint kompiuterizuotu būdu aptikti tokį dėsningumą, reikia euristiniais metodais atpažinti grafiko formą arba naudoti vizualų vaizdą klasifikuojančius algoritmus. Taigi, dėl sudėtingo automatinio aptikimo šiame darbe tokio tipo dėsningumai plačiau nenagrinėjami.

1.4. Fraktaliniai parametrai

Pagrindinis būdas simuliuoti atsitiktinį judėjimą yra naudotis Brauno judesiais [Cro95]. Brauno judesių idėja yra labai paprasta: norint gauti kitą sekos elementą, reikia prie prieš tai buvusio elemento pridėti atsitiktinį skaičių, pasiskirsčiusį pagal normalųjį pasiskirstymą.

$$x_{n+1} = x_n + g_n$$

Čia x_n yra n -asis atsitiktinio Brauno judėjimo elementas, o g_n yra sugeneruotas atsitiktinis skaičius, priklausantis normaliajam skirstiniui. Jei pasirinktume $x_0 = 0$, o $g_i \in \mathcal{N}(0, 1)$, $i = 0, 1, 2, \dots$ ir fiksuotume kokį nors $n \in \mathbb{N}$, aibė $X = \{x_0, x_1, \dots, x_n\}$ būtų standartiškas n elementų Brauno judesys.

Egzistuoja ir kitas Brauno judėjimo generavimo algoritmas. Skirtingai nei pirmasis, šis leidžia generuoti Brauno judesį pasirinktai atkarpai iteratyviai smulkinant tarpinius žingsnius. Tarkime, jog laiko eilutės apibrėžimo sritis yra intervalas $[0, 1]$, tuomet Brauno judėjimo reikšmės bus apibrėžtos funkcija:

$$X(t), \text{ kur } t = \frac{k}{2^n}, k = 0, 1, \dots, 2^n$$

Čia n yra norimas lygių skaičius. Kuo didesnis šis parametras, tuo smulkesni žingsniai tarp aibės elementų. Nesunku pastebėti, kad nuo šio parametro taip pat priklauso elementų skaičius galutinėje aibėje, kuris yra $2^n + 2$.

Pasirenkame pradines reikšmes $X(0) = 0$ ir $X(1) = g_1$. Primename, kad čia $g_i \in \mathcal{N}(0, 1)$, $i \in [0, 1]$ yra nepriklausomas ir atsitiktinai sugeneruotas kiekvienam skirtingam i . Nustačius visus pradinius parametrus, galima pradėti generuoti aibės elementus.

Pirmame lygyje:

$$X\left(\frac{1}{2}\right) = \frac{1}{2}(X(0) + X(1)) + \frac{1}{2}g_{\frac{1}{2}}$$

Antrame lygyje:

$$X\left(\frac{1}{4}\right) = \frac{1}{2}(X(0) + X\left(\frac{1}{2}\right)) + \frac{1}{2^{\frac{3}{2}}}g_{\frac{1}{4}}$$

$$X\left(\frac{3}{4}\right) = \frac{1}{2}(X\left(\frac{1}{2}\right) + X(1)) + \frac{1}{2^{\frac{3}{2}}}g_{\frac{3}{4}}$$

⋮

n -ajame lygyje:

$$X\left(\frac{1}{2^n}\right) = \frac{1}{2}(X(0) + X\left(\frac{1}{2^{n-1}}\right)) + \frac{1}{2^{\frac{n+1}{2}}}g_{\frac{1}{2^n}}$$

⋮

$$X\left(1 - \frac{1}{2^n}\right) = \frac{1}{2}(X\left(1 - \frac{1}{2^{n-1}}\right) + X(1)) + \frac{1}{2^{\frac{n+1}{2}}}g_{1 - \frac{1}{2^n}}$$

Atlikus šią procedūrą nustatytą lygių skaičių, yra gaunamas norimo smulkumo atsitiktinis Brauno judėjimas. Šio algoritmo privalumas yra tai, jog esant poreikiui galima nesunkiai padidinti n reikšmę ir pridėti papildomus lygius neperskaičiuojant jau suskaičiuotų reikšmių.

Pasinaudojus kuriuo nors iš nurodytų metodų galima sugeneruoti atsitiktinę laiko eilutę. Kadangi sugeneruotos laiko eilutės yra panašios į sudarytas iš realių duomenų savo tekstūra, bet ne skaitinėmis vertėmis, jas galima transformuoti, nepakeičiant šios tekstūros.

Yra įmanomos 2 primityvios transformacijos: mastelio pakeitimo ir postūmio. Norint pakeisti mastelį užtektų padauginti kiekvieną Brauno judėjimo sekos elementą iš pasirinktos konstantos. Atliekant postūmį, reikėtų prie kiekvienos aibės vertės pridėti norimą konstantą. Nesunku pastebėti, jog postūmio atlikimas yra analogiškas veiksmas skirtingos pradinės reikšmės parinkimui.

Norint, kad sugeneruota laiko eilutė būtų panaši į siekiamą rezultatą, reikia atsižvelgti į pasirinktus parametrus. Vienas iš parametrų, į kurį galima atsižvelgti derinant realias ir sugeneruotas sekas, yra Hursto koeficientas (H) ir iš jo tiesiogiai išvedamas grafo dimensijos (D) rodiklis. Dvimačiam grafikui formulė yra:

$$D = 2 - H$$

Šį koeficientą galima apskaičiuoti pagal dėžių skaičiavimo metodą [Wah95]. Šis metodas gali būti lengvai užrašomas programiniu kodu [Cro95].

1.5. Indikatorių gyvenimo ciklas

Techninės analizės indikatoriai praranda didelę dalį savo pelningumo po jų publikavimo [MP16]. Tai grindžiama tuo, jog, pradėjus plačiai taikyti indikatorių, yra greičiau išlyginama blogai įvertinta kaina.

Minimą indikatorių savybę galima prilyginti gyvenimo ciklo pabaigos etapui programų sistemų inžinerijoje. Pradžios etapas būtų indikatoriaus atradimas, visas tyrinėjimas ir analizė gali

būti prilyginami jo kūrimui, o publikavimas - išleidimui.

2. Platforma

Norint apdoroti didelius duomenų kiekius, kurie yra reikalingi išsamiai pasirinktų indikatorių analizei, reikia turėti platformą, galinčią tai atlikti. Ši platforma yra atsakinga už visą duomenų gyvavimo ciklą nuo surinkimo iki analizės rezultatų išvedimo.

Tolimesniuose šio skyriaus poskyriuose išvardinami reikalavimai platformai, alternatyvos ir pateikiamas detalus platformos aprašas.

2.1. Reikalavimai platformai

Funkciniai platformos tikslai:

1. Surinkti duomenis analizei iš prieinamų šaltinių.
2. Saugoti duomenis lokaliai (angl.: caching).
3. Atlikti skaičiavimus, reikalingus analizei.
4. Pateikti rezultatus.

Nors lokalus duomenų saugojimas nėra sudėtinga ar didelė dalis palyginus su kitomis, tačiau jis logiškai atskiria duomenų surinkimą ir skaičiavimų atlikimą. Tokiu atveju yra įmanoma lengviau lygiagretinti skaičiavimus, esant reikalui juos pakartoti ir nereikia visko perskaičiuoti pridėjus naujus duomenis. Be to, galima lengviau pasirinkti keletą skirtingų duomenų sluoksnių, kuriais bus analizuojama. Daugiau apie duomenis išdėstyta 3 skyriuje.

Nefunkciniai platformos tikslai:

1. **Lengvas plečiamumas.** Visi programos moduliai turi būti lengvai papildomi ar pakeičiami kitomis, tą pačią funkciją atliekančiomis dalimis.
2. **Rezultatų atkartojamumas.** Rezultatai privalo būti deterministiški ir atkartojami, kad juos būtų galima testuoti ir įsitikinti jų teisingumu. Generuojamų duomenų atveju ši savybė turi būti išlaikyta išsaugant sugeneruotus rinkinius.
3. **Dalinis rezultatų perskaičiavimas.** Ši savybė reikalinga plečiant ar keičiant platformos dalis, nes ji leidžia atskiriems moduliams būti pilnai nepriklausomiems.

Iš čia matoma, kad platformos nefunkciniais tikslais siekiama užtikrinti, kad būtų galima pridėti naujus testavimo duomenis ar analizės metodus. Platforma privalo įgyvendinti visus kečiamus funkcinius ir nefunkcinius tikslus, kitu atveju ji netektų savo prasmės.

2.2. Alternatyvios platformos

Egzistuoja keletas platformos variantų, kurie dalinai įgyvendina reikalingą funkcionalumą. Iš visų alternatyvų pasirinktas naujas platformos įgyvendinimas, nes tik taip galima užtikrinti, jog reikalavimai programinei įrangai bus įgyvendinami. Visos kitos platformų rūšys turi rimtų trūkumų, kurie padaro jas panaudojamas tik tam tikram uždavinių daliai. Primename, jog į uždavinius įeina: duomenų surinkimas, jų saugojimas, analizės atlikimas ir rezultatų pateikimas.

Tolimesniuose poskyriuose yra trumpai apžvelgiami kiekvienos platformos rūšies privalumai ir trūkumai.

2.2.1. Komerčinės platformos

Komerčių platformų pagrindinis tikslas yra finansinių priemonių (valiutų, akcijų, fondų, aukso ir kt.) prekyba. Kartu yra pateikiami įrankiai laiko eilučių analizei.

Privalumai:

1. Greitai pasiekiami techninės analizės metodai.
2. Realus laiko duomenų pasirinkimas skirtingoms finansinių priemonių rūšims.
3. Grafinis laiko eilučių ir indikatorių pateikimas.

Trūkumai:

1. Ribotas techninės analizės metodų pasirinkimas ir galimybės plėsti naujais indikatoriais trūkumas.
2. Uždaras duomenų formatas neleidžiantis laisvai manipuluoti jais. Taip pažeidžiamas analizės kartojimo reikalavimas.

2.2.2. Statistiniai paketai

Statistiniai paketai yra skirti įvairios rūšies duomenų statistinei analizei atlikti.

Privalumai:

1. Įrankiai skirti statistinei analizei atlikti.

Trūkumai:

1. Nėra reikalingų duomenų surinkimo funkcionalumo.
2. Nėra finansiniams duomenims skirto duomenų vizualizavimo.
3. Indikatorių skaičiavimas turėtų būti įgyvendinamas naudojantis specialiu formatu.

2.2.3. Specifinė implementacija

Specialiai šiems reikalavimams sukurta programinė įranga.

Privalumai:

1. Įgyvendinami visi reikalavimai platformai taip, kaip yra reikalinga analizei.

Trūkumai:

1. Implementacija užima daug laiko.

2.3. Naudotos technologijos

Platformos įgyvendinimui pasirinkta Python programavimo kalba. Ši programavimo kalba yra labai aukšto lygio, todėl naudojantis ja greičiausiai sukuriamas galutinis produktas, bet jo vykdymo laikas yra keletą kartų lėtesnis už žemesnio lygio programavimo kalbas. Egzistuoja platus papildinių pasirinkimas, kurie gali atlikti trivalias, su analize nesusijusias, pagalbines užduotis. Dažnas tokių užduočių pavyzdys: papildomas pagalbinis sluoksnis kreipiantis į išorinius API. Python kalba parašytą programinį kodą galima nesudėtingai praplėsti su C++ programavimo kalba parašytais moduliais, jei atskirose platformos dalyse būtų kritiškai svarbus veikimo laikas.

Kiti svarstyti, bet atmesti variantai yra C++ ir C# programavimo kalbos. C++ programavimo kalba atmesta dėl mažo kiekio išorinių modulių ir didelio laiko skirtumo, kurį užtrunka parašyti programinį kodą palyginus su kitomis, aukštesnio lygio, kalbomis.

C# kalba atmesta kaip tarpinis variantas tarp Python ir C++ pagal turimų išorinių modulių kiekį, programavimo laiką ir vykdymo greitį. Šis variantas neturi išskirtinių savybių, kurias turi kitos, aukščiau paminėtos, programavimo kalbos.

Duomenų saugojimui pasirinkta „sqlite“ technologija. Toks pasirinkimas užtikrina labai didelį saugomų duomenų kiekį ir naudojimo paprastumą. Naudojantis „sqlite“ duomenų bazę sudaro vienas failas, kuris gali būti nesunkiai perkeltas iš vienos vietos į kitą. Pasinaudojus šia savybe visa platforma įgyja papildomo portabilumo ir leidžia nesunkiai turėti keletą duomenų bazių.

Kitos duomenų bazių valdymo technologijos, tokios kaip „PostgreSQL“ ar „MySQL“ nėra naudojamos dėl savo dydžio ir perteklinio sudėtingumo. Papildomas sudėtingumas yra visiškai nereikalingas, nes esminė duomenų struktūra yra labai paprasta: laikas, vertė.

Skaičiavimams (veiksmai su matricomis) naudota „numpy“ biblioteka. Ji pasirinkta dėl labai greito veikimo atliekant skaičiavimus su dideliais duomenų kiekiais ir patikimumo. Vizualiam duomenų atvaizdavimui pasirinkta „matplotlib“ biblioteka, dėl labai plačių konfigūravimo galimybių. Abi šios bibliotekos neturi realių alternatyvų savo srityse ir yra naudojamos visur, kur prireikia tokio tipo funkcionalumo.

2.4. Platformos struktūra

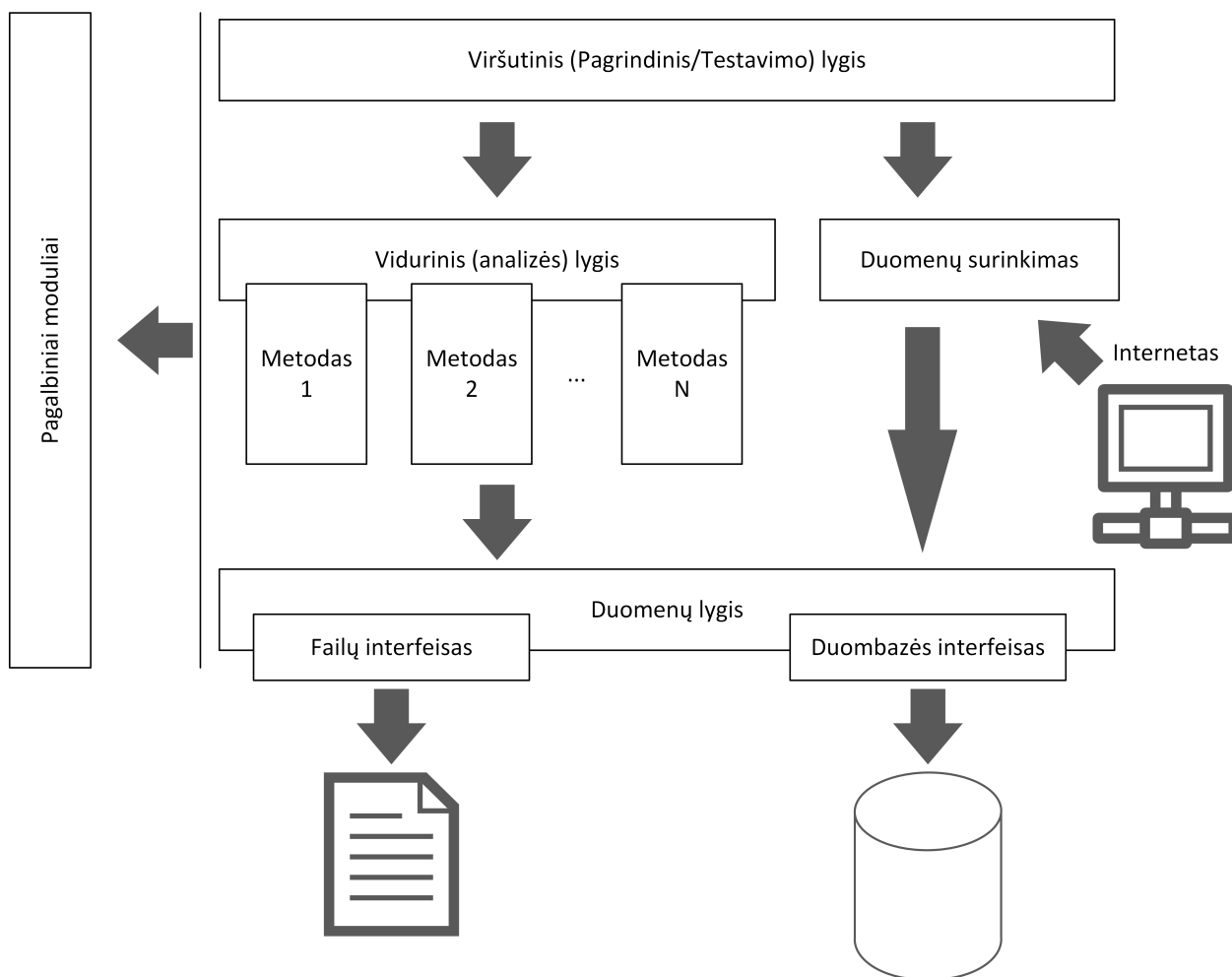
Platforma susidaro iš trijų pagrindinių sluoksnių ir papildomų pagalbinių modulių. **Sluoksniai:**

1. Viršutinis sluoksnis skirtas pagrindiniam iškviečiamajam ir testavimo programiniam kodui.
2. Vidurinis sluoksnis skirtas duomenų analizės scenarijams ir duomenų surinkimui. Nors ir abi sistemos yra tame pačiame sluoksnyje, tačiau jos yra visiškai tarpusavyje nesusijusios.
3. Apatiniame sluoksnyje yra duomenų konteineriai ir jų interfeisai.

Pagalbiniai moduliai:

1. Grafikų braižymo modulis. Atsakingas už įvairiausių grafikų ir lentelių, kurios naudojamos testavimui, vizualiam duomenų pateikimui ir braižymui. Papildomas sluoksnis (angl.: wrapper) virš „matplotlib“ modulio.
2. Progreso sekimo modulis. Atsakingas už analizės ir kitų ilgai vykstančių procesų veikimo progreso pastovų sekimą, jų būsenos pranešimą vartotojui bei atliekamų veiksmų užtrukusio laiko skaičiavimą.
3. Pagalbinės matematinio funkcionalumo klasės. Šiose klasėse sudėti paprasčiausi algoritmai ir matematinės funkcijos, siekiant išvengti programinio kodo pasikartojimų.

Python programavimo kalba palaiko „ančiuko tipus“ (angl.: duck typing) ir neturi įprastos enkapsuliacijos klasėje (visi klasės laukai ir metodai yra pasiekiami iš išorės). Šios savybės leidžia greičiau kurti programinę įrangą, tačiau kartu atsiranda papildomos rizika: nekorektiškas programinio kodo veikimas ir papildomas sudėtingumas plečiant funkcionalumą. Norint išvengti nepageidaujamų pašalinių efektų, būtina apibrėžti architektūrą su aiškiais priklausomybėmis ir praplėtimo taškais.



6 pav. Modulių išdėstymas platformoje

6 pav. parodyta įgyvendintos platformos struktūra, kuri atitinka iškeltus reikalavimus. Rodyklėmis pavaizduotos priklausomybės tarp skirtingų platformos dalių. Pagalbiniai moduliai gali būti iškviesti iš bet kurios programos dalies. Jie yra skirti funkcionalumui, nesusijusiam nei su vienu iš išvardintų modulių, tačiau svarbūs užtikrinant sklandų visos sistemos veikimą.

Kiekvienas modulis gali iškviesti tik žemiau savęs esančius modulius, o pagalbiniai moduliai negali nieko iškviesti. Tokiu reikalavimu yra apsaugoma nuo netinkamo tolimesnio platformos plėtimosi. Pagrindinė platformos plėtimosi vieta yra viduriniame lygyje pridėdant papildomus analizės metodus.

3. Duomenys

Duomenų pasirinkimas yra labai svarbus atliekant empirinius bandymus. Nors ir yra pasirinktos finansinės duomenų eilutės, tačiau egzistuoja skirtingi porūšiai, tokie kaip: akcijų, valiutų, žaliavų (būsimieji sandoriai) ir daugelis kitų. Šiame darbe analizei pasirinkti NASDAQ akcijų biržos S&P500 kompanijų istoriniai kainos rodikliai.

Viešai ir nemokamai yra prieinami tik istoriniai kasdieniniai ir realiu laiku gaunami duomenys. Norint turėti smulkesnio mastelio duomenis, juos reikia surinkti realiu laiku arba nusipirkti. Šiame darbe yra naudojami istoriniai kasdieniniai duomenys paimti iš viešai prieinamų šaltinių, taip pat platformos pagalba yra surinkti dalies S&P500 kompanijų kainos rodikliai realiu laiku, t.y. pakitimai fiksuojami kas sekundę. Realus laiko duomenys yra surinkti nuo 2017-02-27 iki 2017-04-28.

Nors realaus laiko duomenys yra surinkti kas sekundę, tačiau daugeliu atveju akcijų kaina nekinta taip dažnai. Kadangi iš smulkaus mastelio laiko eilutės galima nesudėtingai padaryti stambesnio mastelio laiko eilutę, tai čia iš sekundinės realaus laiko eilutės yra padaroma laiko eilutė, kurios mastelis yra 10 sekundžių. Ši laiko eilutė yra naudojama kaip mažiausio mastelio duomenys.

Akcijoms yra būdingi smulkinimai, kurie iškreipia nominalią vertę. Smulkinimas yra kiekvienos akcijos pakeitimas į kelias dalis, taip padidinant turimų akcijų skaičių ir analogiškai sumažinant kiekvienos akcijos nominalią vertę. Dėl šios priežasties analizei yra naudojama koreguota vertė, o ne nominalioji, kaip pagrindinis laiko eilutės rodiklis. Sudarant koreguotų verčių laiko eilutę, pati naujausia vertė yra paliekama tokia pati, o kiekvienas smulkinimas yra interpretuojamas kaip smulkinimo santykiu mažinantis visas nominaliąsias vertes nuo pačios prekybos pradžios iki smulkinimo datos.

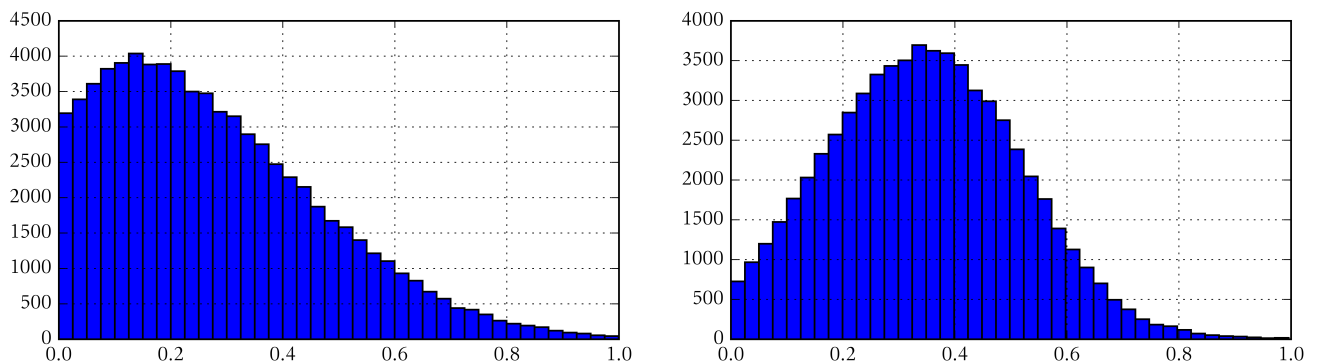
4. Bandymai

4.1. Hursto parametras

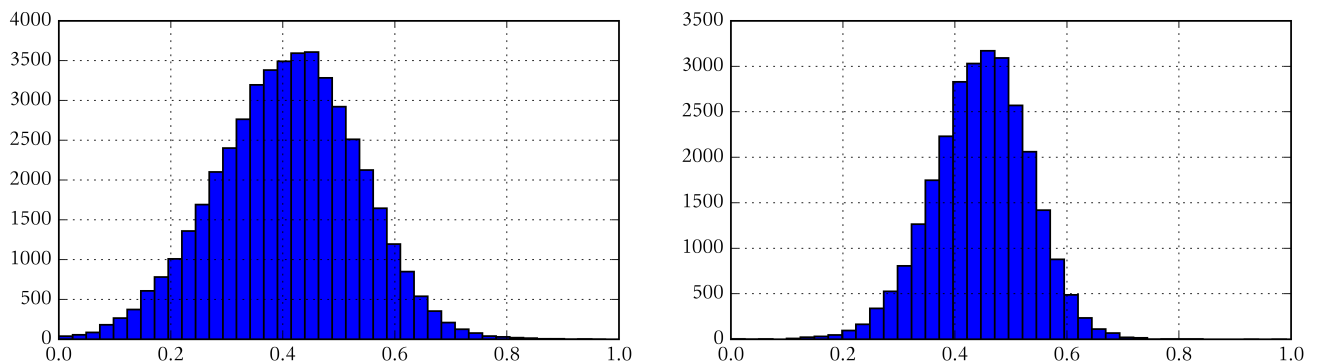
Hursto parametras yra vienareikšmiškai susijęs su grafiko dimensija. Šiame skyriuje nagrinėjama, ar šis parametras yra susijęs su turimomis laiko eilutėmis.

Jei nagrinėjamiems duomenims Hursto parametras būtų apytiksliai lygus 0,5, būtų galima teigti, kad tokie judėjimai yra visiškai atsitiktiniai ir juose yra beprasmiška ieškoti kokių nors dėsningumų. Kadangi dėsningumai yra ieškomi mažuose laiko eilutės gabaluose, Hursto parametras skaičiuosime analogiškuose režiuose. Norint patikrinti, ar parametras gali turėti įtakos kokiam nors dėsningumui, reikia išskirti vieną duomenų režį parametro apskaičiavimui ir kitą duomenų režį, einantį iškart po jo, testavimui. Kadangi turimos laiko eilutės yra labai ilgos, tarp išskirtų režių padarysime tarpus, kad reikšmės būtų labiau nepriklausomos viena nuo kitos. Kiekviename išskirtame režyje suskaičiuosime Hursto parametras, o gautų parametras verčių visumai apžvelgti sudarysime histogramą. Režių ilgiai, kuriuose bus atliekami skaičiavimai, yra 20, 30, 50, 100.

Šios histogramos padarytos naudojant dieninius duomenis. Su mažesnio mastelio duomenimis yra gaunamas toks pats normalusis duomenų pasiskirstymas su tokiu pačiu vidurkiu kiekvienai atitinkamai histogramai.



7 pav. Hursto parametro histograma su režių ilgiais 20 ir 30



8 pav. Hursto parametro histograma su režių ilgiais 50 ir 100

Iš čia matome, kad imant didesnius režius parametro vidurkis artėja prie 0,5, tačiau nepa-

siekia šios vertės ir yra pasiskirstęs kaip normalusis skirstinys. Tai reiškia, kad išskirti režiai laiko eilutėse nėra visiškai atsitiktiniai.

Skaičiuojant Hursto parametrą mažame duomenų rėžyje dažnai pasitaiko situacija, kad visi duomenys yra identiški arba labai panašūs. Kadangi negalima tinkamai nustatyti Hursto parametro tokiose situacijose, jos yra ignoruojamos. Šio tipo situacijose Hursto parametras yra artimas 0.

Toliau bandysime nustatyti, ar Hursto parametras turi kokios nors įtakos laiko eilutės kryptčiai. Kadangi skirtingo ilgio periodams Hursto parametras yra pasiskirstęs skirtingai, pažiūrėsime, ar laiko eilutė pakeičia savo kryptį esant sąlyginai didelei arba mažai reikšmei. Maža reikšmė laikysime pirmojo kvartilio (q_1) reikšmę, o didelė – trečiojo kvartilio (q_3). Šios lentelės taip pat yra sudarytos naudojantis dieniniais duomenimis, nes mažesnio mastelio duomenyse yra matomos labai panašios reikšmės ir pateiktas čia.

S. p.	T. p.	q_1	q_3	pasikeitę ($< q_1$)	prasitęšę ($< q_1$)	pasikeitę ($> q_3$)	prasitęšę ($> q_3$)
20	8	0,1282	0,4029	50,02 %	49,98 %	49,27 %	50,73 %
30	10	0,2260	0,4624	49,29 %	50,71 %	48,88 %	51,12 %
30	15	0,2268	0,4631	49,51 %	50,49 %	49,35 %	50,65 %
30	20	0,2234	0,4613	48,98 %	51,02 %	48,38 %	51,62 %
30	25	0,2254	0,4621	49,5 %	50,5 %	48,54 %	51,46 %
50	10	0,3243	0,4957	49,95 %	50,05 %	49,38 %	50,62 %
50	15	0,3241	0,4954	49,49 %	50,51 %	48,86 %	51,14 %
50	20	0,3247	0,4959	49,13 %	50,87 %	48,66 %	51,34 %
50	25	0,3257	0,4954	49,74 %	50,26 %	46,95 %	53,05 %
50	30	0,3238	0,4943	48,89 %	51,11 %	48,12 %	51,88 %
100	10	0,3920	0,5054	48,59 %	51,41 %	49,28 %	50,72 %
100	20	0,3925	0,5051	48,44 %	51,56 %	47,3 %	52,7 %
100	30	0,3929	0,5047	48,49 %	51,51 %	47,38 %	52,62 %
100	40	0,3928	0,5056	46,96 %	53,04 %	48,52 %	51,48 %

1 lentelė. Hursto parametro įtaka laiko eilutės kryptčiai

Šioje lentelėje stebėjimo periodai (S. p.) rodo, kokio ilgio periode yra skaičiuojama Hursto reikšmė, testavimo periodai (T. p.) rodo, kokio ilgio periodas yra stebimas po Hursto stebėjimo periodų. q_1 ir q_3 yra atitinkamos pirmo ir trečio kvartilų reikšmės. „pasikeitę“ ir „prasitęšę“ vertės lentelėje parodo, kokia dalis išskirtų laiko eilučių pakeitė savo judėjimo kryptį testavimo periode palyginus su stebėjimo periodu, kai Hursto parametras stebėjimo periodu buvo mažesnis už pirmo kvartilio ($< q_1$) arba didesnis už trečio kvartilio ($> q_3$) reikšmę. Rėžiai, kuriuose stebėjimo arba testavimo periodo metu nebuvo pokyčio tarp pradinės ir galinės laiko eilutės reikšmės, buvo ignoruoti.

Iš šitų duomenų galima spręsti, kad santykinai didelis ar mažas Hursto parametras neturi jokios įtakos tokio tipo laiko eilučių tolimesnei vertei. Tokią išvadą galima padaryti matant, jog laiko eilutė pakeičia savo judėjimo kryptį arba išlieka tos pačios krypties beveik vienodą skaičių kartų.

Gautus rezultatus galima sugrupuoti ne pagal reliatyvias Hursto parametro reikšmes, o pagal absoliučias.

S. p.	T. p.	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
20	8	50,24	49,61	50,27	49,59	50,18	49,6	49,53	50,08	50,6	48,87
30	10	47,92	49,69	47,82	49,66	50,42	50,02	49,85	49,08	50,25	49,04
30	15	49,12	48,79	49,75	49,05	49,73	51,14	49,46	49,81	49,27	49,3
30	20	49,33	48,38	50,32	48,29	49,15	49,23	49,21	47,95	49,68	49,36
30	25	50,53	50,05	48,69	49,26	49,18	49,4	49,23	48,98	47,77	48,54
50	10	-	51,81	46,33	49,93	50,3	50,29	49,42	48,86	49,01	49,31
50	15	-	51,39	50,48	48,19	50,0	49,06	50,12	48,26	48,08	49,54
50	20	-	52,45	49,84	50,0	50,31	48,29	48,5	47,97	48,98	47,0
50	25	-	50,0	50,34	50,05	49,71	50,71	48,69	47,86	47,81	46,89
50	30	-	51,55	46,43	48,98	48,4	48,75	49,18	47,86	47,66	48,12
100	10	-	-	-	-	51,12	47,74	48,78	48,27	48,17	48,99
100	20	-	-	-	-	52,36	47,61	47,65	48,9	47,23	46,74
100	30	-	-	-	-	50,57	49,43	48,1	48,07	47,0	47,52
100	40	-	-	-	-	48,1	49,41	45,52	47,24	46,79	47,65

2 lentelė. Hursto parametro įtaka laiko eilutės kryptčiai (1)

S. p.	T. p.	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1
20	8	48,24	50,33	48,83	49,83	48,74	49,33	47,89	46,04	44,77	-
30	10	49,74	47,99	48,0	47,02	48,73	50,45	53,98	-	-	-
30	15	50,22	50,02	47,9	47,5	49,64	53,33	43,4	-	-	-
30	20	47,77	47,27	48,34	51,4	49,13	48,64	47,79	-	-	-
30	25	48,64	47,88	49,06	48,44	48,99	45,53	42,21	-	-	-
50	10	49,17	49,31	50,0	48,95	49,78	-	-	-	-	-
50	15	48,47	47,31	50,87	48,03	51,09	-	-	-	-	-
50	20	49,88	48,43	47,52	47,61	47,95	-	-	-	-	-
50	25	46,68	47,56	47,68	48,47	43,24	-	-	-	-	-
50	30	47,55	47,68	49,09	48,05	49,19	-	-	-	-	-
100	10	49,21	49,39	51,93	49,66	-	-	-	-	-	-
100	20	46,08	48,16	49,84	46,61	-	-	-	-	-	-
100	30	46,39	48,68	47,97	50,0	-	-	-	-	-	-
100	40	48,02	48,63	45,34	56,78	-	-	-	-	-	-

3 lentelė. Hursto parametro įtaka laiko eilutės kryptčiai (2)

Šiose lentelėse stebėjimo periodai (S. p.) ir testavimo periodai (T. p.) turi tokią pačią reikšmę kaip ir prieš tai buvusioje lentelėje. Visi kiti stulpeliai nurodo, kokia dalis (procentais) laiko eilučių rėžių pakeitė savo kryptį, lyginant stebėjimo ir testavimo periodus, turinčius ne didesnę nei nurodytą ir ne mažesnę nei prieš tai buvusią Hursto reikšmę. Langeliuose, kuriuose vietoje reikšmės įrašytas „-“, imtis buvo mažesnė nei 100, todėl tos vertės nebuvo įtrauktos kaip nereikšmingos.

Gautuose rezultatuose matoma, jog nei viena reikšmė nėra ženkliai nutolusi nuo 50% ribos. Pačios mažiausios vertės, kurių reikšmė mažesnė nei 45%, yra ($S.p. = 30; T.p. = 15; 0,8 < H \leq 0,85$), ($S.p. = 30, T.p. = 25, 0,8 < H \leq 0,85$) ir ($S.p. = 20, T.p. = 8, 0,9 < H \leq 0,95$)¹.

¹H - hursto parametras

Pirmosios dvi išvardintos reikšmės yra visiškai netolygios palyginus su gretimomis reikšmėmis, be to, turi mažą imtį, atitinkamai 136 ir 154, kuri yra daug mažesnė, negu kitų reikšmių (vidutiniškai nuo 2000 iki 6000). Dėl šių priežasčių šias reikšmes laikysime nereikšmingomis.

Iš rezultatų, gautų išanalizavus Hursto parametą, matome, kad jis nėra tinkamas prognozavimui.

4.2. Slenkantis vidurkis

Slenkantys vidurkiai (SV) yra naudojami laiko eilutės charakterizavimui. Pasinaudoję slenkančiu vidurkiu patikrinsime, ar jis gali turėti įtakos laiko eilučių prognozavimui.

Šiame darbe yra aprašyti keli skirtingi slenkantieji vidurkiai: paprastasis slenkantis vidurkis, eksponentinis slenkantis vidurkis ir tiesiškai išlygintas slenkantis vidurkis. Išbandysime visas slenkančių vidurkių variacijas ir palyginsime skirtumus tarp jų. Pagal [Val08] aprašytą pastovaus žingsnio slankiojo vidurkio strategiją, slenkančio vidurkio buvimas virš arba žemiau esamos laiko eilutės vertės yra indikatorius. Laiko eilutės reikšmė praėjus fiksuotam periodų skaičiui turėtų priklausyti nuo šio rodiklio.

Tirdami šį indikatorių pasinaudosime tokiais pat duomenų režiais kaip ir atlikdami bandymus su Hursto parametru. Laiko eilutėje išskiriami režiai tolygiais atstumais, kuriuose suskaičiuojama slenkančio vidurkio reikšmė, o nuo šio režio nutolus fiksuotu periodų skaičiumi, tikrinama laiko eilutės reikšmė.

S. p.	T. p.	Vidurkis didesnis už reikšmę		Vidurkis mažesnis už reikšmę	
		didesnė	mažesnė	didesnė	mažesnė
20	8	56,17 %	43,82 %	54,52 %	45,47 %
30	10	56,01 %	43,98 %	55,11 %	44,88 %
30	15	57,26 %	42,73 %	56,27 %	43,72 %
30	20	58,01 %	41,98 %	57,23 %	42,76 %
30	25	58,46 %	41,53 %	57,61 %	42,38 %
50	10	56,20 %	43,79 %	55,00 %	44,99 %
50	15	57,39 %	42,60 %	56,40 %	43,59 %
50	20	58,56 %	41,43 %	57,33 %	42,66 %
50	25	59,05 %	40,94 %	58,12 %	41,87 %
50	30	59,71 %	40,28 %	58,70 %	41,29 %
100	10	56,08 %	43,91 %	54,63 %	45,36 %
100	20	56,83 %	43,16 %	57,16 %	42,83 %
100	30	57,94 %	42,05 %	58,20 %	41,79 %
100	40	61,81 %	38,18 %	60,24 %	39,75 %

4 lentelė. Pastovaus žingsnio slankiojo vidurkio strategija su paprastu slenkančiu vidurkiu

S. p.	T. p.	Vidurkis didesnis už reikšmę		Vidurkis mažesnis už reikšmę	
		didesnė	mažesnė	didesnė	mažesnė
20	8	56,33 %	43,66 %	54,37 %	45,62 %
30	10	56,44 %	43,55 %	54,77 %	45,22 %
30	15	57,44 %	42,55 %	56,12 %	43,87 %
30	20	57,95 %	42,04 %	57,27 %	42,72 %
30	25	58,56 %	41,43 %	57,52 %	42,47 %
50	10	56,40 %	43,59 %	54,84 %	45,15 %
50	15	57,45 %	42,54 %	56,35 %	43,64 %
50	20	58,45 %	41,54 %	57,39 %	42,60 %
50	25	59,13 %	40,86 %	58,05 %	41,94 %
50	30	59,60 %	40,39 %	58,75 %	41,24 %
100	10	55,92 %	44,07 %	54,69 %	45,30 %
100	20	57,32 %	42,67 %	56,87 %	43,12 %
100	30	58,27 %	41,72 %	58,00 %	41,99 %
100	40	62,15 %	37,84 %	59,97 %	40,02 %

5 lentelė. Pastovaus žingsnio slankiojo vidurkio strategija su tiesiškai svoriniu slenkančiu vidurkiu

S. p.	T. p.	Vidurkis didesnis už reikšmę		Vidurkis mažesnis už reikšmę	
		didesnė	mažesnė	didesnė	mažesnė
20	8	56,17 %	43,82 %	54,53 %	45,46 %
30	10	56,13 %	43,86 %	55,02 %	44,97 %
30	15	57,39 %	42,60 %	56,18 %	43,81 %
30	20	57,97 %	42,02 %	57,27 %	42,72 %
30	25	58,52 %	41,47 %	57,56 %	42,43 %
50	10	56,14 %	43,85 %	55,04 %	44,95 %
50	15	57,56 %	42,43 %	56,30 %	43,69 %
50	20	58,53 %	41,46 %	57,35 %	42,64 %
50	25	59,16 %	40,83 %	58,05 %	41,94 %
50	30	59,65 %	40,34 %	58,74 %	41,25 %
100	10	56,04 %	43,95 %	54,65 %	45,34 %
100	20	56,96 %	43,03 %	57,09 %	42,90 %
100	30	57,94 %	42,05 %	58,20 %	41,79 %
100	40	61,58 %	38,41 %	60,38 %	39,61 %

6 lentelė. Pastovaus žingsnio slankiojo vidurkio strategija su eksponentiškai slenkančiu vidurkiu

Aukščiau pateiktose lentelėse S.p. (skaičiavimo periodai) – periodų skaičius, kuriame apskaičiuojama slenkančio vidurkio reikšmė. T.p. (testavimo periodai) – testuojamoji laiko eilutės reikšmė. Stulpelyje „Vidurkis didesnis už reikšmę“ pateikta, kokioje dalyje visų rėžių testavimo periodo gale laiko eilutės reikšmė buvo didesnė arba mažesnė už laiko eilutės reikšmę slenkančio vidurkio skaičiavimo periodo gale.

Iš šių (4, 5, 6) lentelių matome, jog būsima laiko eilutės reikšmė nepriklauso nuo slenkančio vidurkio ir laiko eilutės santykio, nes visi turimi rezultatai yra panašūs lyginant rezultatus, kai slenkantis vidurkis didesnis už laiko eilutės reikšmę ir kai slenkantis vidurkis yra mažesnis už

laiko eilutės reikšmę. Visi šių lentelių rezultatai suskaičiuoti naudojantis istoriniais duomenimis.

Taip pat galima pastebėti, kad tarp šių lentelių analogiškų reikšmių beveik nėra jokio prasmingo skirtumo, dėl to laikysime, kad skirtingi slenkančių vidurkių skaičiavimo būdai neturi įtakos ir toliau naudosime tik paprastąjį slenkantį vidurkį.

Laiko eilutės reikšmės dažniau didėjančios, o ne mažėjančios dėl turimų duomenų specifikos, dėl to testavimo periodo pabaigoje didesnė yra kylančių reikšmių dalis.

Nagrinėjant slenkančius vidurkius išskirtinis dėmesys yra skiriamas taškams, kuriuose laiko eilutės reikšmė kertasi su slenkančio vidurkio reikšme. Pakartojame tą patį eksperimentą vietoje tolygiai išsidėsčiusių režių surasdami taškus, kuriuose laiko eilutės reikšmė kertasi su slenkančio vidurkio reikšme.

S. p.	T. p.	Vidurkis didesnis už reikšmę		Vidurkis mažesnis už reikšmę	
		didesnė	mažesnė	didesnė	mažesnė
20	8	54,06 %	45,93 %	55,17 %	44,82 %
30	10	54,90 %	45,09 %	55,42 %	44,57 %
30	15	56,24 %	43,75 %	56,58 %	43,41 %
30	20	57,44 %	42,55 %	57,33 %	42,66 %
30	25	58,08 %	41,91 %	58,16 %	41,83 %
50	10	55,02 %	44,97 %	55,19 %	44,80 %
50	15	56,28 %	43,71 %	56,54 %	43,45 %
50	20	57,35 %	42,64 %	57,33 %	42,66 %
50	25	58,07 %	41,92 %	58,31 %	41,68 %
50	30	58,93 %	41,06 %	59,13 %	40,86 %
100	10	54,17 %	45,82 %	55,53 %	44,46 %
100	20	56,95 %	43,04 %	57,35 %	42,64 %
100	30	58,32 %	41,67 %	58,76 %	41,23 %
100	40	59,63 %	40,36 %	59,85 %	40,14 %

7 lentelė. Laiko eilutės reikšmės kitimas po slenkančio vidurkio ir laiko eilutės susikirtimo

Lentelėje 7 stulpelių reikšmės yra tokios pačios kaip ir lentelėse 4, 5, 6, todėl jos nebus detalizuojamos. Kadangi šios reikšmės yra itin panašios į 4, 5 ir 6 lentelių reikšmes, tai rodo, jog susikirtimo taškai neturi įtakos laiko eilutės kryptim.

Visos šiame skyriuje parodytos lentelės yra sukurtos naudojantis istoriniais (vienos paros) duomenimis. Pakartosime šiuos bandymus su smulkesnio mastelio duomenimis.

S. p.	T. p.	Vidurkis didesnis už reikšmę		Vidurkis mažesnis už reikšmę	
		didesnė	mažesnė	didesnė	mažesnė
20	8	46,65 %	53,34 %	53,67 %	46,32 %
30	10	46,45 %	53,54 %	53,71 %	46,28 %
30	15	46,91 %	53,08 %	53,16 %	46,83 %
30	20	47,23 %	52,76 %	52,83 %	47,16 %
30	25	47,54 %	52,45 %	52,65 %	47,34 %
50	10	46,31 %	53,68 %	53,90 %	46,09 %
50	15	46,73 %	53,26 %	53,41 %	46,58 %
50	20	47,03 %	52,96 %	52,96 %	47,03 %
50	25	47,34 %	52,65 %	52,85 %	47,14 %
50	30	47,50 %	52,49 %	52,66 %	47,33 %
100	10	46,26 %	53,73 %	54,30 %	45,69 %
100	20	47,05 %	52,94 %	53,43 %	46,56 %
100	30	47,38 %	52,61 %	53,02 %	46,97 %
100	40	47,99 %	52,00 %	52,43 %	47,56 %

8 lentelė. Laiko eilutės reikšmės kitimas po slenkančio vidurkio ir laiko eilutės susikirtimo (realaus laiko duomenys)

8-oje lentelėje matome visiškai priešingą situaciją negu 7-oje lentelėje. Laiko eilutės reikšmė po slenkančio vidurkio ir laiko eilutės susikirtimo taško dažniau leidžiasi, kai vidurkis tampa didesnis ir dažniau kyla, kai slenkantis vidurkis tampa mažesnis.

Kyla klausimas, ar šis dėsningumas yra būdingas tik duomenims su mažais masteliais, kadangi toks dėsningumas negalioja kasdieniniams duomenims. Tam patikrinti naudosime duomenų rinkinį, kurio mastelis yra minutė.

S. p.	T. p.	Vidurkis didesnis už reikšmę		Vidurkis mažesnis už reikšmę	
		didesnė	mažesnė	didesnė	mažesnė
20	8	48,28 %	51,71 %	52,07 %	47,92 %
30	10	48,13 %	51,86 %	51,80 %	48,19 %
30	15	48,88 %	51,11 %	51,16 %	48,83 %
30	20	49,26 %	50,73 %	51,16 %	48,83 %
30	25	49,26 %	50,73 %	50,99 %	49,00 %
50	10	48,03 %	51,96 %	52,13 %	47,86 %
50	15	48,56 %	51,43 %	51,52 %	48,47 %
50	20	48,92 %	51,07 %	51,45 %	48,54 %
50	25	48,94 %	51,05 %	51,15 %	48,84 %
50	30	49,20 %	50,79 %	51,31 %	48,68 %
100	10	47,57 %	52,42 %	51,53 %	48,46 %
100	20	48,61 %	51,38 %	51,38 %	48,61 %
100	30	48,84 %	51,15 %	51,15 %	48,84 %
100	40	49,17 %	50,82 %	51,13 %	48,86 %

9 lentelė. Laiko eilutės reikšmės kitimas po slenkančio vidurkio ir laiko eilutės susikirtimo (minutiniai realaus laiko duomenys)

Lentelėje 9 yra akivaizdžiai mažesnė atskirtis tarp pakilusios ir nusileidusios laiko eilutės

reikšmės dalių po testavimo periodo pabaigos. Lentelėje 8 yra matomas nuo 8,35% iki 18,84% (vidutiniškai apie 13%) polinkio į vieną pusę skirtumas, o lentelėje 9, kurioje mastelis yra didesnis šis skirtumas svyruoja nuo 2,98% iki 10,19% (vidutiniškai apie 5–6%). Iš šių rezultatų galima aiškiai matyti, jog toks dėsningumas galioja tik smulkesnio mastelio duomenims.

Toliau panagrinėsime kaip šis rodiklis kinta priklausomai nuo periodų ilgių kuriuose yra skaičiuojamas slenkantis vidurkis ir stebima kaip keičiasi reikšmė.

100	54.23	53.92	53.73	53.29	52.95	52.81	52.62	52.20	52.01	51.81	51.78	51.58	51.66	51.49	51.44	51.28	51.16
90	54.24	53.92	53.77	53.29	52.89	52.79	52.55	52.18	52.04	51.78	51.75	51.58	51.69	51.53	51.48	51.29	51.14
80	54.19	53.80	53.68	53.34	53.03	52.91	52.53	52.10	52.03	51.82	51.71	51.58	51.62	51.48	51.43	51.22	51.07
70	54.26	53.87	53.65	53.32	52.92	52.70	52.49	52.05	51.95	51.70	51.62	51.56	51.47	51.38	51.32	51.09	50.97
60	54.18	53.90	53.82	53.42	53.08	52.76	52.50	52.18	52.06	51.81	51.65	51.55	51.45	51.40	51.35	51.20	51.09
50	54.25	53.85	53.69	53.27	52.96	52.66	52.50	52.12	51.93	51.80	51.63	51.53	51.46	51.33	51.33	51.21	51.05
40	54.29	53.93	53.75	53.28	52.94	52.60	52.46	52.06	51.97	51.77	51.62	51.52	51.53	51.29	51.26	51.15	51.00
30	54.02	53.64	53.55	53.08	52.77	52.45	52.26	51.88	51.75	51.61	51.52	51.44	51.35	51.16	51.12	50.99	50.92
20	53.80	53.34	53.17	52.76	52.54	52.21	52.05	51.70	51.59	51.52	51.40	51.41	51.32	51.13	51.05	50.99	50.87
	6	8	10	15	20	25	30	35	40	45	50	55	60	70	80	90	100

9 pav. Spalvinė duomenų lentelė parodanti kokioje dalyje (%) laiko eilučių reikšmė sumažėja testavimo periode (horizontali ašis), kai slenkantis vidurkis pasidaro didesnis už laiko eilutės reikšmę ir yra apskaičiuojamas nurodytu periodų skaičiumi (vertikali ašis). Rėžiai sudaryti laiko eilutės ir slenkancio vidurkio susikirtimo taškuose naudojant realaus laiko duomenis.

Iš 9 pav. galima aiškiai matyti, jog kuo mažesnis testavimo periodo ilgis tuo didesnė dalis laiko eilučių rėžių elgiasi pagal nustatytą dėsningumą. Taip pat, matome, kad esant mažam periodų skaičiui (≤ 30) su kuriais yra skaičiuojamas slenkantis vidurkis pastebimai sumažėja laiko eilučių dalis tenkinanti šį dėsningumą.

Apibendrinant slenkančių vidurkių analizę galima pastebėti, jog nėra ryškaus skirtumo tarp skirtingų slenkančių vidurkių. Taip pat slenkantis vidurkis bendruoju atveju nesuteikia informacijos laiko eilutės prognozavimui. Tačiau naudojant smulkaus mastelio duomenis taškuose, kuriuose slenkantis vidurkis susikerta su laiko eilute, galima pasinaudoti pastebėtu tendencingumu prognozavimui.

4.3. Kanalai

Kanalai yra naudojami norint apibrėžti kryptį, kuria juda laiko eilutė. Jei laiko eilutės reikšmė išeina iš kanalo ribų tuomet tai gali rodyti pasikeitimus laiko eilutės kryptyje.

Kanalas susidaro iš dviejų ribų – viršutinės ir apatinės, tarp kurių paprastai būna laiko eilutės reikšmės. Egzistuoja daugybė įvairių kanalų variacijų. Seniau dažniau būdavo naudojami iš tiesių linijų sudaryti ir pagal diskrečius rodiklius parinkti kanalai, nes skaičiavimai būdavo atliekami rankomis. Naudojantis šiuolaikinėmis technologijomis galima daug greičiau atlikti paprastus skaičiavimus, todėl populiariesni yra tokie kanalai kaip Bollingerio juosta, kuri susidaro iš judancio vidurkio ir juostų nukrypusių nuo šio vidurkio per tam tikrą standartinių nuokrypų skaičių.

Kompiuteriui yra palankesni tie metodai, kurių skaičiavimus yra paprasta automatizuoti. Dauguma senų metodų, kurie remiasi iš akies brėžiamais rėžiais yra netinkami, nes tai yra neapibrėžta ir netikslu. Taigi, norint išanalizuoti didelį duomenų kiekį reikia griežtų ir tikslių kriterijų. Toliau pabandydysime apibrėžti tokius kriterijus ir išanalizuoti, kaip kanalais galima charakterizuoti laiko eilutes.

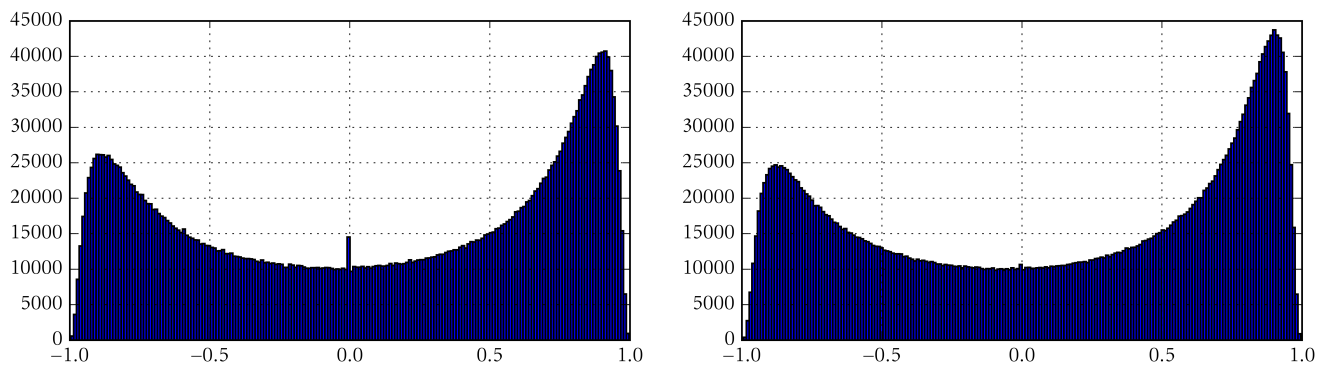
4.3.1. Koreliacijos koeficientas

Didžiausia problema yra nuoseklios srities laiko eilutėje pasirinkimas, kuriai galima būtų suformuoti kanalą. Šiai problemai išspręsti pasitelksime koreliacijos koeficientą. Pearsono koreliacijos koeficientas (*PCC*) parodo tiesinę priklausomybę tarp dviejų kintamųjų X ir Y . Nagrinėjame atvejuje kintamieji būtų laiko eilutės reikšmė ir laikas.

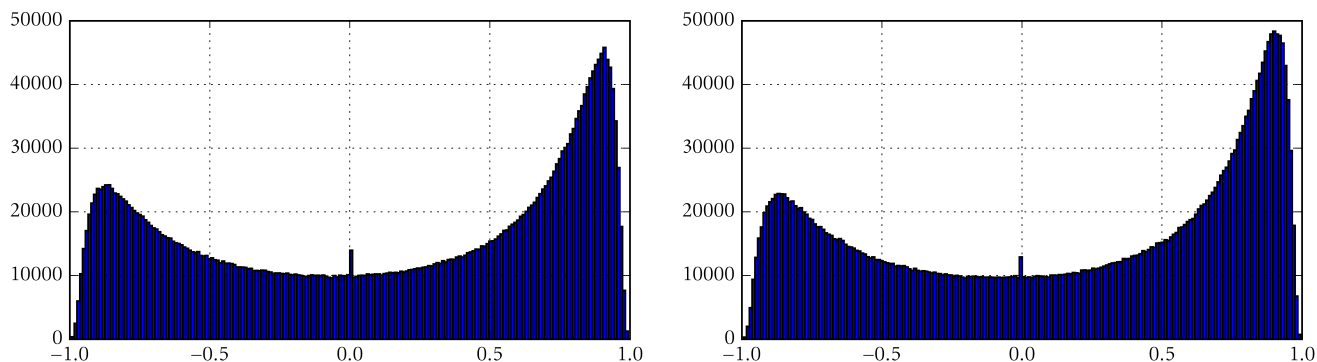
$$PCC_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Šioje išraiškoje $cov(X,Y)$ yra kovariacija, o σ_X yra standartinis nuokrypis. Ši reikšmė gali turėti vertes nuo -1 iki 1 imtinai. 1 reiškia, jog dydžiai X ir Y yra vienas nuo kito pilnai tiesiškai priklausomi, o -1 rodo, jog šie dydžiai yra pilnai atvirkščiai priklausomi. Tačiau šis koeficientas neparodo šio ryšio krypties, tik tai ar jis yra tiesiškas ar atvirkščiai tiesiškas.

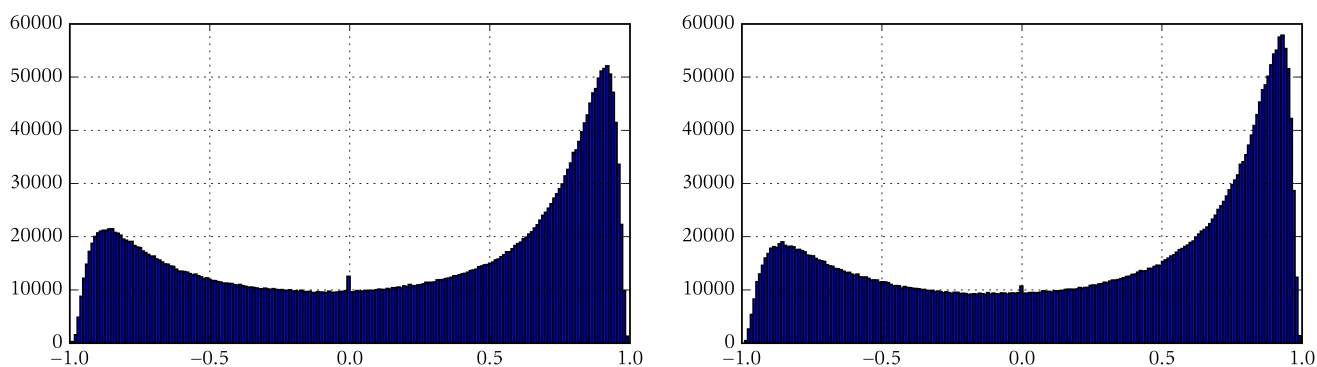
Kita problema pasirenkant nuoseklią sritį laiko eilutėje yra srities ilgis. Norint nustatyti kokie sričių ilgiai yra prasmingi koreliacijos parametro skaičiavimui padarysime histogramas skirtingiems rėžių ilgiams.



10 pav. Koreliacijos parametro histogramos su rėžių ilgiais 15 ir 25



11 pav. Koreliacijos parametro histogramos su rėžių ilgiais 30 ir 45



12 pav. Koreliacijos parametro histogramos su rėžių ilgiais 60 ir 100

Šiose (10, 11 ir 12) diagramose parodytas tam tikro koreliacijos koeficiento pasikartojimo dažnumas visuose turimuose duomenyse. Horizontali ašis žymi koreliacijos koeficiento reikšmę, o vertikali pasikartojimų dažnumą.

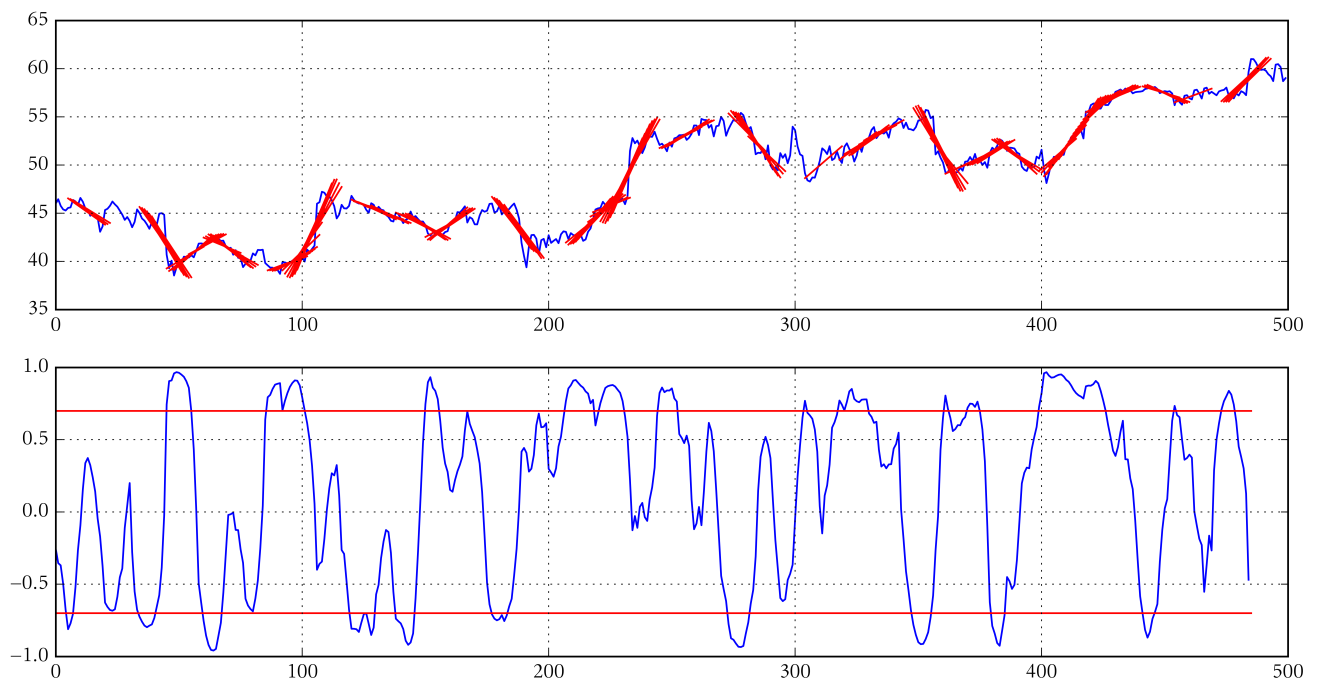
Iš šių duomenų matome, kad pasirinkus didesnę rėžio ilgį, kuriame skaičiuojamas šis koeficientas, daugiau rėžių turės didelį (artimą vienetui) koreliacijos koeficientą negu neigiamą. Šį faktą, jog yra daugiau teigiamų koreliacijos koeficiento reikšmių negu neigiamų galima paaiškinti turimų duomenų laiko eilučių reikšmių polinkiu kilti į viršų, o ne leistis žemyn ilguoju laikotarpiu.

Histogramose taip pat matosi vienas išskirtinis dažnių paaugimas kai koreliacijos koeficiento reikšmė yra 0. Ši išimtis yra gaunama esant nei didėjančioms, nei mažėjančioms laiko eilutės reikšmėms. Tokiu atveju koreliacijos koeficientas yra neapibrėžtas, o turimoje platformoje jam yra priskiriama reikšmė - 0.

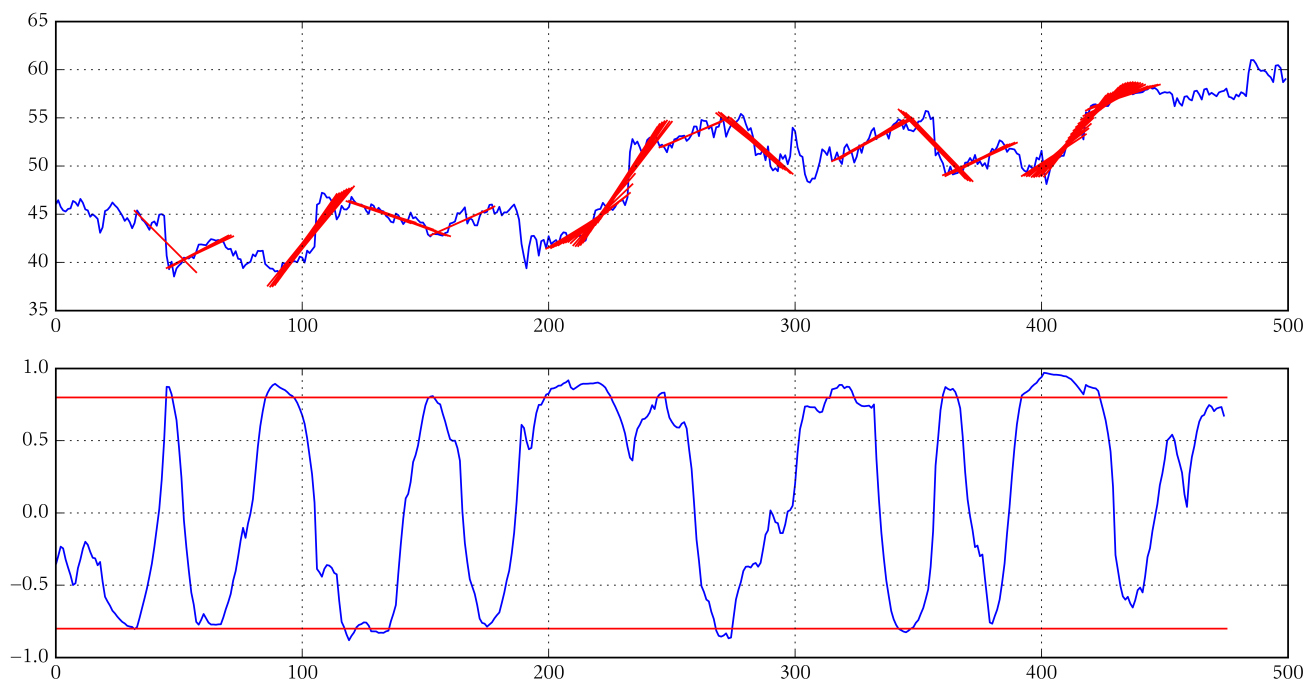
Kadangi imant didelį rėžio ilgį yra gaunama daug daugiau teigiamų koreliacijos reikšmių nei neigiamų, tolimesniuose skaičiavimuose naudosime rėžių ilgius iki 50. Taip pat negalima naudoti mažų rėžių ilgių skaičiavimams, kadangi rezultatai gauti apskaičiavus koreliacijos koeficientą mažam duomenų kiekiui yra netikslūs. Mažiausias rėžių ilgis, kuris bus naudojamas toliau - 20.

Taip pat reikia apibrėžti mažiausią absoliučią koreliacijos koeficiento reikšmę, kuriai esant laikysime, kad turimas laiko eilutės rėžis koreliuoja. Šią reikšmę vadinsime *ribiniu koreliacijos koeficientu*. Toliau pateikiamuose grafikuose parodoma, kaip atrodo koreliacijos koeficiento iš-

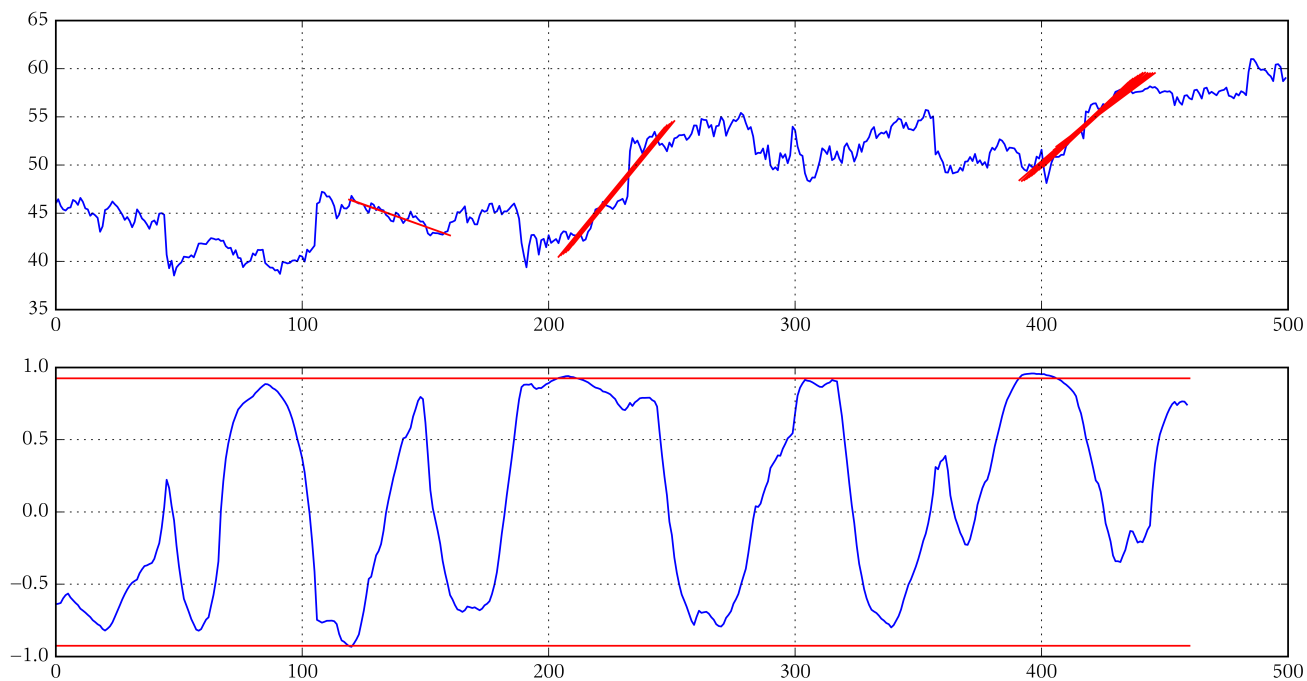
sidėstymas laike kartu su pavyzdiniuose duomenyse pažymėtomis zonomis, kurios yra aptiktos kaip koreliuojančios.



13 pav. Viršuje: pavyzdiniai duomenys su pažymėtomis koreliuojamomis zonomis. Apačioje: koreliacijos koeficiento reikšmė esamu laiko periodu. Rėžio ilgis 15, ribinis koreliacijos koeficientas 0,7



14 pav. Viršuje: pavyzdiniai duomenys su pažymėtomis koreliuojamomis zonomis. Apačioje: koreliacijos koeficiento reikšmė esamu laiko periodu. Rėžio ilgis 25, ribinis koreliacijos koeficientas 0,8



15 pav. Viršuje: pavyzdiniai duomenys su pažymėtomis koreliuojamomis zonomis. Apačioje: koreliacijos koeficiento reikšmė esamu laiko periodu. Rėžio ilgis 40, ribinis koreliacijos koeficientas 0,925

Šiose (13, 14 ir 15) diagramose horizontali ašis atitinka laiko eilutės momentą. Viršutinėje diagramos dalyje yra parodyta laiko eilutės reikšmė tam tikru laiko momentu kartu su nubrėžta geriausiai duomenis aproksimuojančia linija. Apatinėje diagramos dalyje yra koreliacijos koeficiento reikšmė tuo pačiu laiko periodu kartu su pažymėtu ribiniu koreliacijos koeficientu.

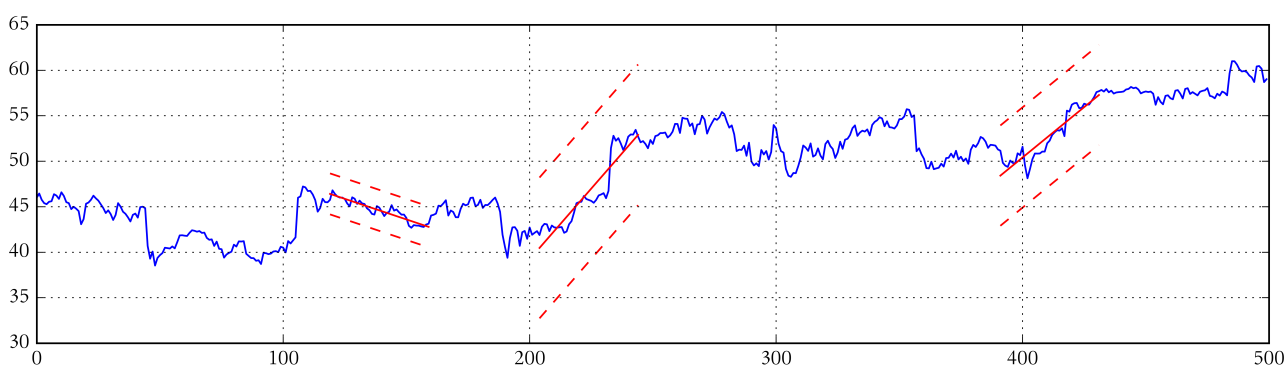
Iš pateiktų diagramų matome, jog kai ribinis koreliacijos koeficientas yra 0,7, fiksuojamos beveik visos laiko eilutės sritys, todėl ši reikšmė yra akivaizdžiai per maža. Diagramoje 14 su ribiniu koreliacijos koeficientu 0,8 tokių sričių yra pakankamai daug, bet į jas neįeina visos laiko eilutės reikšmės. Diagramoje 15 su ribiniu koreliacijos koeficientu 0,925 liko tik 3 tokios sritys. Remdamiesi apdorotais duomenimis nustatėme, kad ribinio koreliacijos koeficiento reikšmė neturėtų būti mažesnė nei 0,8.

Pastebėjome, kad kuo didesnis yra pasirenkamas režio ilgis, tuo koreliacijos koeficiento grafikas banguoja mažiau.

Koreliacijos koeficientas parodo ar laiko eilutės sritis yra koreliuojanti. Nustatydami laiko eilutės regresijos tiesę naudojame mažiausių kvadratų metodą. Tokiu būdu sudarytą tiesės lygtį vadinsime *regresijos linija*.

4.3.2. Paprastas kanalas

Patys primityviausi kanalai buvo sudaromi vizualiai pasirenkant režį ir jame nubrėžiant ribas į kurias patektų daugelis tos laiko eilutės reikšmių. Naudodami koreliacijos koeficiento reikšmes ir nustčius koreliuojančias sritis laiko eilutėje, galima sudaryti kanalą. Paprasto kanalo ribos sudaromos prie regresijos linijos pridendant ir atimant tam tikrą kiekį standartinių nuokrypių. Standartinis nuokrypis skaičiuojamas iš laiko eilutės reikšmių esančių režyje.

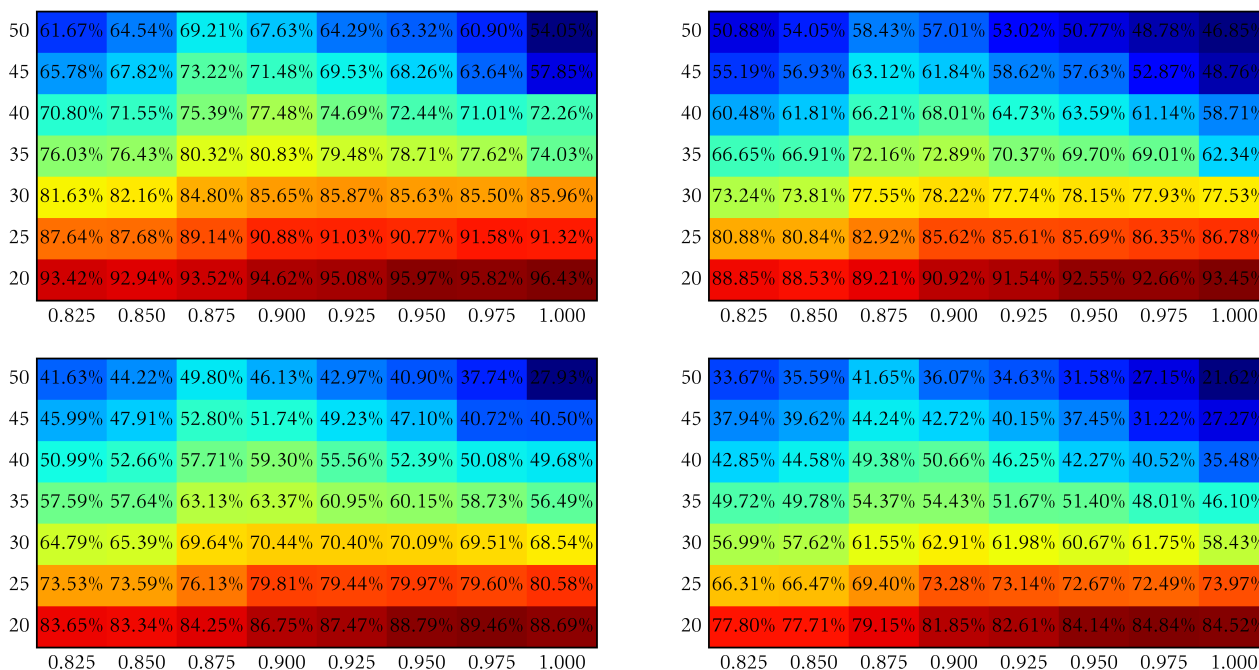


16 pav. Pavyzdiniai duomenys su sudarytais kanalais. Rėžio ilgis ir koreliacijos koeficientas atitinka 15 pav. (40 ir 0,925)

Kanalo pagrindinis tikslas yra numatyti laiko eilutės tendenciją apibrėžiant ribas, kuriose gali būti laiko eilutės reikšmės. Reikšmės išėjimas iš kanalo (lūžis) paprastai rodo laiko eilutės tendencijos pokytį. Statistiškai apskaičiuosime kokiuose kanaluose ir kada įvyksta lūžiai.

Tai patikrinsime pratęsdami kanalą tam tikrą periodų kiekį. Kadangi jau apibrėžėme, jog režių ilgiai yra nuo 20 iki 50 periodų, tai pratęsimė kanalą minimalaus režio ilgiu, kuris neturėtų būti nei per didelis, nei per mažas periodas. *Kanalo pločiu* (KS) vadinsime intervalo, kuriame bus

fiksuojamose laiko eilutės reikšmės, ilgį. Šis kanalas bus sudaromas apie regresijos liniją. Kanalo plotį matuosime standartiniu nuokrypiu. Kalbant tiksliau jo reikšmę parinksime artimą 2 standartiniams nuokrypiams, kadangi ši reikšmė yra naudojama standartiniame Bollingerio kanale. Jei laiko eilutės pokyčių reikšmės būtų pasiskirsčiusios pagal normalųjį skirstinį, tuomet statistiškai į tokį kanalą patektų 95% visų reikšmių ir pralaužimo taškai būtų tik išskirtys.



17 pav. Spalvinės duomenų lentelės parodančios kokioje dalyje laiko eilučių kanalas pralaužiamas pratęsus jį 20 periodų po koreliuojančio režio radimo. Kanalų dydžiai einant iš eilės $KS = 1,5; 1,75; 2; 2,25$ standartinių nuokrypių

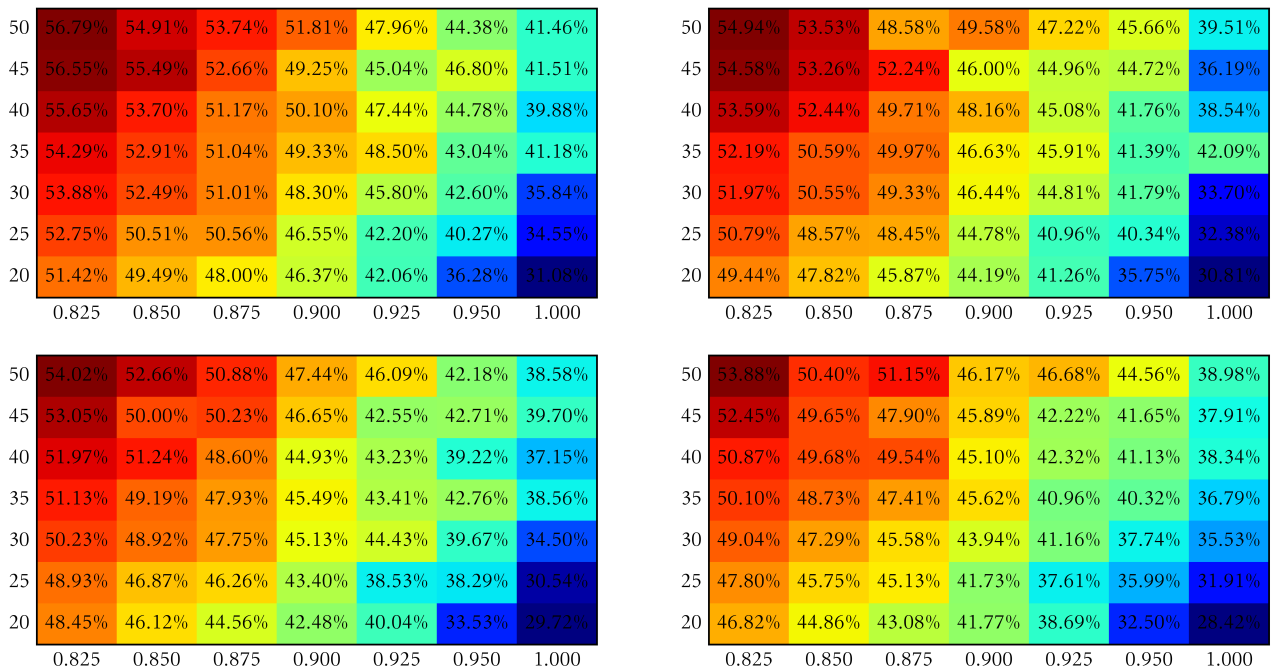
Diagramose (17 pav.) parodyta kokia dalis laiko eilučių reikšmių išeina iš kanalo ribų (toliau: įvyksta kanalo lūžis) pratęsus kanalą 20-ia periodų už koreliuojančio režio radimo taško. Horizontali ašis parodo laiko eilutės režio koreliacijos koeficiento ribas, tai yra vertė 0,825 reiškia, kad tame stulpelyje yra režiai, kurių koreliacijos koeficientas (KK) yra $0,8 \leq KK < 0,825$, 0,850 atitinka $0,825 \leq KK < 0,85$ ir taip toliau. Vertikali ašis yra laiko eilutės režio ilgis. Kiekviename langelyje nurodyta laiko eilučių dalis (procentais), kada įvyksta kanalo pralaužimas. Spalvos parinktos kiekvienai lentelei atskirai nuo tamsiai mėlynos mažiausiai reikšmei iki tamsiai raudonos aukščiausiai reikšmei.

Pastebima, jog kai režio dydis yra 20 beveik visos laiko eilutės užsibaigia pralaužimu. Taip pat pastebime, kad kai kanalo plotis yra $KS = 1,5$, tuomet labai didelė laiko eilučių dalis užsibaigia pralaužimais, kas rodo, jog toks kanalo plotis yra per mažas prasmingiems rezultatams išgauti.

Šios lentelės parodo, kad kanalų lūžio dažnis nepriklauso vienareikšmiškai nuo koreliacijos koeficiento ar režio ilgio, o greičiau nuo šių ir gal dar kitų parametų kombinacijos. Taip pat galima daryti keletą kitų išvadų:

- Esant stabiliam koreliacijos koeficientui ir didinant rėžio ilgį mažėja sudaryto kanalo pralaužimų.
- Esant stabiliam koreliuojančio rėžio ilgiui ir didinant koreliacijos koeficientą kanalo pralaužimų skaičius nėra pastovus.
- Kuo didesnis yra kanalo plotis, tuo rečiau pralaužiamas kanalas. Tai matoma iš to, kad 17 pav. kiekviename lentelės langelyje esanti reikšmė yra mažesnė už analogišką reikšmę tame pačiame langelyje prieš tai buvusioje lentelėje.

Laiko eilutės reikšmė gali sugrįžti (toliau: grįžimas) į kanalą lygiai taip pat kaip ir buvo iš jo prasilaužta, todėl statistiškai pažiūrėsime kokia dalis laiko eilutės rėžių grįžta į kanalą jį pralaužus.



18 pav. Spalvinės duomenų lentelės parodančios kokioje dalyje laiko eilučių reikšmė po kanalo pralaužimo grįžta į kanalą pratęsus jį 20 periodų po koreliuojančio rėžio radimo. Kanalų dydžiai einant iš eilės $KS = 1,75; 2; 2,25; 2,5$ standartinių nuokrypių

Diagramose (18 pav.) parodyta kokia dalis laiko eilučių reikšmių grįžta į kanalą po jo pralaužimo tuose pačiuose 20 periodų, kuriems buvo pratęstas kanalas. Taip pat kaip ir 17 pav. vertikali ašis parodo rėžio ilgį, horizontali - koreliacijos koeficientą. Kiekviename langelyje dalis laiko eilučių (procentais), kuriose įvyksta grįžimas. Spalvos taip pat parinktos analogiškai.

Kadangi laiko eilutės reikšmės grįžimas į kanalą gali įvykti tik po pralaužimo, todėl duomenų, iš kurių sudarytos šios lentelės yra mažiau. Dėl sumažėjusios imties yra matomas išsibarstymas reikšmėse tarp lentelės langelių reikšmių. Taip pat nebėra atskiro stulpelio $0,975 \leq KK < 1$ vertei, nes tokios situacijos yra labai retos ir imties neužteko atskirų verčių suformavimui.

Palyginus 17 pav. ir 18 pav. diagramų grupes galima pastebėti keletą dėsningumų:

- Priešingai nei kanalo pralaužimo statistikoje, grįžimų diagramos yra išsidėsčiusios tolygiai: kuo didesnis režio dydis, tuo daugiau laiko eilučių reikšmių grįžtančių į savo pradinę vertę; taip pat, kuo didesnis koreliacijos koeficientas, tuo mažesnė laiko eilučių reikšmių dalis grįžta į kanalą.
- Priešingai nei kanalo pralaužimo lentelėse, šios reikšmės nėra taip smarkiai priklausomos nuo kanalo pločio. Analogiškų režimų rezultatai 17 pav. lyginant tarpusavyje reikšmes gretimuose kanalų pločių lentelėse vidutiniškai skyrėsi apie 8%, o 18 pav. tokie skirtumai siekia daugiausiai 5% tarp $KS = 1,75$ ir $KS = 2,5$ lentelės.

Pralaužimų statistika parodo tam tikrus dėsningumus laiko eilutėse. Toliau nagrinėjama kada įvyksta kanalų lūžiai. Šie atvejai nagrinėjami kanalą pratęsus 30-ia periodų, kad aiškiau matytųsi dėsningumai grafikuose.

0.950-1.000	0.36	0.93	2.69	4.37	5.66	5.73	7.89	7.46	6.45	7.10	6.38	6.24	5.77	5.23	4.84	4.37	3.15	2.76	2.11	2.33	1.76	1.33	1.11	0.82	0.90	0.86	0.47	0.18	0.36	0.39
0.925-0.950	0.79	2.08	3.44	5.34	6.88	6.58	7.16	6.98	7.28	6.44	6.81	5.96	5.43	4.50	3.56	2.96	2.82	2.75	2.10	1.78	1.46	1.29	1.06	1.16	0.95	0.58	0.58	0.46	0.30	0.53
0.900-0.925	1.16	3.18	4.72	6.13	6.34	7.16	6.58	6.78	6.81	6.00	6.25	4.69	4.35	3.88	3.46	3.26	3.00	2.74	2.09	1.88	1.81	1.47	1.42	1.05	0.75	0.79	0.70	0.65	0.40	0.49
0.875-0.900	1.66	3.99	5.42	5.62	6.63	7.06	6.33	6.52	6.50	6.10	5.63	4.89	3.90	3.90	3.27	2.97	2.79	2.57	2.19	1.98	1.44	1.45	1.46	1.33	1.03	0.70	0.83	0.69	0.56	0.59
0.850-0.875	1.92	3.85	5.30	5.93	6.51	6.31	6.36	5.75	6.03	6.03	5.33	4.97	4.45	3.65	3.16	3.09	2.74	2.70	2.22	2.09	1.72	1.71	1.50	1.51	1.27	1.01	0.71	0.81	0.76	0.60
0.825-0.850	2.11	4.22	5.18	6.02	5.98	5.89	6.21	5.85	5.78	5.45	5.11	4.47	4.24	4.32	3.46	3.27	2.97	2.70	2.25	2.08	1.94	1.77	1.66	1.40	1.33	1.08	0.92	0.95	0.76	0.65
0.800-0.825	2.65	4.66	5.52	6.19	6.26	6.16	5.92	5.61	5.57	5.26	4.87	4.54	4.36	3.99	3.43	3.09	2.94	2.65	2.46	2.08	1.88	1.56	1.55	1.40	1.25	1.05	0.98	0.80	0.76	0.57
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29

19 pav. Spalvinės duomenų lentelės parodančios koku periodu, skaičiuojant nuo koreliuojančio režio pabaigos, kanalas pralaužiamas. Kanalo plotis $KS = 2,0$, režio ilgis 20.

0.950-1.000	0.24	0.78	0.71	1.18	2.12	2.04	2.59	2.90	3.76	5.25	4.70	5.02	5.02	4.86	5.80	4.23	5.96	4.86	4.86	3.45	4.31	4.00	3.53	3.61	2.66	2.12	2.74	2.90	1.25	2.59
0.925-0.950	0.42	0.94	1.94	3.12	2.79	3.45	4.15	4.34	3.97	4.63	4.96	4.58	4.67	4.86	4.53	3.78	4.86	4.25	4.15	3.64	3.68	2.36	2.97	2.97	2.88	2.64	2.03	2.55	2.46	1.42
0.900-0.925	0.61	1.42	2.20	3.17	3.21	3.98	4.67	4.55	4.47	4.96	4.51	5.28	4.51	4.84	4.27	3.98	3.94	3.33	3.54	3.33	3.62	2.76	2.60	3.29	2.60	2.11	1.83	2.48	2.28	1.63
0.875-0.900	1.22	2.04	3.11	3.26	3.48	4.11	5.26	4.85	4.66	5.29	4.44	4.07	4.52	4.37	4.07	3.07	3.96	3.74	3.89	3.33	2.52	3.63	2.96	2.63	2.11	2.48	1.96	1.93	1.67	1.37
0.850-0.875	1.47	2.92	3.03	4.08	3.84	4.58	5.21	5.32	4.74	4.03	4.05	4.58	3.90	4.16	3.95	4.05	3.42	3.95	3.08	3.21	3.19	2.53	2.47	3.00	2.24	2.26	1.55	1.90	1.50	1.79
0.825-0.850	1.28	2.35	2.96	3.57	3.95	4.18	4.08	4.40	4.33	4.73	4.42	4.26	4.15	4.11	3.62	3.82	3.44	3.49	3.47	3.32	3.29	3.31	3.00	2.53	2.45	2.50	2.40	2.25	2.32	1.99
0.800-0.825	1.67	2.68	3.36	3.79	3.84	4.19	4.23	4.44	4.58	4.40	4.52	4.11	4.06	3.93	3.79	3.81	3.48	3.56	3.16	3.28	3.10	2.95	2.85	2.93	2.72	2.30	2.30	2.13	2.02	1.83
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29

20 pav. Spalvinės duomenų lentelės parodančios koku periodu, skaičiuojant nuo koreliuojančio režio pabaigos, kanalas pralaužiamas. Kanalo plotis $KS = 2,0$, režio ilgis 35.

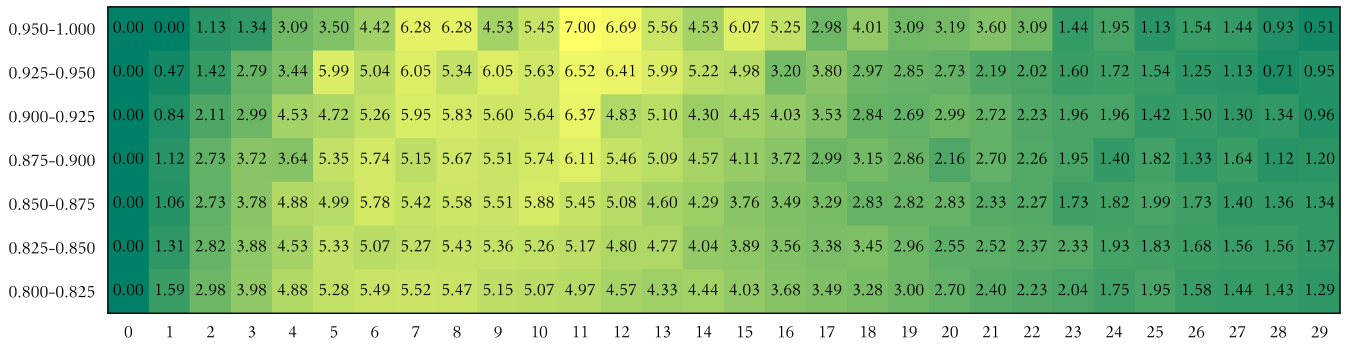
0.950-1.000	0.13	0.13	0.78	0.65	1.69	1.17	2.35	2.87	2.09	3.26	4.04	4.04	3.91	5.35	3.65	3.78	5.08	3.52	4.43	5.48	5.08	4.56	4.17	3.91	3.65	5.61	4.56	3.26	4.17	2.61
0.925-0.950	0.32	0.87	1.03	1.97	2.44	2.60	2.52	2.44	2.68	3.55	3.55	4.34	3.94	4.73	3.94	4.73	5.28	3.79	3.71	4.26	3.63	4.34	4.50	4.57	4.26	3.39	4.02	3.00	2.60	3.00
0.900-0.925	0.48	0.90	2.00	2.42	2.90	2.49	2.97	4.49	4.08	3.80	2.97	3.80	4.28	4.56	4.28	4.35	3.73	4.08	3.73	3.52	4.49	3.94	3.32	3.59	3.94	3.18	2.97	2.90	2.76	3.04
0.875-0.900	0.71	1.49	2.73	2.47	3.51	3.25	2.34	4.74	3.51	3.31	3.96	4.74	4.87	4.42	4.09	2.79	4.55	4.55	3.57	4.09	3.38	3.70	3.31	2.79	2.73	3.51	3.18	2.73	2.34	2.66
0.850-0.875	1.07	2.07	2.73	3.32	3.97	4.09	4.45	4.03	3.38	3.97	4.39	4.62	4.45	3.50	4.15	4.15	3.56	3.50	3.08	4.39	2.79	3.32	2.96	2.85	3.38	3.38	1.96	2.43	1.90	2.19
0.825-0.850	1.27	2.48	2.72	2.72	2.85	3.05	4.11	3.78	4.14	4.32	3.47	3.65	3.85	4.04	4.27	3.36	3.70	3.70	3.13	4.27	3.72	3.41	3.13	3.52	3.26	3.16	3.08	2.87	2.30	2.69
0.800-0.825	1.42	2.22	2.74	2.98	3.34	3.24	3.62	3.77	3.75	3.92	3.59	3.88	3.63	3.88	3.87	3.75	3.44	3.77	3.77	3.70	3.63	3.71	3.49	3.22	3.03	3.07	3.11	2.99	2.75	2.72
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29

21 pav. Spalvinės duomenų lentelės parodančios kokių periodu, skaičiuojant nuo koreliuojančio režio pabaigos, kanalas pralaužiamas. Kanalo plotis $KS = 2,0$, režio ilgis 50.

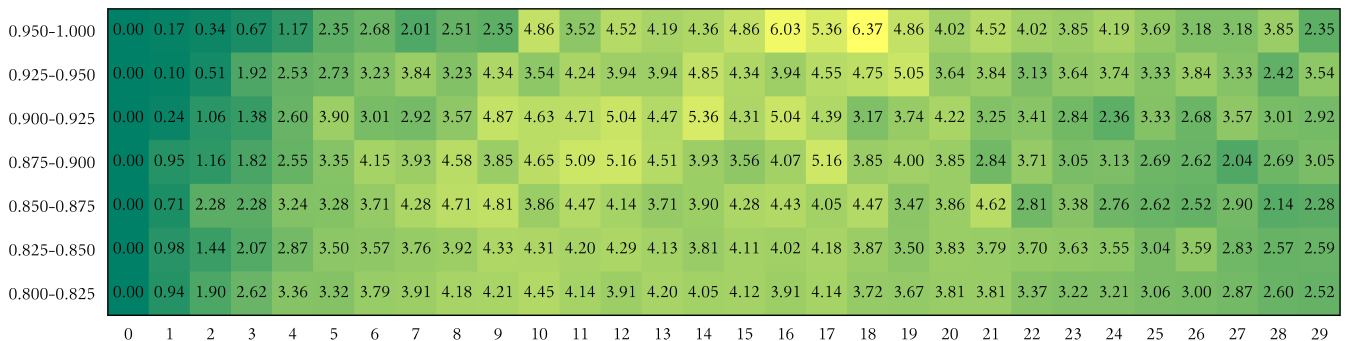
Šiuose (19, 20 ir 21 pav.) grafikuose yra parodyta kada įvyksta laiko eilutės kanalo lūžis. Vertikali ašis nusako koreliacijos koeficiento režius, o horizontali – pralaužimo laiką skaičiuojant momentais nuo koreliuojančio režio pabaigos, čia 0 žymi patį pirmą periodą, kada įvyksta pralaužimas. Visi dydžiai lentelės viduje yra pateikti procentais, skaičiuojant santykinai kokią dalį visų pralaužimų sudaro režiai su nurodytais parametrais neįskaičiuojant pralaužimų, kurie įvyko už maksimalaus periodų skaičiaus. Visi grafikai čia yra paskaičiuoti su vienodu kanalo pločiu $KS = 2,0$, režio ilgiai yra skirtingi, tai yra: 19 pav. – 20 periodų, 20 pav. – 35 periodai ir 21 pav. – 50 periodai.

Žiūrint į pirmą duomenų lentelę aiškiai matoma struktūra, kaip laiko eilutės su režio ilgiu 20 dažniausiai prasilaužia pradėdant 1-3 momentu ir užbaigiant 13-15 momentu priklausomai nuo koreliacijos koeficiento. Kuo didesnis koreliacijos koeficientas, tuo labiau koncentruotas yra lūžimo periodas. Šią tendenciją galima išvelgti ir kituose dviejuose grafikuose, tačiau esant didesniai koreliuojančio režio ilgiui kanalo pralaužimas vėluoja ir pasiskirsto per didesnę laiko eilutės momentų kiekį. Rėžio ilgiui esant 50 ir koreliacijos koeficientui $> 0,85$ imtis kiekvienam lentelės langeliui yra pakankamai maža (< 100), todėl yra matomas didelis spalvų išsibarstymas. Tačiau, net ir tokiomis sąlygomis matome kaip forma, kurią galima aiškiai apibrėžti kai režio ilgis yra 20 (19 pav.), didėjant režio ilgiui plėtėja išlaikydama savo kontūrus.

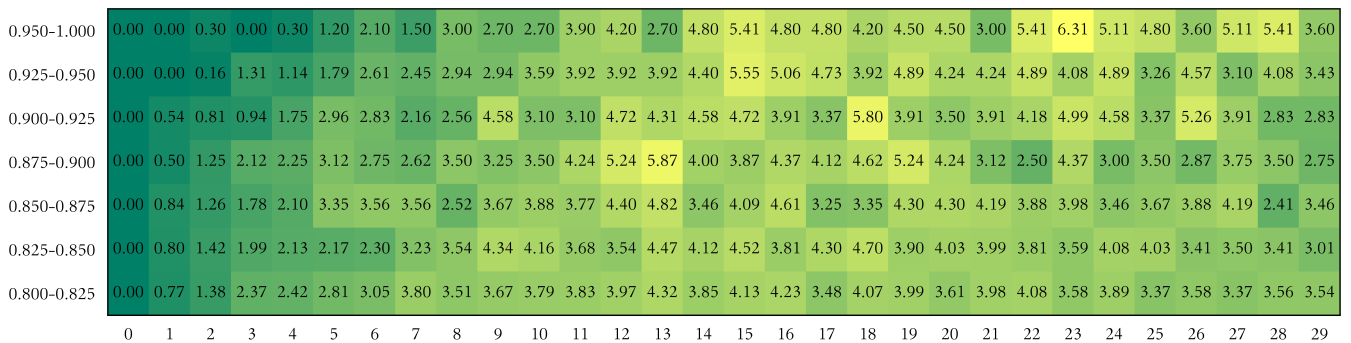
Panagrinėsime kada laiko eilutės vertė grįžta į kanalą po pralaužimo. Sukonstruosime analogiškas lenteles kaip jau nagrinėtais atvejais, kuriose vietoje pralaužimo momentu bus parodomas grįžimo laikas.



22 pav. Spalvinės duomenų lentelės parodančios kokių periodu, skaičiuojant nuo koreliuojančio rėžio pabaigos, laiko eilutės reikšmė grįžta į kanalą. Kanalo storis $KS = 2,0$, rėžio ilgis 20.



23 pav. Spalvinės duomenų lentelės parodančios kokių periodu, skaičiuojant nuo koreliuojančio rėžio pabaigos, laiko eilutės reikšmė grįžta į kanalą. Kanalo storis $KS = 2,0$, rėžio ilgis 35.



24 pav. Spalvinės duomenų lentelės parodančios kokių periodu, skaičiuojant nuo koreliuojančio rėžio pabaigos, laiko eilutės reikšmė grįžta į kanalą. Kanalo storis $KS = 2,0$, rėžio ilgis 50.

Analogiškai kaip ir 19-21 pav., 22-24 pav. parodyta kokių momentu, skaičiuojant nuo koreliuojančio rėžio pabaigos, įvyksta laiko eilutės vertės grįžimas į suformuotą kanalą. Akivaizdu, kad 0-iu laiko momentu yra lygiai 0 grįžimų. Taip pat kaip ir prieš tai buvusioje lentelėje, visi dydžiai lentelės viduje yra pateikti procentais, skaičiuojant santykinai kokią dalį visų lūžių sudaro rėžiai su nurodytais parametrais neįskaičiuojant lūžių, kurie įvyko už maksimalaus periodų kiekio. Visi grafikai čia yra paskaičiuoti su vienodu kanalo pločiu $KS = 2,0$.

Šiuose grafikuose (22, 23 ir 24 pav.) matome panašią struktūrą į prieš tai buvusią, ypač kai režio ilgis yra 20. Tačiau dėl mažesnės imties yra matomas didesnis išsibarstymas visuose grafikuose.

Grįžimo duomenis analizavome analogiškai kaip ir pralaužimo duomenis. Apdoroti duomenys pateikti analogiška forma (22 pav.). Taip analizuojant grįžimo duomenis nėra matoma jokia aiški ir išskirianti struktūra. Todėl šiuos duomenis analizuosime reliatyviai ne koreliuojančio režio pabaigos periodui, o pralaužimo tašką laikysime kaip pradinį atskaitos tašką. Tam, kad rezultatai nebūtų iškreipti, pralaužimo momentais laikysime tas laiko eilutės reikšmes, kurios išeina iš kanalo ribų per pirmas 20 laiko eilutės verčių nuo koreliuojančio režio pabaigos, o grįžimui taikysime 30 laiko eilutės reikšmių intervalą nuo koreliuojančio režio pabaigos. Tokiu atveju, jei pralaužimas įvyksta arti 20 laiko eilutės momento, grįžimas į kanalą nebus garantuotas tik keliems pirmiesiems momentams ir neiškreips duomenų.

0.950-1.000	41.30	19.38	9.03	6.50	6.28	3.41	3.08	2.75	1.32	2.20	1.10	0.77	0.66	0.55	0.33	0.22	0.22	0.44	0.11	0.22	0.11
0.925-0.950	44.37	18.67	10.29	7.23	4.99	3.84	2.24	2.05	1.02	1.15	1.28	0.77	0.51	0.45	0.19	0.13	0.19	0.19	0.06	0.32	0.06
0.900-0.925	42.35	18.26	10.46	5.92	5.21	3.33	2.50	2.13	1.96	0.83	0.96	1.38	0.67	0.92	0.75	0.63	0.42	0.29	0.38	0.58	0.08
0.875-0.900	41.47	17.17	10.09	7.05	4.96	3.24	2.45	2.37	2.14	1.61	1.07	1.52	0.79	0.51	0.70	0.45	0.37	0.59	0.37	0.87	0.20
0.850-0.875	42.10	16.85	10.00	6.95	4.38	4.10	2.55	1.96	1.95	1.60	1.14	0.93	0.76	0.83	0.72	0.56	0.50	0.47	0.44	0.95	0.28
0.825-0.850	40.44	18.07	10.28	6.66	4.39	3.54	2.80	2.32	2.05	1.70	1.34	1.05	0.74	0.62	0.76	0.61	0.50	0.47	0.41	0.94	0.32
0.800-0.825	40.67	17.33	9.96	6.78	4.67	3.67	2.84	2.48	1.93	1.53	1.23	1.01	0.91	0.77	0.70	0.64	0.47	0.40	0.41	1.25	0.36
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20-24	25-29

25 pav. Spalvinės duomenų lentelės parodančios koku periodu, skaičiuojant nuo kanalo **pralaužimo**, laiko eilutės reikšmė grįžta į kanalą. Kanalo plotis $KS = 2,0$, režio ilgis 20.

0.950-1.000	48.34	17.66	10.82	5.52	3.75	2.87	2.21	1.77	2.43	0.66	0.66	1.10	1.10	0.00	0.66	0.22	0.00	0.00	0.00	0.22	0.00
0.925-0.950	42.29	18.12	10.67	6.17	5.01	3.21	3.21	2.44	1.54	1.03	1.41	1.03	1.16	0.51	0.39	0.26	0.90	0.26	0.26	0.13	0.00
0.900-0.925	40.74	18.73	10.66	6.37	4.98	2.29	3.19	1.59	2.69	1.69	1.39	0.70	0.70	0.80	0.80	0.90	0.50	0.10	0.20	0.70	0.30
0.875-0.900	41.82	17.77	11.76	5.48	5.31	3.71	3.27	1.77	1.33	1.59	1.06	1.06	0.88	0.35	0.88	0.35	0.09	0.18	0.53	0.71	0.09
0.850-0.875	39.42	18.49	10.95	7.37	4.82	3.57	2.84	1.93	1.53	1.36	1.36	1.25	0.85	1.08	0.68	0.51	0.45	0.34	0.17	0.91	0.11
0.825-0.850	39.99	17.00	9.64	6.80	5.48	3.75	2.75	2.43	2.36	1.77	1.50	1.05	0.91	0.86	0.48	0.64	0.39	0.55	0.34	1.09	0.23
0.800-0.825	40.03	17.94	10.04	6.96	5.43	3.61	3.05	2.13	1.68	1.62	1.21	1.03	1.04	0.79	0.66	0.50	0.33	0.44	0.32	0.97	0.23
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20-24	25-29

26 pav. Spalvinės duomenų lentelės parodančios koku periodu, skaičiuojant nuo kanalo **pralaužimo**, laiko eilutės reikšmė grįžta į kanalą. Kanalo plotis $KS = 2,0$, režio ilgis 35.

0.950-1.000	40.81	17.49	12.11	8.52	4.93	5.38	3.14	2.24	1.35	0.45	0.45	0.90	0.00	0.45	0.45	0.45	0.00	0.00	0.45	0.45	0.00
0.925-0.950	43.12	21.67	11.29	5.87	4.29	2.71	3.16	1.81	1.13	1.35	0.90	0.45	0.23	0.23	0.45	0.45	0.00	0.00	0.23	0.45	0.23
0.900-0.925	44.62	15.59	10.57	5.91	2.87	5.02	2.69	1.79	2.33	1.25	2.15	1.08	0.72	0.72	0.72	0.90	0.18	0.00	0.00	0.90	0.00
0.875-0.900	44.72	17.24	9.01	6.99	4.19	3.26	1.86	2.02	2.02	1.71	1.09	1.09	1.09	0.62	0.31	0.47	0.47	0.31	0.16	1.24	0.16
0.850-0.875	38.74	15.68	8.96	7.51	5.27	4.48	1.98	3.95	2.77	2.24	1.19	1.45	0.79	0.40	1.45	0.40	0.53	0.66	0.66	0.92	0.00
0.825-0.850	39.86	17.22	10.37	6.85	5.18	3.57	3.34	2.19	2.19	1.27	1.15	1.38	0.98	0.58	1.09	0.86	0.46	0.35	0.23	0.75	0.12
0.800-0.825	40.27	18.02	9.66	6.40	5.08	3.60	2.88	2.49	2.13	1.64	1.14	1.18	1.07	0.71	0.68	0.71	0.43	0.35	0.23	1.15	0.18
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20-24	25-29

27 pav. Spalvinės duomenų lentelės parodančios kokių periodu, skaičiuojant nuo kanalo **pralaužimo**, laiko eilutės reikšmė grįžta į kanalą. Kanalo plotis $KS = 2,0$, režio ilgis 50.

Aukščiau pateiktuose grafikuose (25, 26 ir 27 pav.) horizontali ašis yra laiko eilutės reikšmių skaičius skaičiuojant nuo kanalo pralaužimo momento, kada laiko eilutė grįžta į kanalą. Vertikali ašis skirta koreliacijos koeficientams. Kiekviename lentelės langelyje parodyta kokia dalis visų grįžimų į kanalą įvyksta nurodytu laiko momentu po kanalo pralaužimo ir koks buvo koreliuojančio režio koreliacijos koeficientas. Visi dydžiai lentelės langeliuose pateikti procentais. Taip pat atkreipsime dėmesį į sugrupuotu paskutinius laiko eilučių reikšmių skaičius (horizontali ašis). Šių reikšmių tikimybė yra itin maža ir į jas neatsižvelgiama.

Galima pastebėti, kad nepriklausomai nuo režio ilgio ar koreliacijos koeficiento laiko eilutės reikšmė po pralaužimo dažniausiai iškart grįžta į kanalą. Pirmajame laiko momente grįžta apie 40% visų grįžusių reikšmių, antrame periode šis skaičius svyruoja apie 17–18%. Tokie duomenys rodo, kad įvykus kanalo pralaužimui galima labai greitai spręsti ar laiko eilutės reikšmė pasiliks kanale ar išeis iš suformuoto kanalo.

Kadangi šie rezultatai nepriklauso nuo pagrindinių parametrų (režio ilgio, koreliacijos koeficiento), patikrinsime, kaip jie priklauso nuo kanalo pločio.

0.950-1.000	46.33	17.09	8.34	6.53	4.52	3.22	2.01	1.51	1.61	2.21	1.21	1.21	1.01	0.10	0.30	0.90	0.50	0.30	0.50	0.30	0.30
0.925-0.950	42.41	17.89	10.47	5.93	4.95	3.34	3.22	2.47	1.55	1.38	0.98	0.86	0.92	0.40	0.35	0.69	0.46	0.35	0.12	1.04	0.23
0.900-0.925	41.74	16.97	10.14	7.24	4.37	3.39	2.83	2.56	2.07	1.28	1.28	1.06	0.87	0.53	0.57	0.49	0.60	0.34	0.57	0.94	0.15
0.875-0.900	41.62	17.81	9.58	6.27	4.20	3.90	2.59	2.30	1.68	1.48	1.53	1.21	0.84	0.62	0.84	0.54	0.49	0.37	0.47	1.21	0.44
0.850-0.875	41.15	17.61	9.69	6.38	4.63	3.77	2.67	2.24	1.95	1.33	1.39	1.09	0.85	0.83	0.71	0.61	0.47	0.43	0.53	1.23	0.45
0.825-0.850	40.03	17.06	10.30	6.86	5.28	3.45	2.65	2.15	1.81	1.77	1.31	1.21	0.97	0.66	0.71	0.66	0.44	0.59	0.41	1.30	0.40
0.800-0.825	40.18	17.43	9.89	6.70	4.68	3.44	2.83	2.25	2.04	1.72	1.25	1.14	0.98	0.74	0.72	0.67	0.53	0.56	0.40	1.40	0.46
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20-24	25-29

28 pav. Spalvinės duomenų lentelės parodančios kokių periodu, skaičiuojant nuo kanalo **pralaužimas**, laiko eilutės reikšmė grįžta į kanalą. Kanalo plotis $KS = 1,5$, režio ilgis 20.

0.950-1.000	44.53	22.63	10.22	9.49	4.38	0.73	3.65	1.46	0.73	0.73	1.46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
0.925-0.950	44.21	16.84	9.47	6.67	5.26	3.16	1.75	2.11	1.05	1.40	2.46	1.40	0.35	1.40	0.35	0.00	1.05	0.35	0.00	0.70	0.00
0.900-0.925	41.18	17.91	9.89	8.29	4.55	2.67	2.14	2.41	2.94	1.34	1.87	1.34	1.34	0.53	0.00	0.53	0.00	0.53	0.53	0.00	0.00
0.875-0.900	42.47	19.51	11.60	5.68	3.21	2.22	1.73	2.72	3.70	1.98	0.74	0.99	0.49	0.49	0.49	0.49	1.23	0.00	0.25	0.00	0.00
0.850-0.875	40.53	19.32	10.51	7.32	4.88	2.44	2.25	1.88	2.25	1.69	2.06	0.75	0.56	0.38	1.31	0.19	0.19	0.00	0.56	0.94	0.00
0.825-0.850	39.03	17.26	10.32	7.11	4.77	2.86	3.12	2.25	2.34	1.82	2.25	1.21	1.21	0.69	0.95	0.61	0.61	0.26	0.09	1.13	0.09
0.800-0.825	40.84	17.12	10.74	7.25	4.57	4.13	2.75	2.13	2.15	1.51	1.14	1.22	0.74	0.72	0.52	0.70	0.41	0.33	0.23	0.72	0.08
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20-24	25-29

29 pav. Spalvinės duomenų lentelės parodančios koku periodu, skaičiuojant nuo kanalo **pralaužimas**, laiko eilutės reikšmė grįžta į kanalą. Kanalo plotis $KS = 2,5$, režio ilgis 50.

Šie (28 ir 29 pav.) grafikai yra analogiški prieš tai buvusiam. 28 pav. kanalo plotis yra 1,5 ir režio ilgis 20, tai yra minimalūs parametrai su kuriais atliekami bandymai. 29 pav. kanalo plotis yra 2,5 ir režio ilgis 50, o tai yra maksimalūs parametrai su kuriais yra atliekami bandymai.

Iš šių grafikų galima spręsti, jog grįžimo į laiko eilutės kanalą momentas nepriklauso ir nuo kanalo pločio, tai galima daryti prielaidą apie laiko eilutės fraktališkumą.

4.4. Modeliavimas

Šiame skyrelyje trumpai aptarsime kaip būtų galima naudojant programinius resursus modeliuoti pasirinktas laiko eilutes remiantis kitomis laiko eilutėmis.

Daugybę indikatorių galime apskaičiuoti kiekvienam pasirinktos laiko eilutės taškui ir išreikšti skaitine reikšme taip sudarydami naują laiko eilutę. Tokiu atveju, nors modeliuojama priklausomybė tik nuo kitų laiko eilučių, bet šias laiko eilutes gali sudaryti ir apskaičiuoti indikatoriai.

Šiam tikslui pasiekti naudosime daugialypės regresijos lygtį [Vyd02]:

$$K = (A^T A)^{-1} * A^T * Y$$

Čia K yra regresijos lygties parametrų matrica, A - regresorių matrica, Y - laiko eilutės, kuriai ieškome priklausomųjų, matrica. Su šia lygtimi vienas kintamasis išreiškiamas kitais nepriklausomais kintamaisiais (regresoriais).

Pasinaudoję determinacijos koeficientu (r^2) patikrinsime ar šis modelis yra prasmingas.

$$r^2 = \frac{SSR}{SST}$$

Šioje lygtyje SSR - regresijos kvadratų suma, o SST - visų kvadratų suma. Jie yra apskaičiuojami:

$$SSR = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$SST = \sum_{i=1}^n (y^*(x_i) - \bar{y})^2$$

Šiose lygtyse y_i yra reali modeliuojamo kintamojo reikšmė i -uoju laiko momentu, \bar{y} - modeliuojamo kintamojo vidurkis, $y^*(x_i)$ - sumodeliuota modeliuojamo kintamojo reikšmė i -uoju laiko momentu.

Apskaičiuosime determinacijos koeficientą atsitiktinai pasirinktoms laiko eilutėms kai modeliuojama reikšmė priklauso nuo 5 ir 10 regresorių.

regresorių kiekis	vidurkis	1-as kvartilis	mediana	3-ias kvartilis
5	0,5605	0,3904	0,5816	0,7267
10	0,7044	0,6649	0,7267	0,8263

10 lentelė. Determinacijos koeficiento pasiskirstymas atsitiktinai parinktose laiko eilutėse

Iš čia matome, jog daugeliu atvejų, net ir atsitiktinėms laiko eilutėms, determinacijos koeficiento vertė yra daug aukštesnė už minimalią (0,25) ir daugeliu atvejų yra gana didelė ($> 0,8$). Iš čia galime daryti išvadą, jog toks modeliavimo būdas yra prasmingas.

Rezultatai ir išvados

Šiame darbe atliktoje literatūros apžvalgoje yra pasirenkami indikatoriai ir techninės analizės metodai, kurie toliau įgyvendinami ir analizuojamas jų tinkamumas laiko eilučių prognozavimui. Pasirinktų indikatorių testavimui reikalinga platforma, kuri suprogramuota naudojantis Python programavimo kalba.

Sukurta platforma panaudota reikalingų duomenų surinkimui ir visų scenarijų analizei. Platforma gali būti nesunkiai plečiama ir papildoma naujais duomenų šaltiniais, naujais analizės scenarijais ar nauju išvesties pateikimu. Visi kiekvieno indikatoriaus analizės žingsniai yra atkartojami ir rezultatai gali būti suskaičiuojami iš naujo pasirinktoms vietoms.

Pasinaudojus sukurta platforma į duomenų bazes surinkti istoriniai kiekvienos dienos ir realiu laiku renkami kiekvienos sekundės akcijų kainų duomenys. Sudarytos duomenų bazės naudojamos visų indikatorių analizei.

Išanalizuotas Hursto parametras neturi jokios įtakos laiko eilutės prognozavimui, tačiau gali būti panaudotas kaip įrankis laiko eilučių generavimui. Taip pat juo patvirtinama, jog laiko eilutė nėra atsitiktinis judesys.

Slenkančio vidurkio analizė parodė, kad nėra jokio skirtumo tarp skirtingų slenkančio vidurkio tipų. Slenkantis vidurkis nėra naudingas prognozavimui bendroju atveju. Vis dėlto, naudojant smulkaus mastelio duomenis taškuose, kuriuose slenkantis vidurkis susikerta su laiko eilutės reikšmių linija, nedideliame periodų kiekiui galima nustatyti pakankamai reikšmingą laiko eilutės ateities polinkį į pusę, priklausančią nuo slenkančio vidurkio ir laiko eilutės reikšmės santykio.

Analizuojant kanalus pasinaudota rėžio koreliacijos koeficientu, siekiant automatizuotu būdu išgauti kanalą, kuris atitiktų „pieštą iš akies“. Šie kanalai nepasako nieko apie tai, kokiu periodu nuo jų sudarymo jie bus pralaužiami, tačiau prognozavimui ir laiko eilučių generavimui yra naudingas pastebėtas dėsningumas, jog pralaužus kanalą į jį bus grįžtama labai greitai (t.y. dažniausiai iškart po pralaužimo) arba išvis negrįžtama.

Visus išanalizuotus indikatorius galima apjungti į vieną modelį naudojant daugialypę regresiją. Vis dėlto, šiam metodui reikia daugiau išanalizuotų indikatorių, kurie galėtų būti naudingi laiko eilučių prognozavimui. Sėkmingas pirminis modelio testavimas statistinės analizės metodais leidžia manyti, jog laiko eilučių prognozavimas yra galimas daugialypės regresijos modeliu.

Literatūra

- [Bul03] Thomas N Bulkowski. *Trading classic chart patterns*. Tom. 176. John Wiley & Sons, 2003, p. 407–426.
- [CGB⁺14] Subhabrata Choudhury, Subhajyoti Ghosh, Arnab Bhattacharya, Kiran Jude Fernandes ir Manoj Kumar Tiwari. A real time clustering and svm based price–volatility prediction for optimal trading strategy. *Neurocomputing*, 131:419–426, 2014.
- [CH02] M.P. Clements ir D.F. Hendry. *A companion to economic forecasting*. *Blackwell Companions to Contemporary Economics*. Wiley, 2002, p. 1–18. ISBN: 9780631215691.
- [Cro95] R.M. Crownover. *Introduction to fractals and chaos*. Of *Jones and Bartlett books in mathematics*. Jones and Bartlett, 1995, pp. 229–270. ISBN: 9780867204643.
- [ET05] David Enke ir Suraphan Thawornwong. The use of data mining and neural networks for forecasting stock market returns. *Expert systems with applications*, 29(4):927–940, 2005.
- [Hut84] Jack K Hutson. Filter price data: moving averages vs. exponential moving averages. *Technical analysis of stocks & commodities*, 2, 1984.
- [Kau13] P.J. Kaufman. *Trading systems and methods*. *Wiley Trading*. Wiley, 2013. ISBN: 9781118222249.
- [KD10] C.D. Kirkpatrick and J. Dahlquist. *Technical analysis: the complete resource for financial market technicians*. Pearson Education, 2010, pp. 131–162. ISBN: 9780132599627.
- [Kel60] C.W. Keltner. *How to make money in commodities*. Keltner Statistical Service, 1960.
- [LCH15] Tsung-Hsun Lu, Yi-Chi Chen ir Yu-Chin Hsu. Trend definition or holding strategy: what determines the profitability of candlestick charting? *Journal of banking & finance*, 61:172–183, 2015.
- [LFH⁺08] William Leigh, Cheryl J Frohlich, Steven Hornik, Russell L Purvis ir Tom L Roberts. Trading with a stock chart heuristic. *Ieee transactions on systems, man, and cybernetics-part a: systems and humans*, 38(1):93, 2008.
- [LMP⁺02] William Leigh, Naval Modani, Russell Purvis ir Tom Roberts. Stock market trading rule discovery using technical charting heuristics. *Expert systems with applications*, 23(2):155–159, 2002.
- [MP16] R David McLean ir Jeffrey Pontiff. Does academic research destroy stock return predictability? *The journal of finance*, 71(1):5–32, 2016.
- [SWZ07] Gene Savin, Paul Weller ir Jānis Zvingelis. The predictive power of “head-and-shoulders” price patterns in the us stock market. *Journal of financial econometrics*, 5(2):243–265, 2007.
- [Val08] Eimutis Valatkevičius. *Investavimas finansų rinkose*. Technologija, 2008, p. 289–312.

- [Vyd02] Gediminas Murauskas Vydas Čėkanavičius. *Statistika ir jos taikymai, ii knyga*. TEV, 2002, p. 151–179.
- [Wah95] B. Wahl. *Exploring fractals on the macintosh*. Addison-Wesley Publishing Company, 1995. part 4. ISBN: 9780201626308.