

KAUNO TECHNOLOGIJOS UNIVERSITETAS
MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS

Andrius Popovas

**Trumpų finansinių laiko eilučių algebrinių dėstinių
parametrų analizė**

Baigiamasis magistro projektas

Vadovė

Doc. Dr. K. Lukoševičiūtė

Kaunas, 2017

KAUNO TECHNOLOGIJOS UNIVERSITETAS
MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS

Trumpų finansinių laiko eilučių algebrinių dėstinių parametrų analizė

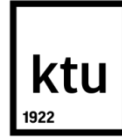
Baigiamasis magistro projektas
Taikomoji matematika (kodas 621G10003)

Vadovė
Doc. Dr. Kristina Lukoševičiūtė
2017

Recenzentė
Doc. Dr. Liepa Bikulčienė
2017

Projektą atliko
Andrius Popovas
2017

Kaunas, 2017



KAUNO TECHNOLOGIJOS UNIVERSITETAS

MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS

(Fakultetas)

Andrius Popovas

(Studento vardas, pavardė)

Taikomoji matematika, 621G10003

(Studijų programos pavadinimas, kodas)

Baigiamojo projekto „Trumpų finansinių laiko eilučių algebrinių dėstinių parametrų analizė“

AKADEMINIO SAŽININGUMO DEKLARACIJA

20 17 m. Gegužės 19 d.
Kaunas

Patvirtinu, kad mano, Andriaus Popovo, baigiamasis projektas tema „Trumpų finansinių laiko eilučių algebrinių dėstinių parametrų analizė“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

(vardą ir pavardę įrašyti ranka)

(parašas)

Popovas, Andrius. Trumpų finansinių laiko eilučių algebrinių dėstinių parametru analizė. Magistro baigiamasis projektas, vadovė Doc. Dr. Kristina Lukoševičiūtė; Kauno technologijos universitetas, Matematikos ir gamtos mokslų fakultetas. – Kaunas, 2017. 90p.

Mokslo kryptis ir sritis: fiziniai mokslai, matematika (01 P).

Reikšminiai žodžiai: finansinės laiko eilutės, prognozavimas, algebrinis skeletas.

SANTRAUKA

Šiame tyrime nagrinėjama algebrinių dėstinių parametru analizė skirta prognozavimo uždaviniui. Nagrinėjamas ganėtinai naujas algebrinio skeleto metodas pateiktas straipsnyje [1]. Pagrindinis tyrimo tikslai – nustatyti optimalų šio modelio parametru parinkimą ir ištirti jo elgseną, prognozuojant finansinę laiko eilutę dviem finansų rinkos savaitėms į priekį. Tyrime sudaryta metodika parametru parinkimui, bei nustatyta, kad modelis nepriklauso nuo prognozuojamų reikšmių kiekio, o jo paklaidos pasiskirstę pagal normalųjį skirstinį, turintį nulinį vidurkį. Modelis buvo pritaikytas finansinėms laiko eilutėms prognozuoti ir atlikta USD/RUB kurso prognozė. Taip pat modelis buvo integruotas į ARIMAX modelį, kaip egzogeninio kintamojo prognozavimo metodas.

Popovas, Andrius. Parameter Analysis of Short Term Financial Time Series Algebraic Sequences: Master's thesis in Applied Mathematics, supervisor assoc. Doc. Dr. Kristina Lukoševičiūtė. The Faculty of Mathematics and Natural Sciences, Kaunas University of Technology. – Kaunas, 2017. 90p.

Research area and field: natural sciences, mathematics (01 P).

Key words: financial time series, forecasting, algebraic skeleton.

SUMMARY

Parameter analysis for algebraic sequences in forecasting problem is considered in this paper. Research is for relatively new algebraic skeleton model proposed in article [1]. Research targets are to find optimal method for selection of model parameters and evaluation of model behavior in financial time series forecasting for two financial weeks. Methodology for optimal parameter selection was designed during research. Additionally found that model is not dependent on forecasting horizon and its errors has normal distribution with zero average. Model was fitted to financial time series and forecasting of USD/RUB was done. Moreover model was integrated in ARIMAX model as exogenous variable forecasting method.

TURINYS

Lentelių ir paveikslų sąrašas.....	7
1. Įvadas.....	9
2. Literatūros apžvalga	10
2.1. Laiko eilučių metodų apžvalga	11
2.2. Optimizavimo algoritmų apžvalga.....	15
3. Metodų aprašymas.....	19
3.1. Algebrinio skeleto metodas.....	19
3.1.1. Hankelio matricos rango sąvoka	19
3.1.2. Prognozavimo strategija	20
3.2. Genetinis algoritmas.....	22
3.3. ARIMA ir ARIMAX.....	25
3.4. Pearsono ir Spearmano koreliacijos koeficientai	27
3.5. Shapiro-Wilk ir Anderson-Darling normalumo tyrimai	28
3.6. Paklaidų metrikos.....	29
4. Tyrimas.....	30
4.1. Prognozavimas iš skirtingų laiko momentų ir rango parinkimas.....	32
4.2. Genetinio algoritmo parametrų parinkimas	37
4.3. Testinės eilutės prognozavimas.....	41
4.4. Testinės eilutės prognozavimo paklaidų analizė.....	44
4.5. Finansinės laiko eilutės paruošimas	46
4.6. Finansinės laiko eilutės prognozavimas	48
4.7. Finansinės laiko eilutės prognozavimo paklaidų analizė	51
4.8. Modelio taikymas su kitais prognozavimo modeliais.....	53
5. Išvados ir rekomendacijos	57

6.	Literatūros sąrašas	58
7.	Priedai.....	65
7.1.	Išorinių regresorių dinamika.	65
7.2.	DJIA ir MICEX regresorių sudėtis.	69
7.3.	Modelio veikimas iš skirtingų laiko taškų – eksperimentų grafikai	71
7.4.	Išorinių regresorių prognozės.....	78
7.5.	ARIMA modelių adekvatumo tyrimai	82
7.6.	Ankstesnių rango tyrimo rezultatai	90

Lentelių ir paveikslų sąrašas

Paveikslai

4.1 pav. Testinės eilutės fragmento dinamika.....	30
4.2 pav. USD/RUB eilutės dinamika	30
4.1.1 pav. Laiko eilutės analizė pradedant nuo 32-os bendros sekos reikšmės. Trijų ašių grafikas.	33
4.1.2 pav. Laiko eilutės analizė pradedant nuo 32-os bendros sekos reikšmės. Dviejų ašių grafikas.	33
4.1.3 pav. Laiko eilutės analizė pradedant nuo 113-os bendros sekos reikšmės. Trijų ašių grafikas.	34
4.1.4 pav. Laiko eilutės analizė pradedant nuo 113-os bendros sekos reikšmės. Dviejų ašių grafikas.	34
4.1.5 pav. Laiko eilutės analizė pradedant nuo 255-os bendros sekos reikšmės. Trijų ašių grafikas.	35
4.1.6 pav. Laiko eilutės analizė pradedant nuo 255-os bendros sekos reikšmės. Dviejų ašių grafikas.	35
4.2.1 pav. Testinės eilutės prognozės paklaidų RMSE priklausomybė nuo genetinio algoritmo parametrų.....	38
4.2.2 pav. Testinės eilutės prognozės paklaidų RMSE priklausomybė nuo genetinio algoritmo parametrų mažiausių paklaidų zonose.....	39
4.3.1 pav. Minimali, vidutinė ir maksimali RMSE reikšmė, kai rangas 14-18.	41
4.3.2 pav. Prognozės paklaidų RMSE priklausomybė nuo rango ir genetinio algoritmo parametrų.....	42
4.4.1 pav. Testinės eilutės procentinių paklaidų histograma ir normaliojo skirstinio kreivė ...	45
4.6.1 pav. Finansinės eilutės, RMSE ir rango priklausomybės grafikas vieno žingsnio prognozei.	48
4.6.2 pav. Finansinės eilutės prognozės paklaidų RMSE priklausomybė nuo genetinio algoritmo parametrų mažiausių paklaidų zonose.....	49

4.6.3 pav. Finansinės eilutės, USD/RUB kursas, prognozės ir tikrosios reikšmės grafikas.	49
4.6.4 pav. Finansinės eilutės, USD/RUB kursas, prognozavimo nuokrypių grafikas.	50
4.7.1 pav. Finansinės eilutės, USD/RUB kursas, procentinių paklaidų histograma ir normaliojo skirstinio kreivė.	52

Lentelės

3.4.1 lentelė. Koreliacijos koeficiento stiprumo skalė.....	28
4.2.1 lentelė. Genetinio algoritmo parametrų mažiausių paklaidų zonos.	39
4.4.1 lentelė. Testinės eilutės procentinių paklaidų pasiskirstymas pagal paklaidos dydį kiekvienai prognozuojamai reikšmei.....	44
4.5.1 lentelė. Eilutės modifikavimo tyrimo rezultatai.	47
4.7.1 lentelė. Finansinės eilutės, USD/RUB kursas, procentinių paklaidų sklaida kiekvienam iš prognozuojamų žingsnių.	51
4.8.1 lentelė. Išoriniai regresoriai ir jų žymėjimas.	53
4.8.2 lentelė. Pearsono ir Spremano koreliacijos koeficientai išoriniams regresoriams.	54
4.8.3 lentelė. Išorinių regresorių prognozavimui naudotų modelių parametrai.....	55
4.8.4 lentelė. Išorinių regresorių prognozių RMSE ir procentinės paklaidos.	55
4.8.5 lentelė. ARIMAX prognozių RMSE ir procentinės paklaidos.	56

1. Įvadas

Daugelyje mokslo sričių laiko eilučių prognozavimas yra nepaprastai svarbus uždavinys. Ekonomikoje ir finansuose – tai vienas iš pagrindinių dalykų, leidžiančių numatyti verslo ar investavimo ateitį. Visų laiko eilučių prognozavimo modelių esminis tikslas yra kuo tiksliau nustatyti istoriniuose duomenyse užkoduotas ar pasislėpusias tendencijas ir pasinaudojant jomis įvertinti labiausiai tikėtiną ateities scenarijų.

Laiko eilučių uždaviniai dažniausiai skiriami į dvi pagrindines grupes: trumpos ir ilgos eilutės. Ilgos laiko eilutės – tai eilutės, kurioms galima įvertinti pakankamai daug istorinių duomenų iš tolimos praeities, o trumpos, priešingai nei ilgos, neturi ilgos praeities. Tačiau vis dažniau yra vengiama naudoti labai senus istorinius duomenis apie ekonomines ar finansines eilutes, nes finansų rinkos ir ekonominė situacija labai dažnai kinta, yra nestabili ir dažniausiai priklauso nuo naujausių įvykių ir duomenų kaitos. Darbe tyrinėjama tik naujausia istorinių duomenų dalis, tai yra, iš ilgos laiko eilutės paimama tik naujausia informacija ir tai laikoma trumpa laiko eilute.

Darbe nagrinėjamas neseniai pasiūlytas ir nedaug ištyrinėtas algebrinio skeleto modelis, kuris pritaikytas trumpoms laiko eilutėms ir skirtas prognozuoti vienam laiko žingsniui į ateitį. Pagrindiniai darbo tikslai: ištirti modelio parametrus, nustatyti jų parinkimo metodiką bei patikrinti, kaip kokybiškai šis modelis gali prognozuoti daugiau nei vieną reikšmę. Darbe naudojama Matlab ir R programinė įranga.

Darbo pradžioje apžvelgiami šiuo metu naudojami populiarūs metodai, kurie pateikti literatūros apžvalgos skyrelyje. Metodologinėje dalyje apžvelgiamas nagrinėjamas algebrinio skeleto modelio veikimo principas. Tyrimo dalyje tiriamas modelio adekvatumas, jį taikant iš skirtingų laiko eilutės taškų, ištiriami modelio parametrai ir nustatomas jų ryšys su prognozavimo paklaidomis. Modelis yra ganėtinai naujas ir, norint patikrinti, ar jis turi pranašumo prieš šiuo metu populiariausius modelius, jis lyginamas su autoregresiniais slenkančio vidurkio modeliais. Taip pat bandoma integruoti šį modelį į autoregresinį slenkančio vidurkio modelį kaip egzogeninio kintamojo prognozavimo strategiją.

2. Literatūros apžvalga

Literatūroje pateikiama labai daug metodų, skirtų analizuoti ir prognozuoti įvairiausias finansinius ir ekonominius duomenis. Bendroji siūloma strategija susideda iš trijų esminių etapų: duomenų apžvalgos, duomenų paruošimo ir duomenų prognozavimo.

Dirbant su ekonominiiais duomenimis, labai dažnai pasitaiko situacija, kai nagrinėjami dydžiai yra matuojami skirtingais mato vienetais ir tampa sunku juos tarpusavyje palyginti. Šiose situacijose atliekama duomenų modifikavimo procedūra. Populiariausios tokio modifikavimo procedūros: standartizavimas – kai duomenų vidurkis paverčiamas nuli, o dispersija tampa lygi vienam; normalizavimas – kai duomenų vidurkis paverčiamas nuli; normavimas – kai duomenų eilutės reikšmės yra suspaudžiamos į intervalą $[0, 1]$, tai yra minimali reikšmė virsta nuli, o maksimali virsta vienetu [2, 3].

Kitas svarbus aspektas ekonominiams duomenims yra normalumo prielaida. Nemažai grupei klasikinių duomenų analizės ir prognozavimo metodų galioja reikalavimas, kad atsitiktinis kintamasis tenkintų arba iš dalies tenkintų normalumo dėsnį, todėl tenka tikrinti hipotezę apie normalųjį dėsnį [4]. Šią prielaidą apytiksliai patikrinti galima pasinaudojant histogramomis, tankio ar kvantilių grafikais. Norint prielaidą įvertinti tiksliau, naudojami Shapiro-Wilk, Anderson-Darling testai arba Pirsono, Kolmogorovo-Smirnovo kriterijai. Šie testai plačiau aprašyti 3.5 skyrelyje.

Dažniausiai literatūroje laiko eilutė yra apibrėžiama kaip atsitiktinis procesas. Tuomet dažniausiai vertinamas proceso stacionarumas, kuris kartais nulemia modelio pasirinkimą, nes kai kurie populiariausi modeliai, tokie kaip autoregresijos ir slenkančio vidurkio, reikalauja proceso stacionarumo. Dažnu atveju tenka procesą integruoti, logaritmuoti ar kitaip modifikuoti, siekiant proceso stacionarumo. Apytiksliai stacionarumo patikrinimui galima naudoti autokoreliacijos ir dalinės autokoreliacijos grafikus, tačiau gerokai plačiau naudojami testai: Dickey-Fuller, papildytas Dickey-Fuller, KPSS, Phillips-Perron [5, 6, 7].

Normalumo ir stacionarumo testai yra įtraukti į populiariausius matematinės programinės įrangos paketus, tokius kaip Matlab, R, SAS. Šie testai ganėtinai dažnai būna įtraukti į automatizuotas prognozavimo ar duomenų analizės funkcijas ir procedūras.

2.1. Laiko eilučių metodų apžvalga

Nuo pat laiko eilučių termino įvedimo sukurta gausybė įvairių metodų, skirtų joms analizuoti ir prognozuoti. Šiame darbe nagrinėjamos finansinės, ekonominės laiko eilutės prognozuoti populiariausius literatūroje metodus galima suklasifikuoti taip [8]:

- Autoregresiniai ir slenkančio vidurkio modeliai
AR, MA, ARMA, ARIMA, ARIMAX, SARIMA modeliai [9, 10];
- Sąlyginiai heteroskedastiškumo modeliai
ARCH, GARCH, Integruotas GARCH, GARCH-M, EGARCH, CHARMA modeliai [11, 12, 13];
- Netiesiniai modeliai
TAR, STAR, neuroninių tinkų modeliai [14, 15];
- Didelio dažnio duomenų analizės modeliai, paremti rinkos mikrostruktūra
Probit, dekompozicijos, ACD modeliai [16, 17, 18];
- Tęstinio laiko modeliai
Wiener ir Ito procesais paremti modeliai, Black-Scholes metodas [19, 20];
- Daugiamačiai modeliai
VAR, VARMA, Kointegruoti VAR modeliai [21, 22];
- Pagrindinių komponentų ir faktoriai modeliai
BARRA, Fama-French, PCA, vienfaktoriniai ir daugiafaktoriniai faktorių išskyrimo modeliai [23, 24];
- Daugiamačio nepastovumo modeliai
V-GARCH, D-VEC, BEKK modeliai [25, 26];
- Būvio būsenų modeliai
Lokalaus trendo, CAPM, būsenos glodinimo modeliai [27, 28];
- Markovo grandinių Monte Carlo metodai
Paprastasis Monte Carlo, papildytas ribojimais Monte Carlo metodai [29, 30].

Visos klasifikacijoje pateiktos nuorodos apžvelgia pagrindinius, daugiausiai ištyrinėtus modelius ir metodus, kuriais tyrinėjamos finansinės ir ekonominės laiko eilutės. Ši klasifikacija yra daugiau bendrinė ir ji negali būti labai griežta, nes pasaulio ekonominei situacijai kintant, atsiranda vis daugiau naujų modelių ir metodų, kurie gali būti integruoti vienas į kitą, taip apjungiant gerąsias dviejų ar daugiau metodų savybes.

Stacionariems procesams dažniausiai taikoma bendroji tiesinio prognozavimo teorija [31]. Tarkime, laiko eilutė yra duomenų vektorius $X = (x_1, x_2, \dots, x_n)^T$, tuomet jo prognozę galime aprašyti:

$$\hat{Y} = a + b_1x_1 + b_2x_2 + \dots + b_nx_n = a + b^tX,$$

čia a ir b yra koeficientai, kurie parenkami taip, kad atsitiktinė paklaida ε turėtų vidurkį lygų nuliui ir būtų nekoreliuota su duomenų vektoriumi X . Šis metodas gerai žinomas kaip tiesinės regresijos modelis.

Iš bendrosios tiesinio prognozavimo teorijos atsirado ir šiuo metu populiariausia naudojamų metodų grupė, tai autoregresiniai ir slenkančio vidurkio metodai, bei išvestiniai jų modeliai. Pagrindiniai modeliai AR, MA, ARMA, ARIMA, ARIMAX yra aprašyti 3.3 skyrelyje. Jie naudojami prognozuoti vienmačius atsitiktinius dydžius. Iš šių pagrindinių modelių gimė išvestiniai daugiamačiai metodai, tokie kaip VAR, VARMA, bei specifikuoti eilutėms su tam tikromis savybėmis, tokie kaip ARCH, GARCH, TAR, STAR bei kiti.

Išvestiniai modeliai atsirado dėl poreikio apdoroti sudėtingas, labai nepastovias, finansines ar ekonomines laiko eilutes, tokias kaip akcijų kaina arba akcijų grąža. Geras pavyzdys – apibendrintas autoregresinis sąlyginio heteroskedastiškumo modelis, arba trumpai GARCH [32]. Šis modelis taikomas tada, kai laiko eilutę prognozuojant su ARMA modeliu liekanos yra nekoreliuotos, tačiau turi nepastovią dispersiją, dėl kurios standartinis ARMA tampa nekorektiškas.

Vieni dažniausių tyrimų ekonomikoje yra daugiamačių dydžių priklausomybės tyrimai ir prognozavimas. Tam naudojami daugiamačiai prognozavimo modeliai, tokie kaip vektorinis autoregresijos ir slenkančio vidurkio modelis, arba trumpai VARMA, kuris gali būti užrašomas panašiai kaip ir ARMA [33]. Esminis skirtumas nuo ARMA modelio yra tas, kad regresorius daugiamatis, o koeficientai aprašomi ne vektoriais, o matricomis.

Pradėjus nagrinėti daugiamačius laiko eilučių atvejus, tapo reikalinga GARCH modelio atmaina, skirta daugiamačiams dydžiams. Tokiu būdu atsirado įstrižasis VEC modelis, dar kitaip žinomas kaip DVEC [34]. Tai – tam tikras eksponentiškai pasvertas slenkančio vidurkio modelis. Šis modelis dažnai naudojamas portfelio parinkimo arba pinigų paskirstymo ilguoju periodu uždaviniams spręsti.

Pagrindinis laiko eilučių tipas ekonomikoje ir finansuose yra dideliu svyravimo dažniu pasižyminčios eilutės (angliškai fuzzy). Nagrinėjant šias eilutes, tokias kaip valiutų kursas, akcijų kaina ar akcijų grąža, kartais naudojami netiesiniai arba rinkos mikrostruktūra paremti modeliai. Gana plačiai naudojamas autoregresijos modelis su ribiniais kintamaisiais arba trumpiau TAR. Dažniausiai naudojamas dviejų ribinių kintamųjų TAR modelis [35]. Nors pats modelis ir yra netiesinis, tačiau TAR modelis iš esmės yra sudarytas iš kelių tiesinių AR modelių. Dviejų ribinių kintamųjų atveju jis sudarytas iš keturių AR modelių. Rinkos analizės uždaviniuose taip pat gana populiarios ir šio modelio atmainos. Keletas atmainų: STAR – papildomai įtraukiamas glodinimas, LSTAR – papildomai įtraukiamas glodinimas ir logistinė perėjimo funkcija, ESTAR – papildomai įtraukiamas glodinimas ir eksponentinė perėjimo funkcija [36], STPAR – papildomai įtraukiamas periodiškumas [37].

Alternatyvūs metodai darbui su dideliu svyravimo dažniu pasižyminčiomis eilutėmis yra įvairiausi išvestiniai metodai. FARIMA modelis – modifikuotas ARIMA pritaikytas specialiai tik šioms didelio dažnio eilutėms [38]. FARIMA gali pakankamai gerai numatyti akcijų kainų kitimą, tačiau nepasiekia didelio tikslumo kainos numatymo prasme, tai yra padeda nuspėti ar kaina kils, ar kris, tačiau nėra pakankamai tiksliai nusakoma koks didelis bus pokytis. Kitas labai panašiomis savybėmis pasižymintis metodas yra GAARIMA, kuris yra genetinio algoritmo ir ARIMA modelio mišinys [39].

Labai sudėtingas, didelį svyravimų dažnį ar didelį matų skaičių turinčias eilutes taip pat puikiai prognozuoja ir kitas netiesinis modelis – neuroninių tinklų modelis. Ši metodologija puikiai pritaikoma ir vienmatėms, ir daugiamatėms laiko eilutėms. Dėl metodo sudėtingumo dažniausiai naudojamas vieno vidinio sluoksnio neuroninis tinklas [40], kiek rečiau naudojami ir gerokai sudėtingesni, daug sluoksnių ir įėjimų turintys neuroniniai tinklai arba neuroniniai tinklai, apjungiantys kelis kitus modelius, pavyzdžiui NARX, kuris papildomai išnaudoja AR modelį ir prideda egzogeninį kintamąjį (išorinį regresorių) [41]. Neuroniniai tinklai taip pat puikiai gali būti taikomi įvairaus pobūdžio, net ir sudėtingiausiomis laiko eilutėms, tokioms kaip: bankroto – lūžio [42], kvaziperiodinėms [43], medžiagų sąmatų [44], proceso atnaujinimo [45] ir daugeliui kitų [46]. Dėl savo sudėtingumo ir ilgos skaičiavimo trukmės šie modeliai taikomi retai ir dažniausiai mokslinei arba komercinei veiklai. Trumpai šio modelio veikimo principą galima paaiškinti schema:

$$\text{apmokymas : } \begin{array}{l} x_1 \rightarrow \\ x_2 \rightarrow \\ \dots \rightarrow \\ x_{n-k} \rightarrow \end{array} \begin{array}{l} \rightarrow \\ \rightarrow \\ \rightarrow \\ \rightarrow \end{array} \begin{bmatrix} w_{i,j} \\ b_{i,j} \end{bmatrix} \rightarrow x_{n-k+1}, \quad \text{progozavimas: } \begin{array}{l} x_1 \rightarrow \\ x_2 \rightarrow \\ \dots \rightarrow \\ x_{n+k} \rightarrow \end{array} \begin{array}{l} \rightarrow \\ \rightarrow \\ \rightarrow \\ \rightarrow \end{array} \begin{bmatrix} w_{i,j} \\ b_{i,j} \end{bmatrix} \rightarrow x_{n+k+1}.$$

Modeliui pateikiame duomenis x_1, x_2, \dots, x_{n-k} , tuomet modelis yra apmokomas ir nusistato ryšių tarp atskirų neuronų svorius $w_{i,j}$, ir leistinus nuokrypius $b_{i,j}$, taip, kad x_{n-k+1} būtų mažiausi. Tuomet modeliui naujai pateikus duomenis x_1, x_2, \dots, x_{n+k} ir apmokymo metu gautus svorius $w_{i,j}$, $b_{i,j}$, gaunama prognozė x_{n+k+1} .

Kartais laiko eilutėms prognozuoti ar numatyti jų elgseną ateityje naudojama visai kita strategija. Faktorišės analizės metodai, dažniausiai finansiniams ir ekonominiams duomenims, taikomi, siekiant vertinti duomenų eilutes, atspindinčias kokią nors riziką. Vykdam bendrąją faktorišę analizę daroma prielaida, kad stebimi duomenys priklauso nuo kokių nors nestebimų parametru, vadinamų faktoriais. Bendrasis modelis nusako tiesinę priklausomybę tarp stebimų ir nestebimų kintamųjų [31]. Tuomet gauname teisinės regresijos uždavinį su tam tikru faktoriu rinkiniu, kurį galime prognozuoti, pasinaudojant bendrąją tiesinio prognozavimo teorija. Vienas įdomesnių šioje grupėje esančių modelių yra BARRA modelis, kuris naudoja daugiafaktorišę principą, bei yra komerciniais tikslais sukurtas modelis, kuris yra skirtas tik toms eilutėms, kurios atspindi kokią nors riziką [47].

Būvio būsenų modeliai suteikia galimybę rasti neišmatuojamų kintamųjų įverčius, o svarbiausia – kad daugelis dinamiškų laiko eilučių modelių gali būti užrašyti būvio būsenų metodu. Pavyzdžiui, standartinis Gauso tiesinis modelis užrašytas būvio būsenų būdu įgyja formą [27]:

$$s_{t+1} = d_t + T_t s_t + R_t \eta_t, \quad y_t = c_t + Z_t s_t + e_t,$$

čia s_t yra m -dimensinis būvio vektorius, y_t yra k -dimensinis stebėjimų vektorius, d_t ir c_t yra m - ir k -dimensiniai deterministiniai vektoriai, T_t ir Z_t yra $m \times m$ ir $k \times m$ koeficientu matricos, R_t yra $m \times n$ matrica, η_t ir e_t yra n ir k eilės Gauso baltas triukšmas. Taip pat lengvai galime pakeisti ARMA modelį ir yra net trys būdai tai padaryti [8]. Pagrindiniu laikomas Harvey keitimas [48]. Šios metodikos pagrindinis plusas ARMA modeliui yra tas, kad AR ir MA modelių koeficientai yra tiesiogiai panaudoti sistemos matricose.

2.2. Optimizavimo algoritmų apžvalga

Apžvelgus keletą laiko eilučių modelių iš kiekvienos ankščiau minėtų grupių, svarbu paminėti, kad dauguma sudėtingų metodų, taip pat ir pagrindinis darbe nagrinėjamas metodas papildomai naudoja optimizavimo algoritmus. Sudėtingi prognozavimo metodai dažnu atveju neturi vienareikšmio rezultato ir kas kartą bandoma pasiekti geriausią galimą reikšmę, kuri tenkintų prognozavimo metodo keliamas sąlygas. Tai galioja tiems prognozavimo metodams, kuriose aprašytos įvairios taisyklės ar sąlygos, kurioms galiojant prognozė bus kiek norima tiksli. Šie metodai bendru atveju turi be galo daug sprendinių, todėl kas kartą prognozuojant siekiama kiek įmanoma priartėti prie pačio tiksliausio, tai yra siekiama priartėti prie kokios nors teorinės reikšmės, su kuria prognozė būtų pati tiksliausia.

Optimizavimo algoritmai bendru atveju yra skirstomi į šias grupes:

- Kombinatoriniai metodai
Apsikeitimo, tiesinio programavimo, matroidų, srautinio tinklo, keliaujančio pirklio metodai [49, 50, 51, 52, 53, 54];
- Dinaminio programavimo metodai
MDP, POMDP, rekursinės ekonomikos, rekursinės makroekonomikos metodai [55, 56, 57, 58, 59, 60];
- Genetiniai (evoliuciniai) algoritmai
Genų ekspresijos programavimo, evoliucijos simuliacijos metodai bei gamtos evoliucija paremtos metaeuristikos [61, 62, 63, 64, 65, 66, 67];
- Gradientiniai metodai
Gradientinio nusileidimo, Landweber iteracijų, Frank-Wolfe, sujungtojo gradiento metodai [68, 69, 70, 71];
- Metaeuristikos
Colonial atrinkimo, MIMIC, dalelių spiečiaus, atsitiktinės paieškos, atkaitinimo modeliavimo, Tabu paieškos, skruzdžių kolonijos, HUMANT, šikšnosparnių grupės, bičių spiečiaus metodai [72, 73, 74, 75, 76, 77, 78, 79];
- Stochastiniai metodai
Bayeso, BRST, CMA-ES, OCBA, lygiagreto užgrūdinimo, kvantinio atkaitinimo metodai [80, 81, 82, 83, 84, 85];

Visos klasifikacijoje pateiktos nuorodos apžvelgia pagrindinius, dažniausiai literatūroje minimus ir analizuojamus optimizavimo modelius bei metodus. Ši klasifikacija yra daugiau bendrinė ir ji nėra labai griežta, nes atsiranda vis daugiau naujų modelių ir metodų, kurie naudoja visiškai naujas idėjas ir ne visada gali būti tiksliai priskirti, kuriai nors iš aukščiau paminėtų grupių.

Kombinatorinio optimizavimo metodai gana retai įtraukiami į prognozavimo modelius. Jie daugiausia skirti surasti kokį nors optimalų rinkinį ir pagrindinis jų panaudojimas prognozavimo modeliuose yra: nustatyti tinkamiausią išorinių veiksnių rinkinį, kai turima labai daug reikšmingų egzogeninių kintamųjų ir pilno perrinkimo metodas užima labai daug laiko. Populiariausias iš jų ekonominiams duomenims yra srautinio tinklo modelis, kuris paremtas grafų teorija. Įvedus tinkamą tikslo funkciją per šį grafą pereina maksimalus srautas. Integracijos į prognozavimo modelį metu grafas dažniausi žymi išorinių regresorių reikšmingumą, o tikslo funkcija siekia, kad bendrasis kelių regresorių reikšmingumas būtų maksimalus [86].

Ekonominiuose tyrimuose pagrindiniai yra rekursinės ekonomikos arba rekursinės makroekonomikos metodai, kurie priklauso dinaminio programavimo grupei. Šie metodai naudojami tais atvejais, kai siekiama maksimizuoti ilgojo periodo investicijų grąžą, tai yra dažniausiai siekiama suformuoti investicinį portfelį, kurio grąža būtų maksimali net ir įvertinus finansinį diskontavimą [87, 88, 89]. Pavyzdžiui, prognozuojame visas rinkos akcijas ir vertiname jų grąžą šiai dienai, tai yra papildomai taikome diskonto procedūrą. Tuomet sprendžiamas klausimas kokias akcijas ir kokiam laikui pirkti norint gauti maksimalią grąžą. Pilnas perrinkimas taip pat galimas, tačiau labai neefektyvus, tuo tarpu rekursinės ekonomikos optimizavimo algoritmas gana nesunkiai susitvarko su šia užduotimi [90, 91].

Genetiniai algoritmai yra vieni pagrindinių algoritmų, kurie naudojami prognozavimo modelių paruošimui. Šie metodai bando simuliuoti natūralioje gamtoje vykstančią evoliuciją. Vienas iš geriausių rezultatus pateikiančių prognozavimo modelių yra neuroninis tinkas, kuris beveik visada yra apmokomas genetinio algoritmo pagalba [92]. Dažnai tokia kombinacija vadinama neuroevoliuciniais algoritmais [93]. Kita šių algoritmų panaudojimo sritis – kokių nors pagalbinių duomenų paruošimas, tai yra prognozavimo metodas reikalauja kokių nors sąlygų, kurios nėra pasiekiamos vienareikšmiškoje situacijoje. Tuomet pagal tikslo funkciją

genetinis algoritmas evoliucionuoja tol, kol tenkina prognozavimo modelio sąlygą ties maksimalia ar minimalia tikslo funkcijos reikšme. Plačiau genetinis algoritmas aprašytas 3.2 skyrelyje.

Gradientiniai metodai naudojami labai specifinėse ekonominės analizės dalyse. Viena iš jų yra eilutės, pasižyminčios dažniais pokyčiais ir didele svyravimų amplitude, vertinimas. Šiuo atveju duomenys yra prilyginami signalui ir bandoma atkurti teigiant, kad signalas buvo paveiktas kelių išorinių veiksnių [68, 94]. Šią situaciją gana gerai iliustruoja valiutų kursų ar akcijų kainų duomenų eilutės, nes jos yra nemažai įtakojamos išorinių veiksnių. Dar šie algoritmai naudojami kai siekiama nustatyti minimalius kaštų srautus modeliuojant su transportavimu susijusius duomenis [95].

Tiriant ekonominius ar finansinius duomenis metaeuristikos yra antra pagal populiarumą optimizavimo metodų grupė. Šios grupės algoritmai, taip pat kaip genetiniai algoritmai, yra integruojami į prognozavimo modelius. Daugiausiai literatūroje sutinkamas dalelių spiečiaus algoritmas, dar žinomas kaip PSO. Ši metaeuristika yra populiari, nes jai reikalinga labai nedaug arba išvis nereikalingos prielaidos [96]. Kita svarbi šios metaeuristikos savybė – galima nagrinėti labai dideles galimų sprendinių aibes. Esminis PSO algoritmo trūkumas – nėra garantijos rasti optimalų sprendinį. Taip pat svarbu paminėti, kad šis algoritmas kaip ir evoliucinis algoritmas, kiekvieną kartą gali generuoti visiškai skirtingus rezultatus [97]. Kita metaeuristika kuri literatūroje sutinkama beveik taip pat dažnai kaip PSO, yra modeliuojamojo atkaitinimo metodika, dar žinoma kaip SA [98]. Modeliuojamojo atkaitinimo metodika pasižymi galimybe nagrinėti labai dideles sprendinių aibes, tačiau šis metodas plačiausiai taikomas kai dirbama su diskrečiais duomenimis, pavyzdžiui, sprendžiant kompleksinius strateginio prognozavimo uždavinius [99].

Stochastinio optimizavimo grupė taip pat plačiai naudojama darbui su finansiniais ir ekonominiais duomenimis. Šioje grupėje yra labai įvairios paskirties metodų. Vienas iš jų OCBA yra tiksliai pritaikytas spręsti biudžeto paskirstymo uždavinius, tačiau šis metodas taip pat yra gana dažnai kombinuojamas su prognozavimu ir sprendžiamas atvirkštinis – biudžeto sudarymo uždavinys [100, 101, 102]. Kitas iš dažnesnių metodų yra kvantinio atkaitinimo metodas, kuris buvo sukurtas pasitelkiant SA metodo idėją [103]. Šis metodas, kaip ir SA metodas, naudojamas sprendžiant kompleksinius strateginio prognozavimo uždavinius, kai duomenys yra diskretūs.

Be aukščiau minėtų optimizavimo algoritmų panaudojimo prognozavime, jie ekonomikoje naudojami įvairioms situacijoms kontroliuoti ir įvertinti. Moderni optimizavimo teorija leidžia spręsti ekonominės pusiausvyros pasiekimo uždavinį, padeda nustatyti optimalias kaštų panaudojimo, lėšų paskirstymo ar tikslų pasiekimo strategijas. Plačiai šis optimizavimo panaudojimas žinomas kaip kontrolės teorija, kuri padeda įvertinti rinką bei pasirinkti optimalią strategiją. Tačiau kaip ir ankščiau minėta, šiuos optimizavimo modelius naudojant kartu su prognozavimo metodais galima dar tiksliau nustatyti ir įvertinti ekonominę situaciją bei sukurti strategiją.

Šiame darbe nagrinėjamas naujas ir dar gana nedaug ištirtas, Hankelio matricomis paremtas, M. Ragulskio, K. Lukoševičiūtės, Z. Navicko ir R. Palivonaitės pasiūlytas algebrinio skeleto modelis [1]. Plačiau šis modelis aprašytas 3.1 skyrelyje. Šis modelis priklauso metodų grupei, kuri dar nėra plačiai žinoma ir taikoma. Tai metodai paremti algebrinėmis taisyklėmis. Algebrinio skeleto modelis tai pat naudoja optimizavimo algoritmą. Tyrime naudojamas genetinis optimizavimo algoritmas plačiau aprašytas 3.2 skyrelyje.

3. Metodų aprašymas

3.1. Algebrinio skeleto metodas

3.1.1. Hankelio matricos rango sąvoka

Algebrinio skeleto modelis yra pilnai paremtas Hankelio matricos rangų, todėl pradžioje apibrėžiame rango sąvoką [104, 105]. Pažymėkime realiųjų skaičių seką S :

$$(x_0, x_1, x_2, \dots) = (x_k; k \in N_0).$$

Atskirkime kokį nors jos posekį $S_j = (x_j, x_{j+1}, x_{j+2}, \dots)$, $j = \overline{0, m}$ ir šiam sekos S posekiui S_j suformuokime Hankelio matricą H :

$$H = \begin{bmatrix} x_0 & x_1 & x_2 & \dots \\ x_1 & x_2 & x_3 & \dots \\ x_2 & x_3 & x_4 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}.$$

Apibrėžiame Hankelio matricos H minorus $H_j^{(m)}$:

$$H_j^{(m)} = \begin{bmatrix} x_j & x_{j+1} & \dots & x_{j+m-1} \\ x_{j+1} & x_{j+2} & \dots & x_{j+m} \\ \dots & \dots & \dots & \dots \\ x_{j+m-1} & x_{j+m} & \dots & x_{j+2m-2} \end{bmatrix},$$

čia m – kvadratinės Hankelio matricos eilė. Šių Hankelio matricų minorų determinantų seką pažymėsime $d_j^{(m)} = \det H_j^{(m)}$, $m \geq 1$.

Tuomet sekos S posekio S_j rangas yra toks $m_j = Hr(x_j; j \in N_0)$, kuris tenkina sąlygą:

$$d^{(m_j+k)} = 0, \quad \forall k \in N, \quad \text{kai } d_j^{(m_j)} \neq 0.$$

Svarbu paminėti, kad $m_j < \infty$, o jei $m_j = \infty$, tuomet seka rango neturi.

Kai posekis S_j yra $Hr(S_j) = m$ ir $m < \infty$, tuomet įmanoma sukonstruoti charakteristinį posekio S_j determinantą [106]:

$$\Delta^{(m)} S_j(\rho) = \begin{vmatrix} x_j & x_{j+1} & \dots & x_{j+m} \\ x_{j+1} & x_{j+2} & \dots & x_{j+m+1} \\ \dots & \dots & \dots & \dots \\ x_{j+m-1} & x_{j+m} & \dots & x_{j+2m-1} \\ 1 & \rho & \dots & \rho^m \end{vmatrix}$$

ir charakteristinę šio posekio lygtį:

$$\Delta^{(m)} S_j(\rho) = 0.$$

Charakteristinės lygties šaknys gali būti apskaičiuojamos, kai $\rho_k \in C$, $k = \overline{1, r}$, čia šaknų rekurentiniai indeksai yra n_k , $n_k \in N$ ir tenkina sąlygą $n_1 + n_2 + \dots + n_r = m$. Tuomet gauname x_n išraišką:

$$x_n = \sum_{k=1}^r \sum_{l=0}^{n_k-1} \mu_{kl} \binom{n}{l} \rho_k^{n-l}, \quad n = j, j+1, j+2, \dots,$$

čia koeficientai $\mu_{kl} \in C$, $k = \overline{1, r}$ ir $l = 0, 1, \dots, n_k - 1$ gali būti apskaičiuojami iš tiesinių lygčių sistemos. Taip pat šios sistemos sprendinys turi vieną vienintelį sprendinį, įrodymas pateiktas [7].

3.1.2. Prognozavimo strategija

Tarkim, kad žinomos $2n + 1$ istorinės laiko eilutės reikšmės, kurias naudosime sukonstruoti bendrąjį proceso modelį [1, 104], kuriuo pasinaudojant bus galima prognozuoti laiko eilutės dinamiką ateityje:

$$x_0, x_1, x_2, \dots, x_{2n-1}, x_{2n},$$

čia x_{2n} yra dabartinio momento reikšmė. Imame nelyginį stebėjimų skaičių ir sukonstruojame Hankelio matricos minorą $H_0^{(n+1)}$:

$$H_0^{(n+1)} = \begin{bmatrix} x_0 & x_1 & \dots & x_n \\ x_1 & x_2 & \dots & x_{n+1} \\ \dots & \dots & \dots & \dots \\ x_n & x_{n+1} & \dots & x_{2n} \end{bmatrix}.$$

Jei $\det(H_0^{(n+1)}) = 0$, tai seka yra algebrinė progresija ir x_{2n+1} galima rasti tiesiogiai pagal formulę:

$$x_n = \sum_{k=1}^r \sum_{l=0}^{n_k-1} \mu_{kl} \binom{n}{l} \rho_k^{n-l}, \quad n = j, j+1, j+2, \dots.$$

Dirbant su realaus pasaulio laiko eilutėmis rangas gali ir neegzistuoti – tai reiškia, kad seka $x_0, x_1, x_2, \dots, x_{2n-1}, x_{2n}$ nėra algebrinė progresija. Todėl daroma papildoma prielaida, kad nagrinėjama eilutė buvo paveikta kokio nors išorinio triukšmo. Pagal šią prielaidą teigiame, kad eilutė yra algebrinė progresija plius tam tikras atsitiktinis triukšmas, kitais žodžiais tariant, formuluojama hipotezė, kad [1]:

$$x_k = \tilde{x}_k + \varepsilon_k, \quad k = 0, 1, 2, \dots, 2n,$$

čia ε_k yra pridėtas atsitiktinis triukšmas. Tuomet determinantas:

$$\tilde{H}_0^{(n+1)} = \begin{vmatrix} x_0 - \varepsilon_0 & x_1 - \varepsilon_1 & \dots & x_n - \varepsilon_n \\ x_1 - \varepsilon_1 & x_2 - \varepsilon_2 & \dots & x_{n+1} - \varepsilon_{n+1} \\ \dots & \dots & \dots & \dots \\ x_n - \varepsilon_n & x_{n+1} - \varepsilon_{n+1} & \dots & x_{2n} - \varepsilon_{2n} \end{vmatrix} = 0,$$

čia

$$\tilde{H}_j^{(m)} = [x_{r+s-2+j} - \varepsilon_{r+s-2+j}]_{1 \leq r, s \leq m}.$$

Taip pat papildomai daroma prielaida, kad ši algebrinė progresija \tilde{x}_k , $k = 0, 1, 2, \dots$ yra tam tikra algebrinė skeletinė seka, išsilaikanti laiko eilutės artimiausioje ateities dinamikoje.

Vienas iš svarbiausių algebrinio skeleto metodo uždavinių yra triukšmo ε_k identifikavimas. Nesunku pastebėti, kad šis uždavinys neturi vienareikšmio sprendinio. Tuomet iškeliamas tikslas – minimizuoti visus įmanomus nuokrypius nuo tikrosios laiko eilutės. Triukšmo sekos generavimui $\{\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}\}$ buvo pasiūlyta tokia tikslo funkcija [1, 104], kurią maksimizuojame:

$$F_1(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}) = \frac{1}{a_e |\det(\tilde{H}_0^{(n+1)})| + \sum_{k=0}^{2n} \lambda_k |\varepsilon_k|}, \quad a_e > 0,$$

čia

$$\lambda_k = \frac{\exp(b_\lambda(k+1))}{\sum_{j=0}^{2n} \exp(b_\lambda(j+1))}, \quad k = 0, 1, \dots, 2n, \quad b_\lambda > 0.$$

Akivaizdu, kad $\sum_{k=0}^{2n} \lambda_k = 1$, $0 < \lambda_0 < \lambda_1 < \dots < \lambda_{2n-1} < \lambda_{2n}$.

Jeigu $\det(\tilde{H}_0^{(n+1)}) = 0$, tuomet tikslo funkcijos reikšmė artės į begalybę, kai visi triukšmo sekos elementai artės į nulį [1], tai yra jei $\varepsilon_0 \rightarrow 0$, $\varepsilon_1 \rightarrow 0$, ..., $\varepsilon_{2n} \rightarrow 0$, tuomet $F(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}) \rightarrow \infty$. Tikslo funkcijos išraiškoje parametras a_e apibrėžia proporcijos tarp determinanto reikšmės ir triukšmo su svoriais sumos baudą (abi baudos turi tą pačią reikšmę, kai $a_e = 1$). Koefficientai $\lambda_0, \lambda_1, \dots, \lambda_{2n}$ nustato tolerancijos koridorių triukšmo sekai $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2n}$. Visų triukšmo sekos elementų svoriai bus vienodi, jei $b_\lambda = 0$. Kuo didesnis b_λ imamas, tuo didesnis svoris suteikiamas triukšmo sekos dabarties momentui, lyginant su triukšmo sekos svoriais praeities momentais. Šis parametras tyrime nekeičiamas ir naudojamas toks koks buvo nurodytas straipsnio [1] autorių.

Kitaip tariant, kuo koreguojamasis taškas yra toliau praeityje, tuo labiau toleruojama tame taške esančios reikšmės korekcija triukšmo sekos elementu. Taip daroma prielaida, kad kuo duomenys yra arčiau dabarties, tuo jie svarbesni [104].

Taip pat reikia pabrėžti, kad šis uždavinys neturi vienintelio sprendinio. Žinoma, triukšmas turi būti kiek įmanoma mažesnis, bet pagal F_1 išraišką taip pat yra leidžiama, kad duomenų, pakoreguotų triukšmu, Hankelio matricos determinantas gali ir nebūti tiksliai lygus nuliui. Taigi uždavinys – maksimizuojant tikslo funkciją rasti tokią triukšmo ε_k seką, kuria paveikus tikrųjų duomenų seką, gaunama duomenų seka, iš kurios suformuojamas Hankelio matricos determinantas yra apytiksliai lygus nuliui. Taip pat reikalaujama, kad ir visų triukšmo sekos elementų svorinė suma būtų kiek įmanoma mažesnė. Visa tai leidžia išskirti apytikslį algebrinį dėsnį – algebrinę progresiją, kuri aprašo analizuojamos duomenų sekos modelį [104].

Pakoregavus seką triukšmo seka, toliau ieškomos duomenų sekos ateities reikšmės pagal x_n formulę, net jei pakoreguotos sekos determinantas yra tik apytiksliai lygus nuliui [1].

Galima pastebėti, kad pristatytas prognozavimo modelis – tai netiesinis algebrinis modelis, nes Hankelio matricos rango apibrėžimas ir idėja apima netiesinius algebrinius sąryšius. Tai yra, šis modelis naudoja netiesinį algebrinį skeletinės sekos identifikavimo metodą. Ypač svarbi šio modelio savybė yra būtinybė fiksuoti parametą n , tai yra identifikuoti sekos pagrindinį fragmentą. Kadangi koreguoti galima skirtingo ilgio sekas, keliamas esminis tikslas: parinkti tokio ilgio sekos fragmentą, kuriam užtektų panaudoti patį mažiausią triukšmą [104].

3.2. Genetinis algoritmas

Genetinius algoritmus, dar vadinamus evoliuciniais, pirmą kartą 1950-aisiais metais pasiūlė Alanas Turingas [107]. Dėl šių A. Turingo pirmųjų darbų, genetiniai algoritmai buvo sėkmingai išbandyti sprendžiant įvairiausias optimizavimo uždavinius. Genetinių algoritmų veikimas yra pagrįstas evoliucija, vykstančia natūralioje gyvojoje gamtoje, t.y., natūraliosios atrankos proceso imitavimu, kitaip tariant evoliucija. Priešingai nei Tabu paieška ir atkaitinimo modeliavimo metodai, genetiniai algoritmai nėra paremti lokaliąja paieška.

Genetinis algoritmas prasideda parenkant atsitiktinį sprendinių sąrašą:

$$\{s_1^0, s_2^0, \dots, s_n^0\}$$

kurie yra vadinami populiacija pradiniu laiko momentu arba populiacija laiku 0.

Esminė algoritmo procedūra yra ciklas kuris sukuria naujas populiacijas

$$\{s_1^{t+1}, s_2^{t+1}, \dots, s_n^{t+1}\}$$

laiko momentui $t + 1$ arba kitaip tariant populiaciją laiku $t + 1$ atsižvelgiant į tai, kad populiacija prasidėjo laiku t . Šiam tikslui reikia kiekvieną kartą perskaičiuoti tikslo funkciją kiekvienam sprendiniui s_i^t esančiam sąrašė.

Pasinaudojant svorine randomizacija, t laiku n elementų yra parenkama iš populiacijos. Akivaizdu, keletas sprendinių gali būti parenkami ne kartą. Svorinė randomizacija paremta tikslo funkcijos reikšme ir tikimybėmis. Idėja yra paprasta – sprendiniai, turintys geresnes tikslo funkcijos reikšmes, įgyja didesnes tikimybes būti parinktais. Šiuo būdu geresni sprendiniai įgyja daugiau kopijų, o blogesni sprendiniai, populiacijai augant, išnyksta dėl per mažų tikimybių [108].

Naudojame vieno taško modifikuotą β – kryžminimo metodą, kai kryžminamo taško vieta kiekvienai kryžminamai porai yra atsitiktinė [109]. Šis metodas ne tik apkeičia dalį sprendinio segmentų, bet ir netrukdo naujai sugeneruotiems sprendiniams tapti daugiau ar mažiau panašiais į juos sukūrusius sprendinius. Šiame metode sprendiniai dažnai vadinami chromosomomis, sprendinio segmentai – genais, o juos sukūrę sprendiniai – tėvais.

β – kryžminimo algoritmą galime aprašyti šiomis lygtimis:

$$\tau_{j,L}^{k+1}|_- = \text{round} \left(\frac{\beta}{\beta + 1} \tau_{j,L}^k + \frac{1}{\beta + 1} \tau_{j,R}^k \right),$$

$$\tau_{j,R}^{k+1}|_- = \text{round} \left(\frac{1}{\beta + 1} \tau_{j,L}^k + \frac{\beta}{\beta + 1} \tau_{j,R}^k \right),$$

$$\tau_{j,L}^{k+1}|_+ = \text{round} \left(\frac{1}{\beta + 1} \tau_{j,L}^k + \frac{\beta}{\beta + 1} \tau_{j,R}^k \right),$$

$$\tau_{j,R}^{k+1}|_+ = \text{round} \left(\frac{\beta}{\beta + 1} \tau_{j,L}^k + \frac{1}{\beta + 1} \tau_{j,R}^k \right),$$

čia k – generacijos numeris, $\tau_{j,L}^k$ yra j -tasis genas iš kairės tėvų chromosomos, $\tau_{j,R}^k$ yra j -tasis genas iš dešinės tėvų chromosomos, $\tau_{j,L}^{k+1}$ yra j -tasis genas iš kairės dukterinės chromosomos, $\tau_{j,R}^{k+1}$ yra j -tasis genas iš dešinės dukterinės chromosomos, minuso ženklas apačioje (−) žymi, kad j -tasis genas yra virš kryžminimo taško, o pluso ženklas apačioje (+) žymi, kad j -tasis genas yra žemiau nei kryžminimo taškas. Taigi, bet kuriuo atveju, dukterinių chromosomų j -

tieji genai tilps į intervalą nuo $\tau_{j,L}^k$ iki $\tau_{j,R}^k$ imtinai. Taip pat yra žinoma, kad β – kryžminimo algoritmas konverguoja į klasikinį kryžminimo algoritmą, kai $\beta \rightarrow 0$ arba $\beta \rightarrow \infty$ [1].

Kryžminimo procedūra nėra naudojama visiems sugeneruotiems sprendiniams. Šiai procedūrai kontroliuoti naudojamas kryžminimo koeficientas p_R , kuris yra tikimybė, kad kryžminimo operacija bus vykdoma sprendinių porai.

Dar yra naudojama mutacija, kuri kryžminimo metu atsitiktinai keičia dalį segmento. Ji svarbi genetiniam algoritmui, nes sukuria galimybę algoritmui neužsistovėti lokaliame sprendinyje ir potencialiai nukeliauti į kitą lokalų ar net globalų sprendinį. Mutacijos koeficientas aprašomas algoritmo pradžioje ir žymimas p_M . Mutacijos koeficientas yra tikimybė, kad sprendiniui bus taikoma mutacijos procedūra. Tyrime naudojama mutacijos procedūra: keliaujame per visus genus kiekvienoje chromosomoje iš tuo metu sugeneruotų sprendinių ir generuojame atsitiktinį skaičių, tolygiai pasiskirsčiusį intervale $[-0.2 ; 0.2]$.

Egzistuoja daugybė įvairių būdų, kaip konkrečiai realizuoti kryžminimo, mutacijos bei populiacijos koeficientų atnaujinimo ar parinkimo procedūras, nemažai jų pateikta D. S. Johnsono darbuose [49].

Genetinis algoritmas yra stabdomas pagal vieną iš trijų iš anksto nustatytų sąlygų:

- kai geriausias sprendinys pasiekia kažkokią iš anksto apibrėžtą tikslo funkcijos reikšmę;
- kai algoritmas sugeneruoja tam tikrą kiekį populiacijų;
- kai tam tikrą laiką algoritmas nepasiekia jokio progreso tikslo funkcijos atžvilgiu.

Pagrindiniai genetinio algoritmo kontrolės parametrai yra:

- populiacijos dydis n ;
- kryžminimo tikimybė p_R ;
- mutacijos tikimybė p_M ;
- generacijų skaičius n_G ;
- algoritmo stabdymo ribojimai.

3.3. ARIMA ir ARIMAX

ARIMA – integruotas autoregresinis slenkančio vidurkio modelis, kuris yra vienas iš dažniausiai naudojamų prognozavimo modelių. Šis modelis susideda iš dviejų procesų modelių ir integravimo operacijos. Pirmasis modelis – autoregresinis modelis $AR(p)$, kuris aprašomas lygtimi [31]:

$$P(L)\xi_t = \varepsilon_t ,$$

čia $P(z) = 1 - a_1z - \dots - a_pz^p$, vidurkis $\mu = P(1)\mathbb{E}\xi_t$, ε_t žymi baltą triukšmą, o L yra postūmio atgal operatorius, tai yra $L^k\xi_t = \xi_{t-k}$. Antrasis modelis – slenkančio vidurkio modelis $MA(q)$, kuris yra aprašomas lygtimi:

$$\xi_t = Q(L)\varepsilon_t ,$$

čia $Q(z) = 1 + b_1z + \dots + b_qz^q$, vidurkis $\mu = \mathbb{E}\xi_t$. Tuomet sujungę šiuos modelius gauname $ARMA(p, q)$ modelį kuris aprašomas taip:

$$P(L)\xi_t = Q(L)\varepsilon_t .$$

Vienas iš svarbiausių *ARMA* modelio reikalavimų yra stacionarumas prognozuojamai laiko eilutei, todėl pridėjama integravimo operacija, kurios pagalba dažnu atveju galima pasiekti stacionarumą. Šiuo atveju integruodami galime paversti laiko eilutę d eilės stacionaria eilute, jei jos d eilės pokyčiai yra stacionarūs, o $d - 1$ eilės pokyčiai nėra stacionarūs. Tuomet galima aprašyti $ARIMA(p, d, q)$ modelį lygtimi:

$$P(L)(1 - L)^d\xi_t = Q(L)\varepsilon_t.$$

Svarbu pastebėti, kad jei eilutė yra stacionari tuomet $d = 0$ ir gauname standartinį *ARMA* modelį.

Turėdami laiko eilutę ξ_t kuri tenkina *ARIMA* procesą, pažymim $\eta_t = (1 - L)^d\xi_t$. Tuomet randam $\eta_d, \eta_{d+1}, \dots, \eta_n$ ir iš šios imties įvertiname $P(z)$ ir $Q(z)$ polinomų koeficientus bei apskaičiuojame kovariacinę funkciją. Pritaikę bendrąją tiesinio prognozavimo teoriją, nesunkiai gauname prognozes $\hat{\eta}_t$, $t = n + 1, n + 2, \dots$. Tuomet $\hat{\xi}_t$ apskaičiavimas tampa rekurentiškas pasinaudojant formulėmis:

$$\eta_t = (1 - L)^d\xi_t = \sum_{k=0}^d (-1)^k \binom{d}{k} \xi_{t-k} ,$$

$$\xi_t = \eta_t - \sum_{k=0}^d (-1)^k \binom{d}{k} \xi_{t-k},$$

čia $\binom{d}{k} = \frac{d!}{k!(d-k)!}$.

Šiam modeliui ypač svarbus eilių p ir q parinkimas. Tam tikslui dažniausiai naudojama autokoreliacinės funkcijos, Akaike informacinis kriterijus arba Bajeso informacinis kriterijus. Autokoreliacinė funkcija aprašoma [31, 110, 111]:

$$acf(\tau) = cor(\xi_{t+\tau}, \xi_t) = \frac{cov(\xi_{t+\tau}, \xi_t)}{\mathbb{D}\xi_t},$$

o dalinės autokoreliacijos funkcija aprašoma:

$$pacf(\tau) = cor(\xi_{t+\tau} - E_{t+1,t+\tau-1}\xi_{t+\tau}, \xi_t - E_{t+1,t+\tau-1}\xi_t).$$

Gerokai dažniau naudojamas Akaike ar Bajeso informacinis kriterijus, kurie remiasi didžiausio tikėtimumo metodu. Šis kriterijus aprašomi:

$$AIC(k) = -2 \ln p(X, \theta) + 2k,$$

$$BIC(k) = -2 \ln p(X, \theta) + k \ln(n),$$

čia $AIC(k)$ yra Akaike informacinis kriterijus, $BIC(k)$ yra Bajeso informacinis kriterijus, $X = (\xi_1, \dots, \xi_n)$ yra imtis, kurios skirstinio tankis yra $p(X, \theta)$, o θ yra k -matis parametras, n yra imties dydis. Pasinaudojant Akaike ar Bajeso kriterijumi pasirenkamas modelis su kuriuo $AIC(k)$ ar $BIC(k)$ yra mažiausi. Yra nemažai ir kitų būdų nustatyti *ARIMA* modelio parametrus, dalis jų pateikiama matematinės programinės įrangos paketuose [112].

Nagrinėjant makroekonomiką buvo pastebėta [113], kad dažnu atveju galima pagerinti *ARIMA* modelio rezultatus pridėdant egzogeninį kintamąjį. Tuomet gauname integruotą slenkančio vidurkio modelį su išoriniu regresoriumi X_t , kuris žymimas *ARIMAX* ir aprašomas:

$$P(L)(1-L)^d \xi_t = \Theta(L)X_t + Q(L)\varepsilon_t.$$

Bendroji *ARIMA* ar *ARIMAX* modelio schema aprašoma Box-Jenkins metodologija [114, 115]:

1. integruotumo eilės nustatymas (parametro d radimas);
2. stacionarios laiko eilutės modelio parametų p ir q parinkimas;
3. modelio testavimas.

Trečiajame Box-Jenkins metodologijos etape tikrinama ar liekanos ε_t yra nekoreliuotos. Tam naudojama Ljung-Box statistika kuri apskaičiuojama pagal formulę:

$$Q = n(n + 2) \sum_{\tau=1}^m \frac{acf(\tau)^2}{n - \tau}.$$

Ši statistika turi skirtinį χ_m^2 , jei liekanos yra nepriklausomos.

3.4. Pearsono ir Spearmano koreliacijos koeficientai

Koreliacijos koeficientai statistikoje aprašo statistinį ryšį tarp dviejų kintamųjų. Šių koeficientų yra daugybė, tačiau populiariausi yra Pearsono ir Spearmano.

Pearsono koreliacijos koeficientui taikomos šios prielaidos [116]:

- duomenys sutvarkyti taip, kad turėtų kiek įmanoma panašesnius matavimus, tai yra, jei vienas duomuo matuojamas valiuta su centų tikslumu, tai ir antrasis duomuo matuojamas taip pat;
- duomenys turi tiesinį sąryšį;
- duomenyse nėra reikšmingų išskirčių;
- duomenys yra pasiskirstę pagal normalųjį skirstinį.

Pearsono koreliacijos koeficientas randamas iš formulės:

$$r_p(X, Y) = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2} \sqrt{\sum(y - \bar{y})^2}},$$

čia \bar{x} yra X vidurkis, \bar{y} yra Y vidurkis, $x \in X$ ir $y \in Y$, $cov(x, y)$ yra kovariacijos koeficientas tarp X ir Y . Duomenų normalumui patikrinti dažniausiai naudojami Shapiro-Wilk arba Anderson-Darling metodai.

Tais atvejais kai duomenys netenkina Pearsono koreliacijos koeficientui keliamų prielaidų, naudojamas Spearmano koreliacijos koeficientas [117]. Šis koeficientas yra neparametrinis ir atspindi ryšį tarp kintamųjų kurį galima nusakyti monotone funkcija. Šis koeficientas randamas iš formulės:

$$r_s(X, Y) = \frac{\text{cov}(rg_x, rg_y)}{\sigma_{rg_x} \sigma_{rg_y}},$$

čia rg_x yra x rangas, rg_y yra y rangas σ yra standartinis nuokrypis.

Abu šie koeficientai yra ribose nuo -1 iki 1 , ir kuo didesnė absoliutinė reikšmė, tuo koreliacija yra stipresnė. Standartinis šių koeficientų stiprumo aprašymas pateiktas 3.4.1 lentelėje.

3.4.1 lentelė. Koreliacijos koeficiento stiprumo skalė

r_p reikšmė	Koreliacijos stiprumas
$ r_p = 1$	Labai stipri
$0.7 \leq r_p < 1$	Stipri
$0.5 \leq r_p < 0.7$	Vidutinė
$0.2 \leq r_p < 0.5$	Silpna
$0 < r_p < 0.2$	Labai silpna
$ r_p = 0$	Nėra ryšio

3.5. Shapiro-Wilk ir Anderson-Darling normalumo tyrimai

Gana dažnai statistiniuose tyrimuose reikia nustatyti duomenų normalumą. Pavyzdžiui ankstesniame skyrelyje minėtam Pearsono koreliacijos koeficientui keliamas reikalavimas, kad duomenys priklausytų normaliajam skirstiniui. Šiam tikslui dažniausiai naudojamas Shapiro-Wilk arba Anderson-Darling tyrimas.

Shapiro-Wilk tyrimas tikrina ar duomenys priklauso normaliajam skirstiniui. Statistika aprašoma [118]:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{\frac{1}{2}}},$$

$$m = (m_1, \dots, m_n)^T,$$

čia $x_{(i)}$ yra i -tasis mažiausias skaičius duomenyse, \bar{x} yra vidurkis, m_1, \dots, m_n yra kitų kintamųjų vidurkiai, o V yra kovariacijos matrica. Ši statistika, žymima W nusako, kad duomenys yra pasiskirstę pagal normalųjį skirstinį, jei p – value $> \alpha$, čia α – parinktas reikšmingumo lygis,

dažniausiai 0.05 . Kitais žodžiais tariant jei parinktas $\alpha = 0.05$, o $p - value = 0.04$, tai duomenys nepriklauso normaliajam skirstiniui.

Andreson-Darling tyrimas naudojasi pasiskirstymo funkcija ir yra aprašomas statistika [118, 119]:

$$A^2 = -n - S ,$$

$$S = \sum_{i=1}^n \frac{(2i - 1)}{n} (\ln F(Y_i) + \ln(1 - F(Y_{n+1-i}))) ,$$

čia F yra duomenų kumuliacinė pasiskirstymo funkcija. Taip pat svarbu, kad čia Y_i yra surikiuoti duomenys. Ši statistika, žymima A nusako, kad duomenys yra pasiskirstę pagal normalųjį skirstinį, jei $p - value > \alpha$, čia α – parinktas reikšmingumo lygis.

3.6. Paklaidų metrikos

Prognozavimo tikslumui vertinti yra įvedamos įvairios paklaidų matavimo metrikos. Įvairiuose šaltiniuose naudojamos net ir kelios metrikos. Kartais metrikos būna gana paprastos, pavyzdžiui, paprasčiausias paklaidų vidurkis. Tačiau prognozavimo tikslumui gerokai dažniau naudojamos metrikos, įvertinančios absoliučias paklaidas arba jų kvadratus. Tokios metrikos yra: vidutinė kvadratinė paklaida MSE ir šaknis iš vidutinės kvadratinės paklaidos RMSE:

$$MSE = \frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_t)^2 ,$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_t)^2} ,$$

čia x_t – tikroji reikšmė, o \hat{x}_t – prognozavimo metu gautas įvertis. Taip pat dažnai naudojama procentinė ρ arba absoliutinė procentinė paklaida ρ_A , ypač kai norima palyginti skirtingų matavimų paklaidas arba vertinami finansiniai, ekonominiai duomenys:

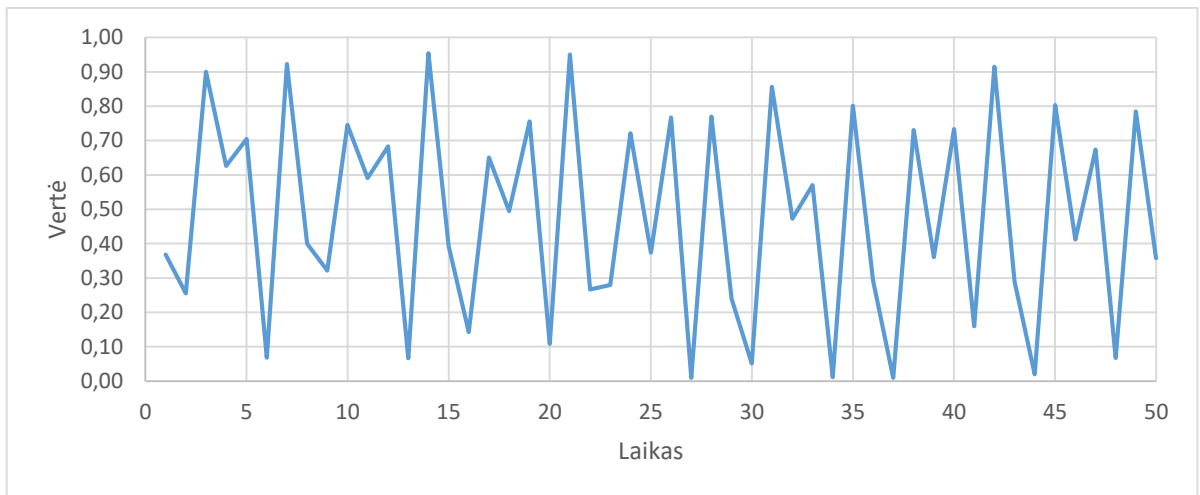
$$\rho = \frac{1}{n} \sum_{t=1}^n \frac{x_t - \hat{x}_t}{x_t} ,$$

$$\rho_A = \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - \hat{x}_t}{x_t} \right| .$$

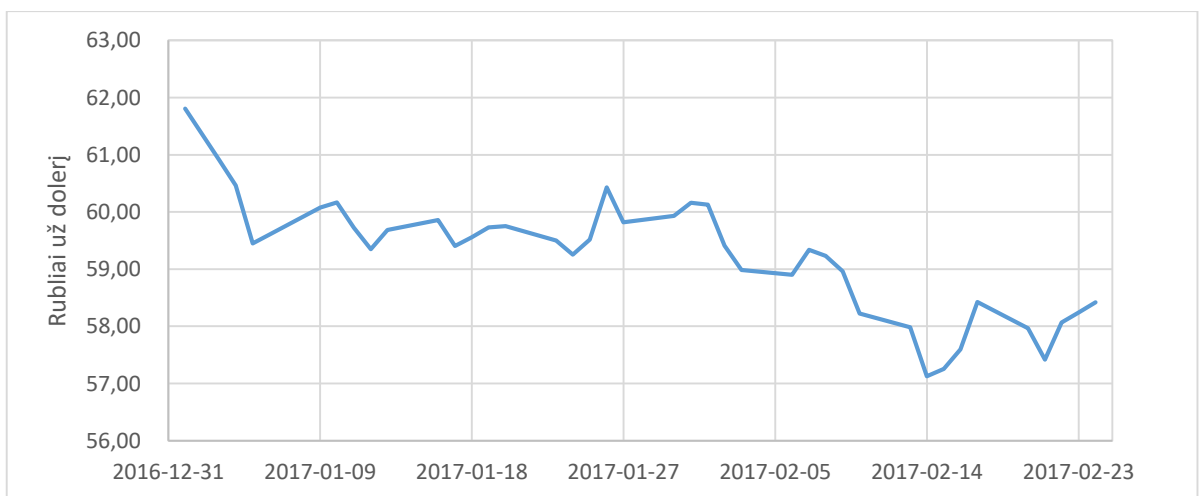
4. Tyrimas

Analizuojamas modelis buvo pritaikytas vieno nario prognozavimui į ateitį. Patobulinus algoritmą metodas buvo pakeistas į slenkantį prognozavimą, tai yra, suprognozavus vieną žingsnį į priekį prognozė įrašoma į darbinį duomenų masyvą ir naudojama tolimesnių žingsnių prognozavimui. Papildžius darbinį duomenų masyvą nauja reikšme iš naujo kartojamas visas algoritmas, tai yra, pilnai įvykdomas ir genetinis algoritmas, ir Hankelio matricių metodas.

Tyrimo naudojama testinė eilutė ir finansinė eilutė – USD/RUB kursas. Testinė eilutė yra dirbtinė eilutė, tai yra rekurentinė seka iš elementų 0.5, 0.7, 0.1, 0.9, 0.3, 0.2, 0.8, kuri kartojama keletą kartų, ir prie kiekvieno elemento pridedamas atsitiktinis triukšmas iš normaliojo skirstinio $[-0.15 ; 0.15]$.



4.1 pav. Testinės eilutės fragmento dinamika



4.2 pav. USD/RUB eilutės dinamika

Tyrime taip pat naudojami išoriniai regresoriai, kurie naudojami tiriant analizuojamo modelio taikymą su kitais prognozavimo modeliais, tai yra 4.8 skyriuje šie išoriniai regresoriai bus prognozuojami su analizuojamu modeliu ir tuomet įtraukiami į ARIMAX modelį, taip patikrinant, ar analizuojamas modelis gali būti integruotas į kitus prognozavimo modelius. Šie regresoriai aprašyti 4.8 skyrelyje, jų laiko eilučių dinamika pateikta 7.1 priede. Regresorių, kurie yra akcijų rinkų indeksai, sudėtis pateikta 7.2 priede.

Pagrindiniai tyrimo tikslai:

- ištirti modelio veikimą analizuojant laiko eilutę skirtingais laiko momentais;
- optimizuoti genetinio algoritmo parametrų parinkimą;
- optimizuoti rango parinkimą;
- pritaikyti modelį finansinėms eilutėms prognozuoti dešimt žingsnių (dvi finansų rinkos savaites) į priekį;
- pabandyti integruoti modelį į dažniausiai naudojamus, nesudėtingus modelius;
- palyginti optimalios modelio prognozės tikslumą su dažniausiai literatūroje naudojamais nesudėtingais modeliais bei ištirti prognozės kokybę, kombinuojant analizuojamus modelius.

Pagrindiniai sprendžiami uždaviniai:

- atlikti eksperimentą taikant modelį skirtingais laiko eilutės momentais;
- nustatyti optimalius modelio parametrus arba sukurti optimalių parametrų parinkimo metodologiją;
- eksperimentiniu būdu nustatyti modelio tinkamumą finansinėms laiko eilutėms;
- algebrinio skeleto modelio integracija į ARIMAX;
- algebrinio skeleto modelio palyginimas su ARIMA.

4.1. Prognozavimas iš skirtingų laiko momentų ir rango parinkimas

Pirmajame tyrimo žingsnyje buvo nutarta patikrinti, kaip modelis prognozuoja skirtingais laiko momentais, tam, kad išvengtų atsitiktinai gerų ar atsitiktinai blogų rezultatų bei pabandyti pastebėti bendrąsias modelio tendencijas ir išvelgti optimalią rango parinkimo metodiką.

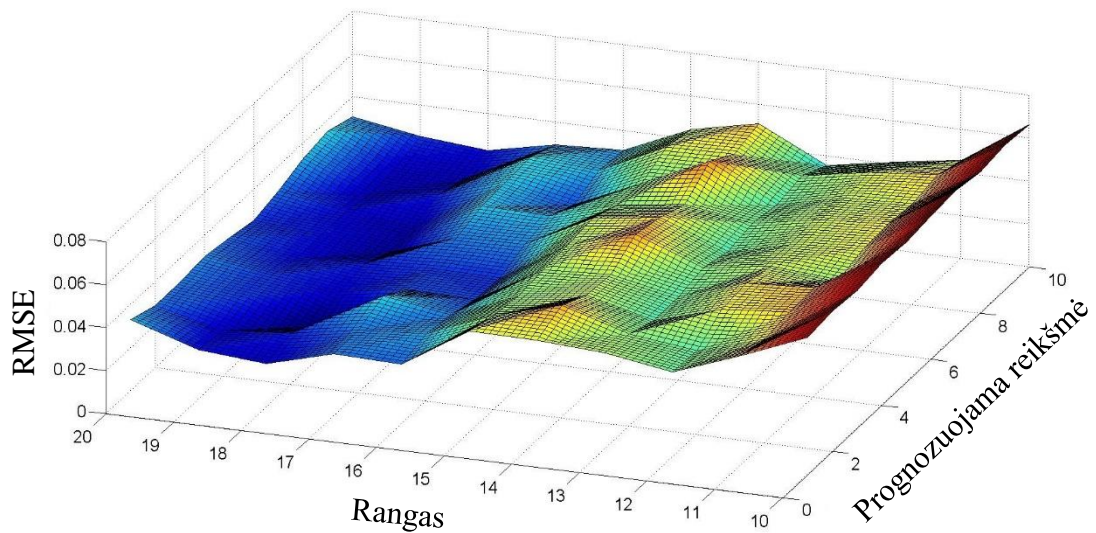
Šiame tyrimo etape naudojama testinė eilutė. Ši eilutė prognozuojama dešimt žingsnių į priekį, parenkant rangus nuo 10 iki 20. Rangai neatsitiktinai parinkti nuo 10. Ankstesnių tyrimų metu buvo pastebėta tendencija, kad rangas turi būti ne mažesnis nei prognozuojamų narių kiekis. Ankstesnių tyrimų rezultatai pateikti 7.6 priede. Tyrimas kartotas dešimt kartų atsitiktiniais laiko momentais. Prognozės tikslumui vertinti buvo pasirinkta RMSE (ang. root mean square error) metrika. RMSE metrika apskaičiuojama pagal formulę $RMSE = \sqrt{\frac{\sum_{i=1}^n (\tilde{y}_i - y_i)^2}{n}}$, čia \tilde{y} – prognozės metu gautas įvertis. Ši metrika pasirinkta, nes ji yra viena dažniausiai naudojamų metrikų.

Pasinaudojant MS Excel paketo funkcija *RANDBETWEEN*(1; 497) atsitiktinai parinkti laiko eilutės pradiniai prognozės momentai:

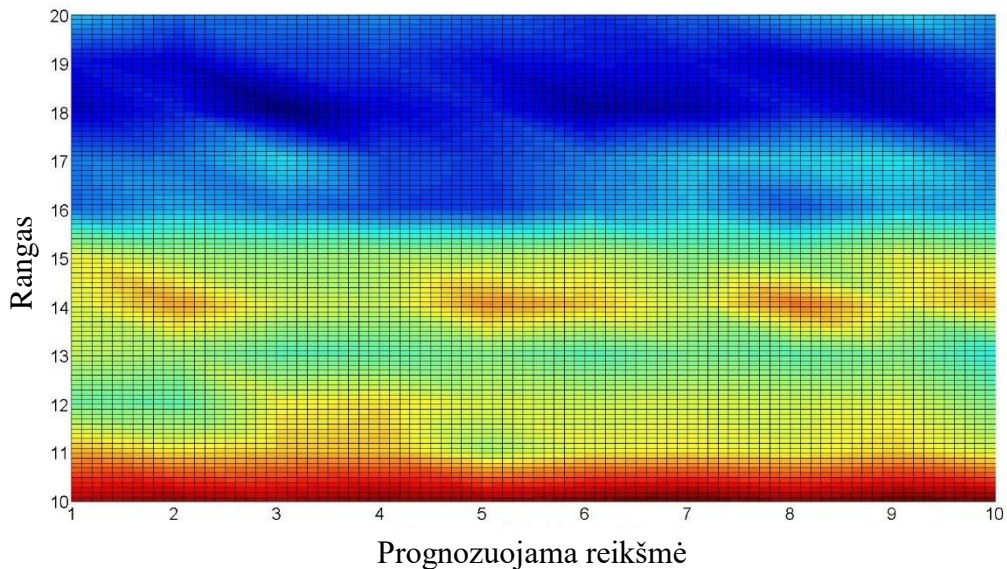
$$\{ 32, 113, 255, 212, 83, 281, 374, 148, 330, 8 \}.$$

Kai laiko eilutės prognozavimo modelis konstruojamas nuo 32-osios reikšmės, rezultatai pateikti paveiksluose 4.1.1-4.1.2, nuo 113-osios reikšmės paveiksluose 4.1.3-4.1.4, nuo 255-osios reikšmės paveiksluose 4.1.5-4.1.6. Kai modelis konstruojamas nuo kitų pradinių reikšmių rezultatų paveikslai pateikti 7.3 priede.

Pateikiama po du paveikslus kiekvienam iš bandymų. Pirmasis paveikslas – trijų ašių RMSE grafikas, antrasis paveikslas – dviejų ašių RMSE grafikas. Abiejuose grafikuose naudojama skalė nuo raudonos, žyminčios blogiausią rezultatą, iki mėlynos, žyminčios geriausią rezultatą, o geltona žymi vidutinį rezultatą.



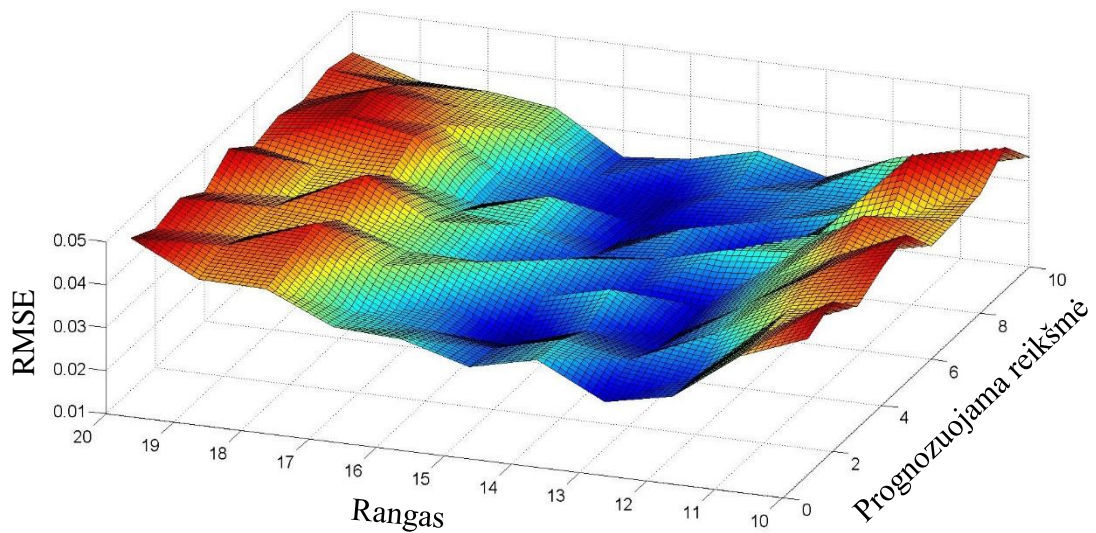
4.1.1 pav. Laiko eilutės analizė pradedant nuo 32-os bendros sekos reikšmės. Trijų ašių grafikas.



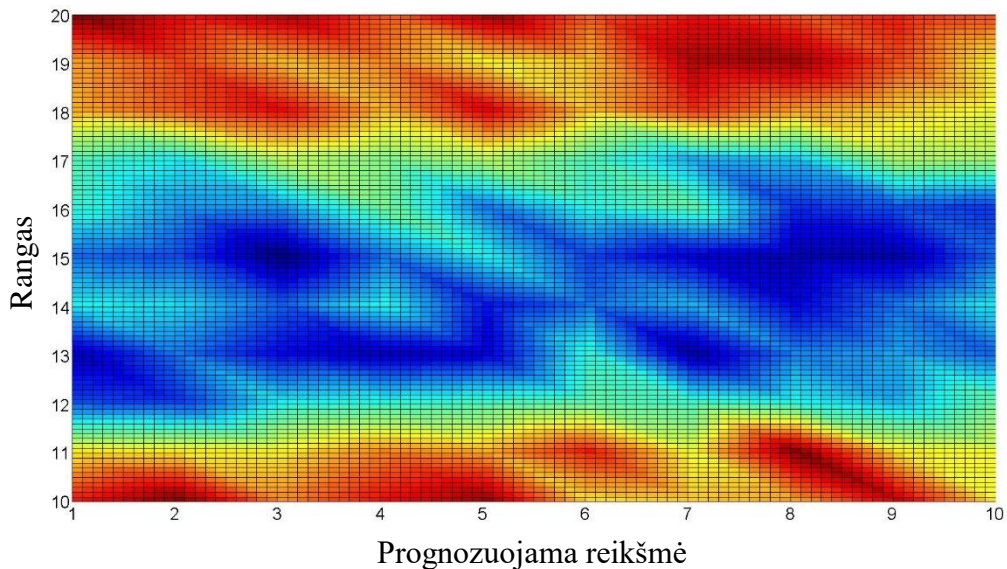
4.1.2 pav. Laiko eilutės analizė pradedant nuo 32-os bendros sekos reikšmės. Dviejų ašių grafikas.

Analizuojant eilutę, kuri prasideda nuo 32-os bendros sekos reikšmės, pastebime, kad:

- yra ryški lūžio linija ties rangu 16;
- geriausia RMSE reikšmė pasiekama ties 18 rangu;
- ties visais rangais, RMSE reikšmė išlieka gana pastovaus dydžio, tai yra kinta labai nežymiai, nepriklausomai nuo prognozuojamų reikšmių skaičiaus.



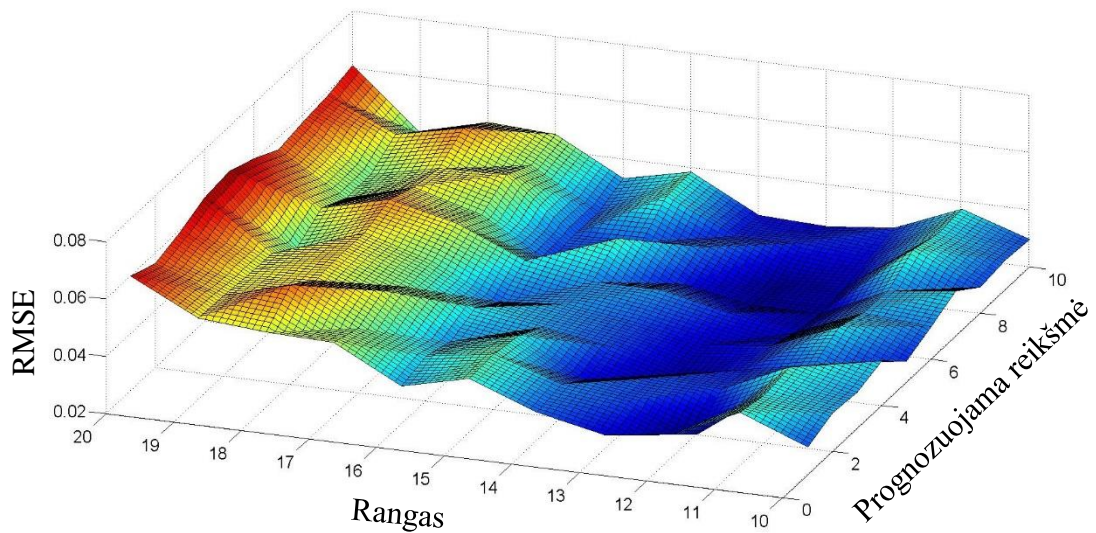
4.1.3 pav. Laiko eilutės analizė pradedant nuo 113-os bendros sekos reikšmės. Trijų ašių grafikas.



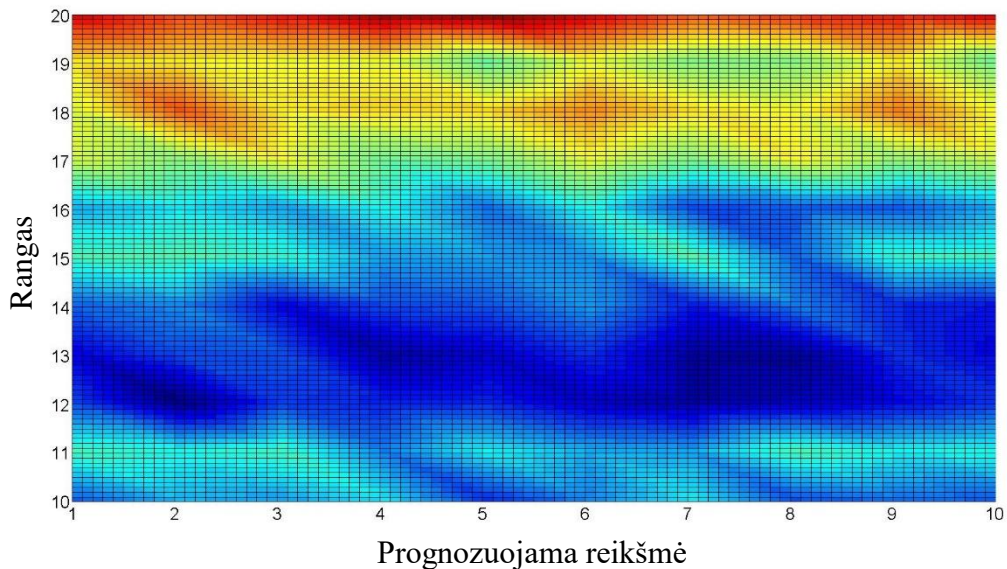
4.1.4 pav. Laiko eilutės analizė pradedant nuo 113-os bendros sekos reikšmės. Dviejų ašių grafikas.

Analizuojant eilutę, kuri prasideda nuo 113-os bendros sekos reikšmės, pastebime, kad:

- yra ryškios lūžio linijos ties rangais 11-12 ir 17-18;
- geriausia RMSE reikšmė pasiekama ties 15 rangu;
- ties visais rangais, RMSE reikšmė išlieka gana pastovaus dydžio, tai yra kinta labai nežymiai, nepriklausomai nuo prognozuojamų reikšmių skaičiaus.



4.1.5 pav. Laiko eilutės analizė pradedant nuo 255-os bendros sekos reikšmės. Trijų ašių grafikas.



4.1.6 pav. Laiko eilutės analizė pradedant nuo 255-os bendros sekos reikšmės. Dviejų ašių grafikas.

Analizuojant eilutę, kuri prasideda nuo 255-os bendros sekos reikšmės, pastebime, kad:

- yra ryški lūžio linija ties rangų 16;
- geriausia RMSE reikšmė pasiekama ties 13 rangų;
- ties visais rangais, RMSE reikšmė išlieka gana pastovaus dydžio, tai yra kinta labai nežymiai, nepriklausomai nuo prognozuojamų reikšmių skaičiaus.

Bendrosios tyrimo dalies išvados:

- visais eksperimento atvejais nustatyta viena arba dvi lūžio linijos, žyminčios RMSE paklaidos staigų sumažėjimą arba padidėjimą;
- eksperimento metu nustatyta, kad rangas kinta priklausomai nuo prognozuojamo eilutės segmento;
- visais eksperimento atvejais pavyko nustatyti optimalų rangą, su kuriuo RMSE paklaida neviršija 0.03, todėl modelis yra pakankamai adekvatus ir gali būti taikomas testinei eilutei;
- atsižvelgiant į tai, kad visais atvejais galima pasiekti ganėtinai mažą RMSE paklaidos lygį, nutarta, kad pakanka tirti rangus iki dvigubai didesnės reikšmės nei prognozuojamų narių kiekis;
- visais eksperimento atvejais, ties visais rangais, RMSE paklaidos reikšmė išlieka gana pastovaus dydžio ir kinta labai nežymiai, tai yra praktiškai nepriklausomai nuo prognozuojamų reikšmių skaičiaus. Tai reiškia, kad rango parinkimui pakanka ištirti tik rangų priklausomybę nuo pirmosios ateities reikšmės prognozės.

4.2. Genetinio algoritmo parametrų parinkimas

Algebrinio skeleto modelis naudoja genetinį optimizavimo algoritmą, kuris yra skirtas sugeneruoti optimalią triukšmo seką. Kaip ir minėta 3.1.2 skyrelyje triukšmo parinkimo uždavinys turi be galo daug sprendinių, todėl optimizavimo algoritmas yra būtinas norint išvengti pilno perrinkimo metodo kuris būtų neefektyvus ir modelį taikant realioms laiko eilutėms pareikalautų didelių laiko resursų.

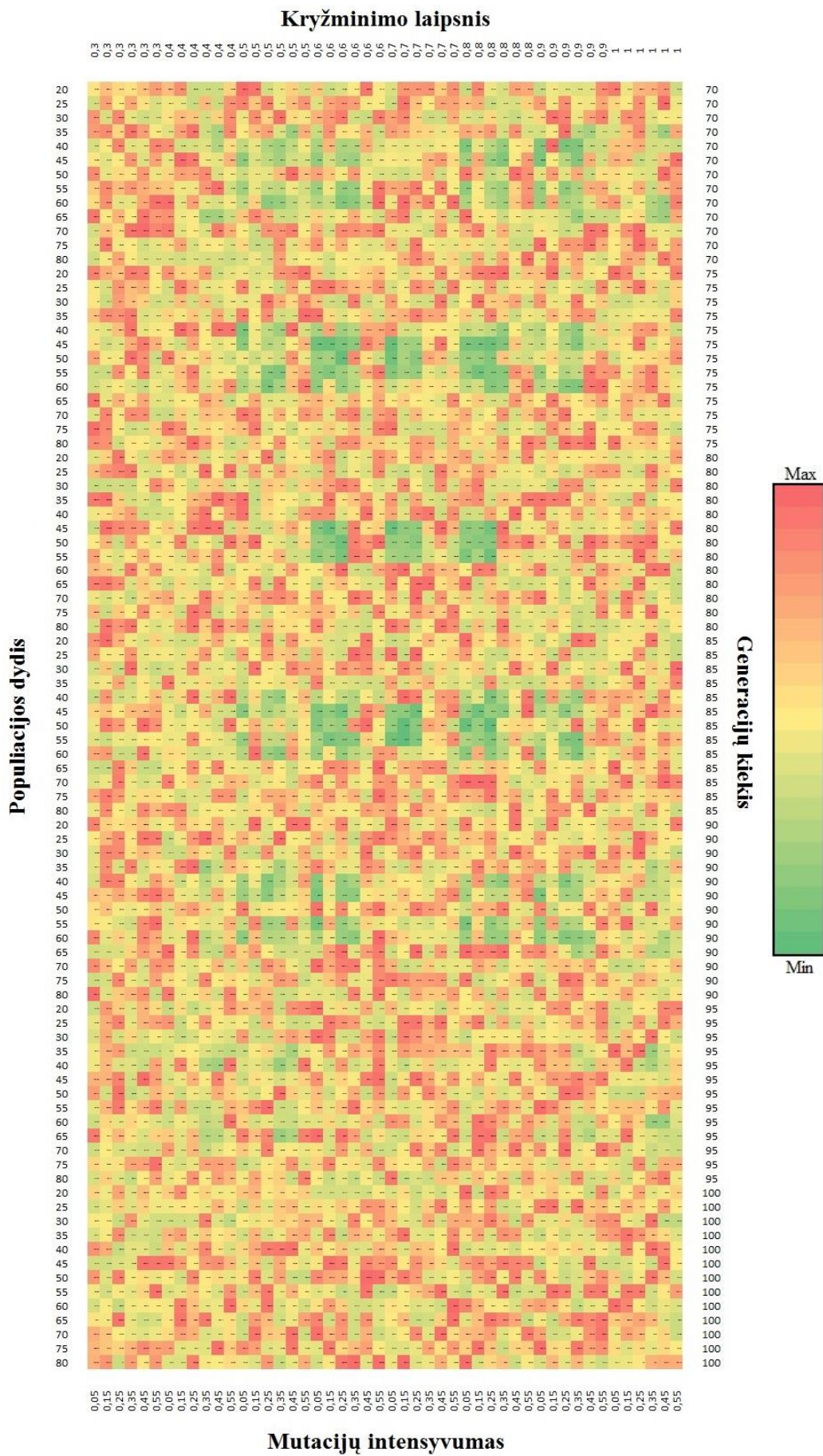
Modelyje naudojamas genetinis algoritmas turi keturis pagrindinius parametrus:

- populiacijos dydį;
- generacijų skaičių;
- kryžminimo laipsnį;
- mutacijos intensyvumą.

Šioje tyrimo dalyje bus analizuojama, kaip modelio rezultatai priklauso nuo genetinio algoritmo parametrų parinkimo. Pirmajame tyrimo žingsnyje buvo atliktas testavimo eilutės – signalo prognozavimas vienam žingsniui į priekį. Modelio rangas kinta nuo 10 iki 20, o genetinio algoritmo parametrai parenkami iš aibių:

- populiacijos dydis: $\{15 + 5 \cdot i \mid i = \overline{1,13}, i \in \mathbb{N}\}$;
- generacijų skaičius: $\{65 + 5 \cdot i \mid i = \overline{1,17}, i \in \mathbb{N}\}$;
- kryžminimo laipsnis: $\{0.2 + 0.1 \cdot i \mid i = \overline{1,8}, i \in \mathbb{N}\}$;
- mutacijos intensyvumas: $\{-0.05 + 0.1 \cdot i \mid i = \overline{1,6}, i \in \mathbb{N}\}$.

Tyrimo rezultatai pateikti keturių ašių RMSE grafiku 4.2.1 paveiksle, kuriame naudojama skalė nuo raudonos, žyminčios blogiausią rezultatą, iki žalios, žyminčios geriausią rezultatą, o geltona žymi vidutinį rezultatą.



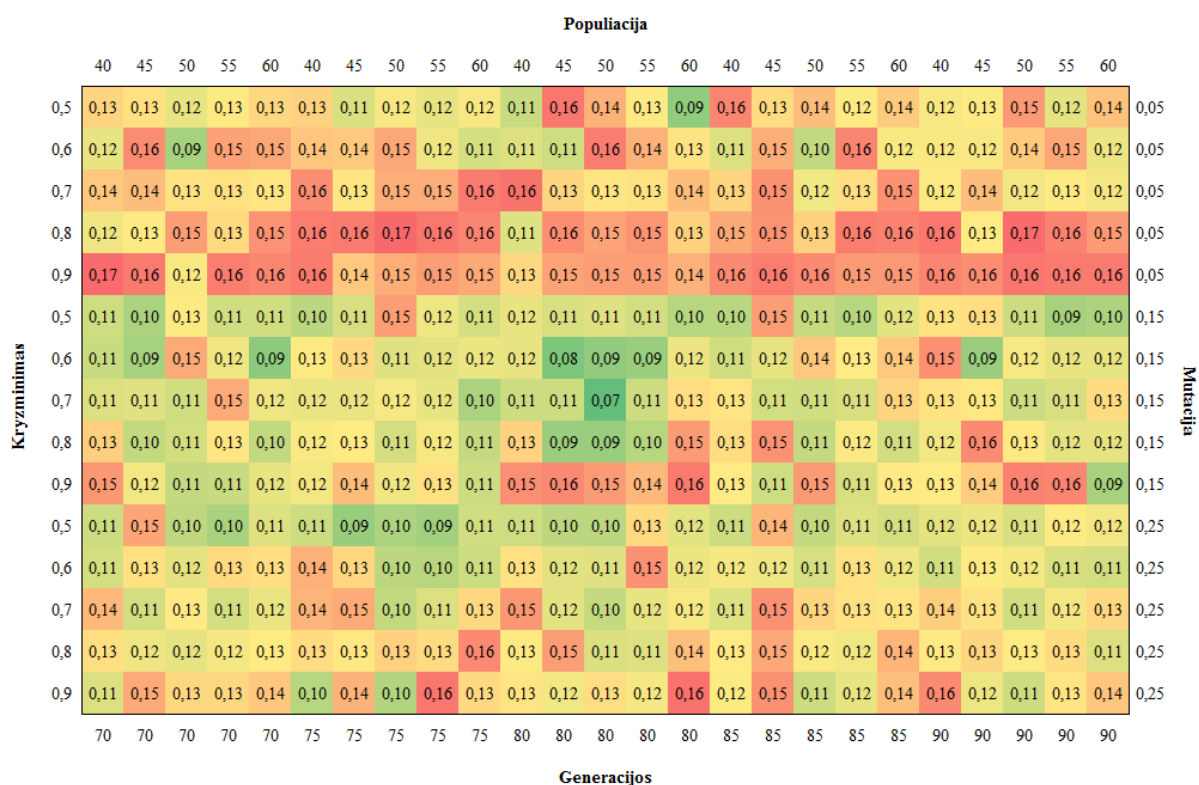
4.2.1 pav. Testinės eilutės prognozės paklaidų RMSE priklausomybė nuo genetinio algoritmo parametru.

Stebint grafiką nesunku pastebėti, kad mažiausios paklaidos gaunamos devyniose grafiko zonose, tai galime vadinti algoritmo minimumais. Šios zonos yra pateiktos 4.2.1 lentelėje.

4.2.1 lentelė. Genetinio algoritmo parametrų mažiausių paklaidų zonos.

Zona	Kryžminimo laipsnis		Generacijų skaičius		Populiacijos dydis		Mutacijų intensyvumas	
	Nuo	Iki	Nuo	Iki	Nuo	Iki	Nuo	Iki
1	0,6	0,6	75	75	45	55	0,05	0,25
2	0,7	0,7						
3	0,8	0,8						
4	0,6	0,6	80	80				
5	0,7	0,7						
6	0,8	0,8						
7	0,6	0,6	85	85				
8	0,7	0,7						
9	0,8	0,8						

Tęsiant tyrimą buvo nuspręsta šias mažiausių paklaidų zonas vienu kiekvienos zonos žingsniu išplėsti į abi puses (išskyrus mutacijų intensyvumą, dėl staigių RMSE šuolių, matomų 4.2.1 paveiksle) ir tik tuomet tirti, prognozuojant dešimt laiko periodų į priekį. Tai yra, tyrime toliau naudoti kryžminimą iš intervalo [0.5 ; 0.9], generacijas iš intervalo [70 ; 90], populiaciją iš intervalo [40 ; 60], o mutaciją iš intervalo [0.05 ; 0.25]. Gautas naujas keturių ašių RMSE grafikas pateiktas 4.2.2 paveiksle.



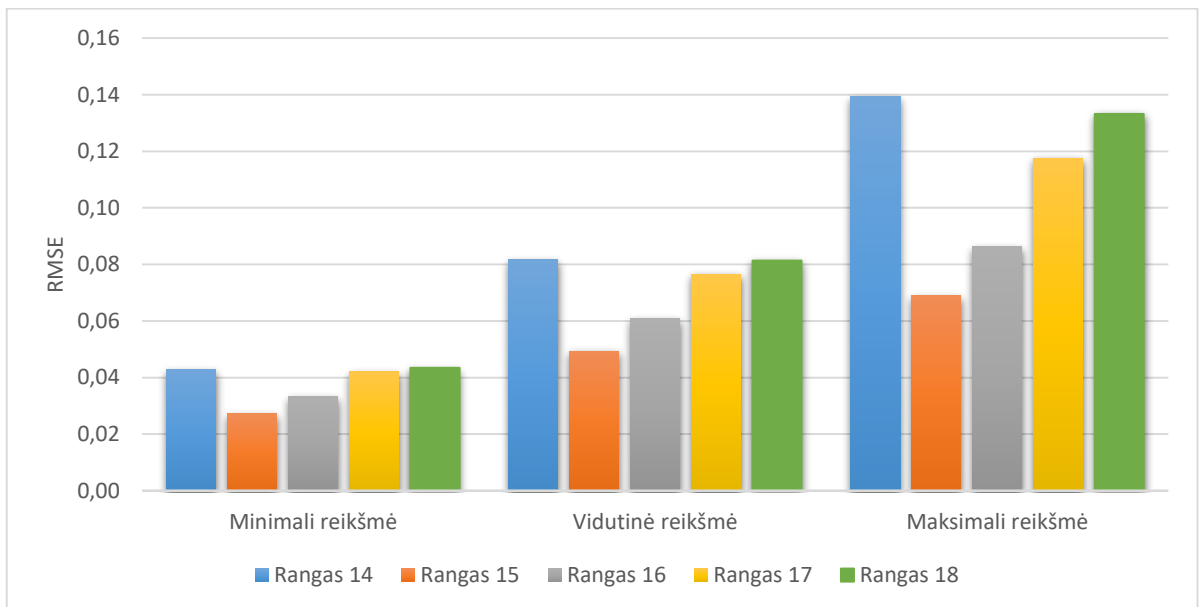
4.2.2 pav. Testinės eilutės prognozės paklaidų RMSE priklausomybė nuo genetinio algoritmo parametrų mažiausių paklaidų zonose.

Stebint rezultatus buvo pastebėta, kad:

- kaskart perrinkimo būdu tirti visus galimus genetinio algoritmo parametrų rinkinius yra neefektyvu, nes sunaudojama labai daug laiko. 4.2.1 paveikslo sudarymas pareikalavo 48048 kartų pakartoti skaičiavimus ir net kai vidutinė vieno skaičiavimo trukmė yra apie 25s, šis tyrimas reikalauja beveik dviejų savaitių nepertraukiamų skaičiavimų (skaičiavimai buvo vykdomi naudojant kompiuterį, kurio techniniai duomenys: procesorius – Intel Core i7-4700MQ 3.40GHz, laikinoji atmintis (RAM) – 32Gb, atmintis – Samsung Pro SSD);
- atsižvelgiant į 4.2.1 paveiksle pastebėtą tendenciją buvo pasirinkta tyrinėti tik šiek tiek praplėstas tam tikras devynias zonas. Tai patvirtina ir 4.2.2 paveikslas kuriame pastebima, kad mažiausios paklaidos išlieka viduriniojoje iš devynių zonų;
- kadangi ryškių tendencijų 4.2.2 paveiksle nepastebima, tolimesnius rango tyrimus buvo nuspręsta vykdyti būtent 4.2.2 paveiksle matomose genetinio algoritmo parametrų ribose, tai yra tyrinėjant tik 375 skirtingus atvejus kiekvienam rangui ir skaičiavimų trukmę sutrumpinant iki vidutiniškai 2 valandų, 36 minučių ir 15 sekundžių;
- optimalus genetinio algoritmo parametrų rinkinys: populiacija – 50, generacijos – 80, kryžminimas – 0.7, mutacija – 0.15.

4.3. Testinės eilutės prognozavimas

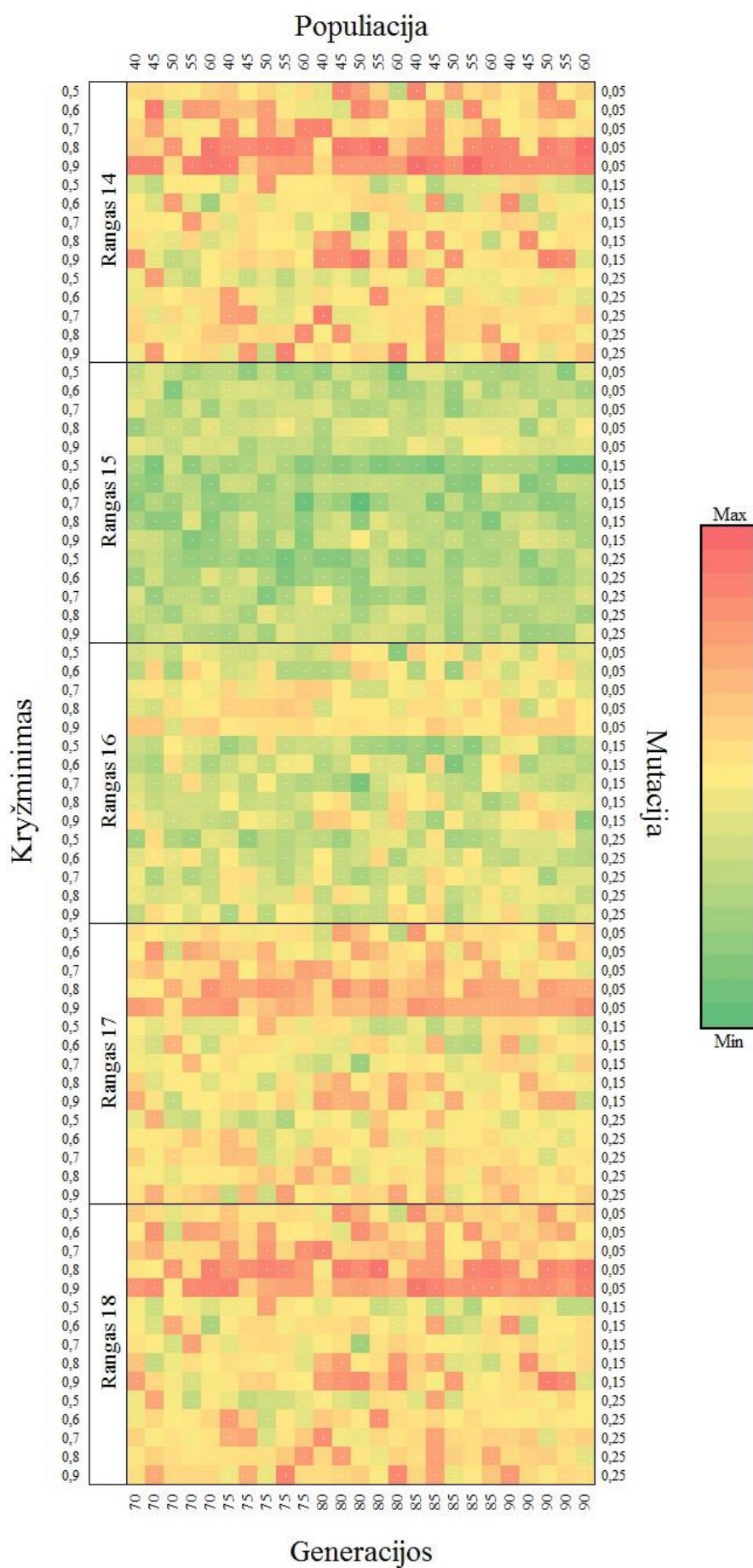
Nustačius optimalų rango parinkimo metodą ir optimizavus genetinio algoritmo parametrų parinkimą, prieš prognozuojant dar reikia įvertinti, ar kintant rangui kinta ir genetinio algoritmo parametrų priklausomybė nuo RMSE paklaidos, tačiau daroma prielaida, kad 4.2 skyrelyje nustatytos optimalios genetinio algoritmo parametrų zonos nekinta. Šioje tyrimo dalyje buvo sudaromi ir analizuojami RMSE ir rango priklausomybių grafikai. Rango parenkamas nuo 10 iki 20. Svarbiausi tyrimo aspektai – maksimali, minimali ir vidutinė RMSE reikšmė kiekviename iš rangų. Tyrimo naudojama testinė eilutė nuo pirmos bendrosios testinės eilutės reikšmės. Atlikus skaičiavimus iškart pastebėta, kad didžiausios paklaidos gaunamos ties rangais 10-13 ir 19-20, todėl šių rangų rezultatai nepateikiami. 4.3.1 paveiksle pateiktos rangų maksimalios, minimalios ir vidutinės RMSE reikšmės.



4.3.1 pav. Minimali, vidutinė ir maksimali RMSE reikšmė, kai rangas 14-18.

Nesunku pastebėti, kad tendencijos išlieka nepriklausomai nuo to, ar vertintume minimalią, vidutinę ar maksimalią RMSE reikšmę kiekviename iš rangų. Galima pastebėti, kad laiko eilutės prognozės paklaidos lieka mažiausios, kai Hankelio matricos rangas lygus 15 arba 16.

4.3.2 paveiksle pateiktas tų pačių 14-18 rangų penkių ašių grafikas, atvaizduojant ir rango, ir 4.2 skyrelyje nustatytų optimalių genetinio algoritmo parametrų kaitą.



4.3.2 pav. Prognozės paklaidų RMSE priklausomybė nuo rango ir genetinio algoritmo parametru.

Stebint rezultatus pastebėta, kad:

- testinės eilutės 10 žingsnių prognozei optimalus rangas yra 15;
- atsižvelgiant į galimybę modifikuoti genetinio algoritmo parametrus ganėtinai artimą rezultatą galima pasiekti ir su 14, 16, 17, 18 rangais;
- stebint 4.3.2 paveikslą pastebėta, kad testinės eilutės geriausiems rangams, 14-18, pastebimas tendencingai blogas 0.05 mutacijos koeficientas;
- testinės eilutės geriausiems rangams optimalus genetinio algoritmo parametrų rinkinys visada yra centrinėje zonoje pagal 4.2 skyrelyje pateiktą grupavimą, tai yra: populiacija $\in \{45, 50, 55\}$, generacijos = 80, kryžminimas $\in \{6, 7, 8\}$, mutacija = 0.15;
- minimali RMSE paklaidos reikšmė 0.02719 pasiekta, kai rangas = 15, populiacija = 50, generacijos = 80, kryžminimas = 0.7, mutacija = 0.15.

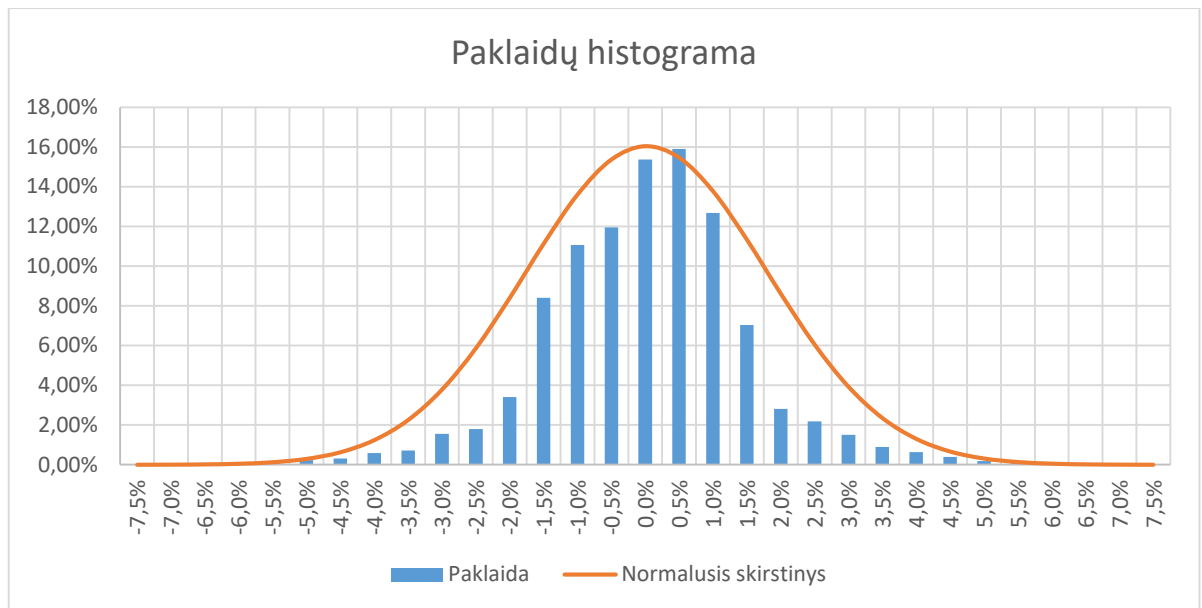
4.4. Testinės eilutės prognozavimo paklaidų analizė

Šioje tyrimo dalyje buvo fiksuojami optimalūs testinei eilutei genetinio algoritmo parametrai (populiacija – 50, generacijos – 80, kryžminimas – 0.7, mutacija – 0.15) ir optimalus rangas 15. Tuomet prognozavimas tam pačiam laiko eilutės segmentui buvo atliekamas 1000 kartų. Pagrindinė naudojama metrika – procentinė paklaida ρ . Ši metrika parinkta, nes ji turi griežtas ribas, nuo –100% iki 100%, ir yra plačiai naudojama paklaidų normalumo tyrimuose.

4.4.1 lentelėje pateikiama procentinių paklaidų sklaida prognozuojant kiekvieną iš dešimties žingsnių. 4.4.1 paveiksle pateikiama bendroji procentinių paklaidų sklaida ir ją atitinkantis normalusis skirstinys.

4.4.1 lentelė. Testinės eilutės procentinių paklaidų pasiskirstymas pagal paklaidos dydį kiekvienai prognozuojamai reikšmei.

Procentinė paklaida	Prognozuojamas žingsnis										Vidurkis
	1	2	3	4	5	6	7	8	9	10	
-7,5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,01%
-7,0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,01%
-6,5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,02%
-6,0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0,07%
-5,5%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0,10%
-5,0%	0%	0%	0%	1%	1%	0%	0%	0%	1%	0%	0,23%
-4,5%	0%	0%	0%	1%	1%	0%	0%	0%	1%	0%	0,31%
-4,0%	0%	1%	1%	1%	2%	1%	0%	0%	1%	0%	0,58%
-3,5%	0%	1%	1%	1%	1%	1%	0%	0%	1%	1%	0,72%
-3,0%	0%	3%	2%	4%	2%	1%	0%	1%	2%	1%	1,56%
-2,5%	0%	1%	1%	7%	2%	1%	1%	1%	2%	1%	1,80%
-2,0%	1%	2%	7%	14%	3%	2%	1%	1%	2%	2%	3,40%
-1,5%	1%	2%	8%	27%	17%	7%	1%	1%	16%	4%	8,40%
-1,0%	1%	8%	9%	22%	33%	9%	1%	2%	18%	7%	11,06%
-0,5%	2%	18%	23%	2%	18%	20%	2%	6%	22%	7%	11,95%
0,0%	3%	27%	21%	8%	10%	20%	8%	22%	18%	17%	15,37%
0,5%	15%	12%	15%	5%	2%	25%	20%	34%	8%	23%	15,90%
1,0%	29%	11%	4%	3%	3%	6%	36%	12%	3%	21%	12,68%
1,5%	19%	5%	3%	1%	2%	3%	20%	12%	2%	4%	7,03%
2,0%	8%	4%	2%	1%	1%	2%	2%	3%	1%	4%	2,81%
2,5%	7%	3%	1%	1%	0%	1%	3%	2%	1%	3%	2,18%
3,0%	6%	1%	1%	1%	0%	1%	2%	1%	1%	2%	1,51%
3,5%	3%	1%	1%	0%	0%	0%	1%	1%	1%	1%	0,89%
4,0%	2%	1%	1%	0%	0%	0%	1%	1%	0%	1%	0,63%
4,5%	1%	0%	0%	0%	0%	0%	1%	1%	0%	1%	0,40%
5,0%	1%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0,18%
5,5%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,11%
6,0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,06%
6,5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,01%
7,0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,01%
7,5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,01%



4.4.1 pav. Testinės eilutės procentinių paklaidų histograma ir normaliojo skirstinio kreivė

Stebint rezultatus pastebėta, kad:

- testinės eilutės prognozių kiekvienam žingsniu procentinės paklaidos yra panašios į normalųjį skirstinį;
- bendrosios testinės eilutės prognozių procentinės paklaidos yra labai panašios į normalųjį skirstinį kurio vidurkis $2.04 \cdot 10^{-4}$, o dispersija $3.16 \cdot 10^{-4}$, todėl buvo atlikti Anderson-Darling ir Shapiro-Wilk normalumo tyrimai, kurie parodė, kad šie duomenys 99% reikšmingumo lygiu atitinka normalųjį skirstinį;
- testinės eilutės procentinės paklaidos telpa intervale $[-7.5\% ; 7.5\%]$.

4.5. Finansinės laiko eilutės paruošimas

Toliau tyrime bus analizuojamos realios laiko eilutės. Jos dažnai būna sezoninės, turi tendą, ar yra labai nestacionarios, todėl pritaikyti mūsų analizuojamą modelį tokiai eilutei prognozuoti yra sudėtinga. Pirmiausia reikia atlikti tam tikrus paruošiamuosius darbus, tai yra, koku nors būdu eilutę apdoroti ar modifikuoti prieš ją pateikiant modeliui. Šiam tikslui darbe buvo pasirinktos šios strategijos:

- eilutė nekeičiama;
- eilutė standartizuojama. Standartizavimui naudojama formulė: $x_i^* = \frac{x_i - \mu}{\sigma}$;
- eilutė normalizuojama. Normalizavimui naudojama formulė: $x_i^* = (x_i - \mu)$;
- eilutė normuojama. Normavimui naudojama formulė: $x_i^* = \frac{x_i - x_{min}}{x_{max} - x_{min}}$;
- eilutė suspaudžiama pagal maksimalią reikšmę. Suspaudimui pagal maksimalią reikšmę naudojama formulė: $x_i^* = \frac{x_i}{x_{max}}$;

čia μ – vidurkis, σ – standartinis nuokrypis, x_i – tikroji reikšmė, x_i^* – nauja reikšmė, x_{min} – minimali x reikšmė, x_{max} – maksimali x reikšmė.

Tyrimui buvo naudojama finansinė eilutė USD/RUB kursas. Tyrimo metu buvo prognozuojama dešimt reikšmių su skirtingais rangais nuo 10 iki 20. Genetinio algoritmo parametrai šio tyrimo metu buvo fiksuoti, darant prielaidą, kad genetinio algoritmo parametru parinkimas gerokai mažiau reikšmingas nei rangas. Parinkti genetinio algoritmo parametrai: populiacijos dydis – 50, generacijų skaičius – 80, kryžminimo laipsnis – 0,7, mutacijos intensyvumas – 0,15.

Tyrimo metu buvo naudojama vidutinė absoliutinė procentinė paklaida, randama pagal formulę $\rho_A = \frac{1}{1100} \sum_{i=1}^{1100} \left| \frac{x_i - \bar{x}}{x_i} \right|$, čia \bar{x} žymi x įvertį, gautą prognozuojant. Procentinė paklaida parinkta neatsitiktinai. Priežastis ta, kad dauguma kitų paklaidų įverčių gali tapti nepalyginami tarp pavyzdžiui normuotos ir nmodifikuotos eilutės. Tyrimo metu buvo atliekama 100 prognozavimų kiekvienai iš minėtų modifikuotų eilučių. Rezultatai pateikti 4.5.1 lentelėje.

4.5.1 lentelė. Eilutės modifikavimo tyrimo rezultatai.

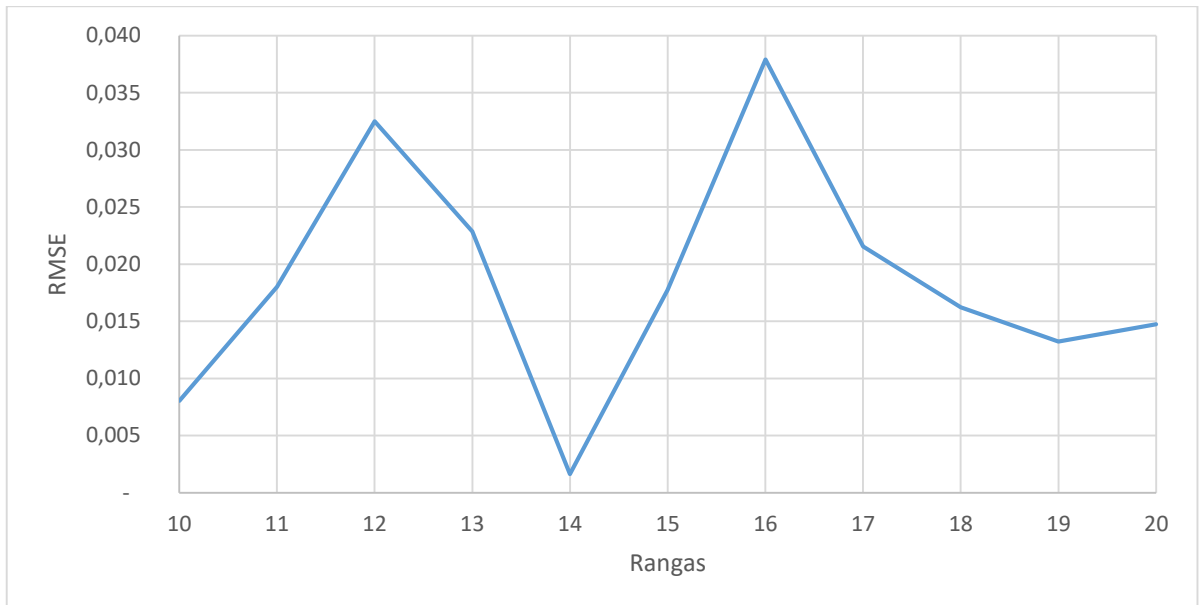
Eilutė	Vidutinė ρ_A	Minimali ρ_A	Maksimali ρ_A
Nekeista eilutė	10,02%	2,23%	38,77%
Standartizuota eilutė	15,87%	8,74%	60,99%
Normalizuota eilutė	6,96%	3,45%	17,23%
Normuota eilutė	4,27%	1,05%	10,28%
Suspausta pagal maksimalią reikšmę eilutė	3,77%	0,15%	14,91%

Stebint rezultatus buvo pastebėta, kad:

- normavimas ir suspaudimas pagal maksimalią reikšmę yra geriausi metodai laiko eilutės paruošimui;
- atsižvelgus į vidutinę procentinę paklaidą buvo pastebėta, kad normuojant ar spaudžiant eilutę pagal maksimalią reikšmę paklaidos pasiskirstę pagal normalųjį skirstinį, tai yra atlikti Anderson-Darling ir Shapiro-Wilk normalumo tyrimai parodė, kad šie duomenys 99% reikšmingumo lygiu atitinka normalųjį skirstinį;
- normavimo ir spaudimo pagal maksimalią reikšmę atvejais paklaidų skirstinių dispersijos beveik lygios, o vidurkiai skiriasi per 0.5%;
- spaudimo pagal maksimalią reikšmę rezultatas dažnu atveju sumažina eilutės svyravimų intervalą gerokai labiau nei normavimas, dėl to galima įvesti tam tikrus ribojimus prognozavimo modeliui, kurių pagalba galima gerokai sumažinti paklaidų ribas;
- pasirinktas eilutės modifikavimas – spaudimas pagal maksimalią reikšmę.

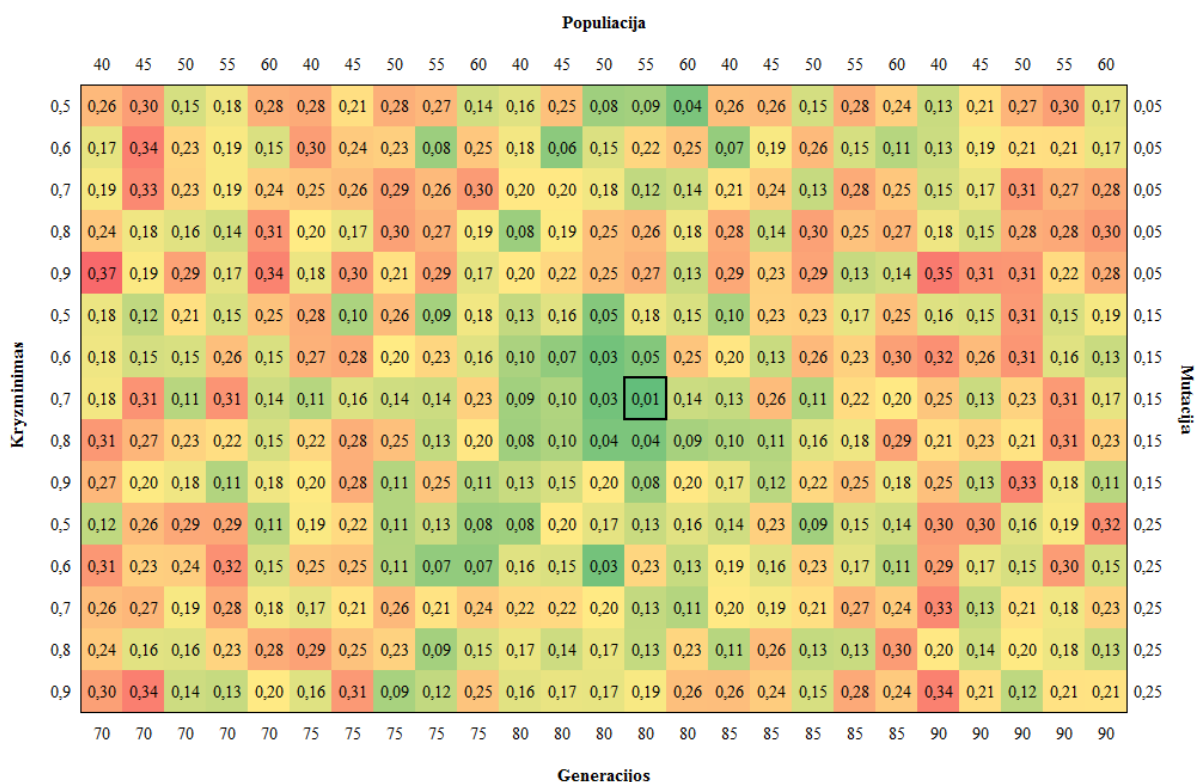
4.6. Finansinės laiko eilutės prognozavimas

Šioje tyrimo dalyje prognozuojama finansinė laiko eilutė USD/RUB kursas. Atsižvelgiant į ankstesniuose skyreliuose gautus rezultatus pradžia buvo nustatytas eilutės rangas, bandant prognozuoti žinomą eilutės segmentą vienu žingsniu į priekį. Rezultatai pateikti 4.6.1 paveiksle.



4.6.1 pav. Finansinės eilutės, RMSE ir rango priklausomybės grafikas vieno žingsnio prognozei.

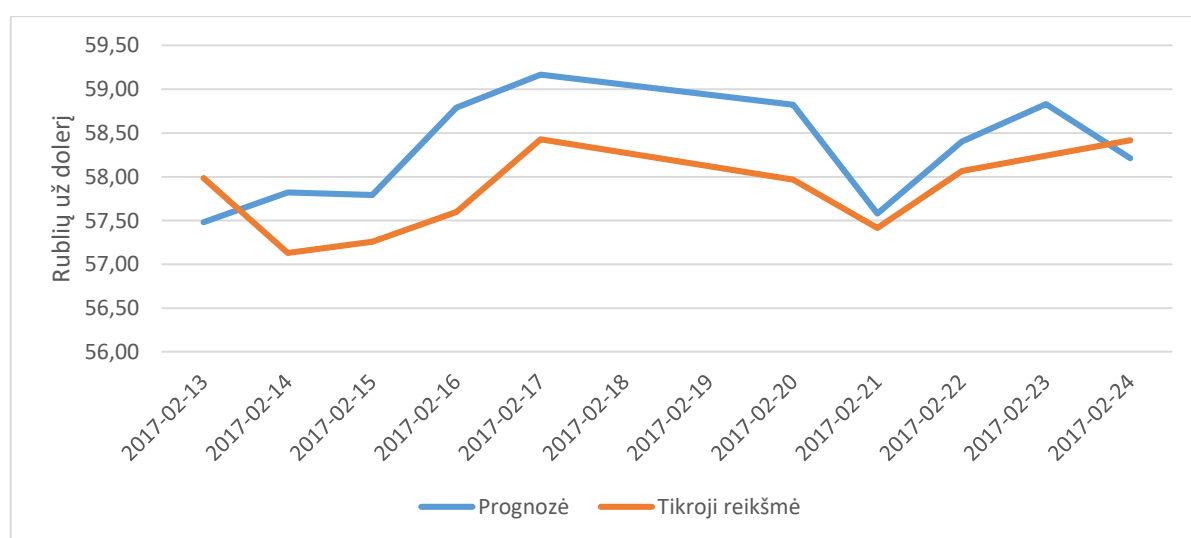
Atsižvelgiant į rezultatus parinktas rangas 14. Buvo bandoma surasti optimalų genetinio algoritmo parametrų rinkinį pasinaudojant 4.2 skyrelyje nagrinėtomis devyniomis išplėstinių parametrų rinkinių grupėmis. Rezultatai pateikti 4.6.2 paveiksle.



4.6.2 pav. Finansinės eilutės prognozės paklaidų RMSE priklausomybė nuo genetinio algoritmo parametrų mažiausių paklaidų zonose.

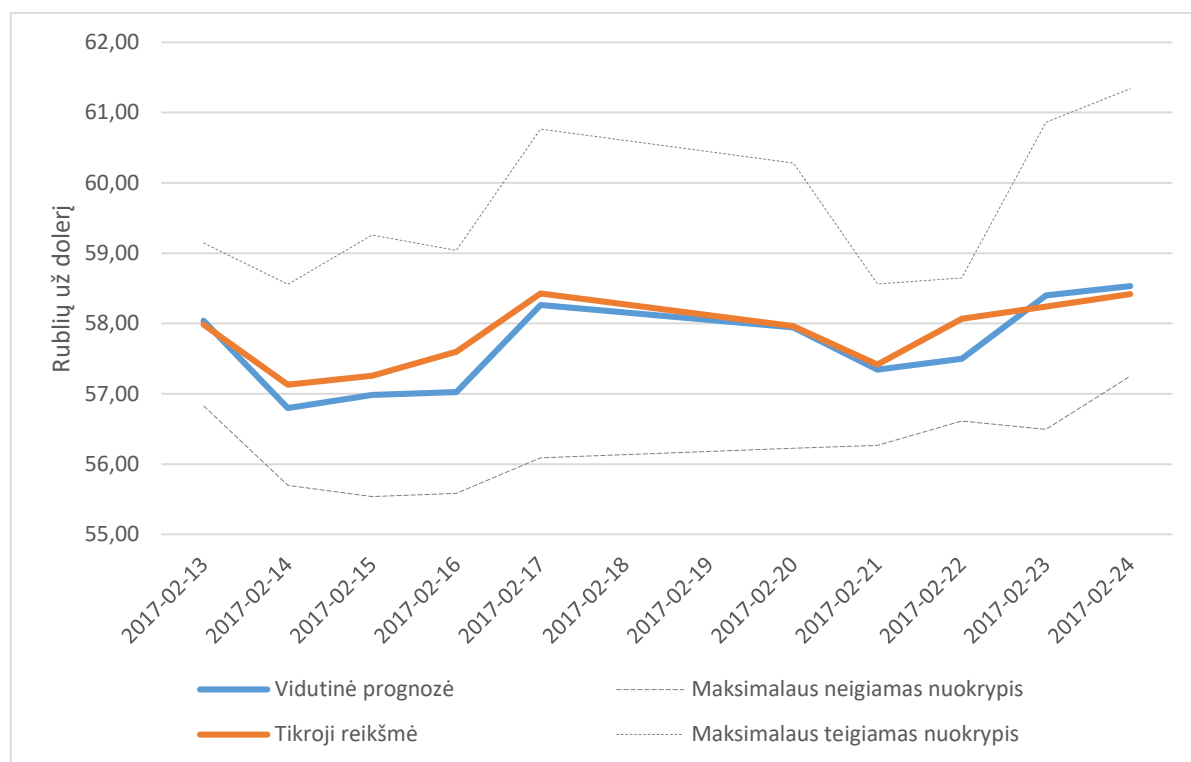
Kaip pastebime iš 4.6.2 paveikslo šiai eilutei geriausi genetinio algoritmo parametrų rinkiniai išsidėstę centrinėje zonoje. Geriausias rinkinys yra populiacija – 55, generacijos – 80, kryžminimas – 0.7, mutacija – 0.15.

Parinkus optimalų rangą ir optimalius genetinio algoritmo parametrus buvo atlikta eilutės prognozė. Ji pateikta 4.6.3 paveiksle.



4.6.3 pav. Finansinės eilutės, USD/RUB kursas, prognozės ir tikrosios reikšmės grafikas.

Kadangi 4.6.3 paveiksle pateikta atsitiktinė prognozė, toliau tyrime to paties segmento prognozavimas buvo atliekamas 1000 kartų, norint nustatyti maksimalius nuokrypius kiekvienam iš prognozuojamų žingsnių. Maksimalaus teigiamo nuokrypio kreivė gaunama, parinkus labiausiai į teigiamą pusę nukrypusias prognozes. Analogišku būdu buvo gauta maksimalaus neigiamo nuokrypio kreivė. Vidutinė prognozės kreivė gauta, apskaičiuojant vidurkį kiekvieno žingsnio prognozėms. Rezultatai pateikti 4.6.4 paveiksle.



4.6.4 pav. Finansinės eilutės, USD/RUB kursas, prognozavimo nuokrypių grafikas.

Stebint rezultatus buvo pastebėta, kad:

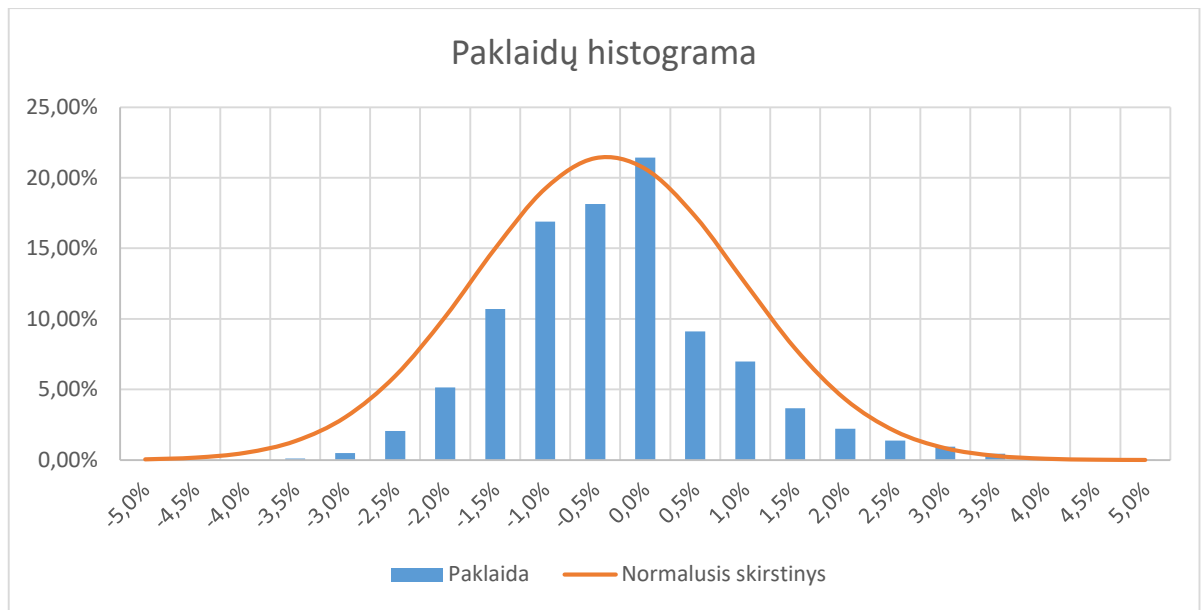
- modelio prognozė kiekvieną kartą yra atsitiktinė;
- vidutinė prognozė yra pakankamai tiksli prognozuoti finansinei eilutei, USD/RUB kursui, o jos nuokrypis nuo tikrosios reikšmės yra intervale $[-0.57 ; 0.16]$;
- nuokrypiai kiekvienam prognozuojamam žingsniui yra individualūs ir svyruoja intervale $[-2.80 ; 2.17]$.

4.7. Finansinės laiko eilutės prognozavimo paklaidų analizė

Šioje tyrimo dalyje buvo nagrinėjamos 4.6 skyrelyje aprašyto prognozių eksperimento, kai kiekvienas iš dešimties žingsnių buvo prognozuojamas 1000 kartų, paklaidos. Pagrindinis rodiklis yra procentinė paklaida ρ . 4.7.1 lentelėje pateikiama procentinių paklaidų ρ sklaida, prognozuojant kiekvieną iš dešimties žingsnių. 4.7.1 paveiksle pateikiama bendroji procentinių paklaidų sklaida ir ją atitinkantis normalusis skirstinys.

4.7.1 lentelė. Finansinės eilutės, USD/RUB kursas, procentinių paklaidų sklaida kiekvienam iš prognozuojamų žingsnių.

Procentinė paklaida	Prognozuojamas žingsnis										Vidurkis
	1	2	3	4	5	6	7	8	9	10	
-5,0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,00%
-4,5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,00%
-4,0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,01%
-3,5%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0,11%
-3,0%	0%	0%	0%	3%	1%	0%	0%	0%	0%	0%	0,50%
-2,5%	0%	0%	1%	9%	5%	2%	0%	1%	3%	0%	2,06%
-2,0%	1%	4%	3%	14%	8%	4%	2%	10%	5%	2%	5,15%
-1,5%	4%	12%	10%	15%	10%	8%	8%	27%	8%	5%	10,70%
-1,0%	14%	27%	20%	16%	12%	13%	17%	30%	10%	9%	16,89%
-0,5%	18%	26%	28%	16%	13%	15%	20%	18%	12%	17%	18,15%
0,0%	14%	23%	30%	19%	21%	26%	17%	7%	22%	34%	21,42%
0,5%	25%	3%	4%	4%	8%	8%	18%	5%	8%	8%	9,11%
1,0%	18%	3%	2%	2%	7%	7%	14%	2%	8%	8%	6,99%
1,5%	5%	1%	1%	1%	5%	6%	4%	0%	7%	6%	3,67%
2,0%	1%	1%	1%	1%	4%	5%	1%	0%	5%	5%	2,21%
2,5%	0%	0%	1%	0%	2%	3%	0%	0%	5%	3%	1,39%
3,0%	0%	0%	0%	0%	2%	2%	0%	0%	4%	2%	0,96%
3,5%	0%	0%	0%	0%	0%	1%	0%	0%	2%	1%	0,45%
4,0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0,18%
4,5%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0,05%
5,0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,01%



4.7.1 pav. Finansinės eilutės, USD/RUB kursas, procentinių paklaidų histograma ir normaliojo skirstinio kreivė.

Stebint rezultatus pastebėta, kad:

- finansinės eilutės, USD/RUB kursas, prognozių kiekvienam žingsniui procentinės paklaidos yra panašios į normalųjį skirstinį;
- bendrosios finansinės eilutės, USD/RUB kursas, prognozių procentinės paklaidos yra labai panašios į normalųjį skirstinį, kurio vidurkis $-2.90 \cdot 10^{-3}$, o dispersija $1.76 \cdot 10^{-4}$, todėl buvo atlikti Anderson-Darling ir Shapiro-Wilk normalumo tyrimai, kuris parodė, kad šie duomenys 99% reikšmingumo lygiu atitinka normalųjį skirstinį;
- finansinės eilutės, USD/RUB kursas, procentinės paklaidos telpa intervale $[-4\% ; 5\%]$.

4.8. Modelio taikymas su kitais prognozavimo modeliais

Finansinėms eilutėms labai dažnai naudojami ARIMAX modeliai, kurie prognozuoja atsižvelgdami į išorinį regresorių. Šioje tyrimo dalyje buvo nagrinėta, ar pritaikius nagrinėjamą modelį išorinių regresorių prognozei galima pagerinti ARIMAX prognozių rezultatus. Tyrime naudojami išoriniai regresoriai pateikti 4.8.1 lentelėje, jų dinamika pateikta 7.1 priede. Akcijų rinkų indeksų rodiklių sudėtis, analizuojamu laikotarpiu, pateikta 7.2 priede.

4.8.1 lentelė. Išoriniai regresoriai ir jų žymėjimas.

Žymėjimas	Reikšmė
Brent	Brent naftos kaina (USD už barelį)
DJIA	Jungtinių Amerikos valstijų akcijų biržos bendrasis rodiklis (Dow Jones Industrial Average)
Gold	Aukso kaina (USD už trojos unciją)
MICEX	Rusijos Federacijos akcijų biržos bendrasis rodiklis
RUS_CPI	Rusijos Federacijos vartotojų kainų indeksas
RUS_EX	Rusijos Federacijos eksporto rodiklis
RUS_IM	Rusijos Federacijos importo rodiklis
RUS_INF	Rusijos Federacijos infliacijos rodiklis
RUS_UE	Rusijos Federacijos nedarbo lygio rodiklis
USA_CPI	Jungtinių Amerikos valstijų vartotojų kainų indeksas
USA_EX	Jungtinių Amerikos valstijų eksporto rodiklis
USA_IM	Jungtinių Amerikos valstijų importo rodiklis
USA_INF	Jungtinių Amerikos valstijų infliacijos rodiklis
USA_UE	Jungtinių Amerikos valstijų nedarbo lygio rodiklis
USA_UE / RUS_UE	Nedarbo santykis
DJIA / MICEX	Biržų santykis
USA_INF / RUS_INF	Infliacijos santykis
USA_CPI/RUS_CPI	CPI santykis (vartotojų kainų indekso santykis)
RUS_IM / RUS_EX	Importo eksporto santykis Rusijos Federacijoje
USA_IM / USA_EX	Importo eksporto santykis JAV
USA_IM / RUS_IM	Importo santykis
USA_EX / RUS_EX	Eksporto santykis

Pradžioje nagrinėsime išorinių regresorių koreliaciją su USD/RUB kursu. Koreliacijai naudojami Pearsono ir Spearmano koreliacijos koeficientai aprašyti 3.4 skyrelyje.

Išorinių regresorių koreliacijos koeficientai pateikti 4.8.2 lentelėje. Priklausomai nuo duomenų normalumo buvo parinktas koreliacijos koeficientas. Normalumui nustatyti buvo naudojami Anderson-Darling ir Shapiro-Wilk normalumo testai. Jei duomenys netenkino normalumo sąlygos, tuomet Pearsono koreliacijos koeficientas jiems nebuvo taikomas, priešingu atveju, buvo vertinami abu koreliacijos koeficientai.

4.8.2 lentelė. Pearsono ir Spremano koreliacijos koeficientai išoriniams regresoriams.

Regresoriaus tipas	Išorinis regresorius	r_p	r_s
Tiesioginis	Gold	-0,8319	-0,8533
Tiesioginis	Brent	-0,4265	-0,3774
Išvestinis	USA_UE / RUS_UE	-	0,8131
Tiesioginis	RUS_UE	-	-0,8031
Tiesioginis	USA_UE	-	0,1338
Išvestinis	DJIA / MICEX	-0,8045	-0,8070
Tiesioginis	MICEX	0,7101	0,7123
Tiesioginis	DJIA	-0,7974	-0,7999
Išvestinis	USA_INF / RUS_INF	-	-0,8131
Tiesioginis	RUS_INF	-	0,7631
Tiesioginis	USA_INF	-	-0,7831
Išvestinis	USA_CPI/RUS_CPI	-	0,8133
Tiesioginis	RUS_CPI	-	-0,8108
Tiesioginis	USA_CPI	-	-0,8008
Išvestinis	USA_IM / RUS_IM	-	0,2716
Išvestinis	USA_EX / RUS_EX	-	0,4991
Išvestinis	RUS_IM / RUS_EX	-	0,6625
Išvestinis	USA_IM / USA_EX	-	0,8131
Tiesioginis	RUS_IM	-	0,2208
Tiesioginis	USA_IM	-	0,7318
Tiesioginis	RUS_EX	-	-0,4417
Tiesioginis	USA_EX	-	0,4573

Kadangi nagrinėjamas dydis yra valiutos kursas, kuris yra santykinis dydis, nenuostabu, kad geriausi išoriniai regresoriai taip pat yra santykiniai dydžiai. Atsižvelgiant į koreliacijos koeficiento reikšmes nutarta tyrinėti aštuonis išorinius regresorius, kurie 4.8.2 lentelėje pateikti paryškintu šriftu.

Toliau tyrime prognozuosime išorinį regresorių su analizuojamu modeliu ir su ARIMA modeliu. ARIMA modelio parametrai parinkti pasinaudojant Akaike ir Bajeso informaciniais kriterijais, kurie aprašyti 3.3 skyrelyje.

Modeliuose naudoti parametrai pateikti 4.8.3 lentelėje. Lentelėje pateikti žymėjimai p – žymi autoregresijos eilę, q – slenkančio vidurkio eilę, d – integravimo eilę, R – rangą, P – populiacijos dydį, G – generacijų skaičių, K – kryžminimo koeficientą, o M – mutacijos koeficientą. Prognozių grafikai pateikti 7.4 priede. ARIMA modelių adekvatumo tyrimai pateikti 7.5 priede.

4.8.3 lentelė. Išorinių regresorių prognozavimui naudotų modelių parametrai.

Išorinis regresorius	ARIMA parametrai			Analizuojamo modelio parametrai				
	p	d	q	R	P	G	K	M
Gold	1	1	1	16	40	80	0,6	0,15
Brent	0	0	3	16	45	85	0,8	0,15
USA_UE / RUS_UE	0	2	0	13	50	85	0,7	0,15
DJIA / MICEX	2	0	0	15	60	80	0,5	0,15
USA_INF / RUS_INF	1	1	0	14	55	85	0,9	0,15
USA_CPI / RUS_CPI	0	2	1	13	40	85	0,5	0,15
RUS_IM / RUS_EX	1	1	1	12	45	85	0,6	0,15
USA_IM / USA_EX	0	2	0	14	55	85	0,9	0,15

Tuomet vertinamos prognozių RMSE ir vidutinės procentinės paklaidos. Procentinė paklaida randama iš formulės: $\rho = \frac{1}{10} \sum_{i=1}^{10} \frac{x_i - \bar{x}_i}{x_i}$. Rezultatai pateikti 4.8.4 lentelėje.

4.8.4 lentelė. Išorinių regresorių prognozių RMSE ir procentinės paklaidos.

Išorinis regresorius	RMSE		$\rho\%$	
	Analizuojamas modelis	ARIMA modelis	Analizuojamas modelis	ARIMA modelis
Gold	7,937850	8,145815	-0,32%	0,68%
Brent	0,368228	0,734412	0,31%	-0,36%
USA_UE / RUS_UE	0,006226	0,000542	-0,13%	-0,05%
DJIA / MICEX	0,042396	0,673218	0,21%	0,42%
USA_INF / RUS_INF	0,003534	0,001440	0,09%	-0,19%
USA_CPI / RUS_CPI	0,002836	0,000000	0,09%	0,00%
RUS_IM / RUS_EX	0,002986	0,006095	-0,22%	0,96%
USA_IM / USA_EX	0,006580	0,000625	0,05%	-0,04%

Dabar įtrauksime prognozuotus išorinius regresorius į ARIMAX modelį ir tyrinėsime, ar analizuojamas modelis pagerina ARIMAX modelio rezultatus. ARIMAX modelio pagrindinio regresoriaus, USD/RUB kursas, modelio parametrai $(p, d, q) = (1,1,1)$. Prognozėms įvertintos RMSE ir vidutinės procentinės paklaidos. Rezultatai pateikti 4.8.5 lentelėje.

4.8.5 lentelė. ARIMAX prognozių RMSE ir procentinės paklaidos.

	Išorinis regresorius prognozuotas su			
	Analizuojamas modelis	ARIMA	Analizuojamas modelis	ARIMA
Išorinis regresorius	RMSE	RMSE	ρ	ρ
Gold	0,389248	0,512105	-0,01%	-0,02%
Brent	0,612673	0,619589	-0,70%	-0,74%
USA_UE / RUS_UE	0,901256	1,367051	0,85%	1,64%
DJIA / MICEX	0,615983	0,649093	0,36%	-0,57%
USA_INF / RUS_INF	0,651553	0,741969	0,45%	0,55%
USA_CPI / RUS_CPI	24,778984	1,557221	-4,27%	1,95%
RUS_IM / RUS_EX	0,736544	0,890046	0,77%	0,90%
USA_IM / USA_EX	1,528585	1,466542	1,69%	1,80%

Stebint rezultatus pastebėta, kad:

- su analizuojamu modeliu prognozuojant išorinius regresorius akivaizdžiai pastebima, kad visais atvejais geriausias genetinio algoritmo parametrų rinkinys išlieka centrinėje zonoje arba labai arti jos, bei pastebima, kad visais atvejais geriausias mutacijos koeficientas išlieka stabilus ir lygus 0.15;
- lyginant ARIMA modelio ir analizuojamo modelio prognozių rezultatus išoriniams regresoriams galima pastebėti, kad keturiais iš aštuonių atvejų yra sumažinama RMSE paklaida ir penkiais iš aštuonių atvejų sumažinama procentinė paklaida;
- pastebimas ir analizuojamo modelio pranašumas prognozuojant dažnai kintančias, svyruojančias eilutes, tokias kaip akcijos indeksas, akcija ar signalas, tačiau tai tampa trūkumu, kai prognozuojama gana stabili, tiesiškumu pasižyminti eilutė;
- nepaisant to, kad analizuojamas modelis prognozavo panašiai kaip ARIMA modelis, panaudojus analizuojamo modelio prognozuotus išorinius regresorius ARIMAX modeliui, rezultatai akivaizdžiai geresni. Šešiais iš aštuonių atvejų sumažinta RMSE paklaida, septyniais iš aštuonių atvejų sumažinta procentinė paklaida. Taip greičiausiai yra todėl, kad dauguma išorinių regresorių yra gana tiesiški, ir tai sąlygoja ARIMAX modelio prognozes būti tiesiškoms. Tuo tarpu analizuojamo modelio prognozės nebuvo tiesiškos, kaip ir nebuvo tiesiška pagrindinio regresoriaus eilutė. Tikėtina, kad įvesti papildomi išorinio regresoriaus svyravimai teigiamai įtakojo pagrindinio regresoriaus prognozę.

5. Išvados ir rekomendacijos

Tyrimo metu gautos išvados:

- ištyrus algebrinio skeleto modelį įvairiais pjūviais nustatyta, kad jis yra adekvatus ir gali būti taikomas laiko eilutėms prognozuoti ilgesniam nei vieno žingsnio periodui, nes nepriklausomai nuo parinkto laiko eilutės segmento, visada galima nustatyti optimalų rangą, tai yra rangas visada egzistuoja, net ir prognozuojant daugiau nei vieną žingsnį į ateitį, be to ištyrus modelio paklaidas, nustatyta, kad modelio paklaidos pasiskirsčiusios pagal normaliojo skirstinio dėsnį, kurio vidurkis apytiksliai lygus nuliui;
- modelio rango parinkimui pakanka ištirti tik rangų priklausomybę nuo pirmosios ateities reikšmės prognozės, nes ties visais rangais, RMSE paklaidos reikšmė išlieka gana stabilus dydžio ir kinta labai nežymiai, tai yra praktiškai nepriklausomai nuo prognozuojamų reikšmių skaičiaus;
- genetinis algoritmas naudojamas triukšmo generavimui, pasiekia mažiausias paklaidas specifinėse devyniose parametrų grupėse ir pakanka kas kartą tirti tik šias parametrų grupes, kurios pateiktos 4.2.1 lentelėje;
- nustatytas analizuojamo modelio pranašumas prognozuojant dažnai kintančias, svyruojančias eilutes, tokias kaip akcijos indeksas ar akcijos kaina, todėl modelis gali būti taikomas finansinėms ir ekonominėms eilutėms prognozuoti, tačiau būtinas pradinės duomenų eilutės modifikavimas – suspaudimas pagal maksimalią reikšmę, nes tai reikšmingai sumažina modelio paklaidas;
- modelis gali būti integruojamas į ARIMAX modelį kaip išorinio egzogeninio kintamojo prognozavimo metodas, nes pusės eksperimentų metu buvo sumažintos prognozavimo paklaidos.

Rekomendacijos:

- taikyti šį algebrinių dėstinių modelį laiko eilutėms, kurios pasižymi dažniais dinamikos pokyčiais (ang. fuzzy time series), nes tuomet pastebimas didelis šio modelio pranašumas;
- išbandyti kitus optimizavimo metodus optimalaus triukšmo parinkimui, nes tai galbūt padės dar labiau optimizuoti modelio parametrų parinkimo trukmę;
- pabandyti integruoti šį modelį į neuroninių tinklų modelį, kaip ryšių tarp atskirų neuronų kūrimo procesą.

6. Literatūros sąrašas

1. M. Ragulskis, K. Lukoseviciute, Z. Navickas and R. Palivonaitė. Short-term time series forecasting based on the identification on skeleton algebraic sequences. *Neurocomputing*, 2011, 74:1735–1747.
2. K.Blind, A.Mangelsdorf. Motives to standardize: Empirical evidence from Germany. *Technovation*, 2016. 48–49: 13–24.
3. Y.Dodge. *The Oxford Dictionary of Statistical Terms*, 2003.
4. T.S.Rao, S.S.Rao, C.Rao. *Time Series Analysis: Methods and Applications*, volume 30 of *Handbook of Statistics*. Elsevier, 2012.
5. D.A.Dickey, W.A.Fuller. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 1979. 74 (366): 427–431.
6. D.Kwiatkowski, P.C.B.Phillips, P.Schmidt, Y.Shin. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 1992. 54 (1–3): 159–178.
7. P.C.B.Phillips, P.Perron. Testing for a Unit Root in Time Series Regression. *Biometrika*, 1988. 75 (2): 335–346.
8. W.A.Shewhart, S.S.Wilks. *Analysis of Financial Time Series*, 2005.
9. G.C.Tiao, R.S.Tsay.. Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Annals of Statistics*, 1983, 11: 856–871.
10. R.S.Tsay, G.C.Tiao. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association*, 1984, 79: 84–96.
11. S.Alizadeh, M.Brandt, F.X.Diebold. Range-based estimation of stochastic volatility models. *Journal of Finance*, 2002, 57: 1047–1092.
12. T.Bollerslev. Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach. *Review of Economics and Statistics*, 1990, 72: 498–505.
13. T.Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 1986, 31: 307–327.
14. K.S.Chan, H.Tong. On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 1986, 7: 179–190.
15. B.Cheng, D.M.Titterington. Neural networks: A review from a statistical perspective. *Statistical Science*, 1994, 9: 2–54.

16. J.Hauseman, A.Lo, C.MacKinlay. An ordered probit analysis of transaction stock prices. *Journal of Financial Economics*, 1992, 31: 319–379.
17. T.H.Rydberg, N.Shephard. Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics*, 2003, 1: 2–25.
18. R.F.Engle, J.R.Russell. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 1998, 66:1127–1162.
19. D.Duffie. *Dynamic Asset Pricing Theory*, 2nd edition. Princeton University Press, Princeton, NJ, 1995.
20. F.Black, M.Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 1973, 81: 637–654.
21. R.A.Johnson, D.W.Wichern. *Applied Multivariate Statistical Analysis*, 4th edition. Prentice Hall, Upper Saddle River, NJ, 1998.
22. W.K.Li, A.I.McLeod. Distribution of the residual autocorrelations in multivariate ARMA time series models. *Journal of the Royal Statistical Society*, 1981, Series B 43: 231–239.
23. C.Alexander. *Market Models: A Guide to Financial Data Analysis*. Wiley, Hoboken, NJ, 2001.
24. E.Zivot, J.Wang. *Modeling Financial Time Series with S-Plus*. Springer-Verlag, New York, 2003.
25. L.Bauwens, S.Laurent, J.V.K.Rombouts. *Multivariate GARCH models: a survey*. 2004.
26. R.E.McCulloch, N.Polson, R.S.Tsay. *Multivariate volatility models*. University of Chicago, 2000.
27. N.H.Chan. *Time Series: Applications to Finance*. Wiley, Hoboken, NJ, 2002.
28. R.H.Shumway, D.S.Stoffer. *Time Series Analysis and Its Applications*. Springer-Verlag, New York, 2000.
29. J.C.Artigas, R.S.Tsay. *Effective estimation of stochastic diffusion models with leverage effects and jumps*. University of Chicago, 2004.
30. W.K.Hasting. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 1970, 57: 97–109.
31. M. Kavaliauskas, R.Rudzkis. *Laiko eilučių analizė. Paskaitų konspektas*, 2015.
32. C.Francq, J.M.Zakoian. *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, 2010.
33. T.S.Rao, S.S.Rao, C.Rao. *Time Series Analysis: Methods and Applications*, volume 30 of *Handbook of Statistics*. Elsevier, 2012.
34. T.Bollerslev, R.F.Engle, J.M.Wooldridge. A capital-asset pricing model with time-varying covariances. *Journal of Political Economy*, 1988, 96: 116–131.

35. H.Chen, T.T.Chong, J.Bai. Theory and applications of TAR model with two threshold variables. *Econometric Reviews*, 2012, 31(2):142-170.
36. D.Van Dijk, T.Teräsvirta, P.H.Franses. Smooth Transition Autoregressive Models—A Survey of Recent Developments. *Econometric Reviews*, 2002. 21 (1): 1–47.
37. L.F.Amaral, R.C.Souza, M.Stevenson. A smooth transition periodic autoregressive (STPAR) model for short-term load forecasting. *International Journal of Forecasting*, 2008. 24: 603–615
38. F.M.Tseng, G.H.Tzeng, H.C.Yu, B.J.C.Yuan, Fuzzy ARIMA model for forecasting the foreign exchange market. *Fuzzy Sets and Systems*, 2001. 118 (1): 9–19.
39. C.S.Ong, J.J.Huang, G.H.Tzeng, Model identification of ARIMA family using genetic algorithms. *Applied Mathematics and Computation*, 2005. 164: 885–912.
40. M.Khashei, M.Bijari. An artificial neural network (p,d,q) model for time series forecasting. *Expert Systems with Applications*, 2010, vol. 37, p. 479-489.
41. J.M.P.Menezes, G.P.Barreto. Long-term time series prediction with NARX network: an empirical evaluation. *Neurocomputing*, 2008, vol. 71, p. 3335-3343.
42. A.E.Celik, Y.Karatepe. Evaluating and forecasting banking crises through neural network models: an application for Turkish banking sector. *Expert Systems with Applications*, 2007. 33: 809–815.
43. Y.Bodyanskiy, S.Popov. Neural network approach to forecasting of quasiperiodic financial time series. *European Journal of Operational Research*, 2006. 175: 1357–1366.
44. M.C.Liu, W.Kuo, T.Sastri. An exploratory study of a neural network approach for reliability data analysis. *Quality and Reliability Engineering International*, 1995. 11: 107–112.
45. T.Wang, S.Chien. Forecasting innovation performance via neural networks – a case of Taiwanese manufacturing industry. *Technovation*, 2006. 26: 635–643.
46. B.K.Wong, S.Vincent, L.Jolie. A bibliography of neural network business applications research: 1994–1998. *Operations Research and Computers*, 2000. 27: 1045–1076.
47. R.C.Grinold, R.N.Kahn. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk*, 2nd edition. McGraw Hill, New York, 2000.
48. A.C.Harvey. *Time Series Models*, 2nd edition. Harvester Wheatsheaf, Hemel Hempstead, UK, 1993.
49. Johnson D.S. Local optimization and the traveling salesman problem. In *Proceedings of the 17th International Colloquium on Automata, Languages and Programming*. Lecture Notes in Computer Science, vol.443, Berlin: Springer, 1990, 446-461.

50. M.J.Todd. Linear and quadratic programming in oriented matroids. *Journal of Combinatorial Theory*, 1985. Series B. 39 (2): 105–133.
51. I.M.Stancu-Minasian. A sixth bibliography of fractional programming. *Optimization*, 2006. 55 (4): 405–428.
52. J.E.Beasley. *Advances in Linear and Integer Programming*. Oxford Science, 1996.
53. A.V.Goldberg, É.Tardos, R.E.Tarjan, *Network flow algorithms*. Stanford University CS Dept., 1989.
54. J.B.Orlin. A polynomial time primal network simplex algorithm for minimum cost flows. *Mathematical Programming*, 1997. 78 (2): 109–129.
55. R.E.Bellman. A Markovian Decision Process. *Journal of Mathematics and Mechanics*, 6, 1957.
56. R.E.Bellman. *Dynamic Programming*. Princeton, NJ: Princeton University Press, 2003.
57. E.J.Sondik. The optimal control of partially observable Markov processes. PhD thesis, Stanford University, 1973.
58. T.H.Cormen, C.E.Leiserson, R.L.Rivest, C.Stein. *Introduction to Algorithms* (2nd ed.), MIT Press & McGraw–Hill, 2001. 327–368.
59. N.Stokey, R.E.Lucas, E.Prescott. *Recursive Methods in Economic Dynamics*. Harvard Univ. Press, 1989.
60. L.Ljungqvist, T.Sargent. *Recursive Macroeconomic Theory*. MIT Press, 2000, 2004, 2012.
61. D.Ashlock. *Evolutionary Computation for Modeling and Optimization*, Springer, 2006.
62. R.Poli, W.B.Langdon, N.F.McPhee. *A Field Guide to Genetic Programming*, 2008.
63. C.Ferreira. Designing Neural Networks Using Gene Expression Programming. *Applied Soft Computing Technologies: The Challenge of Complexity*, 2006. 517–536.
64. Z.Xu, X.Ziye, H.Craig, F.Silvia. Spike-based indirect training of a spiking neural network-controlled virtual insect. *Decision and Control*, 2013, IEEE: 6798–6805.
65. A.Colomi, M.Dorigo, V.Maniezzo. Distributed Optimization by Ant Colonies, *actes de la première conférence européenne sur la vie artificielle*, Paris, France, Elsevier Publishing, 1991, 134-142.
66. D.Karaboga. *An Idea Based on Honey Bee Swarm For Numerical Optimization*, 2005.
67. X.S.Yang. *Nature-Inspired Optimization Algorithms*, Elsevier, 2014.
68. K.L.Clarkson. Coresets, sparse greedy approximation, and the Frank-Wolfe algorithm. *ACM Transactions on Algorithms*, 2010. 6 (4): 1.
69. J.A.Snyman. *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*, 2005.

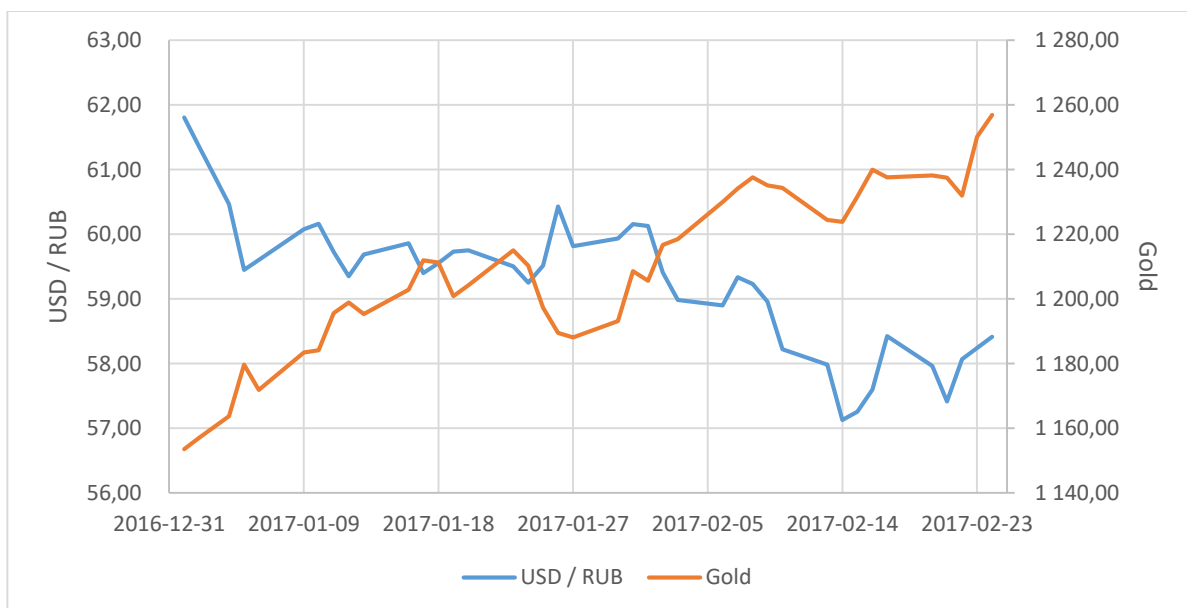
70. E.Polak. Optimization : Algorithms and Consistent Approximations, 1997.
71. G.H.Gene, Y.Qiang. Inexact Preconditioned Conjugate Gradient Method with Inner-Outer Iteration. SIAM Journal on Scientific Computing, 1999. 21 (4): 1305.
72. E.G.Talbi. Metaheuristics – From Design to Implementation, 2009.
73. T.Hendtlass. WoSP: A Multi-Optima Particle Swarm Algorithm. Proceedings IEEE Congress on Evolutionary Computation, 2005, 727–734.
74. F.Glover. Tabu Search - Part 1. ORSA Journal on Computing, 1989. 1 (2): 190–206.
75. F.Glover. Tabu Search - Part 2. ORSA Journal on Computing, 1990. 2 (1): 4–32.
76. S.Kirkpatrick, C.D.Gelatt Jr, M.P.Vecchi. Optimization by Simulated Annealing. Science, 1983. 220 (4598): 671–680.
77. L.A.Rastrigin. The convergence of the random search method in the extremal control of a many parameter system. Automation and Remote Control, 1963. 24 (10): 1337–1342.
78. J.Kennedy, R.C.Eberhart. Swarm Intelligence, 2001.
79. L.N.Castro, F.J.Zuben. Learning and Optimization Using the Clonal Selection Principle. IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems, 2002. IEEE. 6 (3): 239–251.
80. S.Morita, H.Nishimori. Mathematical foundation of quantum annealing, 2008.
81. D.J.Earl, M.W.Deem. Parallel tempering: Theory, applications, and new perspectives, 2005.
82. C.Chun-Hung. Optimal Computing Budget Allocation (OCBA) for Simulation-based Decision Making Under Uncertainty, 2013.
83. N.Hansen. The CMA evolution strategy: a comparing review. Towards a new evolutionary computation. Advances on estimation of distribution algorithms, 2006. 1769–1776
84. C.G.E.Boender, A.H.G.Rinnooy Kan; L.Strougie; G.T.Timmer. A stochastic method for global optimization. Mathematical Programming, 1982. 22: 125–140.
85. J.Mockus. Bayesian Approach to Global Optimization. Kluwer Academic Publishers, Dordrecht, 1989.
86. T.G.Heineman, G.Pollice, S.Selkow. Chapter 8:Network Flow Algorithms. Algorithms in a Nutshell, 2008. 226–250.
87. L.P.Anderson. Business Economics & Finance, 2004. CRC Press.
88. L.P.Anderson. Economics of Business Valuation, 2013. Stanford University Press.
89. M.Miranda, P.Fackler. Applied Computational Economics and Finance, 2002. MIT Press.
90. B.P.Warren. Approximate Dynamic Programming, 2d ed, 2011. Willey.
91. D.Bertsekas. Dynamic Programming and Optimal Control, 2012. Athena Scientific.

92. P.Koehn. Combining Genetic Algorithms and Neural Networks: The Encoding Problem . Master thesis, University of Tennessee, 1994.
93. Y.Kassahun, M.Edgington, J.H.Metzen, G.Sommer, F.Kirchner. Common Genetic Encoding for Both Direct and Indirect Encodings of Networks. Genetic and Evolutionary Computation Conference, 2007, London, UK. 1029-1036.
94. H.J.Trussell, M.R.Civanlar. The Landweber iteration and projection onto convex sets. Signal Process, 1985. 33, 1632–1634.
95. M.Fukushima. A modified Frank-Wolfe algorithm for solving the traffic assignment problem. Transportation Research Part B: Methodological, 1984. 18 (2): 169–153.
96. J.Kennedy, R.Eberhart. Particle Swarm Optimization. Proceedings of IEEE International Conference on Neural Networks IV, 1995. 1942–1948.
97. R.Poli. Analysis of the publications on the applications of particle swarm optimization. Journal of Artificial Evolution and Applications, 2008. 1–10.
98. W.H.Press, S.A.Teukolsky, W.T.Vetterling, B.P.Flannery. Section 10.12. Simulated Annealing Methods. Numerical Recipes: The Art of Scientific Computing (3rd ed.), 2007. New York: Cambridge University Press.
99. V.Vassilev, A.Prahova. The Use of Simulated Annealing in the Control of Flexible Manufacturing Systems. International Journal: Information theories & applications, 6, 1999.
100. C.H.Chen, L.H.Lee. Stochastic simulation optimization an optimal computing budget allocation. Singapore Hackensack, NJ: World Scientific, 2011.
101. C.H.Chen. An Effective Approach to Smartly Allocate Computing Budget for Discrete Event Simulation. Proceedings of the 34th IEEE Conference on Decision and Control, 1995. 2598–2605.
102. L.Trailovic, L.Y.Pao. Computing budget allocation for efficient ranking and selection of variances with application to target tracking algorithms, IEEE Trans Autom Control, 2004. 49:58–67.
103. S.Mukherjee, B.K.Chakrabarti, Multivariable Optimization: Quantum Annealing & Computation. Eur. Phys. J. ST 224 pp 17-24, 2015.
104. K. Lukoševičiūtė. Chaotinių procesų rekonstravimo bei algebrinių sekų modeliai laiko eilučių prognozavime. Daktaro disertacija, Kauno Technologijos Universitetas, 2012.
105. R. Palivonaitė, M. Ragulskis. Skeletinių kreivių panaudojimas su glodinimo procedūra trumpų laiko eilučių prognozei. Lietuvos matematikų draugijos darbai, 2012, ser. B 53t., 90-95.

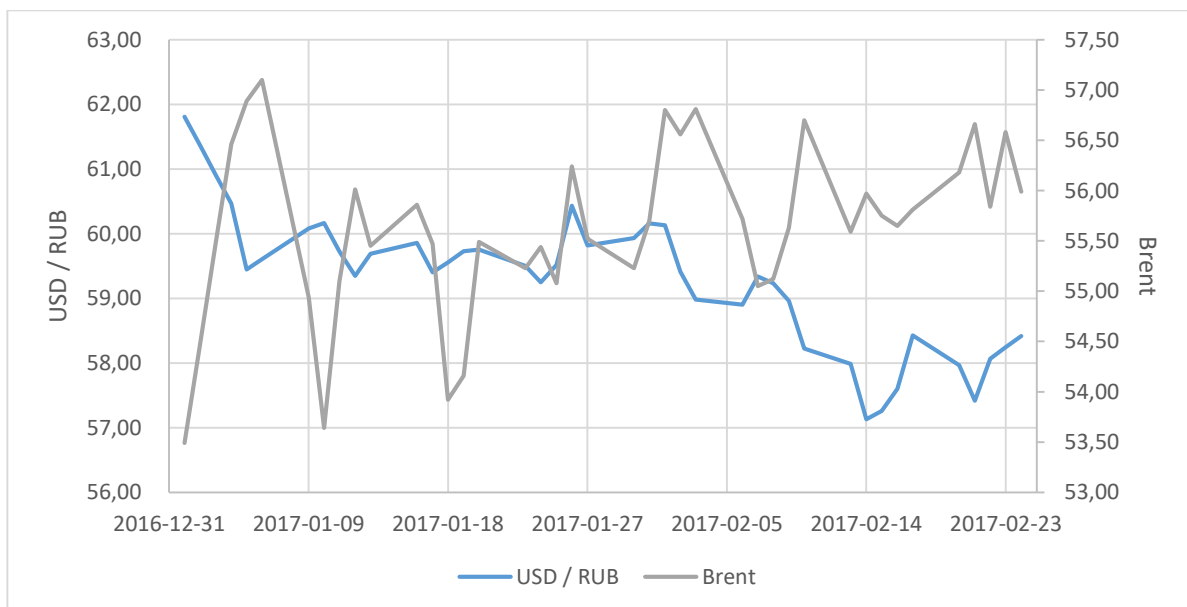
106. Navickas, Z.; Bikulciene, L. Expressions of solutions of ordinary differential equations by standard functions, *Mathematical Modeling and Analysis*, 2006, vol. 11, p. 399–412.
107. A.M.Turing. Computing machinery and intelligence. *Mind*, 1950, LIX (238): 433–460
108. A. Schaerf. A Survey of Automated Timetabling. 1999, vol. 13, 37-38
109. K. Lukoseviciute, M. Ragulskis. Evolutionary algorithms for the selection of time lags for time series forecasting by fuzzy inference systems. *Neurocomputing*, 2010, 73:2077-2088.
110. G. Gonzalez-Rivera, J.Arroyo. A Time series modeling of histogram-valued data: The daily histogram time series of S&P500 intradaily returns. *Internacional Journal of Forecasting*, 2012, 28:20-33.
111. E.Wit, H.Edwin, J.W.Romeyn. All models are wrong...: an introduction to model uncertainty. *Statistica Neerlandica*, 2012, 66 (3): 217–236.
112. C. Paoli, C.Voyant, M.Muselli, M.L.Nivet. Forecasting of preprocessed daily solar radiation time series using neural networks. *Solar Energy* 84:2146-2160, 2010.
113. C.Kongcharoen, T.Kruangpradit. Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) Model for Thailand Export. *International Symposium on Forecasting*, 2013, vol. 33.
114. L.Liu, G.B.Hudak. Forecasting and time series analysis using SCA statistical system. 1994, vol. 1, 92-94.
115. G.E.P.Box, G.M.Jenkins. *Time Series Analysis Forecasting and Control*. 1971, vol. 22, No. 2, 199-201.
116. S.Boslaugh. *Statistics in a Nutshell*, 2012, vol.2.
117. L.Myers, M.J.Sirois. Spearman correlation coefficients, differences between, *Encyclopedia of Statistical Sciences*, 2006.
118. M.R.Nornadiah, W.Y.Bee. Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analytics*, 2011, 2:21-33.
119. T.W.Anderson, D.A.Darling. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, 1952, 23:193–212.
120. <https://www.investing.com/currencies/usd-rub-historical-data>
121. <https://www.investing.com/commodities/brent-oil-historical-data>
122. <https://www.investing.com/commodities/gold-historical-data>
123. <https://www.tradingeconomics.com/russia/indicators>
124. <https://www.tradingeconomics.com/united-states/indicators>

7. Priedai

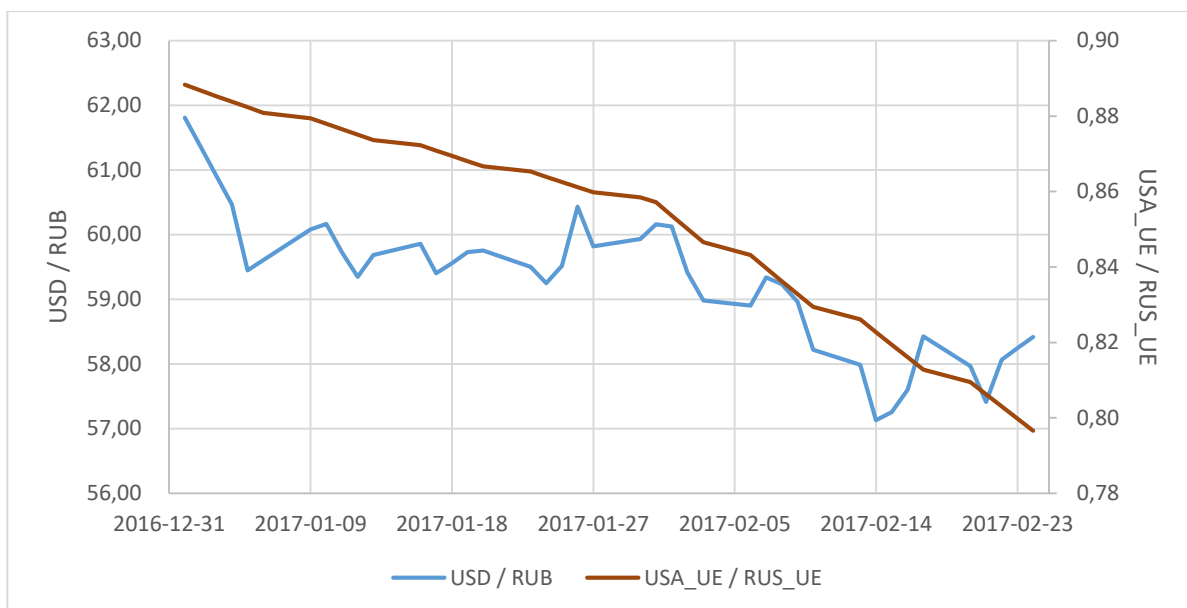
7.1. Išorinių regresorių dinamika.



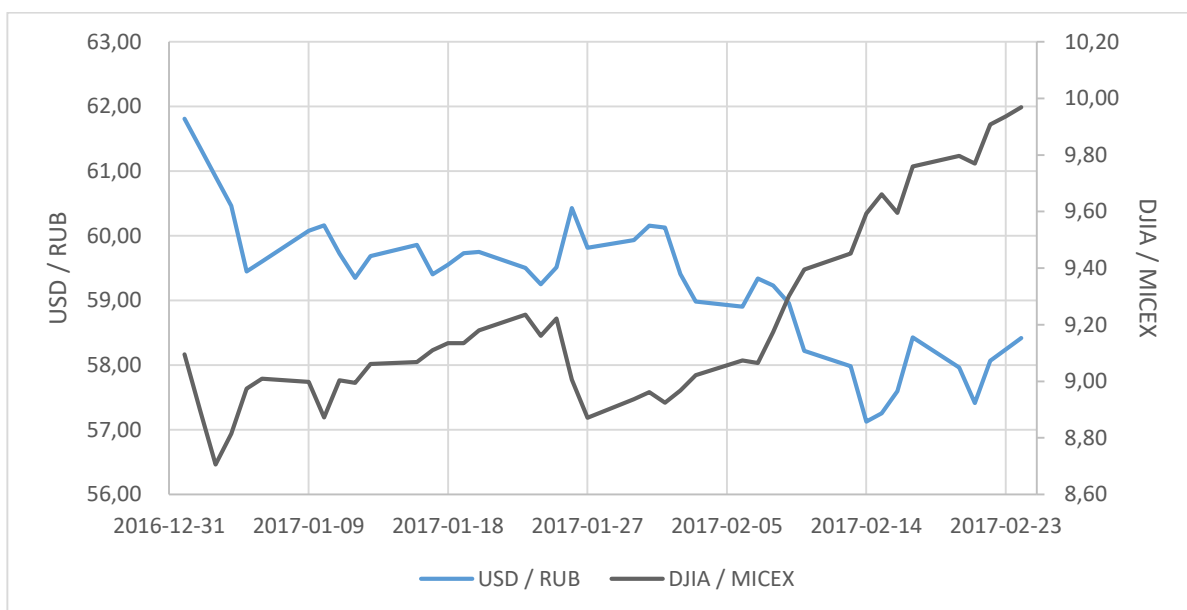
7.1.1 pav. Finansinės eilutės, USD/RUB kursas, ir išorinio regresoriaus Gold dinamika.



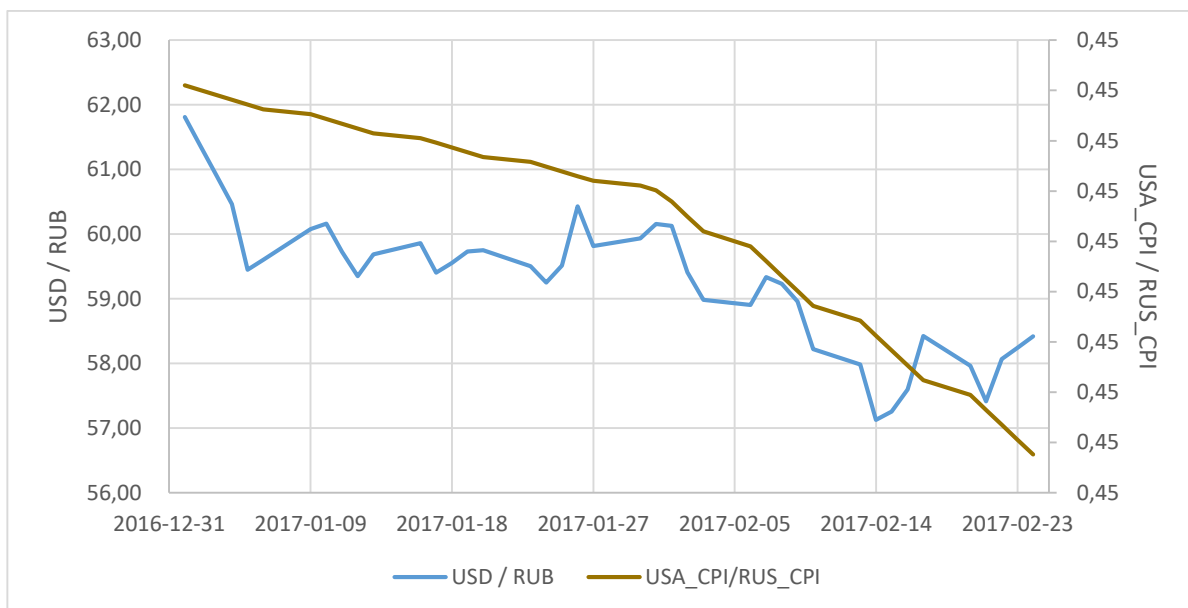
7.1.2 pav. Finansinės eilutės, USD/RUB kursas, ir išorinio regresoriaus Brent dinamika.



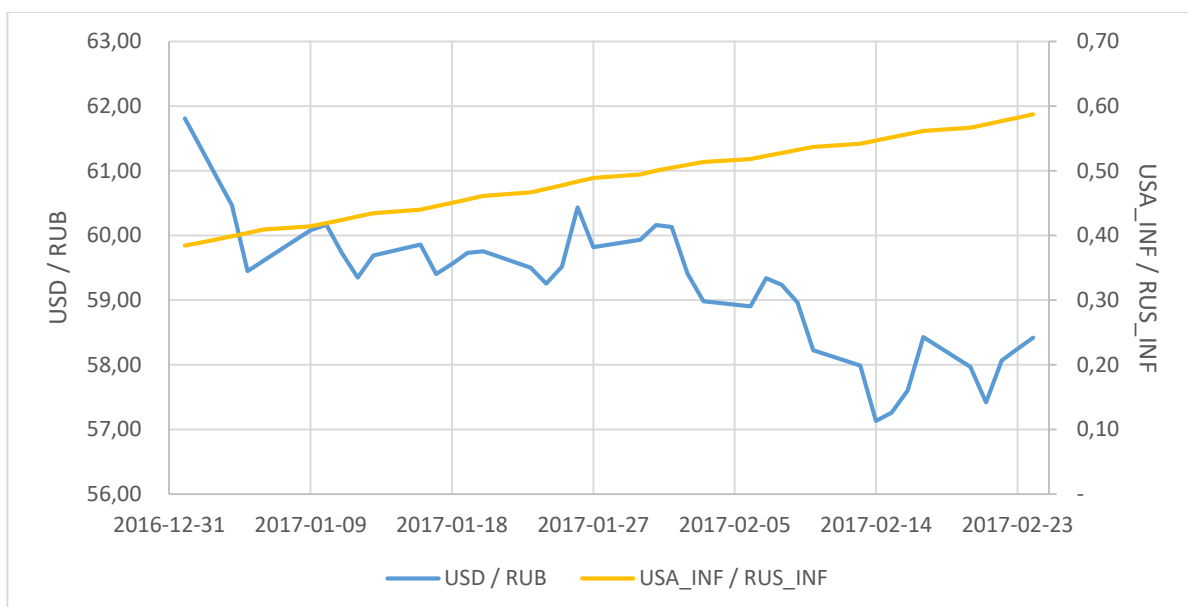
7.1.3 pav. Finansinės eilutės, USD/RUB kursas, ir išorinio regresoriaus USA_UE / RUS_UE dinamika.



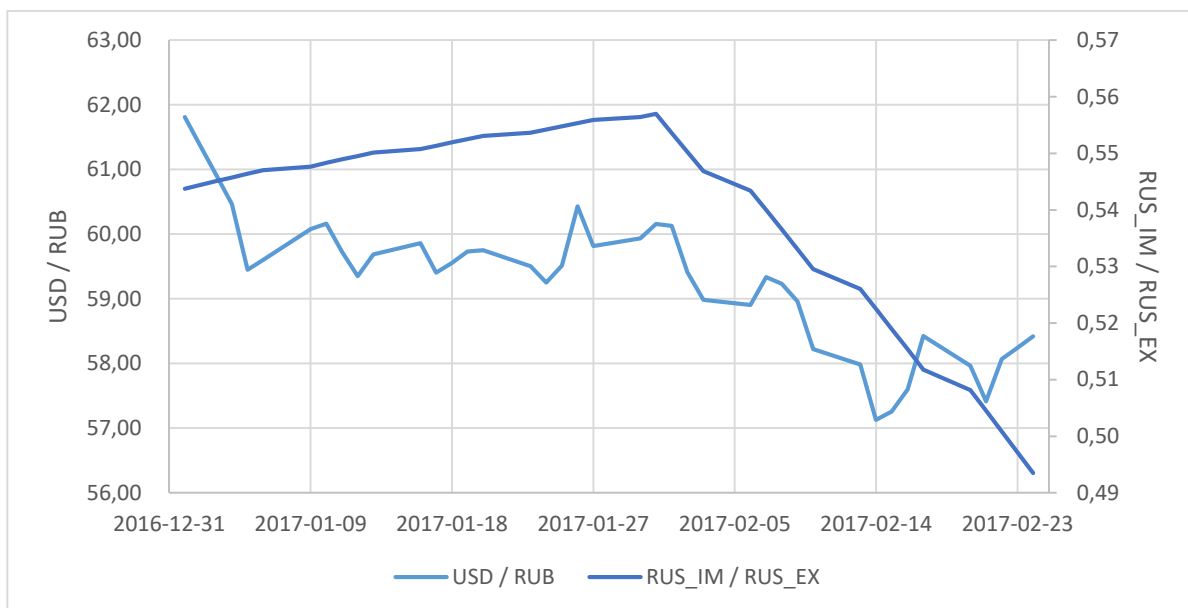
7.1.4 pav. Finansinės eilutės, USD/RUB kursas, ir išorinio regresoriaus DJIA / MICEX dinamika.



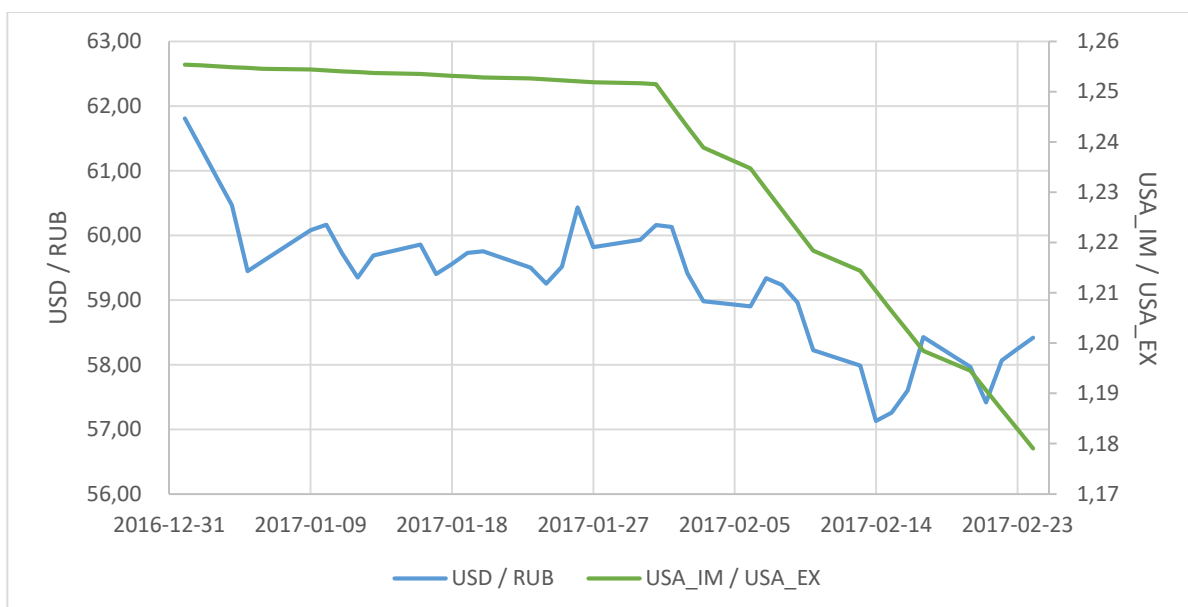
7.1.5 pav. Finansinės eilutės, USD/RUB kursas, ir išorinio regresoriaus USA_CPI / RUS_CPI dinamika.



7.1.6 pav. Finansinės eilutės, USD/RUB kursas, ir išorinio regresoriaus USA_INF / RUS_INF dinamika.



7.1.7 pav. Finansinės eilutės, USD/RUB kursas, ir išorinio regresoriaus RUS_IM / RUS_EX dinamika.



7.1.8 pav. Finansinės eilutės, USD/RUB kursas, ir išorinio regresoriaus USA_IM / USA_EX dinamika.

7.2.DJIA ir MICEX regresorių sudėtis.

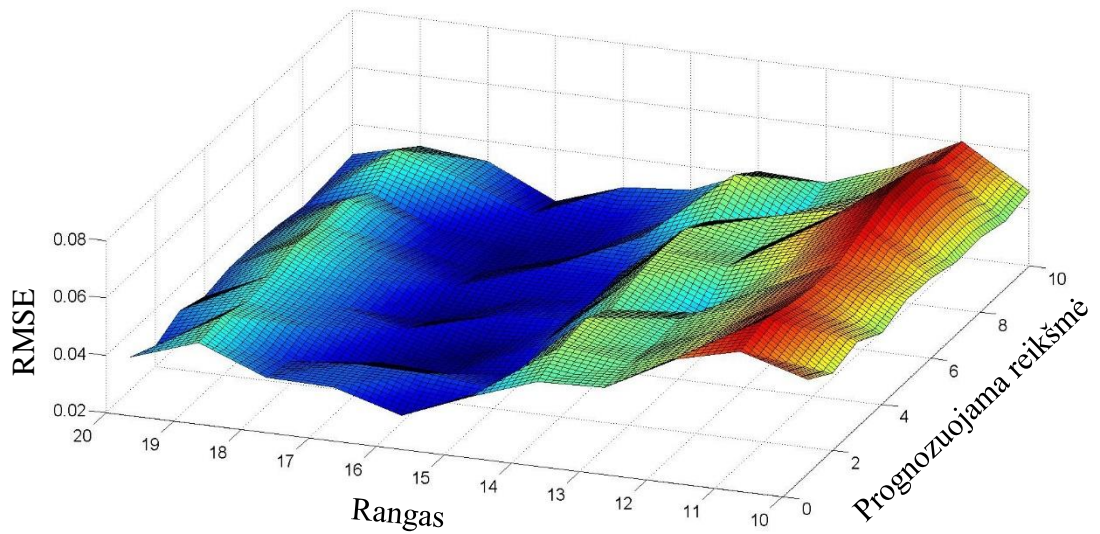
7.2.1 lentelė. DJIA indekso sudėtis.

Kodas	Kompanija
MMM	3M
AXP	American Express
AAPL	Apple
BA	Boeing
CAT	Caterpillar
CVX	Chevron
CSCO	Cisco Systems
KO	Coca-Cola
DD	DuPont
XOM	ExxonMobil
GE	General Electric
GS	Goldman Sachs
HD	The Home Depot
IBM	IBM
INTC	Intel
JNJ	Johnson & Johnson
JPM	JPMorgan Chase
MCD	McDonald's
MRK	Merck
MSFT	Microsoft
NKE	Nike
PFE	Pfizer
PG	Procter & Gamble
TRV	Travelers
UNH	UnitedHealth Group
UTX	United Technologies
VZ	Verizon
V	Visa
WMT	Wal-Mart
DIS	Walt Disney

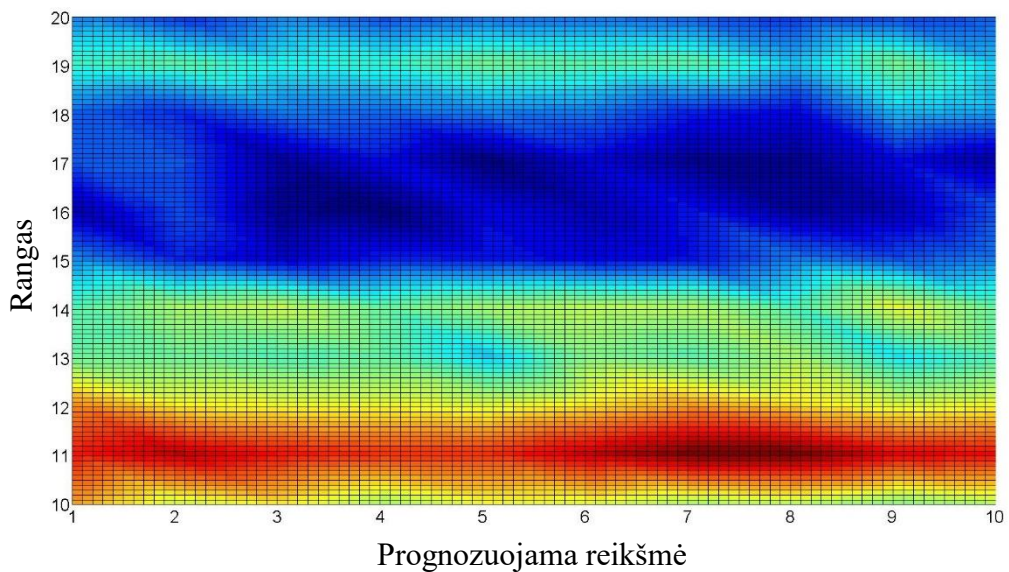
7.2.2 lentelē. MICEX indekso sudētis.

Kodas	Kompanija
AFKS	AFK SISTEMA
AFLT	JSC "AEROFLOT"
AKRN	Acron
ALRS	AC "ALROSA"
BANE, BANEP	OAO ANK "Bashneft"
CHMF	Severstal
DIXY	DIXY Group
EONR	OAO "E.ON Rossiya"
FEES	"FGC UES"
GAZP	GAZPROM
GMKN	"OJSC "MMC "NORILSK NICKEL"
HYDR	JSC "RusHydro"
IRAO	JSC "INTER RAO UES"
LKOH	OAO "LUKOIL"
LSRG	OJSC LSR Group
MAGN	OJSC "MMK"
MGNT	OJSC "Magnit"
MRKH	JSC "IDGC Holding"
MSNG	AO MOSENERGO
MSRS	MOESK
MSTT	OAO "MOSTOTREST"
MTLR, MTLRP	Mechel OAO
MTSS	MTS OJSC
MVID	Open Joint-Stock Company "Company "M.video"
NLMK	NLMK
NMTP	PJSC "NCSP"
NVTK	JSC "NOVATEK"
OGKB	JSC "OGK-2"
PHOR	PhosAgro
PHST	JSC "Pharmstandard"
PIKK	"PIK Group"
RASP	Raspadskaya OJSC
ROSN	Rosneft
RTKM, RTKMP	Rostelecom
RUALR	Rusal, RDR
SBER, SBERP	Sberbank
SNGS, SNGSP	Surgutneftegas
SVAV	Sollers
TATN, TATNP	TATNEFT
TRNFP	Transneft
URKA	Uralkali
VSMO	VSMPO-AVISMA Corporation
VTBR	JSC VTB Bank
VZRZ	V.Bank

7.3. Modelio veikimas iš skirtingų laiko taškų – eksperimentų grafikai



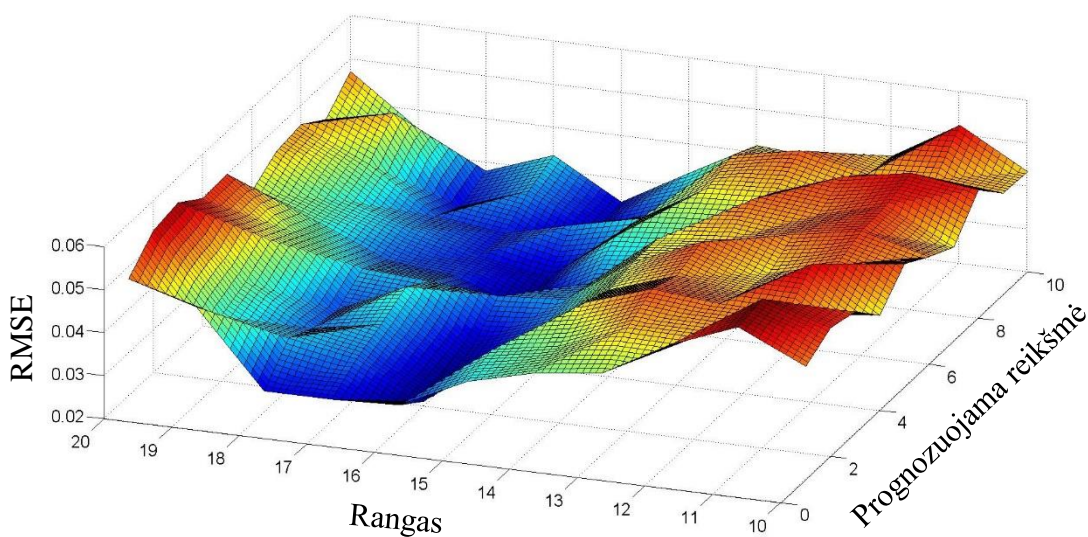
7.3.1 pav. Laiko eilutės analizė pradedant nuo 212-os bendros sekos reikšmės. Trijų ašių grafikas.



7.3.2 pav. Laiko eilutės analizė pradedant nuo 212-os bendros sekos reikšmės. Dviejų ašių grafikas.

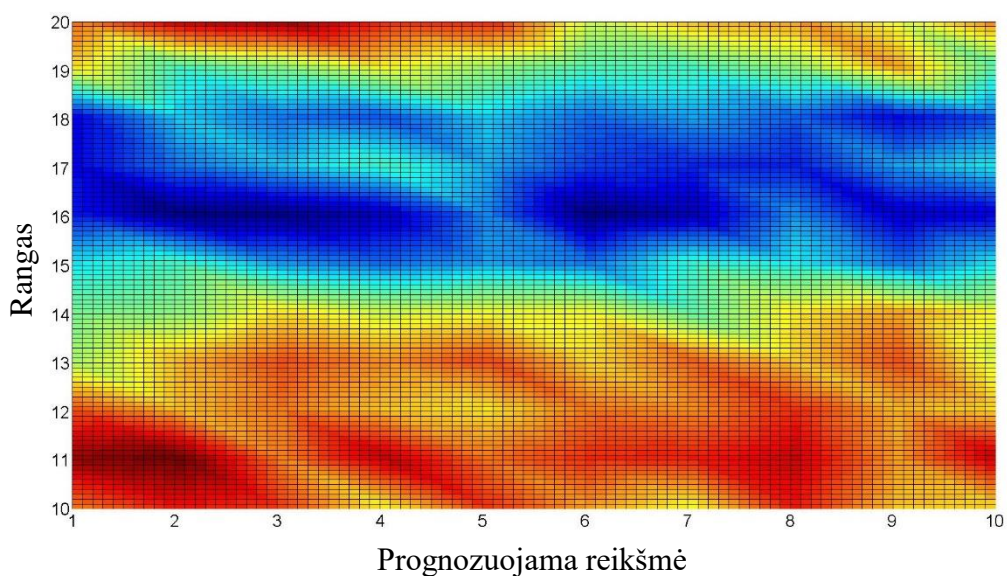
Analizuojant eilutę kuri prasideda nuo 255-os bendros sekos reikšmės pastebime, kad:

- Yra ryškios lūžio linijos ties rangais 15 ir 18.
- Geriausia RMSE reikšmė pasiekama ties 16 rangau.
- Ties visais rangais, RMSE reikšmė išlieka gana pastovaus dydžio, tai yra kinta labai nežymiai, nepriklausomai nuo prognozuojamų reikšmių skaičiaus.



7.3.3 pav. Laiko eilutės analizė pradedant nuo 83-os bendros sekos reikšmės. Trijų ašių grafikas.

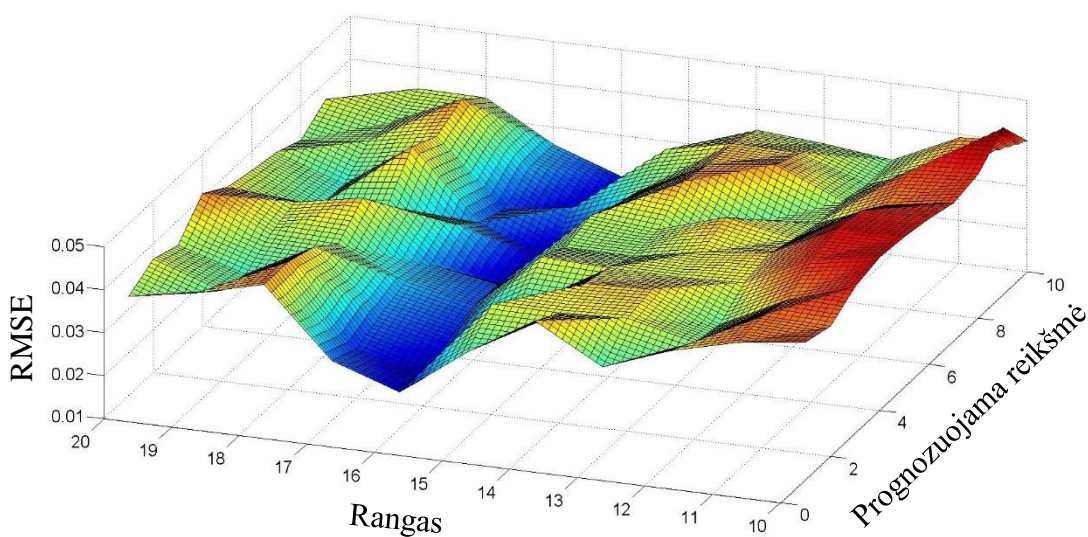
Trijų



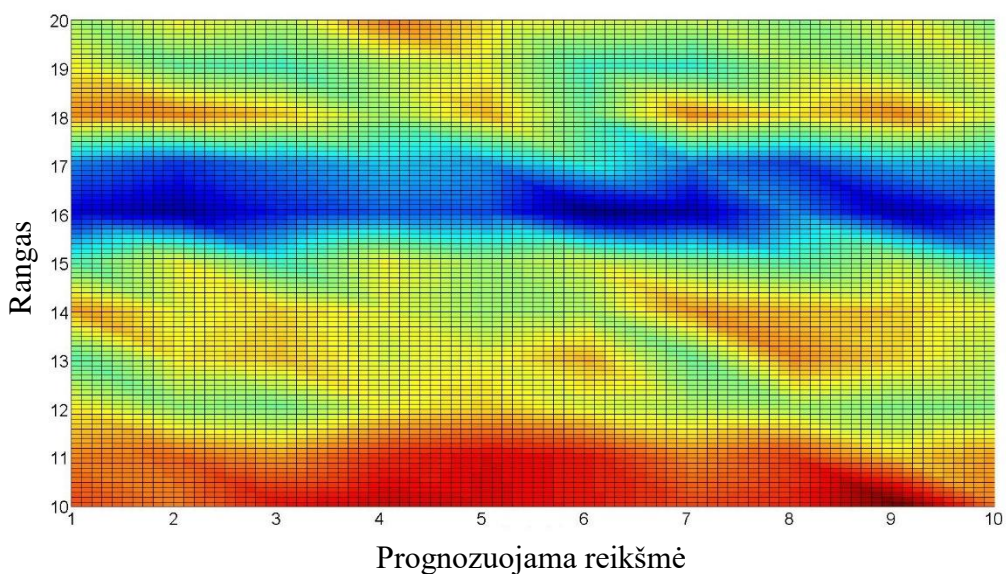
7.3.4 pav. Laiko eilutės analizė pradedant nuo 83-os bendros sekos reikšmės. Dviejų ašių grafikas.

Analizuojant eilutę kuri prasideda nuo 255-os bendros sekos reikšmės pastebime, kad:

- Yra ryškios lūžio linijos ties rangais 15 ir 18.
- Geriausia RMSE reikšmė pasiekama ties 16 rangų.
- Ties visais rangais, RMSE reikšmė išlieka gana pastovaus dydžio, tai yra kinta labai nežymiai, nepriklausomai nuo prognozuojamų reikšmių skaičiaus.



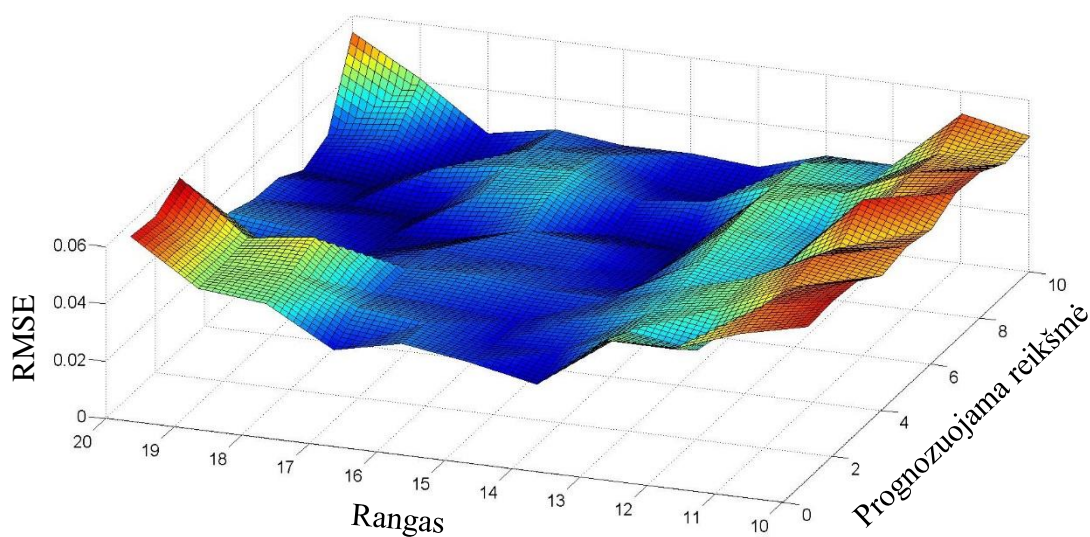
7.3.5 pav. Laiko eilutės analizė pradedant nuo 281-os bendros sekos reikšmės. Trijų ašių grafikas.



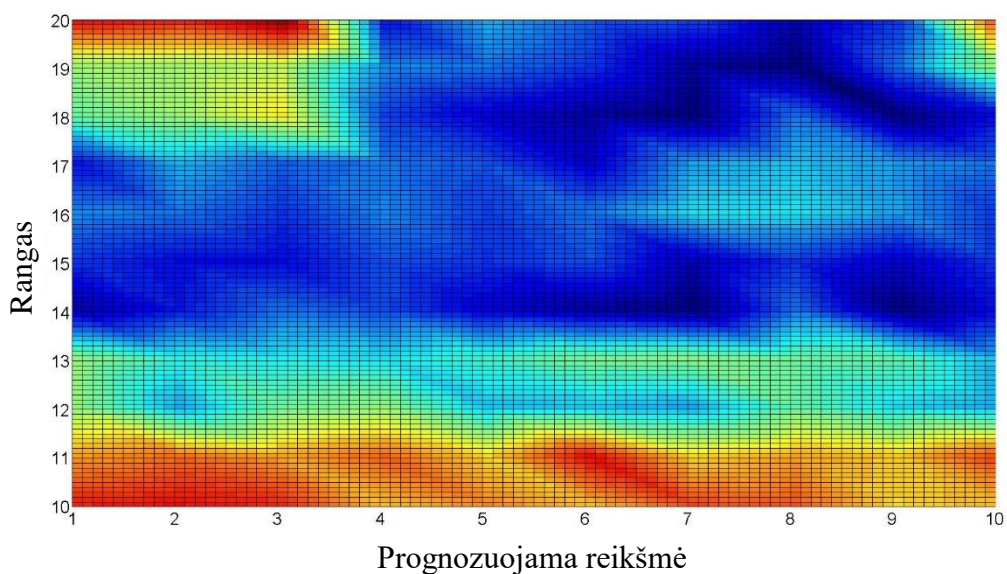
7.3.6 pav. Laiko eilutės analizė pradedant nuo 281-os bendros sekos reikšmės. Dviejų ašių grafikas.

Analizuojant eilutę kuri prasideda nuo 255-os bendros sekos reikšmės pastebime, kad:

- Yra ryškios lūžio linijos ties rangais 15 ir 17.
- Geriausia RMSE reikšmė pasiekama ties 16 rangau.
- Ties visais rangais, RMSE reikšmė išlieka gana pastovaus dydžio, tai yra kinta labai nežymiai, nepriklausomai nuo prognozuojamų reikšmių skaičiaus.



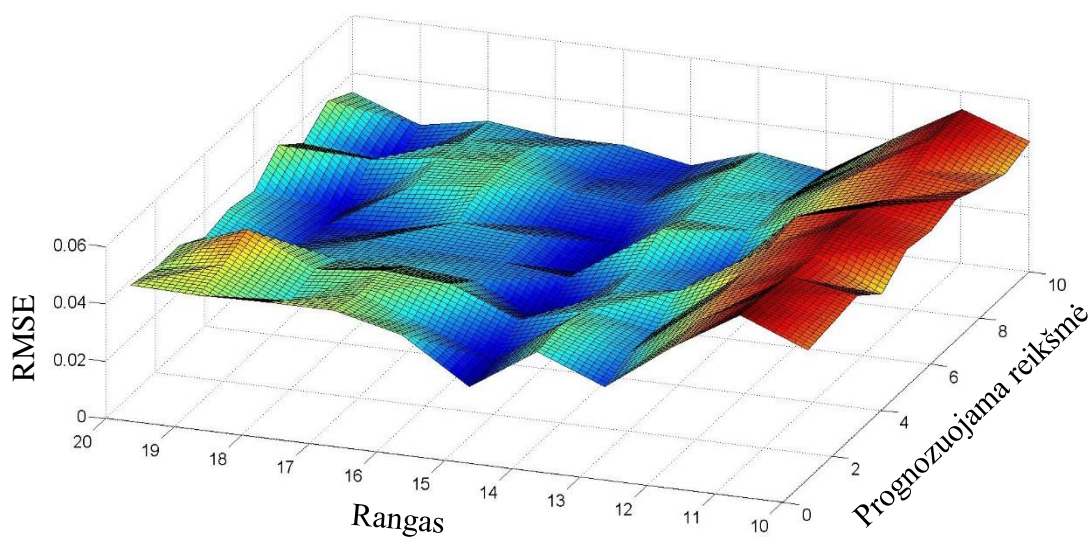
7.3.7 pav. Laiko eilutės analizė pradedant nuo 374-os bendros sekos reikšmės. Trijų ašių grafikas.



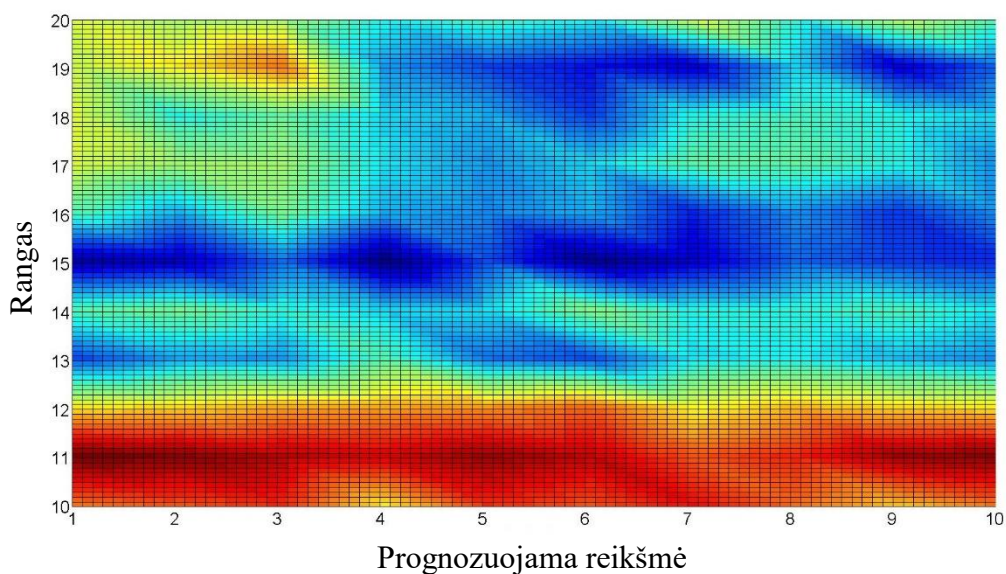
7.3.8 pav. Laiko eilutės analizė pradedant nuo 374-os bendros sekos reikšmės. Dviejų ašių grafikas.

Analizuojant eilutę kuri prasideda nuo 255-os bendros sekos reikšmės pastebime, kad:

- Yra ryški lūžio linija ties rangų 13.
- Geriausia RMSE reikšmė pasiekama ties 14, 19 rangų.
- Ties rangais 10-17, RMSE reikšmė išlieka gana pastovaus dydžio, tai yra kinta labai nežymiai, nepriklausomai nuo prognozuojamų reikšmių skaičiaus.



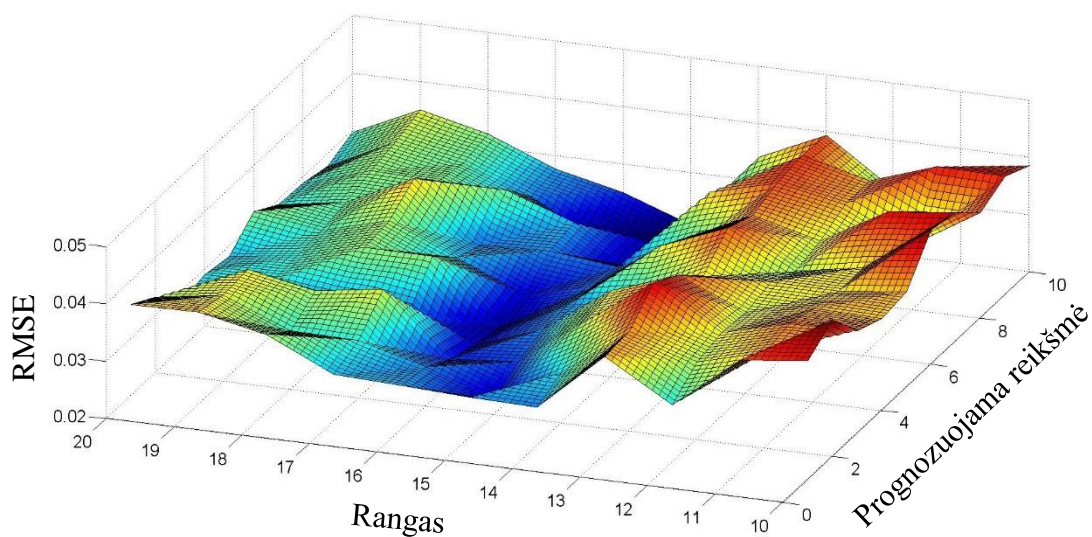
7.3.9 pav. Laiko eilutės analizė pradedant nuo 148-os bendros sekos reikšmės. Trijų ašių grafikas.



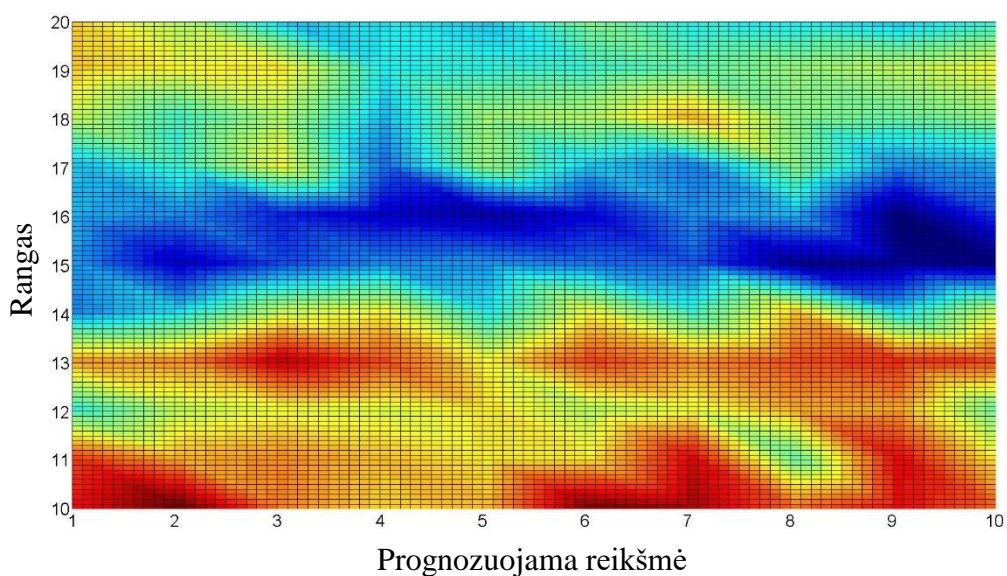
7.3.10 pav. Laiko eilutės analizė pradedant nuo 148-os bendros sekos reikšmės. Dviejų ašių grafikas.

Analizuojant eilutę kuri prasideda nuo 255-os bendros sekos reikšmės pastebime, kad:

- Yra ryški lūžio linija ties rangų 14.
- Geriausia RMSE reikšmė pasiekama ties 15 rangų.
- Ties rangais 10-16, RMSE reikšmė išlieka gana pastovaus dydžio, tai yra kinta labai nežymiai, nepriklausomai nuo prognozuojamų reikšmių skaičiaus.



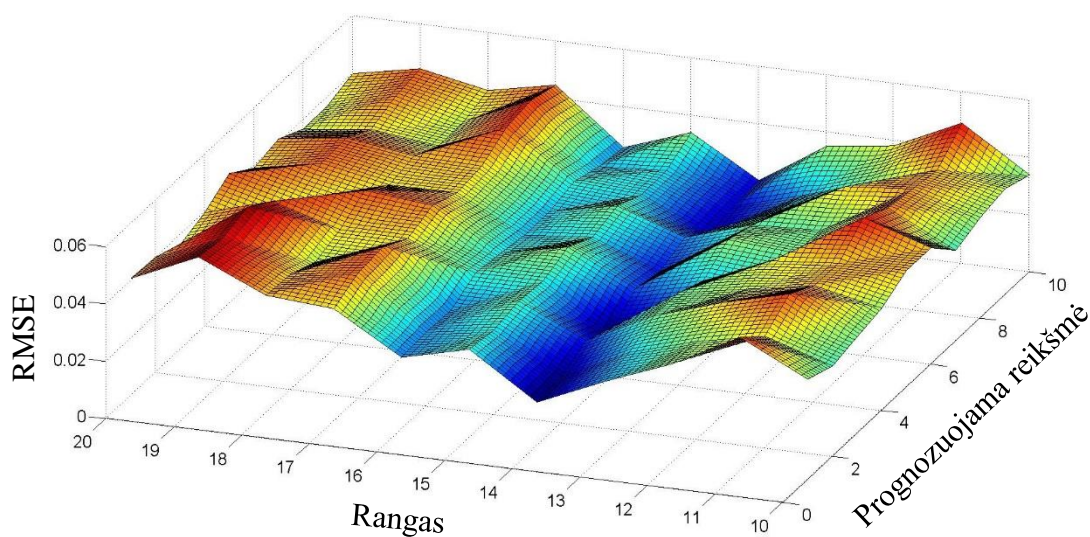
7.3.11 pav. Laiko eilutės analizė pradedant nuo 330-os bendros sekos reikšmės. Trijų ašių grafikas.



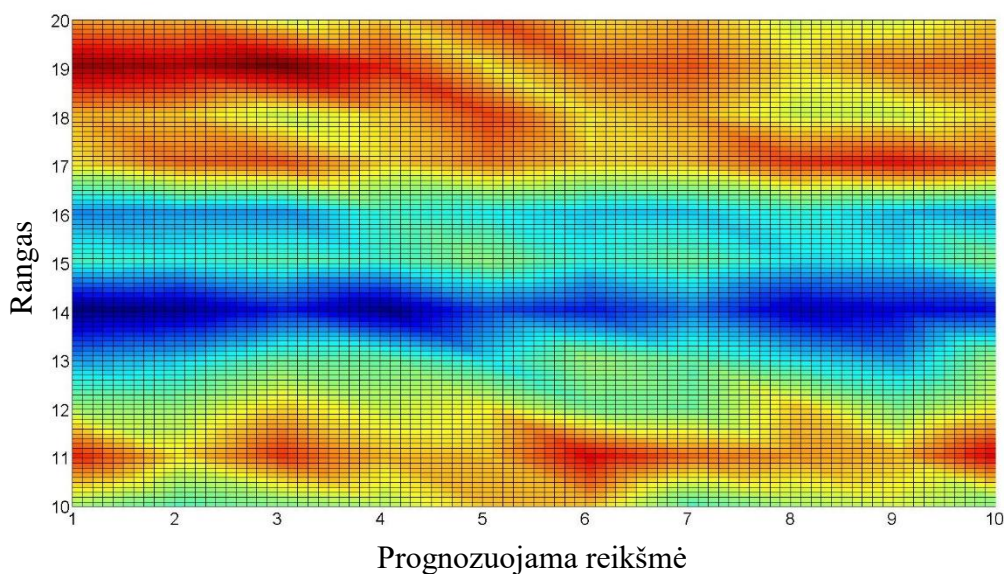
7.3.12 pav. Laiko eilutės analizė pradedant nuo 330-os bendros sekos reikšmės. Dviejų ašių grafikas.

Analizuojant eilutę kuri prasideda nuo 255-os bendros sekos reikšmės pastebime, kad:

- Yra ryškios lūžio linijos ties rangais 14 ir 17.
- Geriausia RMSE reikšmė pasiekama ties 15 rangu.
- Ties visais rangais, RMSE reikšmė išlieka gana pastovaus dydžio, tai yra kinta labai nežymiai, nepriklausomai nuo prognozuojamų reikšmių skaičiaus.



7.3.13 pav. Laiko eilutės analizė pradedant nuo 8-os bendros sekos reikšmės. Trijų ašių grafikas.

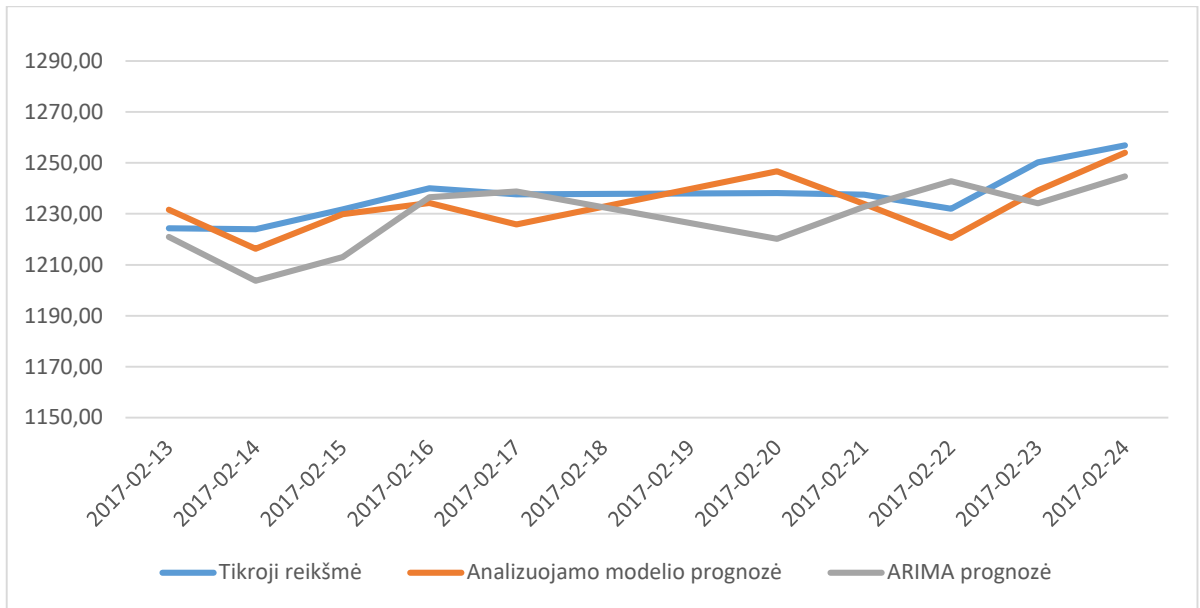


7.3.14 pav. Laiko eilutės analizė pradedant nuo 8-os bendros sekos reikšmės. Dviejų ašių grafikas.

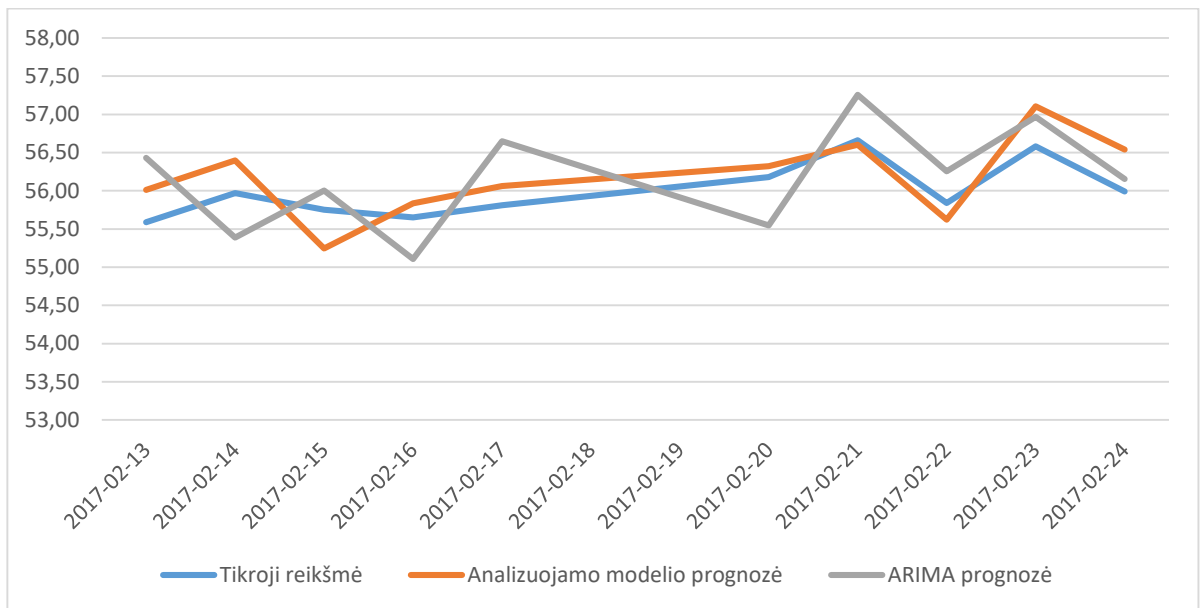
Analizuojant eilutę kuri prasideda nuo 255-os bendros sekos reikšmės pastebime, kad:

- Yra ryškios lūžio linijos ties rangais 13 ir 15.
- Geriausia RMSE reikšmė pasiekama ties 14 rangų.
- Ties visais rangais, RMSE reikšmė išlieka gana pastovaus dydžio, tai yra kinta labai nežymiai, nepriklausomai nuo prognozuojamų reikšmių skaičiaus.

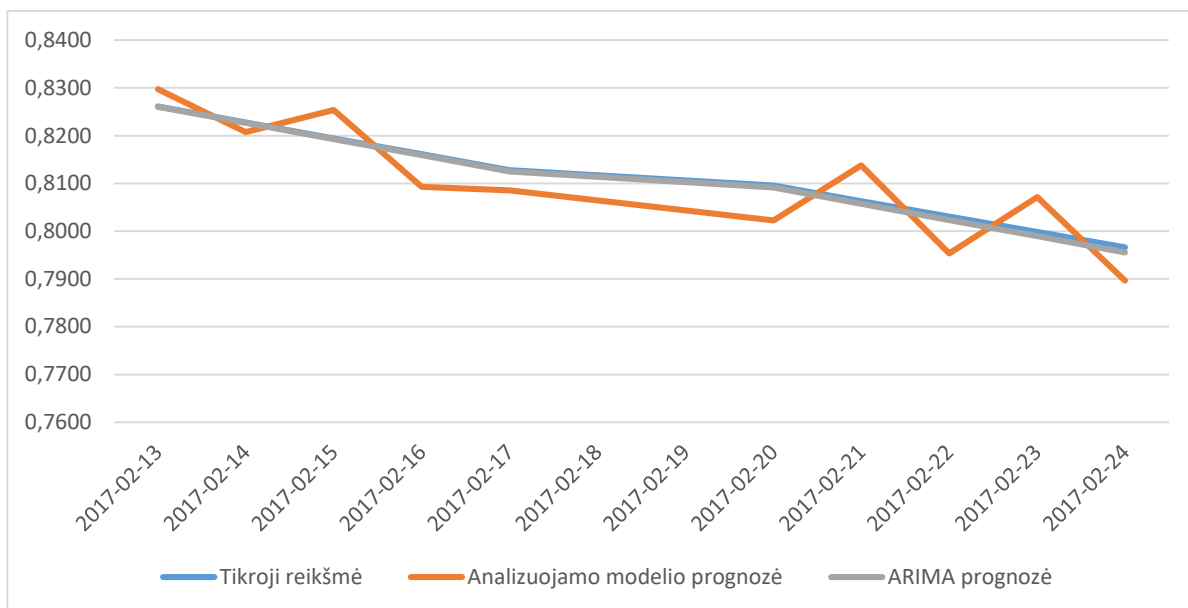
7.4. Išorinių regresorių prognozės



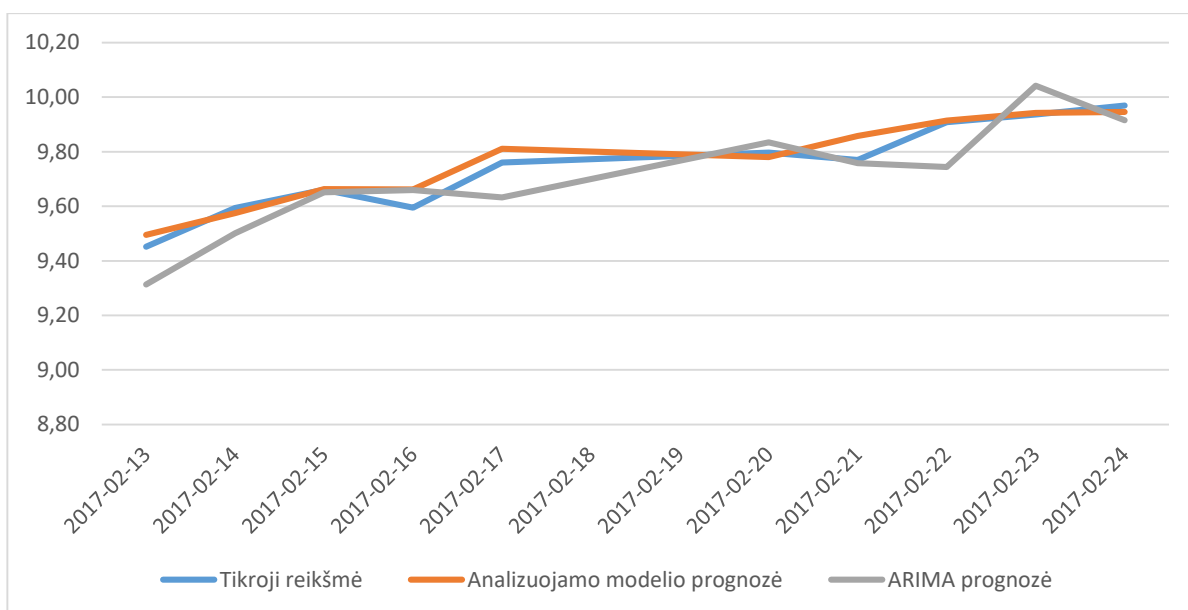
7.4.1 pav. Išorinio regresoriaus Gold prognozė.



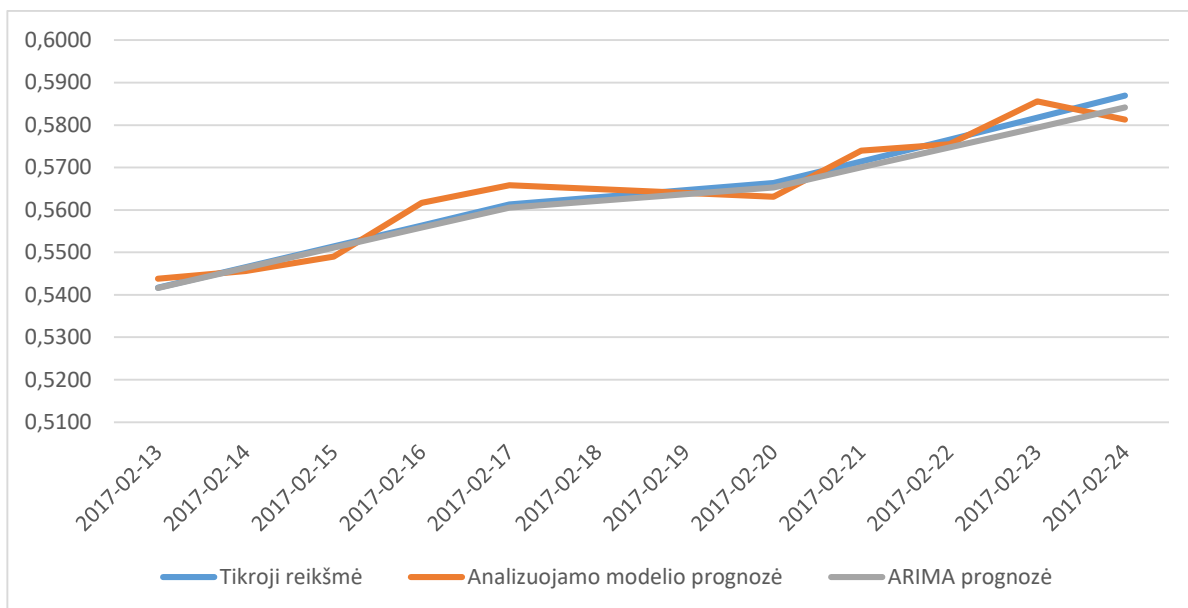
7.4.2 pav. Išorinio regresoriaus Brent prognozė.



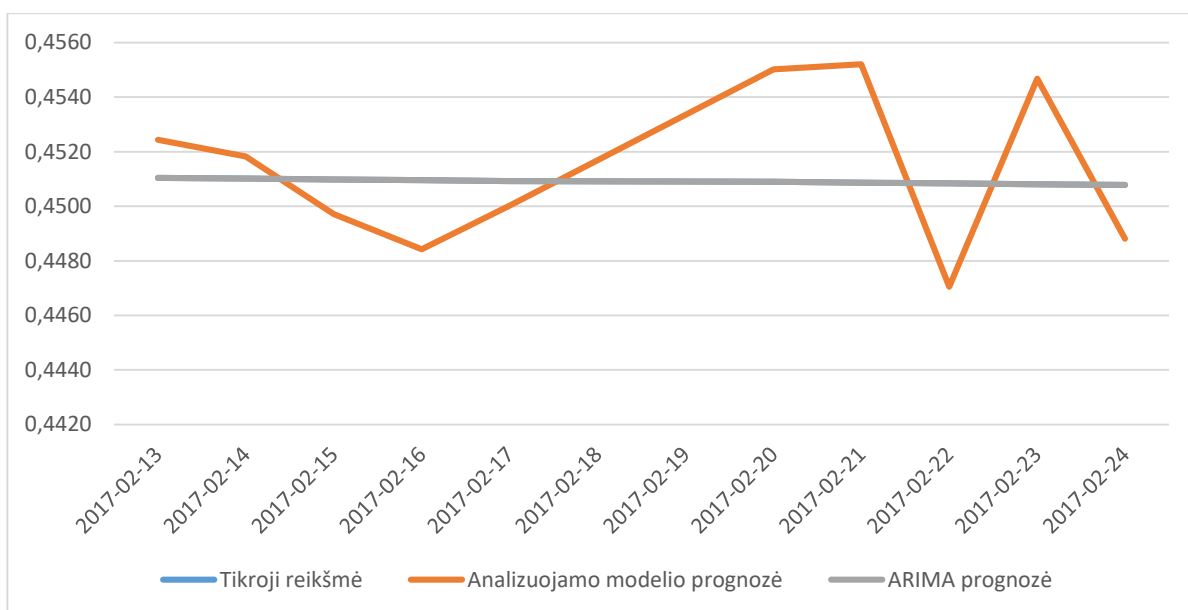
7.4.3 pav. Išorinio regresoriaus USA_UE / RUS_UE prognozė.



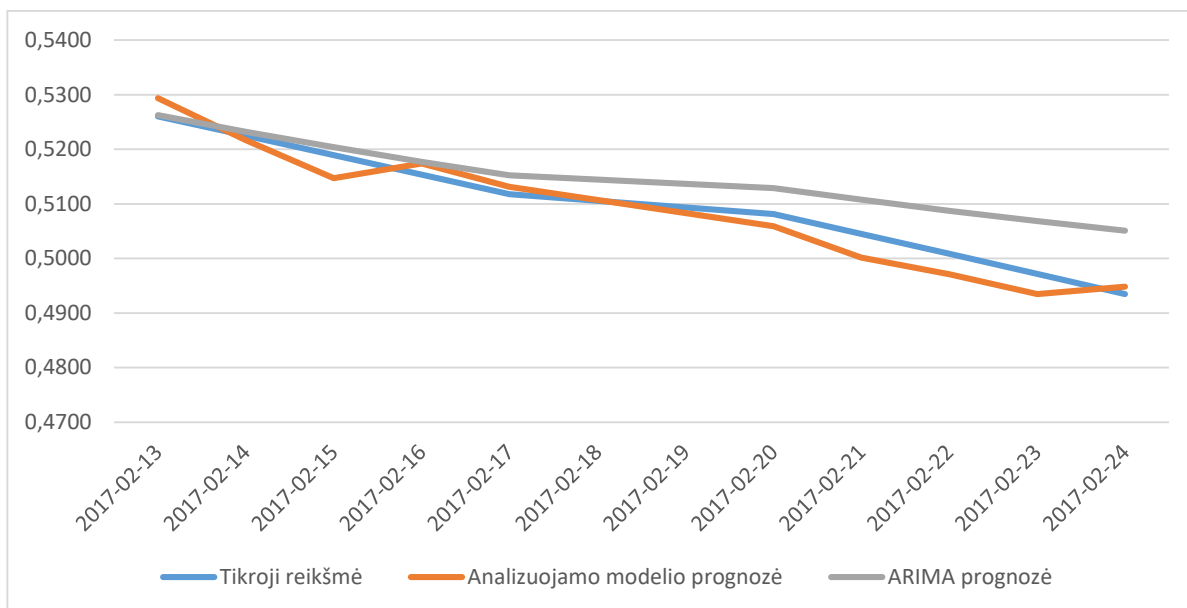
7.4.4 pav. Išorinio regresoriaus DJIA / MICEX prognozė.



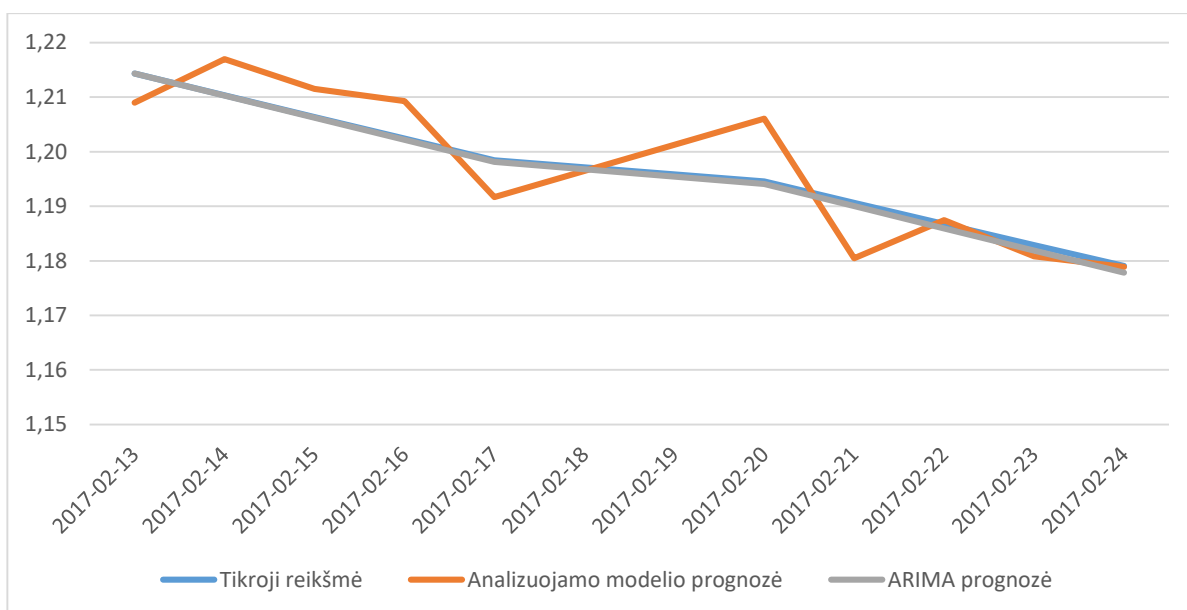
7.4.5 pav. Išorinio regresoriaus USA_INF / RUS_INF prognozė.



7.4.6 pav. Išorinio regresoriaus USA_CPI / RUS_CPI prognozė.



7.4.7 pav. Išorinio regresoriaus RUS_IM / RUS_EX prognozė.



7.4.8 pav. Išorinio regresoriaus USA_IM / USA_EX prognozė.

7.5. ARIMA modelių adekvatumo tyrimai

ARIMA modelio parametrai parinkti pasinaudojant Akaike ir Bajeso informaciniais kriterijais, taip pat atliktas Ljung-Box testas. Buvo naudotas R programinis paketas. Naudotos $AIC(x)$, $BIC(x)$, $Box.test(x, type = "Ljung - Box")$ funkcijos.

7.5.1 lentelė. AIC ir BIC kriterijų reikšmės skirtingiems ARIMA modelio parametrams. Prognozuojamas išorinis regresorius: Brent.

		AIC						BIC							
p	q	d=0	d=1	d=2	d=3	d=4	d=5	p	q	d=0	d=1	d=2	d=3	d=4	d=5
0	0	85	79	91	114	147	170	0	0	90	79	95	118	150	174
0	1	75	79	82	94	123	145	0	1	81	84	87	99	128	150
0	2	75	76	82	89	106	125	0	2	82	82	88	96	113	131
0	3	71	75	82	89	107	112	0	3	79	83	89	97	114	120
0	4	75	75	80	91	104	113	0	4	86	84	89	100	113	122
0	5	74	77	81	90	106	111	0	5	86	87	91	100	116	122
1	0	79	80	91	109	137	156	1	0	85	85	96	114	142	161
1	1	76	81	83	95	119	137	1	1	84	87	89	101	125	143
1	2	73	77	84	90	108	122	1	2	82	84	92	98	115	130
1	3	75	74	81	91	110	114	1	3	86	83	91	100	119	123
1	4	77	76	79	213	106	116	1	4	89	86	90	223	117	126
1	5	75	79	81	89	106	117	1	5	88	91	93	101	118	129
2	0	73	79	89	104	130	148	2	0	81	85	95	110	136	155
2	1	71	74	82	93	114	131	2	1	84	82	90	101	122	138
2	2	77	76	80	90	106	118	2	2	87	85	89	99	115	128
2	3	76	76	84	90	105	113	2	3	88	86	94	101	115	124
2	4	78	77	81	91	105	113	2	4	91	89	93	103	117	125
2	5	74	79	83	91	106	117	2	5	89	92	96	105	119	131
3	0	75	80	89	100	118	131	3	0	84	88	97	108	126	139
3	1	76	76	84	94	111	121	3	1	87	85	93	103	120	130
3	2	79	77	82	211	112	117	3	2	91	88	92	222	123	127
3	3	78	79	85	213	108	117	3	3	91	91	97	225	120	129
3	4	74	79	83	92	109	115	3	4	89	92	96	106	123	129
3	5	76	77	85	94	108	116	3	5	92	92	100	109	123	131
4	0	76	79	91	102	120	130	4	0	87	89	100	111	129	140
4	1	78	76	84	96	113	123	4	1	90	87	95	107	124	133
4	2	73	75	87	97	114	119	4	2	86	87	99	109	126	131
4	3	79	74	213	94	110	119	4	3	93	87	226	108	123	133
4	4	74	78	84	94	112	117	4	4	91	93	99	109	127	132
4	5	78	75	215	94	110	118	4	5	95	92	231	111	127	134
5	0	78	75	83	97	117	128	5	0	90	80	93	108	128	139
5	1	78	72	76	89	109	121	5	1	92	84	88	101	121	133
5	2	78	73	78	86	104	115	5	2	93	87	92	100	118	129
5	3	80	75	79	88	101	112	5	3	96	90	94	103	116	127
5	4	76	75	81	89	104	112	5	4	94	92	98	106	120	129
5	5	78	77	83	93	110	119	5	5	97	95	101	111	128	137

Prognozuojant išorinį regresorių Brent:

- parinktas ARIMA parametų rinkinys $(p, d, q) = (0, 0, 3)$;
- su šiuo parametų rinkiniu Ljung-Box testas parodė, kad liekanos yra nekoreliuotos.

7.5.2 lentelė. AIC ir BIC kriterijų reikšmės skirtingiems ARIMA modelio parametrams.
 Prognozuojamas išorinis regresorius: USA_CPI/RUS_CPI.

		AIC						BIC							
p	q	d=0	d=1	d=2	d=3	d=4	d=5	p	q	d=0	d=1	d=2	d=3	d=4	d=5
0	0	-470	-574	-660	-633	-399	-384	0	0	-465	-571	-657	-630	-395	-381
0	1	-504	-608	-731	-632	-397	-382	0	1	-498	-603	-692	-628	-392	-377
0	2	-535	-730	-729	-637	-395	-380	0	2	-528	-691	-687	-631	-388	-374
0	3	-563	-729	-727	-635	-393	-378	0	3	-554	-687	-682	-627	-385	-370
0	4	-588	-727	-726	-633	-391	-376	0	4	-577	-682	-677	-624	-382	-367
0	5	-605	-726	-658	-631	-389	-374	0	5	-593	-677	-648	-621	-378	-363
1	0	-230	-409	-664	-631	-397	-382	1	0	-224	-404	-659	-627	-392	-377
1	1	-55	-407	-664	-636	-395	-380	1	1	-47	-401	-658	-629	-388	-374
1	2	-294	-405	-662	-635	-393	-378	1	2	-285	-398	-654	-627	-385	-370
1	3	-575	-403	-660	-633	-391	-376	1	3	-564	-394	-651	-624	-382	-367
1	4	-610	-674	-658	-631	-389	-374	1	4	-598	-664	-648	-621	-378	-363
1	5	-678	-672	-656	-629	-387	-372	1	5	-665	-660	-644	-617	-375	-360
2	0	-420	-407	-664	-634	-395	-380	2	0	-412	-401	-657	-627	-388	-374
2	1	-54	-411	-662	-635	-393	-378	2	1	-45	-404	-654	-627	-385	-370
2	2	-52	-676	-660	-633	-391	-376	2	2	-41	-667	-651	-624	-382	-367
2	3	-417	-674	-658	-631	-389	-374	2	3	-405	-664	-648	-621	-378	-363
2	4	-693	-671	-656	-629	-387	-372	2	4	-679	-659	-644	-617	-375	-360
2	5	-412	-670	-654	-627	-385	-370	2	5	-397	-657	-641	-614	-371	-357
3	0	-418	-405	-662	-632	-393	-378	3	0	-409	-398	-654	-624	-385	-370
3	1	-703	-403	-660	-633	-391	-376	3	1	-692	-394	-651	-624	-382	-367
3	2	-696	-674	-658	-631	-389	-374	3	2	-684	-664	-648	-621	-378	-363
3	3	-412	-672	-656	-629	-387	-372	3	3	-672	-660	-644	-617	-375	-360
3	4	-410	-670	-654	-627	-385	-370	3	4	-395	-657	-641	-614	-371	-357
3	5	-693	-668	-652	-626	-383	-368	3	5	-676	-653	-637	-611	-368	-353
4	0	-368	-403	-660	-631	-391	-376	4	0	-358	-394	-651	-622	-382	-367
4	1	-53	-401	-658	-631	-389	-374	4	1	-41	-391	-648	-621	-378	-363
4	2	-731	-672	-656	-629	-387	-372	4	2	-692	-660	-644	-617	-375	-360
4	3	-729	-670	-654	-627	-385	-370	4	3	-677	-657	-641	-614	-371	-357
4	4	-727	-668	-652	-625	-383	-368	4	4	-672	-653	-637	-610	-368	-353
4	5	-726	-666	-650	-623	-381	-366	4	5	-29	-650	-634	-607	-364	-350
5	0	-51	-401	-658	-629	-389	-374	5	0	-39	-391	-648	-618	-378	-363
5	1	-412	-672	-656	-629	-387	-372	5	1	-398	-660	-644	-617	-375	-360
5	2	-731	-670	-654	-627	-385	-370	5	2	-692	-657	-641	-614	-371	-357
5	3	-728	-668	-652	-625	-383	-315	5	3	-678	-653	-637	-611	-368	-300
5	4	-727	-666	-650	-623	-381	-366	5	4	-672	-650	-634	-607	-364	-350
5	5	-726	-664	-648	-621	-379	-364	5	5	-672	-646	-630	-603	-361	-346

Prognozuojant išorinį regresorių USA_CPI/RUS_CPI:

- parinktas ARIMA parametų rinkinys $(p, d, q) = (0, 2, 1)$;
- su šiuo parametų rinkiniu Ljung-Box testas parodė, kad liekanos yra nekoreliuotos.

7.5.3 lentelė. AIC ir BIC kriterijų reikšmės skirtingiems ARIMA modelio parametrams. Prognozuojamas išorinis regresorius: DJIA / MICEX.

		AIC						BIC							
p	q	d=0	d=1	d=2	d=3	d=4	d=5	p	q	d=0	d=1	d=2	d=3	d=4	d=5
0	0	-31	-56	-41	-11	27	56	0	0	-26	-47	-38	-8	31	60
0	1	-56	-54	-47	-33	2	30	0	1	-41	-47	-42	-29	7	35
0	2	-54	-52	-47	-37	-17	9	0	2	-46	-46	-40	-31	-11	15
0	3	-54	-51	-45	-36	-19	-7	0	3	-45	-43	-37	-28	-11	1
0	4	-52	-50	-43	-34	-19	-8	0	4	-41	-41	-34	-25	-10	1
0	5	-50	-47	-41	-33	-17	-7	0	5	-38	-37	-31	-22	-6	4
1	0	-55	-54	-42	-22	10	34	1	0	-48	-47	-38	-18	15	39
1	1	-55	-52	-46	-34	-8	15	1	1	-47	-46	-40	-28	-2	21
1	2	-54	-52	-36	-36	-17	0	1	2	-45	-44	-35	-28	-10	8
1	3	-52	-50	-36	-34	-19	-7	1	3	-41	-41	-35	-25	-9	2
1	4	-50	-48	-36	-33	-17	-7	1	4	-38	-38	-35	-22	-6	4
1	5	-48	-46	-35	-34	-15	-3	1	5	-35	-33	-34	-22	-3	9
2	0	-56	-53	-42	-26	2	23	2	0	-48	-46	-36	-20	8	30
2	1	-54	-51	-45	-34	-11	8	2	1	-45	-43	-37	-26	-3	16
2	2	-52	-51	-43	-34	-16	-2	2	2	-41	-42	-34	-25	-7	7
2	3	-50	-49	-43	-34	-17	-6	2	3	-38	-38	-32	-23	-6	5
2	4	-48	-46	-41	-33	-15	-4	2	4	-35	-34	-29	-21	-3	8
2	5	-48	-46	-39	-31	-14	-3	2	5	-33	-33	-26	-17	-1	11
3	0	-54	-51	-43	-27	-1	19	3	0	-45	-44	-35	-19	7	27
3	1	-53	-50	-43	-34	-11	6	3	1	-42	-41	-34	-25	-2	15
3	2	-50	-49	-41	-33	-16	-2	3	2	-38	-38	-31	-22	-5	9
3	3	-48	-47	-41	-32	-15	-4	3	3	-35	-35	-29	-20	-3	8
3	4	-46	-45	-39	-30	-12	-2	3	4	-31	-31	-26	-17	2	11
3	5	-47	-46	-38	-28	-11	-1	3	5	-30	-31	-23	-13	3	14
4	0	-52	-49	-43	-29	-6	12	4	0	-41	-40	-34	-20	3	21
4	1	-53	-47	-41	-33	-12	2	4	1	-41	-37	-31	-23	-2	13
4	2	-49	-47	-39	-31	-15	-2	4	2	-35	-35	-27	-19	-3	10
4	3	-50	-47	-39	-31	-13	-3	4	3	-35	-34	-26	-17	1	10
4	4	-49	-44	-38	-28	-11	-1	4	4	-32	-29	-23	-13	4	14
4	5	-47	-44	-36	-27	-10	-1	4	5	-29	-27	-20	-11	6	16
5	0	-51	-48	-41	-29	-6	8	5	0	-39	-37	-31	-18	4	18
5	1	-49	-46	-39	-31	-12	2	5	1	-36	-34	-27	-19	0	14
5	2	-47	-44	-38	-32	-13	-1	5	2	-32	-30	-25	-18	1	12
5	3	-48	-44	-36	-34	-12	-1	5	3	-31	-29	-21	-20	3	14
5	4	-46	-44	-36	-33	-10	1	5	4	-29	-27	-19	-16	6	17
5	5	-45	-42	-36	-31	-10	-1	5	5	-25	-24	-18	-13	8	17

Prognozuojant išorinį regresorių DJIA / MICEX:

- parinktas ARIMA parametų rinkinys $(p, d, q) = (2, 0, 0)$;
- su šiuo parametų rinkiniu Ljung-Box testas parodė, kad liekanos yra nekoreliuotos.

7.5.4 lentelė. AIC ir BIC kriterijų reikšmės skirtingiems ARIMA modelio parametrams. Prognozuojamas išorinis regresorius: Gold.

		AIC						BIC							
p	q	d=0	d=1	d=2	d=3	d=4	d=5	p	q	d=0	d=1	d=2	d=3	d=4	d=5
0	0	284	209	220	245	276	298	0	0	289	212	223	248	280	302
0	1	260	210	204	219	248	270	0	1	266	215	214	224	253	275
0	2	243	212	206	208	226	246	0	2	251	218	213	214	233	252
0	3	237	213	208	210	220	227	0	3	247	221	216	218	227	235
0	4	237	213	209	212	222	224	0	4	248	222	218	221	231	234
0	5	235	213	209	212	224	226	0	5	247	224	220	223	234	236
1	0	222	210	213	227	250	265	1	0	228	215	218	232	255	270
1	1	224	203	206	213	233	246	1	1	232	212	213	219	239	252
1	2	225	213	212	210	222	232	1	2	234	221	217	218	230	240
1	3	227	215	210	211	222	225	1	3	237	224	219	220	231	234
1	4	226	214	211	213	223	226	1	4	238	224	221	223	233	237
1	5	226	215	211	214	225	227	1	5	240	227	223	226	237	239
2	0	224	212	214	226	247	258	2	0	232	218	220	232	253	264
2	1	226	213	208	214	232	243	2	1	235	217	216	222	240	251
2	2	226	214	210	212	224	232	2	2	237	223	219	221	233	241
2	3	228	215	209	213	226	227	2	3	240	225	219	224	237	237
2	4	226	215	211	212	225	229	2	4	240	227	223	224	237	241
2	5	228	217	212	214	224	230	2	5	243	230	225	228	237	243
3	0	225	214	209	221	243	255	3	0	234	221	217	228	250	263
3	1	227	215	210	210	228	240	3	1	237	224	219	219	237	250
3	2	227	215	211	211	221	229	3	2	239	226	222	222	232	239
3	3	230	216	212	213	223	226	3	3	243	228	224	225	235	238
3	4	228	217	213	214	225	227	3	4	243	230	227	228	238	240
3	5	230	219	214	216	226	228	3	5	246	234	229	231	241	243
4	0	227	213	209	214	234	247	4	0	238	223	218	223	243	256
4	1	229	215	211	211	222	233	4	1	241	226	221	221	232	243
4	2	228	217	212	212	222	225	4	2	242	229	224	224	234	237
4	3	232	218	213	214	224	226	4	3	246	231	226	228	237	239
4	4	230	219	215	216	226	228	4	4	246	234	230	231	241	243
4	5	232	221	215	218	228	230	4	5	249	237	231	234	244	246
5	0	226	215	211	214	230	241	5	0	238	225	221	224	241	252
5	1	228	216	213	212	222	230	5	1	241	228	225	224	235	242
5	2	230	218	213	214	224	226	5	2	245	231	226	228	238	240
5	3	231	219	215	216	225	228	5	3	247	234	229	231	240	243
5	4	234	220	216	218	228	229	5	4	251	236	233	234	244	245
5	5	234	221	216	219	230	232	5	5	253	239	234	237	247	250

Prognozuojant išorinį regresorių Gold:

- parinktas ARIMA parametų rinkinys $(p, d, q) = (1, 1, 1)$;
- su šiuo parametų rinkiniu Ljung-Box testas parodė, kad liekanos yra nekoreliuotos.

7.5.5 lentelė. AIC ir BIC kriterijų reikšmės skirtingiems ARIMA modelio parametrams.
 Prognozuojamas išorinis regresorius: USA_INF / RUS_INF.

AIC								BIC							
p	q	d=0	d=1	d=2	d=3	d=4	d=5	p	q	d=0	d=1	d=2	d=3	d=4	d=5
0	0	-100	-231	-403	-368	-317	-276	0	0	-95	-228	-400	-364	-313	-272
0	1	-135	-266	-401	-383	-340	-301	0	1	-129	-261	-396	-378	-335	-296
0	2	-167	-296	-399	-381	-348	-320	0	2	-160	-290	-393	-375	-342	-314
0	3	-197	-316	-397	-379	-347	-325	0	3	-188	-308	-390	-371	-339	-317
0	4	-224	-332	-396	-377	-345	-323	0	4	-214	-323	-386	-368	-335	-314
0	5	-247	-341	-394	-375	-343	-321	0	5	-235	-331	-383	-364	-332	-311
1	0	-230	-418	-401	-374	-330	-294	1	0	-224	-412	-396	-369	-325	-289
1	1	-55	-417	-401	-381	-344	-312	1	1	-47	-411	-394	-375	-337	-306
1	2	-294	-416	-399	-379	-347	-323	1	2	-285	-410	-391	-371	-339	-315
1	3	-264	-415	-397	-377	-344	-323	1	3	-254	-409	-387	-368	-335	-314
1	4	-323	-414	-395	-375	-343	-321	1	4	-311	-408	-384	-364	-332	-311
1	5	-300	-413	-393	-373	-341	-319	1	5	-286	-407	-381	-361	-328	-307
2	0	-406	-412	-399	-375	-334	-303	2	0	-412	-401	-393	-368	-328	-296
2	1	-54	-411	-399	-379	-344	-316	2	1	-45	-404	-391	-371	-336	-308
2	2	-52	-410	-43	-377	-345	-322	2	2	-41	-394	-34	-368	-335	-313
2	3	-418	-409	-395	-375	-343	-321	2	3	-405	-391	-384	-364	-332	-311
2	4	-402	-408	-393	-373	-341	-319	2	4	-389	-387	-381	-361	-328	-307
2	5	-412	-407	-391	-371	-339	-317	2	5	-397	-384	-377	-358	-325	-304
3	0	-418	-406	-397	-374	-336	-307	3	0	-409	-398	-390	-367	-328	-299
3	1	-53	-403	-397	-377	-343	-317	3	1	-42	-394	-387	-368	-334	-308
3	2	-50	-401	-395	-375	-343	-321	3	2	-38	-391	-384	-364	-332	-310
3	3	-412	-399	-393	-373	-341	-319	3	3	-412	-387	-381	-361	-328	-307
3	4	-410	-45	-391	-371	-339	-317	3	4	-411	-31	-377	-358	-325	-304
3	5	-408	-395	-389	-369	-337	-315	3	5	-410	-381	-374	-354	-322	-300
4	0	-368	-403	-395	-373	-337	-309	4	0	-410	-394	-386	-364	-327	-300
4	1	-53	-401	-395	-375	-341	-317	4	1	-41	-391	-384	-364	-331	-306
4	2	-49	-399	-393	-373	-341	-319	4	2	-35	-387	-381	-361	-328	-307
4	3	-50	-397	-391	-371	-339	-317	4	3	-35	-384	-378	-358	-325	-304
4	4	-414	-395	-389	-369	-337	-315	4	4	-398	-380	-374	-354	-322	-300
4	5	-47	-393	-387	-367	-335	-313	4	5	-29	-377	-371	-351	-318	-297
5	0	-51	-401	-394	-372	-336	-311	5	0	-39	-391	-383	-361	-326	-301
5	1	-412	-399	-393	-373	-340	-316	5	1	-398	-387	-381	-361	-328	-304
5	2	-47	-397	-391	-371	-339	-317	5	2	-32	-384	-377	-358	-325	-304
5	3	-48	-44	-389	-369	-337	-315	5	3	-31	-29	-375	-354	-322	-300
5	4	-46	-393	-387	-367	-335	-313	5	4	-29	-377	-371	-351	-318	-297
5	5	-404	-42	-385	-365	-333	-311	5	5	-385	-24	-368	-347	-315	-293

Prognozuojant išorinį regresorių USA_INF / RUS_INF:

- parinktas ARIMA parametų rinkinys $(p, d, q) = (1, 1, 0)$;
- su šiuo parametų rinkiniu Ljung-Box testas parodė, kad liekanos yra nekoreliuotos.

7.5.6 lentelė. AIC ir BIC kriterijų reikšmės skirtingiems ARIMA modelio parametrams.
 Prognozuojamas išorinis regresorius: USA_UE / RUS_UE.

		AIC						BIC							
p	q	d=0	d=1	d=2	d=3	d=4	d=5	p	q	d=0	d=1	d=2	d=3	d=4	d=5
0	0	-165	-283	-390	-337	-286	-247	0	0	-160	-280	-370	-334	-283	-243
0	1	-199	-313	-368	-352	-308	-272	0	1	-193	-309	-364	-347	-303	-267
0	2	-231	-332	-366	-350	-315	-289	0	2	-223	-325	-360	-344	-309	-283
0	3	-257	-341	-364	-348	-313	-293	0	3	-248	-333	-357	-340	-305	-286
0	4	-277	-347	-362	-346	-311	-291	0	4	-266	-338	-353	-337	-302	-282
0	5	-293	-352	-360	-344	-309	-289	0	5	-281	-341	-350	-333	-298	-279
1	0	79	80	-368	-343	-298	-265	1	0	85	85	-370	-338	-293	-260
1	1	-293	81	-366	-350	-311	-282	1	1	-285	87	-360	-344	-305	-276
1	2	-339	77	-364	-348	-313	-291	1	2	-330	84	-357	-340	-305	-284
1	3	-277	74	-362	-346	-311	-291	1	3	-266	83	-353	-337	-302	-282
1	4	-344	-370	-360	-344	-309	-289	1	4	-332	-359	-350	-333	-298	-279
1	5	-368	-368	-358	-342	-307	-287	1	5	-354	-356	-346	-330	-295	-275
2	0	73	79	-366	-344	-303	-273	2	0	81	85	-360	-337	-296	-267
2	1	75	-374	-364	-348	-311	-285	2	1	84	-366	-357	-340	-303	-278
2	2	77	-372	-362	-346	-311	-291	2	2	87	-363	-353	-337	-302	-281
2	3	76	-370	-360	-344	-309	-289	2	3	88	-359	-350	-333	-298	-279
2	4	-390	-368	-358	-342	-307	-287	2	4	-369	-356	-346	-330	-295	-275
2	5	-372	-366	-356	-340	-305	-285	2	5	-357	-352	-343	-327	-291	-272
3	0	75	80	-364	-343	-304	-277	3	0	84	88	-357	-336	-296	-269
3	1	76	76	-362	-346	-310	-286	3	1	87	85	-353	-337	-300	-277
3	2	79	-370	-360	-344	-309	-289	3	2	91	-359	-350	-333	-298	-278
3	3	-390	-368	-358	-342	-307	-287	3	3	-367	-356	-346	-330	-295	-275
3	4	-378	-366	-356	-340	-305	-285	3	4	-364	-352	-343	-327	-291	-272
3	5	-376	-364	-354	-338	-303	-283	3	5	-360	-349	-340	-323	-288	-268
4	0	76	79	-362	-342	-304	-279	4	0	87	89	-353	-333	-295	-270
4	1	-337	-370	-360	-344	-308	-286	4	1	-325	-359	-350	-333	-297	-275
4	2	-381	-368	-358	-342	-307	-287	4	2	-367	-356	-346	-330	-295	-275
4	3	-367	-366	-356	-340	-305	-285	4	3	-352	-352	-343	-327	-291	-272
4	4	-374	-364	-354	-338	-303	-283	4	4	-358	-349	-340	-323	-288	-268
4	5	-345	-362	-352	-336	-301	-281	4	5	-328	-345	-336	-320	-285	-265
5	0	78	70	-360	-341	-304	-281	5	0	90	80	-350	-330	-293	-270
5	1	-299	-368	-358	-342	-306	-285	5	1	-286	-356	-346	-330	-294	-273
5	2	-377	-366	-356	-340	-305	-285	5	2	-363	-352	-343	-327	-291	-272
5	3	-377	-364	-354	-338	-303	-283	5	3	-360	-349	-340	-323	-288	-268
5	4	-375	-362	-352	-336	-301	-281	5	4	-357	-345	-336	-320	-285	-265
5	5	-372	-360	-350	-334	-299	-279	5	5	-353	-342	-333	-316	-281	-262

Prognozuojant išorinį regresorių USA_UE / RUS_UE:

- parinktas ARIMA parametų rinkinys $(p, d, q) = (0, 2, 0)$;
- su šiuo parametų rinkiniu Ljung-Box testas parodė, kad liekanos yra nekoreliuotos.

7.5.7 lentelė. AIC ir BIC kriterijų reikšmės skirtingiems ARIMA modelio parametrams.
 Prognozuojamas išorinis regresorius: RUS_IM / RUS_EX.

		AIC						BIC							
p	q	d=0	d=1	d=2	d=3	d=4	d=5	p	q	d=0	d=1	d=2	d=3	d=4	d=5
0	0	-222	-293	-337	-305	-257	-218	0	0	-217	-289	-334	-302	-254	-214
0	1	-254	-316	-335	-320	-281	-244	0	1	-248	-311	-330	-315	-276	-239
0	2	-278	-326	-333	-318	-292	-264	0	2	-270	-320	-327	-312	-285	-258
0	3	-300	-330	-331	-316	-290	-270	0	3	-291	-322	-323	-309	-282	-262
0	4	-307	-333	-329	-314	-288	-268	0	4	-297	-323	-320	-305	-279	-259
0	5	-321	-333	-327	-312	-286	-266	0	5	-309	-322	-317	-302	-275	-255
1	0	-230	-345	-335	-311	-270	-236	1	0	-224	-340	-330	-306	-266	-232
1	1	-55	-678	-333	-318	-286	-255	1	1	-47	-665	-327	-312	-280	-249
1	2	-294	-610	-331	-316	-290	-267	1	2	-285	-598	-323	-309	-282	-259
1	3	-338	-575	-329	-314	-288	-268	1	3	-327	-564	-320	-305	-279	-259
1	4	-336	-338	-327	-312	-286	-266	1	4	-324	-327	-317	-302	-275	-255
1	5	-343	-336	-325	-310	-284	-264	1	5	-337	-324	-313	-298	-272	-252
2	0	-420	-343	-333	-312	-276	-245	2	0	-412	-337	-327	-306	-269	-239
2	1	-54	-342	-331	-316	-286	-260	2	1	-45	-334	-323	-309	-279	-252
2	2	-343	-340	-329	-314	-288	-267	2	2	-332	-331	-320	-305	-279	-258
2	3	-345	-337	-327	-312	-286	-266	2	3	-333	-327	-317	-302	-275	-255
2	4	-343	-336	-325	-310	-284	-264	2	4	-329	-324	-313	-298	-272	-252
2	5	-341	-334	-323	-308	-282	-262	2	5	-326	-320	-310	-295	-268	-248
3	0	-418	-341	-331	-312	-278	-250	3	0	-409	-334	-323	-304	-270	-242
3	1	-347	-340	-329	-314	-285	-261	3	1	-336	-331	-320	-305	-276	-252
3	2	-344	-338	-327	-312	-286	-265	3	2	-332	-328	-317	-302	-275	-255
3	3	-343	-336	-325	-310	-284	-264	3	3	-329	-324	-313	-298	-272	-252
3	4	-338	-334	-323	-308	-282	-262	3	4	-323	-321	-310	-295	-268	-248
3	5	-339	-332	-321	-306	-280	-260	3	5	-323	-317	-306	-291	-265	-245
4	0	-368	-340	-329	-311	-278	-252	4	0	-358	-330	-320	-302	-269	-243
4	1	-345	-338	-327	-312	-284	-261	4	1	-333	-328	-317	-302	-273	-250
4	2	-343	-336	-325	-310	-284	-264	4	2	-329	-324	-313	-298	-272	-252
4	3	-341	-334	-323	-308	-282	-262	4	3	-326	-321	-310	-295	-268	-248
4	4	-339	-332	-321	-306	-280	-260	4	4	-323	-317	-306	-291	-265	-245
4	5	-337	-330	-319	-304	-278	-258	4	5	-319	-314	-303	-288	-261	-241
5	0	-51	-338	-327	-309	-278	-255	5	0	-39	-327	-317	-299	-268	-244
5	1	-343	-336	-325	-310	-282	-260	5	1	-329	-324	-313	-298	-270	-248
5	2	-341	-334	-323	-308	-282	-262	5	2	-326	-321	-310	-295	-268	-248
5	3	-339	-332	-321	-306	-280	-260	5	3	-323	-317	-306	-291	-265	-245
5	4	-337	-330	-319	-304	-278	-258	5	4	-319	-314	-303	-288	-262	-241
5	5	-335	-328	-317	-302	-276	-256	5	5	-316	-311	-299	-285	-259	-238

Prognozuojant išorinį regresorių RUS_IM / RUS_EX:

- parinktas ARIMA parametų rinkinys $(p, d, q) = (1, 1, 1)$;
- su šiuo parametų rinkiniu Ljung-Box testas parodė, kad liekanos yra nekoreliuotos.

7.5.8 lentelė. AIC ir BIC kriterijų reikšmės skirtingiems ARIMA modelio parametrams.
 Prognozuojamas išorinis regresorius: USA_IM / USA_EX.

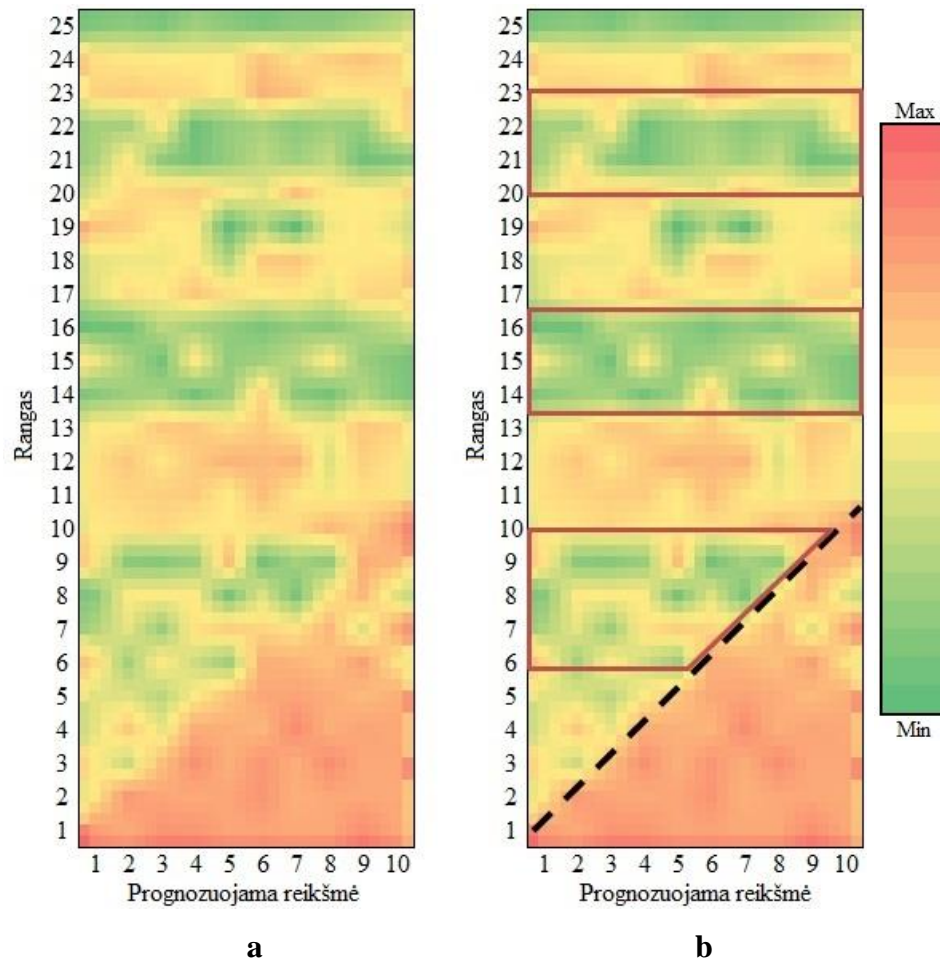
AIC								BIC							
p	q	d=0	d=1	d=2	d=3	d=4	d=5	p	q	d=0	d=1	d=2	d=3	d=4	d=5
0	0	-194	-284	-359	-303	-254	-216	0	0	-189	-280	-332	-300	-251	-212
0	1	-226	-308	-356	-318	-277	-241	0	1	-220	-303	-329	-313	-272	-236
0	2	-252	-320	-354	-316	-285	-260	0	2	-244	-314	-328	-310	-279	-253
0	3	-277	-325	-352	-314	-283	-264	0	3	-268	-317	-326	-307	-276	-257
0	4	-289	-328	-351	-312	-281	-262	0	4	-278	-319	-323	-303	-272	-253
0	5	-302	-328	-349	-310	-279	-260	0	5	-290	-318	-321	-300	-269	-250
1	0	-266	-316	-348	-309	-267	-234	1	0	-261	-311	-319	-304	-262	-229
1	1	-283	-323	-347	-316	-281	-252	1	1	-277	-317	-316	-310	-275	-246
1	2	-282	-321	-345	-314	-283	-262	1	2	-275	-314	-314	-307	-276	-255
1	3	-283	-319	-344	-312	-281	-262	1	3	-274	-310	-311	-303	-272	-253
1	4	-278	-317	-342	-310	-279	-260	1	4	-268	-307	-310	-300	-269	-250
1	5	-279	-315	-341	-308	-277	-258	1	5	-267	-303	-309	-296	-265	-246
2	0	-271	-317	-340	-310	-272	-242	2	0	-264	-311	-306	-304	-265	-236
2	1	-283	-321	-338	-314	-281	-256	2	1	-275	-314	-305	-307	-273	-248
2	2	-280	-319	-336	-312	-281	-262	2	2	-271	-310	-304	-303	-272	-252
2	3	-281	-317	-335	-310	-279	-260	2	3	-271	-307	-302	-300	-269	-250
2	4	-276	-315	-334	-308	-277	-258	2	4	-264	-303	-299	-296	-265	-246
2	5	-274	-313	-332	-306	-275	-256	2	5	-261	-300	-296	-293	-262	-243
3	0	-275	-317	-330	-310	-273	-247	3	0	-268	-309	-294	-302	-266	-239
3	1	-341	-319	-327	-312	-280	-257	3	1	-330	-310	-293	-303	-271	-248
3	2	-336	-317	-326	-310	-279	-260	3	2	-324	-307	-291	-300	-269	-249
3	3	-332	-315	-325	-308	-277	-259	3	3	-319	-303	-288	-296	-265	-246
3	4	-330	-313	-322	-306	-275	-256	3	4	-315	-300	-287	-293	-262	-243
3	5	-327	-311	-319	-304	-273	-254	3	5	-310	-296	-285	-289	-259	-240
4	0	-325	-316	-317	-309	-274	-249	4	0	-309	-307	-284	-300	-265	-240
4	1	-349	-317	-314	-310	-278	-257	4	1	-330	-307	-283	-300	-268	-246
4	2	-337	-315	-312	-308	-277	-258	4	2	-324	-303	-281	-296	-265	-246
4	3	-335	-313	-311	-306	-275	-257	4	3	-320	-300	-279	-293	-262	-243
4	4	-333	-311	-308	-304	-273	-255	4	4	-317	-296	-278	-289	-259	-240
4	5	-331	-309	-305	-302	-271	-253	4	5	-313	-293	-275	-286	-255	-236
5	0	-339	-314	-304	-307	-274	-251	5	0	-327	-304	-274	-297	-263	-240
5	1	-344	-315	-303	-308	-277	-256	5	1	-330	-303	-272	-296	-265	-244
5	2	-335	-313	-302	-306	-276	-256	5	2	-320	-300	-270	-293	-262	-243
5	3	-333	-311	-300	-304	-273	-255	5	3	-317	-296	-269	-289	-259	-240
5	4	-331	-309	-297	-302	-271	-253	5	4	-313	-293	-267	-286	-255	-236
5	5	-329	-325	-315	-300	-269	-251	5	5	-310	-307	-266	-283	-252	-233

Prognozuojant išorinį regresorių USA_IM / USA_EX:

- parinktas ARIMA parametų rinkinys $(p, d, q) = (0, 2, 0)$;
- su šiuo parametų rinkiniu Ljung-Box testas parodė, kad liekanos yra nekoreliuotos.

7.6. Ankstesnių rango tyrimo rezultatai

Ankstesnių tyrimų metu buvo tyrinėjama rango priklausomybė nuo prognozuojamų narių kiekio prognozuojant iki dešimt reikšmių į ateitį su rangais nuo 1 iki 25. Eksperimentai kartoti 100 kartų, pradedant iš laiko eilutės reikšmių intervale [1,10]. Tyrimams buvo naudojama laiko eilutė aprašyta 4 skyrelyje. Bendrasis rezultatas buvo gautas paskaičiuojant vidurkį.



7.6.1 pav. Rango ir prognozuojamų žingsnių priklausomybė išreikšta per modelio paklaidos RMSE reikšmę.

Tyrime buvo pastebėta:

- kad naudojant mažesnę rangą nei prognozuojamų žingsnių kiekis gaunamos didelės RMSE paklaidos (žiūrėti 7.6.1 paveikslo b dalį žemiau juodos brūkšninės linijos);
- modelis veikia pakankamai adekvačiai, nes naudojant dirbtinę eilutę, kurios rangas yra 7, mažiausios RMSE reikšmės pasiektos tie šios sekos kartotinais arba artimais jiems skaičiais, tai yra ties rangų 7, 14, 21 ir jiems artimais skaičiais (žiūrėti 7.6.1 paveikslo b dalį, ir zonas apvestas raudona kontūrine linija).