

**VILNIAUS UNIVERSITETAS**  
**MATEMATIKOS IR INFORMATIKOS FAKULTETAS**

Magistro darbas

**FINANSINIŲ TRUKMIŲ MODELIAVIMAS**

**Fedotenkov Igor**

**VILNIUS 2007**

**Ekonometrinės analizės katedra**

Darbo vadovas **prof. habil. dr. R. Leipus** \_\_\_\_\_  
(parašas)

Darbas apgintas 2007 m. birželio mėn. 04 d.

Registravimo Nr. \_\_\_\_\_  
2007-05-24 \_\_\_\_\_

## Summary

Nowadays, many researches are made in ultra high frequency data series. Considering the data in time intervals as it arrives, helps to understand a variety of issues relating to trading process and market microstructure. The empirical analyses of such data present a number of new and unique statistical challenges.

This work considers the dynamics of time intervals between transactions of one share in London Stock Exchange. The aim of this work is to make a statistically relevant model suitable for analysis of time intervals, and making the forecasts. First, some classical ACD-type models were tried to apply for a description of the data under consideration. After, a more sophisticated model using nonlinear functions was proposed in order to manage the problems arising in the classical models. More over, the model enables different dynamics of intervals between trades, depending on last price change. Comparing with the classical methods, the proposed model much better describes nonlinear dependence existing in the data. It enhances a goodness of fit, and is more applicable for practical use. Theoretical features such as an existence of equilibrium, stationarity and ergodicity were considered under specific functional form. But the idea can be extended for more general classes of models.

## Santrauka

Šiais laikais atliekama daug tyrimų ultra aukšto dažnio duomenų srityje. Nagrinėjant duomenis jų pasirodymo metu, galima suprasti daugelį įvairių procesų, vykstančių prekyboje, bei geriau išnagrinėti rinkos mikrostruktūrą. Be to, ultra aukšto dažnio duomenys pasižymi daugeliu naujų ir unikalių savybių.

Šis darbas yra skirtas laiko intervalų tarp sandorių dinamikai aprašyti. Analizei buvo paimti vienos akcijos, kuria prekiaujama Londono vertybinių popierių biržoje, duomenys. Pagrindinis šio darbo tikslas - sukurti modelį, tinkamą duomenų analizei ir prognozavimui. Iš pradžių buvo bandoma pritaikyti kai kuriuos klasikinius ACD tipo modelius. Atsižvelgus į jų trūkumus, buvo pasiūlytas sudėtingesnis modelis, leidžiantis skirtingą trukmių dinamiką, priklausomai nuo kainų pokyčių. Į modelio specifikaciją įtraukus netiesinę funkciją, modelis geriau aprašo sąryšius, kurie egzistuoja duomenyse, padidėja modelio lankstumas, todėl jis gerai tinka praktiniam panaudojimui. Teorinės modelio savybės, tokios kaip pusiausvyros egzistavimas, stacionarumas ir ergodiškumas, buvo išnagrinėtos konkrečiai modelio funkciniai formai, bet panaudoti metodai gali būti pritaikyti ir bendresnėms modelių klasėms.

# Turinys

1.	Įvadas .....	2
1.1.	Ultra aukšto dažnio laiko eilutės .....	2
1.2.	Trukmių analizės modeliai.....	2
1.3.	Trumpa literatūros apžvalga .....	3
1.4.	Duomenys .....	3
2.	Metodologija .....	4
2.1.	Duomenų paruošimas.....	4
2.2.	ACD modelis .....	4
2.3.	TACD modelis .....	5
2.4.	Modelio vertinimas .....	5
3.	Preliminari duomenų analizė .....	7
3.1.	Aprašomosios duomenų statistikos.....	7
3.2.	Sezoniškumo išskyrimas .....	9
4.	Klasikinių modelių rezultatai .....	12
4.1.	EACD(1,1) .....	12
4.2.	TACD.....	12
5.	Slenkstinis modelis su logaritmine funkcija .....	15
5.1.	Įvertintas modelis.....	15
5.2.	Pusiausvyros egzistavimas ir ergodiškumas .....	16
5.3.	Neigiamų reikšmių neegzistavimas. ....	17
5.4.	Impulso atsakas ir trumpa atmintis .....	18
5.5.	Stacionarumas .....	21
5.6.	Liekanų testavimas.....	21
6.	Prognozavimas .....	23
	Literatūra.....	25
	Priedas 1.....	26
	Priedas 2.....	28

# 1. Įvadas

## 1.1. Ultra aukšto dažnio laiko eilutės

Šiuolaikinės technologijos leidžia stebėti ir kaupti informaciją realiu laiku, t.y. tokiu laiku, kai ji pasirodo. Tokie duomenys yra vadinami ultra aukšto dažnio duomenimis. Pastarieji pasižymi kai kuriomis unikaliomis savybėmis, kuriomis nepasižymi žemo dažnio duomenys. Perėjimas nuo žemo dažnio duomenų analizės prie aukšto dažnio padeda geriau išnagrinėti rinkos mikrostruktūrą, patikrinti rinkos efektyvumo hipotezes bei galbūt sudaryti tikslesnes prognozes.

Egzistuoja keli būdai tokio tipo informacijai tirti: pirma, galima agreguoti duomenis į vienodo dydžio laiko intervalus ir jiems taikyti standartinius laiko eilučių metodus. Tačiau, nagrinėjant duomenis šiuo metodu, prarandama daug svarbios informacijos, be to, prielaida apie laiko intervalų lygybę nėra visada teisinga. Pavyzdžiui, tiriant vertybinių popierių kainų dinamiką natūralu nagrinėti paskutines nagrinėjamų laiko intervalų sandorių kainas. Tačiau, priklausomai nuo paskutinio sandorio momento, laiko trukmė tarp stebėjimų irgi gali būti skirtinga.

Kitas būdas - nagrinėti modelį, kuris aprašytų norimų rodiklių dinamiką realiu laiku. Savime aišku, kad viena pagrindinių komponentų tokia modelyje bus laiko intervalas tarp stebėjimų. Pavyzdžiui, jeigu mes nagrinėjame vertybinių popierių kainas, tada neužtenka nustatyti, kaip pasikeis kaina, taip pat reikia nustatyti ir kada ji pasikeis. Skirtumas tarp to, ar kaina pasikeis po kelių sekundžių, ar po kelių savaičių nėra mažas.

Šis darbas yra skirtas trukmių tarp sandorių analizei. Tai yra sudedamoji projekto dalis, kuriame be trukmių tarp sandorių yra modeliuojamos vertybinių popierių sandorių kainos, geriausios pasiūlos ir paklausos kainos (angl. „ask“ ir „bid“) bei kita užsakovą dominanti informacija. Iš pradžių mes apžvelgsim kai kuriuos klasikinius modelius, skirtus trukmių tarp sandorių modeliavimui, bei pažiūrėsime kokius jie duoda rezultatus mūsų duomenims. Vėliau bus pasiūlytas lankstesnis modelis, geriau pagaunantis netiesinius sąryšius, kurie egzistuoja tarp duomenų, ir išnagrinėtos jo teorinės savybės.

## 1.2. Trukmių analizės modeliai

Šiuo metu yra sukurta daug įvairių metodų, skirtų trukmių tarp transakcijų analizei. Patys populiariausi iš jų yra „Autoregresinių sąlyginių trukmių“ arba ACD modeliai (nuo angl. Autoregressive Conditional Duration). ACD tipo modelius pasiūlė R.F. Engle ir J.R. Russell [1]. Tokio tipo modeliai paprastai skiriasi prielaidomis apie liekanų pasiskirstymą bei netiesiškumo modeliavimo būdais. Kadangi ACD modelių struktūra yra labai panaši į GARCH, todėl dauguma GARCH modelio išplėtimų yra panaudoti ir ACD modeliuose. Šiame darbe didžiausią dėmesį skirsime būtent ACD tipo modeliams.

Alternatyvi modelių grupė yra PCD (angl. Price Change and Duration) modeliai. Šie modeliai yra šiek tiek paprastesni, kadangi jie skirti trukmių tarp kainų pasikeitimų analizei. Šiame modelyje paprastai nagrinėjamas vektoriaus  $\begin{bmatrix} \Delta p_i \\ \Delta t_i \end{bmatrix}$  sąlyginis pasiskirstymas, žinant informaciją iki  $i$ -tojo laikotarpio. Tarp nagrinėjamų rodiklių galimi simultaniniai sąryšiai. Čia  $\Delta p_i$  yra trukmė tarp kainos pasikeitimų,  $\Delta t_i$  - transakcijų kiekis tarp gretimų kainos pasikeitimų,  $\beta$  - kainos pasikeitimo kryptis, jis yra lygus  $-1$ , jei kaina kilo ir  $1$  jei kaina augo,  $\alpha$  - kainos pasikeitimas išreikštas „tikais“ (minimaliomis kainos pasikeitimo reikšmėmis, kurios yra leistinos vertybinių popierių biržoje).

Priklausomai nuo modeliavimo tikslų į analizę gali būti įtraukti ir kiti dominantys rodikliai. Didžiausias modelio privalumas - vieno modelio kontekste nagrinėjami keli pagrindiniai rodikliai, reikalingi finansinėms spekuliacijoms.

Trukmės tarp transakcijų gali būti nagrinėjamos ir ARMA modeliais, užtikrinus, kad trukmės negali įgyti neigiamų reikšmių. Tai gali būti padaryta nagrinėjant trukmių logaritmus arba įvedus apribojimus parametrų bei tinkamai pasirinkus liekanų pasiskirstymą.

Svarbu paminėti ir funkcinės analizės metodus. Pavyzdžiui, vietoj tiesioginės trukmių tarp transakcijų analizės galima nagrinėti, kiek transakcijų įvyko tam tikrais laiko tarpais, pvz. kas 5 minutes. Tokiu būdu įvertinus prekybos intensyvumą dienos bėgyje, galima gauti atsitiktinę funkciją ir nagrinėti tokių atsitiktinių funkcijų sekas. Mano manymu, tai yra perspektyviausias ultra-dažnių duomenų nagrinėjimo metodas.

### 1.3. Trumpa literatūros apžvalga

Kaip jau buvo minėta, 1998 metais pasirodė R. F. Engle ir J. R. Russell straipsnis „Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data“ [1], kuriame jie pritaikė GARCH modelio idėją trukmėms tarp transakcijų aprašyti. Vėliau, 2000 metais, pasirodė kitas Engle straipsnis [2], kuriame jis išnagrinėja teorines modelio savybes, pasiūlo modelio vertinimo algoritmą naudojant kvazitiktinumo funkciją bei programinę įrangą, skirtą GARCH modeliams. Vėliau pasirodė daug modelio išplėtimų ir variacijų. Pavyzdžiui, Zhang, Russell ir Tsay [3] pasiūlė kelių režimų ACD modelį, pavadintą TACD (Threshold ACD), o Meitz ir Terasvirta [4] pasiūlė modelio išplėtimą, kuriame parametrai tolydžiai kinta laike. 2006 metais Maria Pacurar straipsnyje „ACD models in Finance: A Survey of the Theoretical and Empirical Literature“ [5] apžvelgė ACD modelio išplėtimus bei variacijas.

Šiame darbe taip pat buvo panaudota informacija iš R.S. Tsay knygos „Analysis of Financial Time Series“ [14]. Testuojant modelį bei tiriant jo stacionarumo savybes buvo vadovaujamosi W.J. Granger ir T. Terasvirta knyga „Modelling Nonlinear Relationships“ [15]. Parenkant lietuvišką terminologiją paprastai buvo atsižvelgiama į V. Čekanavičiaus ir G. Murausko knygą „Statistika ir jos taikymai“ (1 dalis) [18]. Pilnas panaudotos literatūros sąrašas pateiktas darbo pabaigoje.

### 1.4. Duomenys

Iš viso analizei buvo prieinami duomenys apie 2000 skirtingų vertybinių popierių, kuriais prekiaujama didžiausiose Europos biržose: LSE, DAX, CAC ir kitose. Taupant laiką ir vietą, šiame darbe buvo panaudoti Londono biržos duomenys apie prekybą farmacijos įmonės GlaxoSmithKline (GSK) akcijomis nuo 2005-10-01 iki 2005-10-31. Daugelis kitų vertybinių popierių buvo išnagrinėti analogiškai, su nedideliais pakeitimais, atspindinčiais jų specifiką.

Matuojami rodikliai apima transakcijų įvykdymo ir registravimo laiką, sandorio kainą bei apimtį, geriausias pasiūlos ir paklausos kainas (angl. ask ir bid) bei kitą informaciją. Transakcijų laikas yra fiksuojamas sekundės tikslumu. Taigi, laikas tarp transakcijų yra išreikštas sekundėmis. Išmetus kai kuriuos akivaizdžiai klaidingus įrašus, turima informacija apie 64110 transakcijas. Kadangi Londono biržoje prekyba vertybiniais popieriais yra labai intensyvi, 57.8% trukmių tarp sandorių yra nulinės.

## 2. Metodologija

### 2.1. Duomenų paruošimas

Gerai žinomas faktas, kad prekybos intensyvumas biržoje pasižymi tam tikru dienos periodiškumu. Todėl trukmės tarp transakcijų yra nestacionarios. Tam, kad su duomenimis būtų galima dirbti, juos reikia atitinkamai transformuoti, t.y. eliminuoti šią deterministinę dalį. Daroma prielaida, kad deterministinė eilutės dalis paveikia stebėjimus multiplikatyviai.

$$\boxed{\phantom{0}} \cdot \quad (1)$$

Čia  $\boxed{\phantom{0}}$  yra laikas kada buvo įvykdyta  $i$ -toji transakcija, jis matuojamas sekundėmis nuo biržos atsidarymo momento, tada  $\boxed{\phantom{0}}$  yra trukmė tarp transakcijų,  $\boxed{\phantom{0}}$  yra tam tikra deterministinė dalis, o  $\boxed{\phantom{0}}$  yra transformuoti duomenys, tinkami analizei. Funkcija  $\boxed{\phantom{0}}$  galima įvertinti daugeliais metodais, dažniausiai ji yra vertinama splineais arba neparometriškai. R.S. Tsay knygoje „Analysis of Financial Time Series“ tiesiog pasirenkamos tam tikros deterministinės funkcijos, ir jomis mažiausių kvadratų metodu aproksimuojami duomenys. Šiame darbe funkcija  $\boxed{\phantom{0}}$  bus įvertinta neparometriškai.

### 2.2. ACD modelis

Kad būtų paprasčiau, pažymėkime  $\boxed{\phantom{0}}$ . Tegul  $\boxed{\phantom{0}}$ , čia  $\boxed{\phantom{0}}$  yra informacija, pasiekama iki  $i-1$  laikotarpio. ACD modelis yra apibrėžiamas taip:

$$\boxed{\phantom{0}} \cdot \quad (2)$$

$$\boxed{\phantom{0}} \cdot \quad (3)$$

Toks modelis vadinamas ACD(r,s). Turima omeny, kad paklaidos  $\boxed{\phantom{0}}$  įtakoja duomenis multiplikatyviai. Paprastai yra daroma prielaida, kad  $\boxed{\phantom{0}}$  yra teigiami, nepriklausomi bei vienodai pasiskirstę atsitiktiniai dydžiai.

Tuo atveju, kai daroma prielaida, kad paklaidos  $\boxed{\phantom{0}}$  turi eksponentinį pasiskirstymą, modelis vadinamas EACD(r,s). Jei  $\boxed{\phantom{0}}$  turi Veibulo pasiskirstymą, tada toks modelis vadinamas WACD(r,s). Kartais daroma prielaida, kad liekanos turi apibendrintą Gama pasiskirstymą, tada modelis yra vadinamas GACD(r,s). Teoriškai liekanos gali turėti bet kokią pasiskirstymą, kuris užtikrintų, jog lygtimis (2) ir (3) apibrėžtas procesas negalėtų įgyti neigiamų reikšmių. Tam pačiam tikslui yra padaromos prielaidos, kad koeficientai

$\boxed{\phantom{0}}$   $0, \dots, r, j=1, \dots, s$  yra neneigiami. Be to, tariama, kad  $\boxed{\phantom{0}}$ . Ši prielaida

yra reikalinga tam, kad pradėjus nuo kažkurių baigtinių reikšmių einant laikui procesas neaugtų į begalybę. Kaip ir GARCH modelį, ACD galima suvesti į ARMA pavidalą:





Čia,  $\theta$  yra modelio parametrų vektorius. Tankio funkcijos  $f(x_T|\theta)$  įtaka mažėja didėjant stebėjimų skaičiui. Todėl, paprastumo dėlei šis narys paprastai praleidžiamas. Pavyzdžiui, WACD modelio atveju log-tikėtinumo funkcija atrodo taip:

$$l(x_T|\theta) = \sum_{t=1}^T \ln f(x_t|\theta) \quad (10)$$

Didžiausio tikėtinumo funkcija apibendrinto Gama pasiskirstymo atveju atrodo dar komplikuočiau. Tačiau R.Engle, 2000 metų straipsnyje [2], pasiūlė vietoj šių funkcijų maksimizavimo, maksimizuoti kvazi-tikėtinumo funkciją:

$$l(x_T|\theta) = \sum_{t=1}^T \ln f(x_t|\theta) \quad (11)$$

Jei liekanos turi eksponentinį pasiskirstymą, tada tai yra tikroji log-tikėtinumo funkcija. Iš tikro ši idėja nėra nauja, dar 1984 metais Gourieroux, Monfort ir Trognon [6] įrodė, kad šios funkcijos maksimizavimas duoda suderintus parametrų įverčius, bei parametrų kovariacinę matricą. Modelio vertinimui tinka ir kvazi-tikėtinumo funkcija gauta iš normalinio tankio. Ji yra ypač patogi tuo, kad yra prieinama bet kuriuose paketuose, skirtuose GARCH modelio vertinimui. Prisiminsim, kaip atrodo GARCH(r,s) modelis:

$$\sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (12)$$

$$a_t = \sigma_t z_t \quad (13)$$

Pakėlus pirmą lygybę kvadratu, gauname formą, analogišką ACD modelio formai. Tada

$$\sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

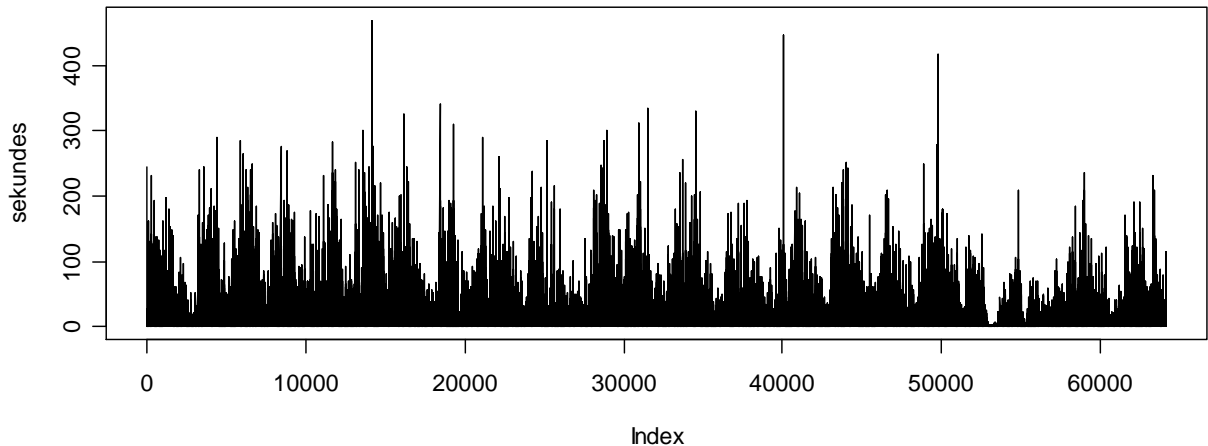
Taigi, pasiėmus  $\sigma_t^2$ , galima tiesiogiai taikyti GARCH modeliams skirtą programinę įrangą. Be to, šaknies panaudojimas garantuoja, kad netgi tuomet kai modelio vertinimo metu pasirinktas pasiskirstymas yra simetrinis 0 atžvilgiu,  $\sigma_t^2$  negalės įgyti neigiamų reikšmių. Jeigu modelis yra teisingai specifiukuotas, tam kad gautumėm suderintus koeficientų įverčius, reikia kad paklaidos turėtų antrą momentą. Tam, kad būtų suderinta koeficientų kovariacinė matrica, reikia kad egzistuotų ketvirtas momentas. (J. R. Magnus 2007 [7])

TACD atveju log-tikėtinumo funkcijų išraiška yra ta pati, tik sumų nariai dauginami iš vienetinio indikatoriaus  $1(x_t > 0)$ . Tiesa, tikėtinumo funkcija dažnai yra netolydi parametrų, nusakančių aibės  $\mathbb{R}^+$  režius, atžvilgiu, todėl tai daro jos maksimizavimą sudėtingesniu. Šios funkcijos maksimizavimui paprastai yra naudojami specialūs optimizavimo algoritmai, galintys maksimizuoti netolydžias funkcijas, arba, kaip savo knygoje pasielgė Tsay [14], galima šiuos režius generuoti atsitiktinai, bei parinkti geriausius.

### 3. Preliminari duomenų analizė

#### 3.1. Aprašomosios duomenų statistikos

Kaip jau buvo minėta, analizei mes pasirinkom 21 darbo dienas GSK akcijų duomenis. Jie yra pavaizduoti paveiksle XXX.

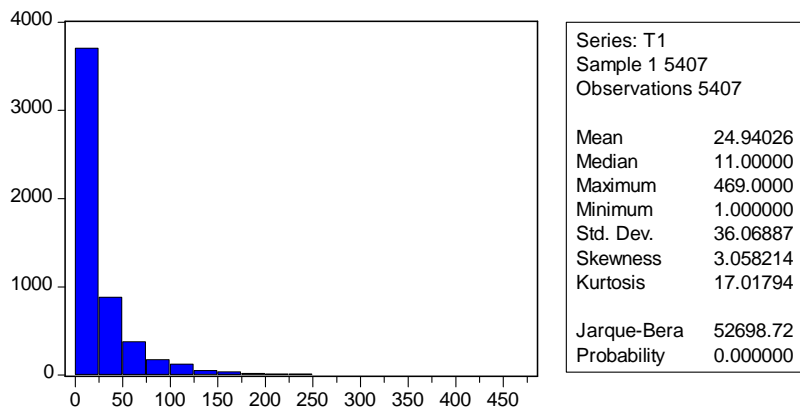


1 pav.: trukmės tarp sandorių (viena po kitos)

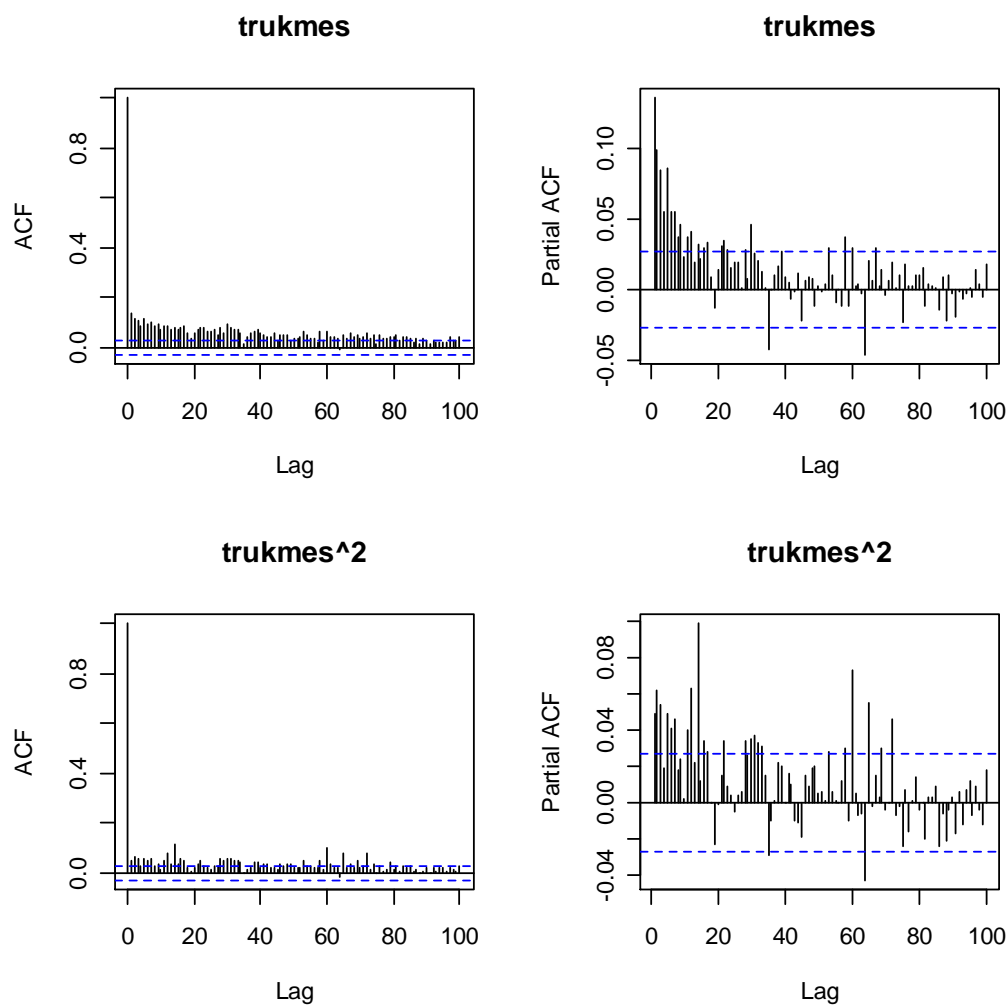
Šiame paveiksle trukmės tarp sandorių yra užrašytos viena po kitos. Kaip matome, netgi „iš akies“ paveiksle galima atskirti, kur pasibaigia viena diena ir prasideda kita. Duomenys turi tam tikrą struktūrą. Dienos bėgyje trukmių grafikas turi apverstą U formą. Dienos pradžioje ir dienos pabaigoje prekyba yra intensyvesnė, o dienos viduryje trukmė tarp sandorių kartais pasiekia net 7 minutes. Tokia struktūra paaiškinama tuo, kad biržos atidarymo metu prekybos dalyviai „atidirba“ informaciją susikaupusią naktį, be to, kai kurie „treideriai“, nenorėdami rizikuoti naktį, perka vertybinius popierius ryte ir parduoda akcijas biržos uždarymo metu. Todėl biržos atidarymo ir uždarymo valandomis prekyba yra labai intensyvi, o dienos viduryje, per pietus, trukmė tarp transakcijų yra sąlyginai didelė. Toks prekybos intensyvumo pasiskirstymas dienos bėgyje yra vadinamas „dienos sezoniškumu“.

Antrajame paveiksle yra pavaizduota nenulinių trukmių histograma. Kaip matome, duomenys yra susikongravę ties nuliu. Pirmas kvantilis yra lygus 3, mediana 11, o trečias kvantilis yra lygus 31 sekundei. Maksimali užregistruota trukmė nagrinėjamu laikotarpiu buvo lygi 469 sekundėm, t.y. truputi mažiau nei 8 minutės.

Trukmių autokoreliacijos ir dalinės autokoreliacijos (3 paveikslas) funkcijos parodo, jog tarp duomenų yra priklausomumas. Labai lėtai gęstanti autokoreliacinė funkcija liudija apie tam tikrą duomenų nestacionarumą.



2 pav.: Trukmių tarp sandorių histograma ir kai kurios aprašomosios statistikos



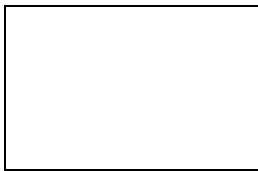
3 pav.: Trukmių tarp sandorių ir jų kvadratų autokoreliacijos ir dalinės autokoreliacijos.

Trukmių kvadratų autokoreliacijos irgi yra reikšmingos, tai parodo, jog duomenų dispersija nėra pastovi, taip pat šis faktas gali liudyti apie tam tikrų netiesinių sąryšių egzistavimą.

### 3.2. Sezoniškumo išskyrimas

Kaip jau buvo minėta, trukmė tarp transakcijų dienos bėgyje pasižymi tam tikru dėsningumu, vadinamu dienos sezoniškumu. Kadangi skirtingos darbo dienos gali pasižymėti skirtingu dėsningumu, skirtingas darbo dienas nagrinėsime atskirai. Užbėgant į priekį, reikėtų paminėti, jog mes nagrinėsime tik tas transakcijas, kurios įvyko skirtingom sekundėm. Todėl sezoniškumą vertinsim tik toms transakcijoms, tarp kurių trukmė nėra nulinė. Iš tikro, nėra didelio skirtumo, ar imtis suskaidoma į nulines ir nenulines transakcijas iki, ar po sezoniškumo koregavimo. Gautų pakoreguotų nenulinių transakcijų imčių koreliacija yra 0.996, o nulinės transakcijos bet kuriuo atveju išlieka nulinėmis.

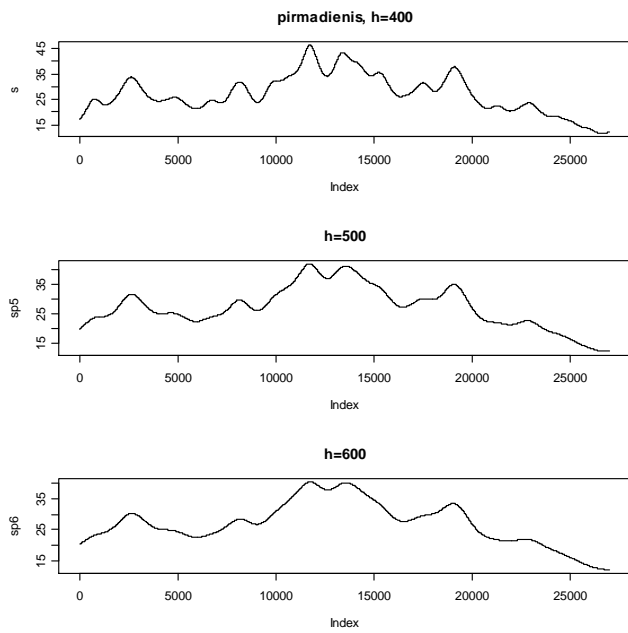
Darbo dieną (nuo 8.00 iki 15.30) suskaidėme į 27000 sekundžių, ir kiekvienai sekundei įvertinom sezoniškumo funkcijos reikšmę. Tai buvo padaryta neparometriškai:



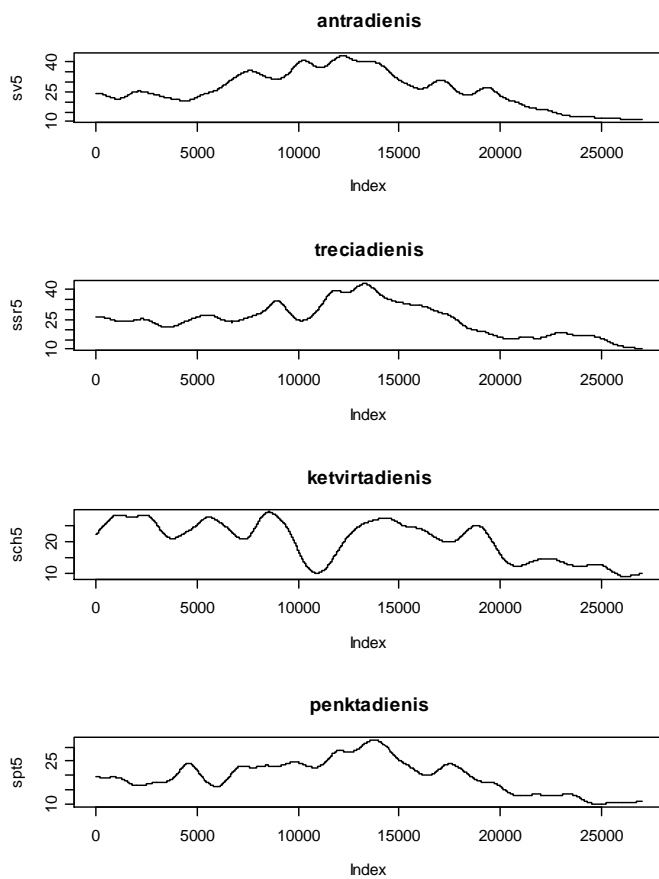
(14)

buvo pasirinkta standartinio normalaus tankio funkcija. Iš tikro, mūsų atveju nėra būtina, kad sezoniškumo funkcija būtų be galo daug kartų diferencijuojama, todėl galima pasirinkti ir kitoki, paprastesnį branduolį. Tačiau nuo pasirinktos branduolinės funkcijos gauti rezultatai beveik nesikeitė, todėl nėra svarbu, kuri branduolį pasirinkti. Daug svarbesnis yra lango pločio  pasirinkimas. Lango plotis buvo pasirinktas bandymų metu. Ketvirtame grafike yra pavaizduotos dienos sezoniškumo funkcijos skirtingiems . Tolimesnei analizei buvo pasirinktas , kadangi, kai , tada atsiranda svyravimai, kurie gali būti atsitiktiniai. Kai , gauta funkcija yra truputi glodesnė, bet ne daug skiriasi nuo  atvejo.

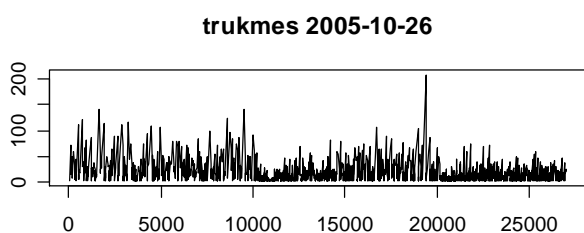
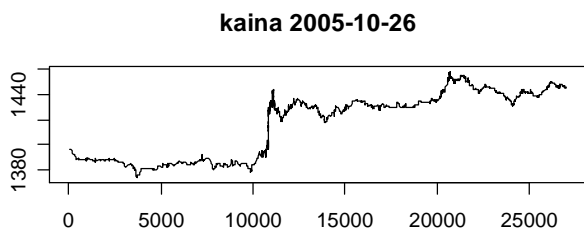
Mūsų duomenyse yra 5 pirmadieniai ir po 4 kitas darbo dienas. Kaip matome visos, išskyrus ketvirtadienių, sezoniškumo funkcijos turi panašų pavidalą. Šiek tiek skiriasi skirtingų darbo dienų mastelis. Intensyviausia prekyba yra ketvirtadienį ir penktadienį, o trečiadieniais ir ketvirtadieniais investuotojai yra mažiausiai aktyvūs. Šiek tiek keistai atrodo ketvirtadienio funkcija. Kadangi vertinimui buvo panaudotos tik 4 dienos, viena „ypatinga“ diena galėjo iškraipyti rezultatus. Iš tikro, jeigu pažiūrėsime į 2005 spalio 26 dieną, galima pastebėti, kad ji elgėsi gana keistai (6 paveikslas). Ji pasižymi labai aktyvia prekyba dienos viduryje, kai pasirodė įmonės veiklos ataskaita, ir akcijos kaina staigiai kilo. Iš tolimesnės analizės ši diena buvo pašalinta.



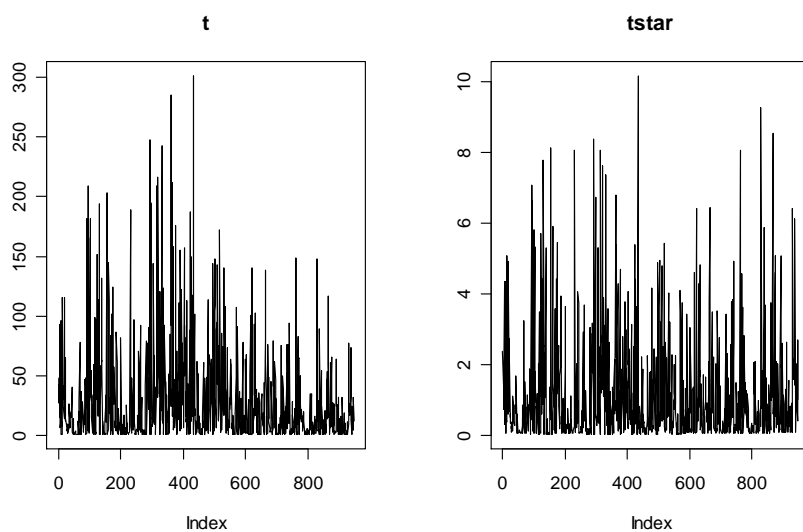
4 pav.: Dienos sezoniškumo funkcijos įvertiniai skirtingiems  $h$



5 pav.: Skirtingų darbo dienų įvertintos sezoniškumo funkcijos.



6 pav.: „Ypatinga“ diena 2005 Spalio 26.



7 pav.: Vienos dienos duomenys iki ir po sezoniškumo išlyginimo.

Pasinaudojus formule (14) buvo išlygintas dienos sezoniškumas. 7 paveiksle yra pavaizduotos vienos dienos spalio 17d.) trukmės tarp transakcijų iki ir po dienos sezoniškumo išlyginimo. Kaip matome, pakoreguoti duomenys jau nebeturi ryškios varpo struktūros.

## 4. Klasikinių modelių rezultatai

### 4.1. EACD(1,1)

Iš 5 paveikslo matome, jog savaitės dienos pasižymi skirtingu dienos sezoniškumu. Galima tikėtis, jog skirtingoms savaitės dienoms galios skirtingi modeliai. Taupydami laiką ir vietą, šiame skyriuje modeliuosime tik pirmadienių duomenis. Iš viso pirmadienių duomenų yra 5407, iš jų 5000 panaudojome modelio vertinimui, o 407 palikome modelio testavimui, kad galėtume patikrinti modelio prognozavimo galimybes. Kitų savaitės dienų modelių sudarymas yra analogiškas. Jų rezultatai pateikti Priede 1. Prieš sudarydami kažkokį sudėtingesnį modelį, naudinga pažiūrėti paprastą ACD(1,1) modelį, pasiūlytą Engle ir Russell 1998 metais [1].

	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.054637	0.012399	4.406576	0.0000
GARCH(-1)	0.867184	0.020740	41.81276	0.0000
SQT(-1)^2	0.078904	0.011652	6.771534	0.0000
R-squared	-2.137808	Mean dependent var		0.823904
Adjusted R-squared	-2.138970	S.D. dependent var		0.563550
S.E. of regression	0.998449	Akaike info criterion		2.802879
Sum squared resid	5386.254	Schwarz criterion		2.806539
Log likelihood	-7573.181	Durbin-Watson stat		0.538604

1 lentelė: ACD(1,1) modelio parametrų įverčiai

Reikia pastebėti, kad beveik visos modelio charakteristikos kurias Eviews apskaičiuoja automatiškai (lentelės apačioje) neturi prasmės, kadangi jos yra apskaičiuotos GARCH modeliui, o ne ACD. Todėl, kitose lentelėse šios statistikos bus praleidžiamos.

$$\boxed{\phantom{0.054637}} \quad (16)$$

$$\boxed{\phantom{0.867184}} \quad (17)$$

Modelio tikslumas nėra labai didelis: koreliacija tarp modelio prognozių (angl. fitted values) ir tikrųjų reikšmių yra tik apie 0.04, be to modelio liekanos išlieka autokoreliuotos: Ljung-Box statistika  $Q(1)=64.413$ ,  $Q(30)=401.28$ ,  $Q(100)=561.02$ . 5% reikšmingumo lygmens kritinės reikšmės atitinkamai yra 3.84, 43.77 ir 124.34. Liekanų kvadratai bei kitos netiesinės funkcijos taip pat yra autokoreliuotos. Tai reiškia, kad iš duomenų visa informacija dar nėra ištraukta.

### 4.2. TACD

Geresnius rezultatus pateikia slenkstinis TACD modelis, pasiūlytas Zhang, Russell ir Tsay [3] Šis modelis yra aprašytas lygybėmis (2), (7) ir (8). Buvo pasirinkta įtraukti 3 režimus. Didžiausia problema yra ta, kad slenksčių  $\square$  ir  $\square$  atžvilgiu didžiausio

tikėtumo funkcija yra netolydi. Todėl modelis buvo vertinamas dviem etapais. Iš pradžių tarus, kad  $\psi$  turi apibendrintą Gama pasiskirstymą, buvo maksimizuota didžiausio tikėtumo funkcija „Weighted Boosted Search“ algoritmu [11]. Algoritmo realizacija R kodu yra pateikta priede 2. Šio algoritmo trūkumas yra tas, kad parametų reikšmingumą galima patikrinti bootstrap'o metodu arba pasinaudoti permutaciniais testais (F. Pesarin [17]), o tai užima labai daug laiko. (Vieno modelio parametų reikšmingumo tikrinimas trunka maždaug naktį.) Todėl šių metodų buvo atsisakyta. Problemą išsprendėm taip: įvertintos slenksčių reikšmės buvo perkeltos į EViews'a, buvo maksimizuota kvazi-tikėtumo funkcija, gauta iš normalaus tankio. Jos pagalba testuojamas parametų reikšmingumas. Apibendrinto Gama pasiskirstymo parametų reikšmingumas nebuvo tirtas, o įvertintų slenksčių reikšmingumas buvo iširtas netiesiogiai. Buvo pritaikytas Valdo testas hipotezei, kad skirtingų režimų atitinkami koeficientai yra lygūs. Ši hipotezė buvo atmesta. Tačiau, buvo priimta hipotezė, kad šių procesų laisvieji nariai bei koeficientai prie  $\psi$  yra lygūs. Todėl ši dalis visiems 3 režimams yra vienoda.

	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.067395	0.020596	3.272272	0.0011
GARCH(-1)	0.402709	0.147008	2.739364	0.0062
GARCH(-2)	0.389486	0.133783	2.911316	0.0036
(RESID(-1)^2)*(SQT(-1)<D1)	0.212509	0.059393	3.578010	0.0003
(RESID(-1)^2)*(SQT(-1)>D1)*(SQT(-1)<D2)	0.130627	0.018934	6.898909	0.0000
(RESID(-1)^2)*(SQT(-1)>D2)	0.084937	0.031366	2.707887	0.0068

2 lentelė: TACD modelio parametų įvertiniai, kai režimai priklauso nuo praeitos stebėjimo reikšmės.

Įvertintas modelis yra:

$$\psi_t = 0.067395 + 0.402709\psi_{t-1} + 0.389486\psi_{t-2} + \begin{cases} 0.212509x_{t-1}, & \text{jei } x_{t-1} < 0.9105 \\ 0.130627x_{t-1}, & \text{jei } 0.9105 \leq x_{t-1} < 2.5862 \\ 0.084937x_{t-1}, & \text{jei } x_{t-1} \geq 2.5862 \end{cases} \quad (18)$$

Pirmajame režime koeficientų prie  $\psi$  ir  $\psi$  suma viršija 1, tai yra atvejais, kai vieno režimo nestabilumas nereiškia viso modelio nestabilumo. Modelio stabilumui užtikrinti pakanka, kad paskutinio režimo atitinkamų koeficientų suma būtų mažesnė už vienetą. Faktiškai, toks rezultatas reiškia, kad po mažų trukmių dažniausia seka didesnės trukmės, o po didesnių trukmių- trumpos.

Šiuo atveju, koreliacija tarp modelio prognozių ir tikrų trukmių padidėja iki 0.177, kas yra pakankamai gerai aukšto dažnio modeliams. Paklaidos nėra autokoreliuotos, Ljung-Box statistika  $Q(1)=0.0152$ ,  $Q(30)=26.318$ ,  $Q(100)=101.21$ , nėra koreliuoti ir liekanų kvadratai. Tačiau liekanų kvadratinės šaknys yra koreliuotos. Pvz., Ljung-Box statistika kvadratinėms šaknims  $Q(1)=4.9524$ . Didesnių eilių liekanų šaknys (kubinė



šaknis, ketvirtos eilės) yra dar labiau autokoreliuotos. Tai liudija, kad liekanos nėra nepriklausomos, ir duomenyse yra kitų netiesinių sąryšių.

Panašūs rezultatai gaunami, kai režimai pasirenkami ne pagal praeitą stebėjimų reikšmę, o priklauso nuo praeito kainos pasikeitimo. T.y kai kaina mažėja, didėja, ar nesikeičia, trukmės tarp sandorių yra modeliuojami skirtingai.

	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.080278	0.018198	4.411344	0.0000
GARCH(-1)	0.320388	0.126814	2.526439	0.0115
GARCH(-2)	0.483986	0.117103	4.133004	0.0000
(RESID(-1)^2)*(DS11(-1)<0)	0.111322	0.028119	3.958921	0.0001
(RESID(-1)^2)*(DS11(-1)=0)	0.099981	0.017695	5.650276	0.0000
(RESID(-1)^2)*(DS11(-1)>0)	0.177271	0.035024	5.061444	0.0000

3 lentelė: TACD modelio parametrų įverčiai, kai režimai priklauso nuo kainos pasikeitimo.

Modelis atrodo taip:

$$\boxed{\hspace{10em}} \quad (19)$$

Čia,  $\Delta p_{t-1}$  yra kainos pokytis  $t-1$  laikotarpiu. Šiuo atveju modelio stabilumui užtikrinti reikia, kad visų režimų atitinkamų koeficientų sumos būtų mažesnės už vienetą. Šiuo atveju šiek tiek pagerėja modelio prognozavimo tikslumas, koreliacija tarp prognozių ir faktinių reikšmių pasiekė 0.183, tačiau netiesinės paklaidų funkcijos liko koreliuotos. Tam, kad pataisytume šią situaciją, buvo nuspręsta įtraukti į  $\Delta p_{t-1}$  išraišką netiesinę funkciją- logaritmą.

## 5. Slenkstinis modelis su logaritmine funkcija

### 5.1. Įvertintas modelis

Kad modelis „pagautų“ kuo daugiau netiesinių sąryšių, kurie egzistuoja tarp duomenų, į modelio specifikaciją buvo įtrauktas logaritmas. Tam, kad supaprastintume teorinių modelio savybių nagrinėjimą, prie duomenų po logaritmo ženklų buvo pridėtas vienetas. Formaliai, tam, kad užtikrintume, jog trukmės būtų neneigiamos, reikia reikalauti, kad visi koeficientai būtų neneigiami. Tačiau tai nėra būtina. Vėliau bus parodyta, kad mūsų atveju jos vis tiek negalės būti neigiamos.

	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.059481	0.017591	3.381414	0.0007
GARCH(-1)	0.840792	0.032388	25.95980	0.0000
(RESID(-1)^2)*(DS11(-1)<0)	0.035324	0.018717	1.887232	0.0591
(RESID(-1)^2)*(DS11(-1)=0)	0.046464	0.013609	3.414306	0.0006
(RESID(-1)^2)*(DS11(-1)>0)	0.042503	0.016010	2.654713	0.0079
LOG(SQRT(-1)^2+1)*(DS11(-1)<0)	0.086216	0.017648	4.885323	0.0000
LOG(SQRT(-1)^2+1)*(DS11(-1)=0)	0.070869	0.014206	4.988722	0.0000
LOG(SQRT(-1)^2+1)*(DS11(-1)>0)	0.057430	0.021887	2.623947	0.0087
LOG(SQRT(-2)^2+1)*(DS11(-1)<0)	-0.069835	0.018296	-3.816898	0.0001
LOG(SQRT(-2)^2+1)*(DS11(-1)=0)	-0.059681	0.014793	-4.034497	0.0001
LOG(SQRT(-2)^2+1)*(DS11(-1)>0)	-0.084615	0.018648	-4.537519	0.0000
1/SPREAD(-1)	0.069646	0.022984	3.030269	0.0024
F1	-0.010271	0.004739	-2.167385	0.0302
F2	-0.008698	0.004399	-1.977440	0.0480

4 lentelė: Slenkstinio modelio su logaritmine funkcija parametrų įverčiai.

Šis modelis yra aprašomas tokiomis lygtimis:

$$\begin{aligned} & \boxed{\phantom{0}} \\ & \boxed{\phantom{0}} \\ & \boxed{\phantom{0}} \end{aligned} \tag{20}$$

Čia, kaip ir anksčiau,  $\boxed{\phantom{0}}$  yra kainos pokytis  $t-1$  laikotarpiu, kintamasis  $\boxed{\phantom{0}}$  yra geriausios pasiūlos ir geriausios paklausos kainų skirtumas, išreikštas „tikais“ (mažiausiais kainos vienetais). Šiuo atveju „tikas“ yra lygus vienam Eurui. Paprastai (97% atvejais) jis yra lygus 1 arba 2, turimuose duomenyse didžiausia šio kintamojo reikšmė yra 10. Šis kintamasis yra paimtas  $1/\boxed{\phantom{0}}$  forma, kad užtikrintume, jog jį

atitinkantis koeficientas būtų teigiamas ir šis narys negalėtų padaryti trukmes neigiamomis. Fiktyvūs kintamieji  $F_1$  ir  $F_2$  atitinka pirmą ir paskutinį pirmadienį, kurie elgiasi šiek tiek skirtingai. Iš tikro, iš pradžių modelis buvo bandytas be jų, tačiau atlikus pasiūlytą Engle 2000 metais testą, netiesiškumui ir struktūriniais vidurkio pasikeitimams aptikti, buvo priimtas sprendimas juos įtraukti. Šį testą vėliau atliksim ir šiam modeliui, todėl dabar jo neaprašinėsim.

Modelio blogumas yra tas, kad dėl logaritminių funkcijų netiesiškumo šiuo atveju yra sunkiau nustatyti formalias stacionarumo sąlygas, todėl, jei nagrinėtume ne vieną, o daug vertybinių popierių, skaičiavimai užimtų daugiau laiko. Bet, šiuo atveju mes parodysime, kad gautas modelis pasižymi tam tikromis geromis savybėmis.

Visų pirma, reikia paminėti, kad koreliacija tarp modelio prognozių ir tikrųjų reikšmių išaugo iki 0.203, liekanos yra neautokoreliuotos:  $Q(1)= 0.5750$ ,  $Q(30)= 32.502$ ,  $Q(100)= 106.93$ . Taip pat nėra autokoreliuotos įvairios liekanų funkcijos, pvz. liekanų kvadratai, šaknys, logaritmai.

## 5.2. Pusiausvyros egzistavimas ir ergodiškumas

Toliau visur nagrinėsime bendrą, visų 3 režimų formą be fiktyviųjų ir egzogeninių kintamųjų, kadangi jie modelio elgesiui nedaro didelės įtakos:

$$\square \quad (21)$$

Netiesiniuose modeliuose pusiausvyros egzistavimas ir stacionarumas dažnai yra papildomi ar net pakeičiami ergodiškumo sąvoka. Dažnai ergodiškumas yra apibrėžiamas kaip tokia sekos savybė, kuomet, kai jos empiriniai momentai, jeigu jie egzistuoja, augant imties dydžiui konverguoja (kažkokia prasme) į teorinius modelio momentus. Clive W.J. Granger ir Timo Terasvirta knygoje [15], ergodiškumas yra apibrėžiamas kaip sekos savybė, kai modelio parametrai įverčiai „gerėja“, augant imties pločiui, su sąlyga, jog modelis turi tinkamą funkcinę formą, o J. Fan ir Q. Yao knygoje [16] pateiktas ergodiškumo apibrėžimas yra ekvivalentus asimptotiniam stacionarumui plačiąja prasme. Mes nesigilinsim į ergodiškumo apibrėžimų ekvivalentumą, ir nagrinėsime ergodiškumą Grangerio ir Terasvirtos prasme.

Tam, kad būtų paprasčiau nagrinėti ergodiškumą netiesiniams modeliams, buvo įvesta „modelio skeleto“ sąvoka. Modeliui  $\square$  yra tariama, kad nėra jokių šokų, t.y. multiplikatyviu atveju ateina tik vienetiniai šokai, o adityviu- visi šokai yra lygūs nuliui. Tada tokio modelio „skeletas“ yra  $\square$ . Mūsų atveju  $\square$ , todėl nėra skirtumo ar mes naudosisim  $\square$  ar  $\square$ . Paprastumo dėlei pasirinksim  $\square$ :

$$\square \quad (22)$$

Yra įrodyta (pvz. B. H. Cline ir Huay-min [9]), kad procesams kurių „skeletas“ turi  $\square$  formą, ergodiškumui įrodyti pakanka, kad funkcija  $\square$  būtų tolydi, ir  $\square$  dideliems  $\square$ .

$$\square \quad (23)$$

Vėliau, bus parodyta, kad neigiamos reikšmės negali egzistuoti, todėl mūsų atveju (23) lygybėje visus dydžius galima imti be modelių. Formaliai, bendrai išraiškai (22) šios

savybės taikyti negalima, kadangi į ją įeina  $\square$  pavėlinimai. Tačiau mūsų atveju, reikia pastebėti, kad  $\square$ , (iš formulės (20)). Todėl

$$\square$$

$$\square$$
(24)

Šiam procesui jau galima taikyti sąlygą (23). Kadangi mūsų atveju  $\square$  nelygybės (24) dešinėje užrašytas procesas yra ergodiškas.

Įrodant proceso ergodiškumą, (23) sąlygos ir funkcijos  $\square$  tolydumo reikia, kad būtų patenkinta Lipšico sąlyga, kad egzistuoūt atitinkamos diferencialinės lygties sprendinys, t.y. pusiausvyros būseną. Be to, įrodomas sprendinio stabilumas pagal Liapunovą. Stabilios pusiausvyros būsenos egzistavimas savo ruožtu lemia ergodiškumą. Taigi, mūsų nagrinėjamas procesas yra iš viršaus aprėžtas procesu, kuris yra ergodiškas, o iš apačios yra aprėžtas 0. Intuityviai aišku, kad mūsų nagrinėjamas procesas turi elgtis pakankamai stabiliai. Lygtį (21) perrašykime taip:

$$\square$$
(25)

Tam, kad suprastume, kaip elgiasi mūsų modelis nestabilumo atveju, tarsime priešingai: augant  $\square$  procesas  $\square$ . Tada  $\square$ , Taigi, pirmas logaritmas, kai trukmės yra pakankamai didelės, jokios įtakos nedaro. Imkime antrą logaritmą:

$$\square$$
(26)

tarkime, kad  $\square$  yra kažkokia prasme aprėžtas dideliems  $\square$  (pavyzdžiui  $\square$  kažkokiam  $\square$ ). Kadangi  $\square$ , kai  $t$  didėja, ir antro logaritminio nario įtaka artėja į 0. Formaliai ši savybė galios ne tik, kai  $\square$  yra aprėžtas, bet ir kai pakankamai lėtai auga į begalybę (tokį procesą galima sukonstruoti!). Taigi, didelėms trukmėms proceso elgesį sąlygoja tik tiesinė modelio dalis. Ir tam, kad procesas visą laiką grįžtų į pusiausvyrinę būseną, užtenka, kad tiesinės dalies parametrų suma būtų mažesnė už 1. Koeficientų prie logaritmų ženklai ir reikšmės yra svarbūs tik sąlyginai mažoms trukmėms.

Pusiausvyros būseną galima surasti tiesiog pakeitus (21) lygtyje  $\square$  ir  $\square$  į pusiausvyrinę būseną  $\square$  ir išsprendus lygtį  $\square$  atžvilgiu. Pvz. atveju, kai  $\square$ :

$$\square$$
(27)

sudėjus panašius narius, ir išsprendus sistemą, gauname šaknį:  $\square = 0.42686$ . Kitoms lygtims  $\square = 0.58213$ , ir  $\square = 0.6312$ . Pastarosios lygtys turi dar vieną šaknį, bet ji yra neigiama ir tai yra nestabilios pusiausvyros taškas. Todėl jis mūsų nedomina.

### 5.3. Neigiamų reikšmių neegzistavimas.

(20) lygtyje neigiamas reikšmės gali duoti tik antras logaritmas su pavėlinimu  $\square$ , kadangi tik šis narys turi neigiamą koeficientą. Todėl tam, kad ištirtume, ar taip aprašytas procesas gali įgyti neigiamas reikšmes, išrašysime sistemą nuo kažkokių pradinių reikšmių  $t-3$  laikotarpiu ir liekanų.

	(28)

Galima tarti, kad iki  $t-2$  laikotarpio sistema buvo pusiausvyroje,  $t-2$  laikotarpiu atėjo impulsas, vėliau impulsai vėl buvo vienetiniai, tada galima apskaičiuoti, koks turi būti impulsas, kad iš pusiausvyros būsenos sistema „išstumtų“ į neigiamas reikšmes. Įstačius į (24) lygtį įvertintus koeficientus ir išsprendus  $\square$  atžvilgiu, gauname, kad pirmai lygčiai (kai kaina didėja)  $\square > 0.98325027 \cdot 10^9$ . Analogiškai, kai kaina nekinta,  $\square > 0.817575312 \cdot 10^9$ , o kai kaina mažėja,  $\square > 0.3067886996 \cdot 10^{10}$ . Tik tokios didelės paklaidos gali sistemą iš pusiausvyros „nustumti“ į neigiamas reikšmes. Tai labai didelės paklaidos, kurių tikimybė yra beveik 0. Faktiškai tokios didelės paklaidos reiškia, kad jeigu 25 metus iš eilės nebuvo nei vienos transakcijos, tada sekančiai transakcijai modelis gali suprognozuoti neigiamą trukmę. Iš tikro, tokia didelė trukmė tarp transakcijų gali įvykti, jeigu dėl kažkokių priežasčių birža nustos prekiauti šiuo vertybiniu popieriumi arba kažkas atsitiks nagrinėjamai įmonei, tačiau tada mūsų modelis tiesiog taps nereikalingu. Todėl tolimesnei analizei galime laikyti, jog trukmės niekada negali būti neigiamos.

#### 5.4. Impulso atsakas ir trumpa atmintis

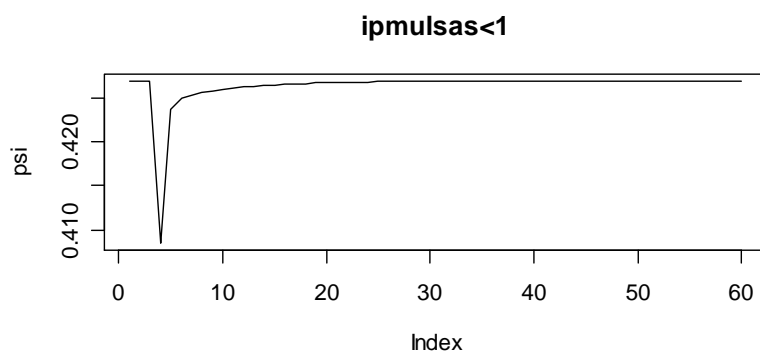
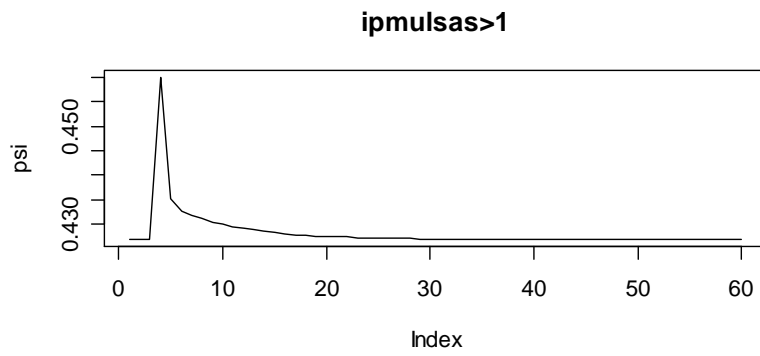
Tam, kad suprastume, kaip elgiasi sistema atėjus naujam impulsui, paimsime vieną iš įvertintų lygčių:

--

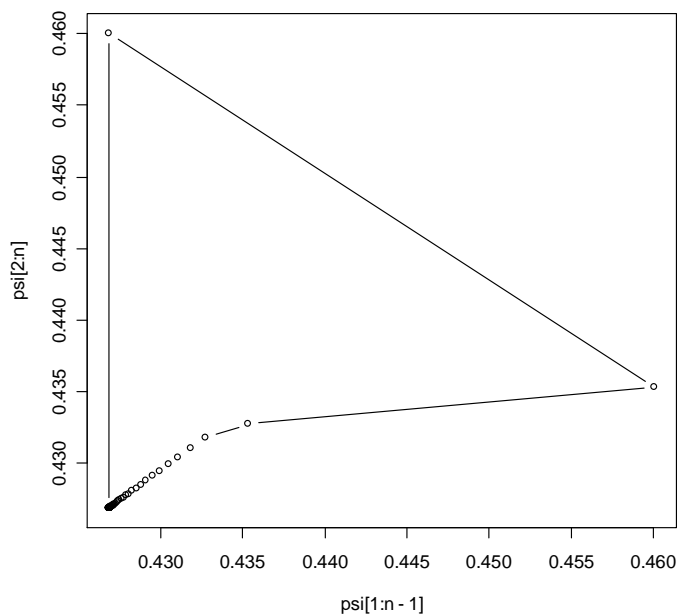
Tarkime, kad iki  $t-1$  momento sistema yra pusiausvyros būsenoje  $\square$ . Momentu  $t$  sistemai duodame teigiamą ar neigiamą impulsą, vėliau į sistemą ateina vien tik vienetiniai šokai. Tada galime stebėti, kaip elgsis sistema grįžtant į pusiausvyrą.

Kaip, matome 7 brėžinyje, išėjusi iš stabilumo būsenos sistema, iš pradžių grįžta prie pusiausvyros labai greitai, po to, jos grįžimas šiek tiek sulėtėja. Tai geriau matosi ( $\square$ ) grafike: išvesta iš pusiausvyros sistema gana staigiai leidžiasi 3 laiko tarpus po išvedimo iš pusiausvyros, vėliau grįžimas į pusiausvyrą sulėtėja, ir elgiasi maždaug kaip laipsninė funkcija.

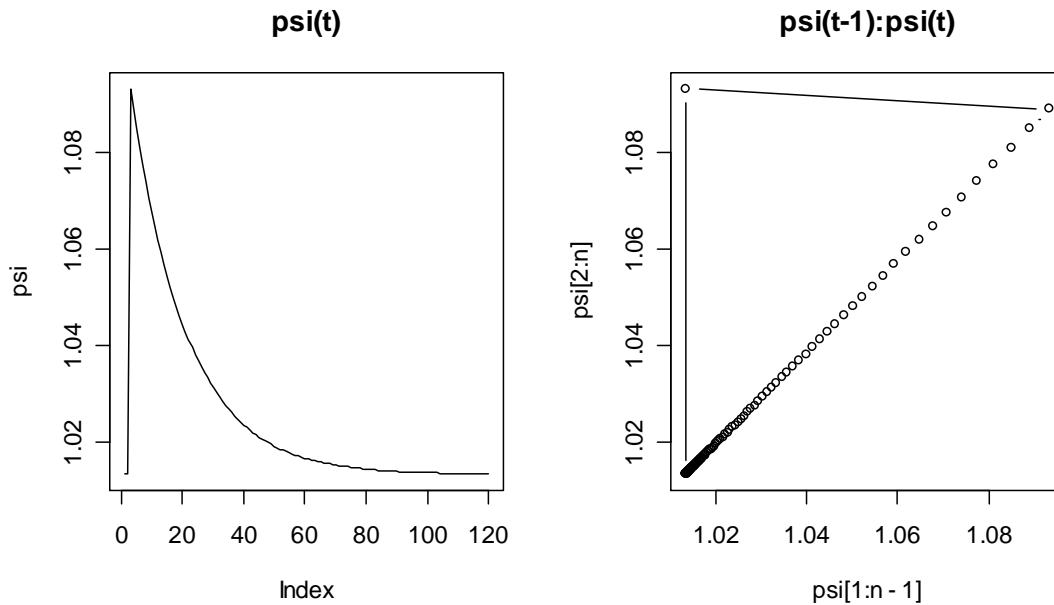
Palyginimui galime pažiūrėti, kaip elgiasi ACD(1,1) modelis aprašytas (16) ir (17) lygtyse. Kaip matome iš 9 paveikslo, jo elgesys panašus į antrą priartėjimo prie pusiausvyros fazę paveiksle 8. Labai tikėtina, kad logaritmas daro didesnę įtaką tik keliems pirmiems pavėlinimams, ir jau po kelių pavėlinimų modelio elgesį nusako tik tiesinė modelio dalis.



7 pav.: Sistemos grįžimas į pusiausvyrą, priklausomai nuo lagų skaičiaus.



8 psv.: Sistemos grįžimas į pusiausvyrą  grafike. Iš pradžių sistema buvo pusiausvyros būsenoje (kairės apatinis kampas), vėliau atėjo šokas ir  išaugo iki 0.460.



9 pav.: Grįžimas į pusiausvyrą ACD(1,1) modelio atveju

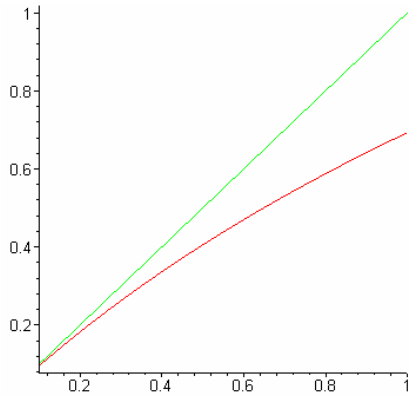
Tam, kad pažiūrėtume, kaip gęsta šokai, natūralu išrašyti lygtį, kuri priklausytų tik nuo paklaidų. Tam pasirinksiame paprastesnę sistemą, vien tik su logaritmu ir laisvuju nariu ir be pridėto vieneto:

$$\begin{aligned}
 & \psi_t = \psi_{t-1} + \alpha(\psi_{t-1} - \psi_t) + \epsilon_t \\
 & \psi_t = \psi_{t-1} + \alpha(\psi_{t-1} - \psi_t) + \epsilon_t \\
 & \psi_t = \psi_{t-1} + \alpha(\psi_{t-1} - \psi_t) + \epsilon_t \\
 & \psi_t = \psi_{t-1} + \alpha(\psi_{t-1} - \psi_t) + \epsilon_t
 \end{aligned}
 \tag{29}$$

Formaliai mes gavome impulso gesimo išraišką, tačiau iš lygties (28) ne kiekvienas gali suprasti, ar šis gesimas yra greitas ar lėtas. Todėl sistemą nagrinėkime kitaip:

$$\begin{aligned}
 & \psi_t = \psi_{t-1} + \alpha(\psi_{t-1} - \psi_t) + \epsilon_t \\
 & \psi_t = \psi_{t-1} + \alpha(\psi_{t-1} - \psi_t) + \epsilon_t \\
 & \psi_t = \psi_{t-1} + \alpha(\psi_{t-1} - \psi_t) + \epsilon_t
 \end{aligned}
 \tag{30}$$

Gauta išraiška rodo, kad impulsai gęsta greičiau negu laipsniškai. Tam, kad konverguotų begalinės sumos išraiškoje (30), reikia, kad  $\alpha < 1$ . Iš tikro, panaudota nelygybė yra gana grubi, ji yra pavaizduota paveiksle 10. Intuityviai aišku, kad, jeigu  $\alpha$  netgi dideliems  $\alpha$ , sistema turės pusiausvyrą ir elgsis pakankamai stabiliai.



10 pav.: Nelygybė  $\alpha < 1$ .

### 5.5. Stacionarumas

Egzogeninis kintamasis *spread* ir fiktyvieji kintamieji nedaro įtakos stacionarumui, galima laikyti, kad jie tiesiog pakeičia laisvąjį narį. Todėl vėl nagrinėsime bendrą režimų išraišką be šių kintamųjų:

$$\alpha = \frac{\sigma^2}{\mu} \quad (31)$$

$$\alpha = \frac{\sigma^2}{\mu} \quad (32)$$

Tada:

$$\alpha = \frac{\sigma^2}{\mu} \quad (33)$$

Paskutinėje išraiškoje pasinaudojome nelygybe, pateikta 10 grafike, bei faktų, kad visose įvertintose lygtyse koeficientas  $\alpha$  buvo neigiamas. Gauta išraiška yra paprastas ACD(1,1) modelis. Jis yra stacionarus, jei  $\alpha < 1$ . Šios nelygybės galioja visiems 3 mūsų režimams, nors ši suma yra gana arti vieneto. Ji atitinkamai yra lygi 0.940725, 0.958125, 0.962332. Atsargumo dėlei buvo patikrinta hipotezė apie vienetinės šaknies egzistavimą visiems duomenims kartu ir atskirai tarp režimų, ji buvo neabejotinai atmesta. Visais atvejais Dickey-Fullerio testo gauta p-reikšmė buvo 0.

### 5.6. Liekanų testavimas

Pagrindinės modelio prielaidos yra liekanų nepriklausomumas bei liekanų vidurkio pastovumas (pagal prielaidą jis yra lygus vienetai, bet užtenka, kad jis būtų teigiama konstanta). Nagrinėjant modelį buvo pateiktos Ljung-Box statistikų reikšmės liekanoms bei kai kurioms netiesinėms liekanų funkcijoms. Į modelį įtraukus logaritmus, ir liekanos ir jų netiesinės funkcijos gavo nekoreliuotos. Iš tikro, tai negarantuoja liekanų nepriklausomumo, tačiau tai yra geras rezultatas, kuriuo galima remtis. Liekanų vidurkio lygybė vienetai galima patikrinti testu, kurį pasiūlė Engle 2000 metais, netiesiniams sąryšiams aptikti. Jo idėja buvo suskaidyti nagrinėjamą laiko tarpą į N intervalų, kiekvienam intervalui sukuriamas fiktyvus kintamasis, kuris yra lygus 1 tame intervale, ir 0 kitur. Po to, mažiausių kvadratų metodu įvertinta liekanų regresija nuo fiktyviųjų



kintamųjų. Valdo ar kitu testu galima patikrinti koeficientų prie fiktyviųjų kintamųjų lygybę vienetui. Nors Engle pasiūlė šį metodą netiesiniams sąryšiams tirti, jis tiesiogiai tiria liekanų vidurkio pastovumą. Mes irgi pasinaudojome šiuo testu, suskaidėme nagrinėjamą laikotarpį į 10 lygių dalių. Testo rezultatai pateikti žemiau 5 lentelėje.

Hipotezė apie liekanų vidurkio pastovumą neabejotinai priimama. Priminsime, būtent dėl šio testo rezultatų, buvo priimtas sprendimas į mūsų modelį įtraukti fiktyviuosius kintamuosius, atitinkančius pirmą ir paskutinę dienas. Tada nagrinėjamas laikotarpis buvo suskaidytas į 5 intervalus.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
I1	0.987966	0.062099	15.90952	0.0000
I2	1.009006	0.061850	16.31371	0.0000
I3	0.950016	0.061850	15.35998	0.0000
I4	0.998920	0.061850	16.15068	0.0000
I5	1.051283	0.061850	16.99728	0.0000
I6	1.028551	0.061850	16.62976	0.0000
I7	1.017238	0.061850	16.44684	0.0000
I8	0.922342	0.061850	14.91255	0.0000
I9	0.937288	0.061850	15.15420	0.0000
I10	0.959638	0.061850	15.51555	0.0000

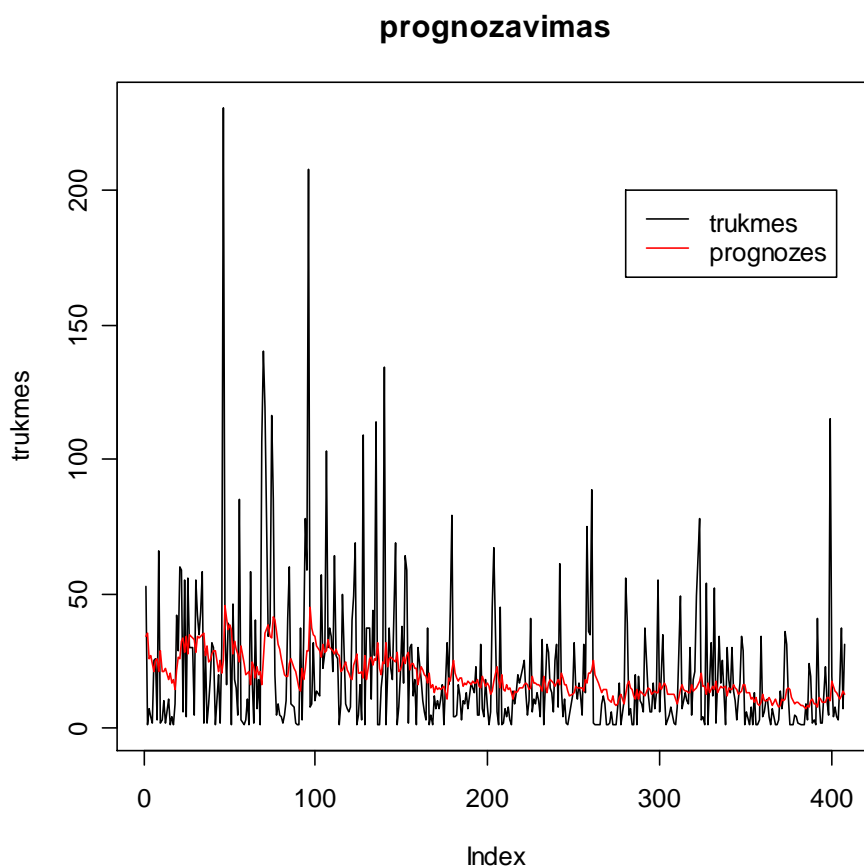
Wald Test:  
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	0.472072	(10, 5393)	0.9089
Chi-square	4.720717	10	0.9090

5 lentelė: Testas liekanų vidurkio pastovumui tikrinti.

## 6. Prognozavimas

Kaip jau buvo minėta, 407 duomenys buvo palikti patikrinti, kaip modelis prognozuoja. Tam mes tarėme, kad žinome viską, kas įvyko iki momento  $t$ , ir prognozavom trukmes vienu žingsniu į priekį. Koreliacija tarp  $X_t$  ir  $X_{t-1}$  yra lygi 0.164251. Vėliau, suprognuotas  $X_t$  buvo padaugintas iš dienos sezoniškumo funkcijos reikšmės taške  $t-1$ . Suprognuotos ir tikros  $X_t$  reikšmės yra pavaizduotos paveiksle XXX. Koreliacija tarp jų yra lygi 0.2793050.



11 pav.: Tikrosios ir prognozuotos trukmės tarp transakcijų.

## 7. Išvados

Šiame darbe buvo išnagrinėta trukmių tarp transakcijų dinamika prekiaujant GSK akcijomis. Pagrindinės išvados apie trukmių dinamiką yra tokios:

- Prekybos intensyvumas pasižymi tam tikru dienos sezoniškumu,
- Skirtingomis darbo dienomis prekybos intensyvumas yra skirtingas,
- Kai kuriomis dienomis prekybos intensyvumas elgiasi ypatingai,
- Trukmių tarp transakcijų dinamika priklauso nuo kainos pasikeitimų,
- Trukmių tarp transakcijų dinamiką nusako tam tikri netiesiniai sąryšiai,
- Trukmė tarp transakcijų priklauso nuo geriausios pasiūlos ir geriausios paklausos kainų skirtumų.

Be to, buvo gauti tam tikri teoriniai rezultatai:

- Gautos stacionarumo bei ergodiškumo sąlygos,
- Gautos pusiausvyros egzistavimo sąlygos,
- Nustatyta, kad esant tam tikrom sąlygom, logaritminiuose sąryšiuose šokai gęsta ne lėčiau nei laipsninė funkcija.
- Nustatyta, kad logaritminio netiesiškumo atveju modelio elgseną didelėms trukmėms sąlygoja tik tiesinė modelio dalis.

Šios savybės negali būti automatiškai apibendrintos kitiems vertybiniams popieriams. Pvz., mobiliojo ryšio bendrovei Vodofon akcijoms trukmių dinamika nepriklausė nuo kainos pasikeitimo, o farmacinei įmonei Astrazeneca koeficientai prie logaritmų buvo nereikšmingi, tačiau modelis gerai veikė kai vietoj trukmių logaritmo buvo paimti trukmių kvadratinės šaknys. Tai reiškia, kad gautas modelis nėra universalus ir negali būti tiesiogiai pritaikytas kitiems vertybiniams popieriams. Ši modelio savybė trukdo automatizuoti modelio parinkimo procesą, kad jį būtų galima panaudoti daugeliui vertybinių popierių (tarkim keliems tūkstančiams). Be to, modelis neatsižvelgia į kitų vertybinių popierių dinamiką ir makroekonominis duomenis. Todėl gali būti ieškomi kiti tikslesni ir paprastesni modeliai trukmėms tarp transakcijų aprašyti.

## Literatūra

(Straipsniai:)

1. R. F. Engle ir J. R. Russell „Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data“, *Econometrica*, **66**, 1127-1162, (1998).
2. R. F. Engle „The Econometrics of Ultra-High-Frequency Data“, *Econometrica*, **68**, 1-22, (2000).
3. M.Y. Zhang, J.R. Russell, R.S. Tsay, „A nonlinear autoregressive conditional duration model with applications to financial transaction data“, *Journal of Econometrics*, **104**, 179-207, (2001).
4. M.Meitz, T. Teräsvirta, "Evaluating Models of Autoregressive Conditional Duration", *Journal of Business and Economic Statistics*, **24**, 104-124, (2006).
5. M. Pacurar, „ACD models in Finance: A Survey of the Theoretical and Empirical Literature“, *working paper*, (2006).
6. C. Gouriéroux, A. Monfort, A. Trognon „Pseudo Maximum Likelihood Methods: Application to Poisson Models“, *Econometrica*, **52**, 681-700, (1984).
7. J. R. Magnus, „The Asymptotic Variance of the Pseudo Maximum Likelihood Estimator“, *CIRGE*, **479**, 1-12, (2007).
8. D. B. H. Cline, „Regular variation of order 1 nonlinear AR-ARCH models“, *Science Direct*, 840-862, (2005).
9. D. B. H. Cline, Huay-min H. Pu „Geometric ergodicity of nonlinear time series“, *Statistica Sinica*, **9**, 1103-1118, (1999).
10. F. Legland, L. Mevel, „Exponential Forgetting and Geometric Ergodicity in Hidden Markov Models“, *Math. Control Signals Systems*, **13**, 63 – 93, (2000).
11. S. Chen, X.X. Wang C. J. Harris, „Experiments with Repeating Weighted Boosting Search for Optimization in Signal Processing Applications“, *IEEE Transactions on Systems, Man and Cybernetics* **35(4)**, 682-693, (2004).
12. G. Koop „Nonlinearity, Structural Breaks or Outliers in Economic Time Series“, *Nonlinear Econometric Modeling in Time Series Analysis*, **4**, 61-78, (1997).
13. M. Meitz, „A Necessary and Sufficient Condition for the Strict Stationarity of a Family of GARCH Processes“, *Econometric Theory*, **22**, 985-988 (2006).

(Knygos:)

14. R.S. Tsay, „Analyses of Financial Time Series“, Wiley-Interscience, Chicago (2001).
15. W.J. Granger ir T. Teräsvirta „Modelling Nonlinear Relationships“, Oxford University Press, New York (1993).
16. J. Fan, Q. Yao, „Nonlinear Time Series, Nonparametric and Parametric Methods“, Springer, (2005).
17. F. Pesarin „Multivariate Permutation Tests with Applications in Biostatistics“, Wiley, Chichester (2001).
18. V.Čekanavičius, G. Murauskas, „Statistika ir jos taikymai 1“ TEV, Vilnius (2000).

## 8. Priedas 1

	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.064677	0.061856	1.045593	0.2957
GARCH(-1)	0.696992	0.231108	3.015867	0.0026
RESID(-1)*(DS(-1)<0)	0.148688	0.042147	3.527867	0.0004
RESID(-1)*(DS(-1)=0)	0.114326	0.061418	1.861436	0.0627
RESID(-1)*(DS(-1)>0)	0.110522	0.048464	2.280489	0.0226
LOG(RESID(-1)+1)*(DS(-1)<0)	0.108921	0.047043	2.315339	0.0206
LOG(RESID(-1)+1)*(DS(-1)=0)	0.161901	0.030333	5.337399	0.0000
LOG(RESID(-1)+1)*(DS(-1)>0)	0.103271	0.047343	2.181336	0.0292
LOG(RESID(-2)+1)*(DS(-1)<0)	-0.052475	0.018151	-2.890977	0.0038
LOG(RESID(-2)+1)*(DS(-1)=0)	-0.098473	0.041553	-2.369816	0.0178
LOG(RESID(-2)+1)*(DS(-1)>0)	-0.094407	0.043898	-2.150625	0.0315
1/SPREAD(-1)	0.087117	0.025692	3.390810	0.0007

6 lentelė: Slenkstinio modelio su logaritmine funkcija parametų įverčiai antradieniams

	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.041055	0.074232	0.553069	0.5802
GARCH(-1)	0.784444	0.035081	22.36103	0.0000
RESID(-1)^2*(DS<0)	0.126307	0.049932	2.529564	0.0114
RESID(-1)^2*(DS=0)	0.134997	0.059520	2.268087	0.0233
RESID(-1)^2*(DS>0)	0.076060	0.038134	1.994524	0.0461
LOG(RESID(-1)^2+1)*(DS<0)	0.063132	0.032505	2.045231	0.0408
LOG(RESID(-1)^2+1)*(DS=0)	0.018547	0.064521	0.287450	0.7738
LOG(RESID(-1)^2+1)*(DS>0)	0.100029	0.042801	2.337077	0.0194
LOG(RESID(-2)^2+1)*(DS<0)	-0.092321	0.044463	-2.076349	0.0379
LOG(RESID(-2)^2+1)*(DS=0)	-0.062632	0.032941	-1.901315	0.0573
LOG(RESID(-2)^2+1)*(DS>0)	-0.084598	0.039206	-2.157786	0.0309
1/SPREAD(-1)	0.135334	0.047120	2.872140	0.0041

7 lentelė: Slenkstinio modelio su logaritmine funkcija parametų įverčiai trečiadieniams

	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.395847	0.151515	2.612600	0.0090
GARCH(-1)	0.614055	0.089694	6.846150	0.0000
RESID(-1)^2*(DS(-1)<0)	0.082580	0.043983	1.877530	0.0604
RESID(-1)^2*(DS(-1)=0)	0.059917	0.023535	2.545840	0.0109
RESID(-1)^2*(DS(-1)>0)	0.070073	0.041149	1.702904	0.0886
LOG(RESID(-1)^2+1)*(DS(-1)<0)	0.104929	0.030207	3.473634	0.0005
LOG(RESID(-1)^2+1)*(DS(-1)=0)	0.087078	0.020889	4.168555	0.0000
LOG(RESID(-1)^2+1)*(DS(-1)>0)	0.120136	0.016827	7.139331	0.0000
LOG(RESID(-2)^2+1)*(DS(-1)<0)	-0.043628	0.024653	-1.769713	0.0768
LOG(RESID(-2)^2+1)*(DS(-1)=0)	-0.079218	0.026018	-3.044718	0.0023
LOG(RESID(-2)^2+1)*(DS(-1)>0)	-0.089150	0.037811	-2.123820	0.0337
1/SPREAD(-1)	0.249374	0.109436	2.278716	0.0227

8 lentelė: Slenkstinio modelio su logaritmine funkcija parametru įverčiai ketvirtadieniams

	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.495007	0.061324	8.071993	0.0000
GARCH(-1)	0.571375	0.045931	12.43982	0.0000
RESID(-1)^2*(DS(-1)<0)	0.106375	0.043180	2.463525	0.0138
RESID(-1)^2*(DS(-1)=0)	0.087899	0.041802	2.102760	0.0355
RESID(-1)^2*(DS(-1)>0)	0.076035	0.038637	1.967946	0.0491
LOG(RESID(-1)^2+1)*(DS(-1)<0)	0.065721	0.027659	2.376105	0.0175
LOG(RESID(-1)^2+1)*(DS(-1)=0)	0.117327	0.016391	7.158103	0.0000
LOG(RESID(-1)^2+1)*(DS(-1)>0)	0.059361	0.029147	2.036634	0.0417
LOG(RESID(-2)^2+1)*(DS(-1)<0)	0.008768	0.032078	0.273340	0.7846
LOG(RESID(-2)^2+1)*(DS(-1)=0)	0.038499	0.017731	2.171227	0.0299
LOG(RESID(-2)^2+1)*(DS(-1)>0)	0.021812	0.036419	0.598910	0.5492
1/SPREAD(-1)	0.142770	0.059827	2.386368	0.0170

9 lentelė: Slenkstinio modelio su logaritmine funkcija parametru įverčiai penktadieniams

**Pastaba:** Kai kurie koeficientai yra nereikšmingi. Tačiau, buvo priimtas sprendimas juos palikti, kad būtų paprasčiau lyginti modelius tarpusavyje.

## Priedas 2

# TACD(3) (Gamma)

# **Didžiausio tiketino f-ja.**

```
l_TG<-function(param,x){
niu<-param[1] #Gama pasiskirstymo "mastelio parametras"
alfa01<-param[2]
alfa11<-param[3]
alfa12<-param[4]
alfa13<-param[5]
beta11<-param[6]
beta12<-param[7]
d1<-param[8]
d2<-param[9]
k1<-param[10] # Gama pasiskirstymo "shape parametrai"
k2<-param[11]
k3<-param[12]

n<-length(x)
nstart<-50
psi<-c(1:n)
psi[1:nstart]<-mean(x[1:nstart])
mx<-min(x)

for (i in nstart:n){
psi[i]<-alfa01+beta11*psi[i-1]+beta12*psi[i-2]
if (x[i-1]<d1) {psi[i]<-psi[i]+alfa11*x[i-1]} else {if (x[i-1]<d2) psi[i]<-psi[i]+alfa12*x[i-1] else psi[i]<-
psi[i]+alfa13*x[i-1]}
}
lambda1<-psi*gamma(k1)/gamma(k1+1/niu)
lambda2<-psi*gamma(k2)/gamma(k2+1/niu)
lambda3<-psi*gamma(k3)/gamma(k3+1/niu)

S<-0
for (i in 2:n){
if (x[i-1]<d1){
S<-S+log(niu)+(k1*niu-1)*log(x[i]/lambda1[i])-log(lambda1[i]*gamma(k1))-(x[i]/lambda1[i])^niu
} else
if (x[i-1]<d2)
{S<-S+log(niu)+(k2*niu-1)*log(x[i]/lambda2[i])-log(lambda2[i]*gamma(k2))-(x[i]/lambda2[i])^niu }
else S<-S+log(niu)+(k3*niu-1)*log(x[i]/lambda3[i])-log(lambda3[i]*gamma(k3))-(x[i]/lambda3[i])^niu
}
}
S
}

#Weghted boosting Search optimizavimo algoritmas
WBST<-function(X){
n<-200 # generuojamu parametru kiekis
N<-300 # Maksimalus iteraciju skaicius
alfa<-0.0001 # Normias tikslumo lygis

U<-matrix(nrow=n,ncol=12)
U[,1]<-runif(n,min=0.01,max=5) #niu

U[,2]<-runif(n,max=2.5) #alfa01
```

```

U[,6]<-runif(n,max=1)          #beta11
U[,7]<-runif(n,max=1-U[,6])    #beta11
U[,3]<-runif(n,max=1)          #alfa11
U[,4]<-runif(n,max=1)          #alfa12
U[,5]<-runif(n,max=1-U[,6]-U[,7]) #alfa13
U[,8]<-runif(n,max=5)          #d1
U[,9]<-runif(n,min=U[,8],max=10) #d2
U[,10]<-runif(n,min=0.01,max=5) #k1
U[,11]<-runif(n,min=0.01,max=5) #k2
U[,12]<-runif(n,min=0.01,max=5) #k2

weight0<-rep(1/n,n)
f<-c(1:n)
Fmax<-F(U[1,],X)
Fmin<-Fmax
Imax<-1
Imin<-1

for (i in 1:n){
  f[i]<-F(U[i,],X)
}
for (j in 1:N) {
  for (i in 1:n){
    if (f[i]<Fmin){
      Fmin<-f[i]
      Imin<-i
    }
  }
}

loss<-f/sum(f)
eps<-sum(weight0*loss)
B<-eps/(1-eps)

if (B<1) weight01<-weight0*B^(loss-1) else weight01<-weight0*B^loss
weight0<-weight01/sum(weight01)

U1<-t(U)%*%weight0
U2<-(U1+U[Imax,])/2

f1<-F(t(U1),X)
f2<-F(t(U2),X)

if (f1>f2) {U[Imin,]<-U1
  f[Imin]<-f1 }
else {U[Imin,]<-U2
  f[Imin]<-f2}

if (f[Imin]>Fmax){
  Fmax<-f[Imin]
  Imax<-Imin
}
if ((Fmax-Fmin)<alfa)break

}
c(U[Imax,], Fmax)
}

```