TARTU ÜLIKOOL

MATEMAATIKA–INFORMAATIKATEADUSKOND

Matemaatilise statistika instituut

Finants- ja kindlustusmatemaatika eriala

Kärt Päll

Eesti elektrienergia hinna analüüs ja ühesammuline prognoosimine ARIMA tüüpi mudelitega

Magistritöö

(30 EAP)

Juhendaja: dotsent Raul Kangro

Tartu 2015

Eesti elektrienergia hinna analüüs ja ühesammuline prognoosimine ARIMA tüüpi mudelitega

Antud magistritöö eesmärgiks on vaadelda järgmise päeva elektrihinna prognoosimist erinevate meetoditega, valida vaadeldud mudelite hulgast välja parim ning võrrelda selle abil saadud tulemusi automaatse koodiga leitud parima mudeli tulemustega. Kasutatud metoodika on tuttav aegridade analüüsi kursusest. Töös antakse ülevaade ühe- ja mitmemõõtmelistest aegridade mudelitest ning nendega seotud definitsioonidest. Teema paremaks lahtimõtestamiseks tutvustatakse lühidalt Eesti elektriturgu ning hinna kujunemist elektribörsil. Töös kasutatakse vaid ARIMA tüüpi mudeleid, sest need on enim kasutatavad mudelid ennustamaks aegridade käitumist tulevikus. Kogu analüüsi teostamiseks on kasutatud reaalseid andmeid.

Märksõnad: *aegridade analüüs, ARIMA tüüpi mudelid, ühemõõtmelised aegridade mudelid, mitme-mõõtmelised aegridade mudelid*

Analysis of Estonian electricity prices and one step ahead forecast using ARIMA models

The aim of this thesis is to look at the next day's electricity price forecast using different methods, to provide the best model and to compare it with the best model found with automatic code. Used methodology is familiar from the course of Time Series Analysis. In this master's thesis the overview of univariate and multivariate time series models and their characteristics are given. This thesis also gives a brief overview of the Estonian power market and describes what the electricity price is influenced by. Models that are used for analysing Estonian electricity price are all ARIMA models. ARIMA models are the most typical class of models for forecasting a time series. All the analysis are carried out using real data.

Keywords: time series analysis, *ARIMA models, univariate time series, multivariate time series*

Sisukord

Sissejuhatus

Käesoleva töö teema on aktuaalne, kuna alates 2013. aastast ei ole elektrihind Eestis enam riiklikult määratud, vaid kujuneb vabalt elektribörsil pakkumise ja nõudluse vahekorrast, kuid mitte ainult. Lõpliku hinna kujunemist mõjutavad oluliselt ka näiteks kliima ja kõiksugused kaablite/jaamade hooldustööd. Kui pikas perspektiivis on elektrihind pigem tõusva trendiga, siis lühiajaliselt võib hind väga palju kõikuda ja ennustamine on juba raskem.

Antud magistritöö eesmärgiks on vaadelda järgmise päeva elektrihinna prognoosimist erinevate meetoditega, valida vaadeldud mudelite hulgast välja parim ning võrrelda selle abil saadud tulemusi automaatse koodiga leitud parima mudeli tulemustega.

Töö koosneb seitsmest põhipeatükist. Esimesed kaks peatükki võtavad lühidalt kokku Eesti elektrisüsteemi ja selle toimimise ning elektribörsi Nord Pool Spot. Selles peatükis kirjeldatakse ka Eesti ja naaberriikide vahel loodud ühendusi ning elektrienergia hinna kujunemist vabaturul. Lisaks on välja toodud Eesti elektrituru kokkuvõte 2013. aastal.

Kolmas ja neljas peatükk kirjeldavad kasutatud metoodika teooriat. Kolmas peatükk keskendub ühemõõtmelistele aegridade mudelitele, mis tähendab, et mudeli leidmisel kasutatakse ainult uuritava aegrea minevikuväärtuseid ning neljas peatükk keskendub mitmemõõtmelistele mudelitele, kus lisaks vaadeldavale aegreale kasutatakse ka teiste tunnuste minevikuväärtuseid.

Viiendas peatükis keskendutakse Eesti elektrienergia hinna analüüsile ühemõõtmeliste mudelitega ja uuritakse lisaks tavalisele elektrienergia hinna aegrea käitumisele ka logaritmitud rida. Selles peatükis leitakse mõlema rea korral parim mudel, esitatakse ühesammulised prognoosid ning võrreldakse mudelite prognoosivigasid tuntuimate mõõdikutega.

Kuues peatükk uurib Eesti elektrienergia hinna aegrida mitmemõõtmeliste mudelitega. Lisaks ajaloolistele elektrihindadele kasutatakse sama perioodi Eesti õhutemperatuuride ning töö- ja puhkepäevi eristavat aegrida. Nii nagu eelnevas peatükis, uuritakse ka siin logaritmitud elektrihinna aegrida. Mõlemal juhul leitakse parima mudeli korral prognoosivead ning võrreldakse neid omavahel. Selles peatükis loodetakse saada viiendas peatükis leitud mudelist parem mudel, st väiksemate prognoosivigadega mudel, kuna antud juhul on kasutada rohkem informatsiooni.

Viimane peatükk võrdleb töös leitud parimat mudelit automaatse aegridade analüüsi koodi abil leitud parima mudeliga.

Kogu teostatud analüüsiks on kasutatud statistikapaketti R.

1 Eesti elektrisüsteem

Elektrijaamad, ülekandevõrgud ning elektritarbijad moodustavad Eesti elektrisüsteemi. Eesti elektrisüsteemi sõltumatu ja iseseisev haldur on Elering, mille ülesandeks on kindlustada Eesti tarbijatele pidev ning kvaliteetne elektrivarustus. Elering eraldus Eesti Energia AS-st 27. jaanuaril 2010. aastal. Eesti Vabariik on Eleringi omanik ning aktsionäri õiguseid teostab Majandus- ja Kommunikatsiooniministeerium. (Eesti Energia kodulehekülg [1], 2015)

Eestis kuulub elektrisüsteemi põhivõrgu osale kokku umbkaudu 5539 km ülekandeliine, millest

- 3479 km on 110 kV liine,
- 1702 km on 330 kV liine,
- 158 km on 220 kV liine,
- 139 km on alalisvooluliine,
- 61 km on 35 kV liine

ning kokku on 146 alajaama. (Elering AS-i kodulehekülg [2], 2015)

1.1 Elektri tootmine Eestis

Eesti on maailma suurim põlevkivikaevandaja ning ainuke riik maailmas, mis kasutab põlevkivi, et katta ära põhiline osa riigi energiavajadusest. Energia tootmiseks kasutatakse aastas ca 15 miljonit tonni põlevkivi ning näiteks ühest tonnist põlevkivist saab 850 kWh elektrit ja 125 kg põlevkiviõli. Näiteks 2011. aastal oli Eesti majapidamiste energiatarbimine inimese kohta aastas 8,15 MWh. Põlevkivi kasutatakse nii vedela sünteetilise õli kui ka elektrienergia tootmiseks. Eesti suurim põlevkivist elektrienergia tootja on Eesti Energia.

Oluline elektritootmisharu on ka taastuvenergia. Eestis toodetakse elektrit tuulest, veest ja biomassist. Esimene tuulegeneraator rajati juba 1997. aastal Tahkunale Hiiumaale, mille võimsus oli $0,15$ MW. Eesti Energia esimene elektrituulik rajati Virtsu 2002. aastal. Hetkel on Eesti Energial viis tuuleparki, mis asuvad Ruhnus, Virtsus, Aulepas, Paldiskis ja Narvas. Taastuvenergia allikatest on Eestis omal kohal ka hüdroenergia. Vesi on hea energiaallikas, sest veevarude kasutamine ei ole Eestis maksustatud ja vesi on meil maa peal olemas. Küll on aga hüdroenergia kasutamise maht piiratud ja kasutatav energiahulk on vaid umbes 10 MW. Eesti Energial on kaks hüdroelektrijaama - Linnamäe ja Keila-Joa, mis ei suuda suuremahulises elektritootmises kaasa aidata. Vett kasutab Eesti Energia elektritootmiseks alates 2002. aastast. Viimaseks kasutatavaks taastuvenergia allikaks on biokütus. Biomass on elusaine mass, mis on toodetud madala kvaliteediga metsamaterjalist, põllumajandusest saadud jäätmetest, võsast ning roostikest. Kuna Eestis on suured biomassi varud, siis sellise kütuse kasutamine on kasulik. Biokütuseid kasutatakse näiteks Paides, kus toodetakse soojust biomassil töötavas katlamajas. Ka Narva elektrijaamad kasutavad oma elektritootmises biokütust. (Eleringi toimetised [3], 2012)

Eestis on kasutusel ka koostootmisjaamad. Koostootmisjaam on elektrijaam, mis toodab nii elektrit kui ka soojust. Koostootmisjaamad on rajatud tarbijate vahetusse lähedusse, kuna saadud soojusenergia suunatakse elamute küttesüsteemidesse või suurtarbijate tootmisprotsessidesse. Sellistes jaamades kasutatakse nii fossiilseid, näiteks maagaasi, põlevkivi, põlevkiviõli, kui ka taastuvaid energiaallikaid. (Eesti Energia kodulehekülg [1], 2015)

Eesti elektrisüsteemi suurimad riskid on looduslikud – näiteks torm, äike, vesi, kuid ka inimtegevus võib ohtu seada reaalajas toimiva elektriühenduse. Tänapäeval on veel kasutusel palju ebatöökindlaid ja halvas seisukorras seadmeid, mis samuti võivad põhjustada rikkeid ja probleeme elektrileviga. Kõiksugused muud välisriskid on juba seotud näiteks avariidega alajaamades või elektriülekandeliinidel väljaspool Eestit.

Eesti elektrisüsteem on ühendatud naaberriikidega ja kuulub suurde sünkroonselt töötavasse ühendsüsteemi BRELL. Eesti elektrisüsteem on ühenduses Venemaa, Läti ja Soomega. Eesti elektrisüsteem on vahelduvvooluliine pidi ühendatud Venemaa ja Lätiga ning need omakorda naabrite Leedu ja Valgevenega. Soomega on loodud alalisvoolu ühendus. Venemaaga on Eesti ühendatud kolme 330 kV liiniga ning Lätiga on kaks 330 kV liini. (Elering AS-i kodulehekülg [2], 2015; Eleringi toimetised [3], 2012; Konkurentsiamet [4], 2013)

1.2 EstLink

Eesti ja Soome vahele on praeguseks loodud kaks alalisvoolu ühendust, mis tagavad Eesti elektri varustuskindluse ja energiajulgeoleku. Avatud elektrituru efektiivse toimimise jaoks on oluline tugevate ülepiiriliste ühenduste olemasolu. Sellest tulenevalt suureneb konkurents ning tarbijatel on võimalik saada parima hinnaga elektrit.

Esimene alalisvooluühendus EstLink 1 avati 4. detsembril 2006. aastal. Selle ühenduse pikkus on 105 km, võimsus 350 MW ning alalisvoolu pinge 150 kV. Pideva elektrituru arenguga tekkis järjest suurenev vajadus lisa ühendusvõimsuste järele, seega otsustati rajada ka EstLink 2. EstLink 2 avamine toimus 6. märtsil 2014. aastal, mille tulemusena kolmekordistusid Balti- ja Põhjamaade vahelised elektri ülekandevõimsused. EstLink 2 on üldpikkusega 170 km, võimsusega 650 MW ning alalisvoolu pingega 450 kV.

EstLink 1 ja 2 omanikeks on Eestis Elering AS ning Soomes Fingrid Oyj, kummalegi kuulub 50% osalusest. Eestis on EstLink 2 ühenduskohaks Püssi- ning Soomes Antilla alajaam, EstLink 1 on aga Harku ja Espoo alajaamade vahel. Näiteks ainuüksi EstLink 2 võimsus katab ära poole Eestis talvekuudel vajaminevast elektrist. (Estlink 2 kodulehekülg [5], 2015)

2 Nord Pool Spot

2.1 Euroopa Liidu avatud elektriturg

Nord Pool Spot on elektrituru kauplemispiirkond ehk elektribörs, kus määratakse elektrihind eraldi igaks tunniks ja kõikidele piirkondadele eraldi. Turg toimib põhimõttel, et elekter liigub madalama hinnaga piirkonnast kõrgema hinnaga piirkonda, et sealset kõrget hinnataset alla tuua. Lisaks Eestile kuuluvad sellesse turu süsteemi Norra, Rootsi, Soome, Taani, Saksamaa, Suurbritannia, Läti ja Leedu. Põhjamaadest oli Norra esimene riik, kus korraldati elektriturg ümber.

Nord Pool Spot'i saab jagada kaheks:

1. Elspot ehk päev ette turg – hinnad ja kogused fikseeritakse järgmise päeva 24 tunniks.

2. Elbas ehk päevasisene turg – võimalik osta/müüa puudujääv/ülejääv osa ehk tasakaalustada täiendavalt bilanssi.

Euroopa Liidus on kehtestatud elektriturgude avamise põhimõte, kus igal tarbijal on õigus ja vabadus valida elektrimüüjat. Kuna vabalt kujunev elektri turuhind ja konkurents on soodne pinnas investeeringute ja efektiivse majanduse jaoks, siis üldjoontes võidavad nii tootjad kui ka tarbijad. EL-i elektriturg avanes täielikult 2007. aastal ning Eestis 1. jaanuaril 2013. aastal. Osaliselt avanes Eesti elektriturg juba 2010. aasta aprillis, kuid ainult suurtarbijatele. Suurtarbijateks nimetatakse tarbijaid, kes vajavad aastas energiat rohkem kui 2 GWh. (Energiatalgute koduleheküljg [6], 2015)

Lisaks valikuvabadusele on avatud elektriturg oluline Eesti energiajulgeoleku seisukohast, et saavutada sõltumatus Venemaa elektrienergiast. Avatud elektrituruga on võimalik luua elektrienergia edastamise välisühendusi kogu Euroopaga ning tänu mitmetele investeeringutele on rohkem erinevaid elektritootmis võimalusi Eestis. (Avatud elektrituru kodulehekülg [8], 2015)

2.2 Elektrienergia hindade kujunemine

Nord Pool Spot'i turul arvutatakse elektri turuhind igaks tunniks eraldi. Kauplemine toimub iga päev ja iga tund. Kauplemine toimub päev ette põhimõttel - järgmise päeva hinnad saadakse teada eelneval päeval kell 15:00.

Elektribörsil kujuneb hind mitmete välistegurite mõjul ja on üpris muutlik. Peamisteks mõjutajateks on:

1. õhutemperatuur, kliima;
2. CO2 kvootide maailmaturu hind;
3. hüdroreservid;
4. pakkumise/nõudluse suhe;
5. riikidevahelised ülekandevõimsused;
6. elektrijaamade või merekaabli hooldus- ja avariitööd.

Tarbija jaoks kujuneb elektrienergia hind eelkõige siiski müüja ja tarbija vahel sõlmitud lepingust. (Imatra Elekter AS-i koduleht [9], 2015; Elering AS-i koduleheküljg [2], 2015)

2.3 Avatud elektriturg Eestis

Eestis on Nord Pool Spot'i omanikuks Elering. Elektrisüsteemihaldur Elering juhib elektrisüsteemi reaalajas ja vastutab selle kvaliteetse toimimise eest.

Eestis tegutseb 12 elektrimüüjat – 220 Energia, VKG Elektrivõrgud, Elektrum Eesti, Imatra Elekter, Eesti Energia, Elektrimüügi AS, AS Sillamäe SEJ, Eesti Gaas AS, Baltic Energy Services, Nordic Power Management, INTER RAO Eesti ja Starman (seisuga 08.03.2015). Elektrimüüjad pakuvad kas fikseeritud hinnaga-, osaliselt muutuva hinnaga- või börsihinnast sõltuvaid pakette. (Avatud elektrituru kodulehekülg [8], 2015)

Tarbijal ei ole kohustust elektrimüüjat valida, kuid sellisel juhul on elekter tagatud üldteenusega ja hind võib olla kallim kui elektrimüüjatel. Üldteenuse hind kujuneb eelmise kuu keskmise kaalutud turuhinna ja elektrimüüja kasumiprotsendi summana. Oluline on teada, et elektrihind ei võrdu elektriarvetel kajastuva hinnaga. Elektriarve koosneb viiest komponendist, milleks on elektriaktsiis, käibemaks, taastuvenergia tasu, võrgutasu ja elektrienergia enda hind. Umbes kolmandik igakuisest elektriarvest on mõjutatud avatud elektriturust. Võrgutasu, taastuvenergia tasu ja makse turu avamine ei puuduta. Enne elektrituru täielikku avamist oli elektrienergia hind riiklikult määratud ja septembris 2012 oli selle hinnaks ligikaudu 31 EUR/MWh. (Avatud elektrituru kodulehekülg [8], 2015; Energiatalgute kodulehekülg [6], 2015)

2.4 Elektrituru 2013. aasta kokkuvõte

NPS Eesti ja Soome turu hinnad ühtisid päev-ette turul 69% ajast ja tänu EstLink 2 kasutuselevõtuga 2014. aasta esimeses kvartalis juba 89% tundidest. Eesti hinnapiirkonna aasta keskmine hind oli 43,14 EUR/MWh (Põhjamaade keskmine oli 38,10 EUR/MWh). Kuna aasta kokkuvõttes ületas elektri tootmine tarbimist 45%, siis netoekspordiks jäi 3,6 TWh. Taastuvenergia allikate toodang oli 1,15 TWh, moodustades kogu elektrieneriga tarbimisest 12,6%. Eesti turuosalised ostsid siseriikliku tarbimise katteks elektrienergiat kokku 91,3% ulatuses tarbimisest ehk 7,3 TWh.

Kuna Läti liitus 3. juunil 2013. aastal Nord Pool Spot'i süsteemiga, siis selle tõttu on keskmine aasta hind teiste riikidega võrreldes kõrgem. Kui 2013. aastal erines Eesti hind 1,98 EUR/MWh võrra Soome piirkonna hinnast, olles ligikaudu 5% Soome hinnast kallim, siis tänu EstLink 2 avamisele oli 2014. aasta esimeses kvartalis hindade erinevus vaid 0,58 EUR/MWh ehk võrreldes 2013. aastaga oli erinevus vähenenud 70% (vt tabel 1).

	Keskmine 2013	Max	Min	Keskmine 2012
NPS süsteem	38,1	109,55	1,38	21,2
NPS Soome	41,16	210,01	1,38	36,64
NPS Eesti	43,14	210,01	5,08	39,2
NPS Läti	52,4	210,01	5,08	-

Tabel 1: 2013. aasta Nord Pool Spot'i hinnad ning võrdlus 2012. aasta keskmise hinnaga (EUR/MWh). Allikas:Eleringi toimetised [10], 2014.

Kogu võrku sisenenud elektrit oli kokku 14 071 GWh, millest 82,8% ehk 11 655 GWh moodustas sisemaine tootmine, millest omakorda 10% taastuvenergia (1151 GWh). Välisliinidelt imporditi elektrit 2416 GWh ulatuses ning eksporditi suuruses 6011 GWh ehk eksport ületas importi 60%. Kogu sisemaine tarbimine koos võrgukadudega oli 8060 GWh. (Eleringi toimetised [10], 2014)

2014. aasta elektrituru täielik analüüs on veel ilmumata.

3 Ühemõõtmelised aegridade mudelid

Selle peatüki koostamisel on kasutatud järgmiseid allikaid: Kangro [11], Shumway, Stoffer [12], Box, Jenkins, Reinsel [13] ja Wikipedia kodulehekülg [16].

Ühemõõtmelisteks aegridade mudeliteks nimetatakse mudeleid, mis kasutavad ainult ühe oleku/tunnuse minevikuväärtuseid. Näiteks, kui soovitakse ennustada Eesti elanike arvu järgmisel aastal, siis kasutatakse selleks eelnevate aastate elanike arvu. Sellisel juhul ei ole oluline, kas elanike arvu mõjutab ka näiteks sisserändajate arv või mõni muu tegur, sest kasutatakse vaid vastava vaatluse ajaloolisi väärtuseid. Kuna minevikuandmed sisaldavad enamasti ka juhuslikkust, mis omakorda võib kaasa tuua riske, siis nende vältimiseks püütakse tulevast käitumist nö ära arvata, et ebasoodsaid olukordi tulevikus vältida. Et sellist analüüsi teostada on vaja rakendada aegridade teooriat. Lihtsalt öeldes nimetatakse aegreaks numbrite jada, mis kirjeldab suuruse ajalist muutumist ja sisaldab teatud juhuslikkuse komponenti.

Definitsioon 3.1. *Aegrida nimetatakse pidevaks, kui mõõtmised toimuvad pidevalt ning vastupidisel juhul diskreetseks.*

Antud töös käsitletakse aegridu, mis on diskreetsed, kuna mõõtmised on tehtud kord päevas. Aegridade kontekstis on oluline, et erinevatele ajamomentidele vastavad väärtused ei pea olema sama jaotusega ning ei saa eeldada, et rea moodustaksid mõne konkreetse jaotusega juhusliku suuruse sõltumatud vaatlused.

Enne statsionaarsuse mõiste kasutamist on vaja selgitada järguga k momentide tähendust.

Definitsioon 3.2. *Juhusliku vektori $(X_1, ..., X_m)$ järguga k momentideks nimetatakse suuruseid*

$$m_{\alpha_1,...,\alpha_m} = E(X_1^{\alpha_1}...X_m^{\alpha_m}),$$

kus $\alpha_i, i = 1, ..., m$ on mittenegatiivsed täisarvud ning $\sum_{i=1}^{m} \alpha_i = k$.

Definitsioon 3.3. *(Tugevalt) statsionaarseks protsessiks nimetatakse protsessi $(Z_t)_{t\in\mathbb{Z}}$, mille korral iga $k = 1, 2, ...$ ja kõikide ajapunktide $t_1, t_2, ..., t_k$ korral väärtuste komplekt*

$$\{Z_{t_1}, Z_{t_2}, ..., Z_{t_k}\}$$

on sama jaotusega kui ajanihkega $h = 0, \pm1, \pm2, ...$ väärtuste komplekt

$$\{Z_{t_1+h}, Z_{t_2+h}, ..., Z_{t_k+h}\}.$$

Seega, tulenevalt definitsioonist 3.3, on juhuslikud vektorid sama jaotusega parajasti siis, kui vastavad tõenäosused on kõikide reaalarvude $c_1, ..., c_k$ korral samad.

Definitsioon 3.4. *K-järku nõrgalt statsionaarseks protsessiks nimetatakse protsessi, mille korral iga $k = 1, 2, ...$ ja kõikide ajapunktide $t_1, t_2, ..., t_k$ väärtuste komplektide $\{Z_{t_1}, Z_{t_2}, ..., Z_{t_k}\}$ ja $\{Z_{t_1+h}, Z_{t_2+h}, ..., Z_{t_k+h}\}$ kõik kuni k-järku momendid on võrdsed.*

Kuna teist järku nõrgalt statsionaarsuse eeldusest järeldub, et suuruste Z_t ja Z_{t+k} korrelatsioon sõltub ainult ajamomentide vahest k, siis saab anda definitsiooni k-ndat järku autokorrelatsioonile.

Definitsioon 3.5. *K-ndat järku autokorrelatsiooniks nimetatakse suurust*

$$\rho_k = \frac{E[(Z_t - \mu)(Z_{t+k} - \mu)]}{\sqrt{E[(Z_t - \mu)^2]E[(Z_{t+k} - \mu)^2]}} = \frac{E[(Z_t - \mu)(Z_{t+k} - \mu)]}{\sigma_z^2}.$$

Autokorrelatsioonide abil saab sobitada autoregressiivseid ehk AR(p) (3.1) mudeleid.

Definitsioon 3.6. *Aegrea mudelit nimetatakse ARIMA tüüpi mudeliks, kui Z_t avaldub lõpliku arvu varasemate $Z_i, i < t$ ja lõpliku arvu sõltumatute sama jaotusega häirituste $A_i, i \leq t$ lineaarkombinatsioonina kujul*

$$Z_t = \phi_0 + \sum_{t=1}^{p} \phi_i Z_{t-i} + A_t + \sum_{t=1}^{q} \psi_i A_{t-i}.$$

ARIMA tüüpi mudelid sobivad juhul, kui võib eeldada, et rida ise või sobivalt teisendatud rida vastab statsionaarsele protsessile. Mittestatsionaarse rea korral võib keskmine ja/või varieeruvus ajas selgelt muutuda, samuti autokorrelatsioonid võivad kahaneda järgu kasvades aeglaselt või esimene autokorrelatsioon on väga suur ehk peaaegu võrdne ühega. Kui juhuslik protsess ei ole statsionaarne, st üks või mitu eelnevalt kirjeldatud tingimust kehtib, siis tuleks aegreast eemaldada trend või rida diferentseerida. Diferentsimise vajadusele saab kinnitust Phillips-Perroni testiga. Phillips-Perroni testi nullhüpoteesiks on see, et esimest järku diferentsimine on vajalik. Alternatiivseks hüpoteesiks on see, et rida on statsionaarne (vt täpsemalt Wikipedia kodulehekülg [14]). Seega, kui Phillips-Perroni testi korral p-väärtus on väikseim kui 0,05, siis ei ole mõistlik rida diferentseerida.

Käesolevas töös hakatakse sobivat mudelit otsima järgmiste protsesside seast.

Definitsioon 3.7. *Olgu $A_t, t \in \mathbb{Z}$ tsentreeritud, mittekorreleeritud ja vähemalt teist järku nõrgalt statsionaarne ning protsess $Z_t, t \in \mathbb{Z}$ omagu konstantset keskväärtust μ. Lõpliku arvu kordajatega lineaarsete protsesside klassid on järgmised:*

- *Järguga p autoregressiivseteks protsessideks ehk* AR(p) *protsessideks nimetatakse protsesse kujul*

$$\tilde{Z}_t = Z_t - \mu = \sum_{i=1}^{p} \phi_i \tilde{Z}_{t-i} + A_t. \tag{3.1}$$

- *Järku q liikuva keskmisega protsessideks ehk* MA(q) *protsessideks nimetatakse protsesse kujul*

$$\tilde{Z}_t = Z_t - \mu = A_t - \sum_{i=1}^{q} \theta_i A_{t-i}. \tag{3.2}$$

- ARMA(p, q) *protsessideks nimetatakse protsesse kujul*

$$\tilde{Z}_t = Z_t - \mu = \sum_{i=1}^{p} \phi_i \tilde{Z}_{t-i} + A_t - \sum_{i=1}^{q} \theta_i A_{t-i}. \tag{3.3}$$

Sobiva mudeli leidmisel uuritakse aegrea autokorrelatsioone ja osaautokorrelatsioone ning vaadeldakse mitu väärtust satub väljapoole 95% veapiire. Selleks, et teada saada, kas terve väljaarvutatud autokorrelatsioonikordajate komplekt võib vastata sõltumatutele juhuslikele suurustele, sobib Ljung-Box test (vt Wikipedia kodulehekülg [15]).

Ljung-Box testi aluseks on järgnev tulemus.

Kui rida Z_t vastab sõltumatutele sama jaotusega juhuslikele suurustele, siis suurus

$$Q = N(N+2)\sum_{p=1}^{m} \frac{r_p^2}{N-p},$$

kus N on vaatluste arv, $r_p, p > 1$ on lõpliku pikkusega aegrea põhjal arvutatud hinnangud ning m näitab, kui mitmest erinevast autokorrelatsioonist koosnevat rühma testitakse, on asümptootiliselt m vabadusastmega χ^2 jaotusega ning selle põhjal saab hinnata tõenäosust, et vaadeldavad kordajad vastavad sõltumatutele juhuslikele suurustele.

Kuna aegridade puhul tekib jääkide aegrida, tuleb testi modifitseerida, võttes arvesse sobitamisel kasutatud parameetrite arvu. Kui leitud mudeli korral on Ljung-Box testi põhjal saadud p-väärtused suuremad kui 0,05, siis loetakse mudel sobivaks.

Kui vaadeldavale aegreale on leitud mitu sobivat mudelit, siis oluliseks võrdluse mõõdikuks suure valimi korral võiks kasutada Akaike informatsiooni kriteeriumit ning väikese valimi korral annab täpsema tulemuse parandatud Akaike informatsiooni kriteerium.

Definitsioon 3.8. *Akaike informatsiooni kriteeriumiks* (AIC) *nimetatakse suurust*

$$\mathrm{AIC} = 2k - 2\ln(L) \tag{3.4}$$

kus k on mudelis kasutatud parameetrite arv ning L on aegrea tõepära sobitatud mudeli korral.

Definitsioon 3.9. *Parandatud akaike informatsiooni kriteeriumiks* (AICc) *nimetatakse suurust*

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2k(k+1)}{n-k-1},$$

kus k on mudelis kasutatud parameetrite arv ning n on aegrea vaatluste arv.

Erinevaid mudeleid saab võrrelda ka kasutades prognoosivigade mõõdikuid. Antud töös kasutatakse kolme suurust, kus n on vaatluste arv, z_i on huvipakkuva aegrea väärtus kohal i ja $\tilde{z}_i$ on aegrea prognoositud väärtus kohal i.

- Keskmine absoluutviga ehk MAD

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|z_i - \tilde{z}_i|$$

- Ruutkeskmine viga ehk RMSE

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_i - \tilde{z}_i)^2}$$

- Keskmine suhteline viga ehk MAPE

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|z_i - \tilde{z}_i|}{z_i}$$

4 Mitmemõõtmelised aegridade mudelid

Eelnevas peatükis käsitletud ühemõõtmelised mudelid võimaldasid ennustada vaid uuritava aegrea minevikuväärtuse abil. Kahjuks ei pruugi sellised prognoosid alati kõige täpsemaid tulemusi anda. Selleks, et prognoose parandada tuleks lisaks vaadeldavale aegreale kasutada ka teisi andmeid, mille käitumine võib mõjutada esialgset rida. Näiteks tarbija ostujõu käitumist mõjutab tema sissetulek või transpordiettevõtete käive sõltub reisijate arvust.

Sageli eeldatakse, et huvipakkuva suuruse väärtus sõltub abitunnustest lineaarselt. Tavalise regressioonmudelite puhul eeldatakse, et nii saadavad prognoosivead on omavahel sõltumatud. Aegridade puhul ei ole aga selline eeldus enamasti õigustatud. Kui eeldada, et rea hetkeväärtuse prognoosimisel regressorite lineaarfunktsiooniga tekkivad vead ε_t, kus mudel on kujul

$$Z_t = \sum_i c_i \mathrm{regressor}_{i,t} + \varepsilon_t,$$

vastavad mingile ARIMA tüüpi protsessile, siis on tegemist lineaarse regressiooniga ARIMA tüüpi vigadega ehk ARIMAX mudeliga. Sellise mudeli leidmise korral sobitatakse andmetele tavalise regressiooni abil mitmese regressiooni mudel ning seejärel analüüsitakse vigade käitumist. Vigade käitumise põhjal valitakse ARIMA mudeli kuju vigade ε_t jaoks ning leitakse ARIMAX mudeli parameetrid. Parameetrite leidmine toimub enamasti kas suurima tõepära meetodi või ühesammuliste prognoosivigade minimiseerimise teel. Kui mudel on leitud, siis headuse mõõtmisel on oluline prognoosivigade sõltumatus.

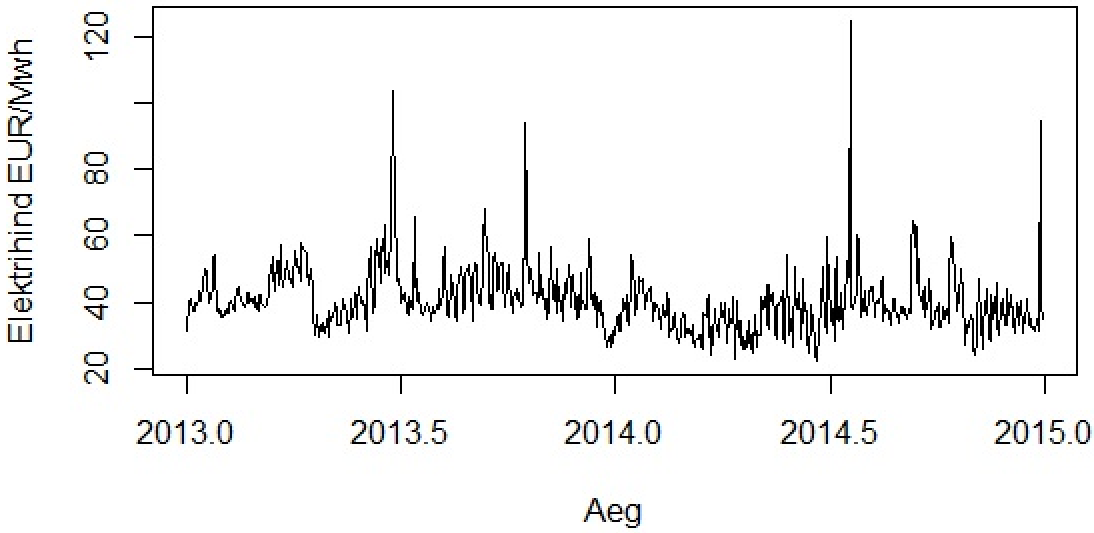
Erijuhul, kui vead ε_t on sõltumatud, siis on tegemist mitmese lineaarse regresiooniga. Kuna aegridade puhul vigade sõltumatuse eeldus tihti ei kehti, siis sellise mudeli põhjal saadud prognoosid võivad anda väga valesid tulemusi.

5 Analüüs ühemõõtmeliste mudelitega

5.1 Eesti elektrienergia hinna aegrea analüüs

Töös on kasutatud Eesti elektrienergia hinna päevaseid andmeid alates 1. jaanuar 2013 kuni 31. märts 2015. [Allikas:[7]] Kokku on andmestikus 820 rida. Andmed on töö koostaja hinnangul korrektsed, st aegreas ei leidu ühtegi puuduvat väärtust. Aegrida jaotatakse kaheks – treening- ja testandmestikuks. Treeningandmestiku peal treenitakse ehk leitakse parim mudel ja tehakse prognoosid ning testandmestikku kasutatakse selleks, et kontrollida, kui hästi leitud mudel suudab uusi (sobitamisel mittekasutatud) andmeid prognoosida. Treenimisel võrreldakse mudeleid Akaike informatsiooni kriteeriumi põhjal, lisaks võrreldakse sobitatud mudelite prognoosivõimet testperioodil mõõdikute MAD, RMSE ja MAPE abil. Treeningaegrea moodustavad andmed alates 1. jaanuar 2013 kuni 31. detsember 2014 ning testaegrea 1. jaanuar kuni 31. märts 2015.

Esmalt uuritakse Eesti elektrienergia hinna graafikut läbi kahe aasta ehk treeningandmestikku.



Joonis 1: Eesti päevased elektrienergia hinnad (EUR/MWh)

Uurides joonist 1, on märgata nelja suurt elektrihinna hüpet, mis eristuvad keskmisest ligikaudu kaks või isegi kolm korda. Esimene hüpe toimus 25. juunil 2013. aastal, kui elektrihind tõusis rekordilisele, 103,85 EUR/MWh, tasemele. Kuna elektriseadmete ja -ühenduste hooldustööd jäetakse üldjuhul suveperioodi, kus tarbimine on väiksem, siis antud juhul langesid naaberriikide hooldused ühele ja samale perioodile, mis ajas elektrihinna erakordselt kõrgeks. Narva elektrijaamades oli neli energiablokki remondis, Venemaa poolt Soome minev ühendus ei töötanud, Soome tuumaelektrijaam seisis, Leedus oli gaasijaam hoolduses, Valgevene-Leedu piiril toimusid hooldustööd ning Estlink 1 oli hoolduses, mille tulemusena tekkis Lätis ja Leedus elektripuudus. (Koppel [17], 2013)

16. oktoobril 2013. aastal, kui elektrihind oli küündinud 93,94 EUR/MW tunnini, tulenes kõrge hind Soome tootmis- ja ülekande probleemidest ning seetõttu liikus Eestist elekter Soome. (Kallas [18], 2013)

Läbi aegade üks kõige kõrgemaid päevaseid elektrihindasid oli 124,77 EUR/MWh 20. juulil 2014. aastal, sest sellel ajal teostati mitmeid elektri- ja tuumajaamade ning EstLink'i hooldustöid. (Ärileht [19], 2014)

Viimane kõrge elektrihinna hüpe 29. detsembril 2014. aastal oli tingitud mitme teguri koosmõjul. Määravaks said külm ilm, Soome ja Rootsi vahelise elektriühenduse vähenenud võimsus ning tööseisak mitmes elektrijaamas. (Õepa [20], 2014)

Joonise 1 põhjal trendi ei leidu ning ei paista ka vastuolusid statsionaarsuse eeldustega. Phillips - Perroni test annab p-väärtuseks 0,01, seega kinnitab, et diferentseerimine antud juhul otstarbekas ei ole. Kuna andmed on sobival kujul, siis saab hakata leidma sobivat ARIMA tüüpi mudelit. Selleks vaadeldakse antud aegrea autokorrelatsioonide ja osaautokorrelatsioonide graafikuid (vt lisa 1 joonis 14).

Lähtudedes aegridade kursusel õpitust, võiks osaautokorrelatsioonide järgi proovida mudeleid AR(1) ja AR(3). Kuna autokorrelatsioone on väga palju piiridest väljas, siis MA mudeleid ei ole mõtet otsida. ARMA mudelite seast võiksid sobida ARMA(1,0,1), ARMA(2,0,1) ning ARMA(1,0,2). Kui vaadelda segamudeleid perioodiga 7, siis mittesesoonsesse osasse tasuks proovida ARMA(1,0,1), AR-

MA(1,0,2) ja ARMA(2,0,1) ning samuti AR(1) ja AR(3). Kuna sesoonsete autokorrelatsioonide hulgas on kolm selgelt piiridest väljas, siis sesoonses osas tasuks otsida sama arvu parameetritega MA(3) mudelit, kuid uurida, kas äkki sobivad ka juba väiksema parameetrite arvuga, näiteks ARMA(1,0,2), ARMA(2,0,1) ja ARMA(1,0,1) mudelid. Samuti tasuks sesoonsete osaautokorrelatsioonide põhjal proovida ka AR(1) mudelit. Mudelite sobitamisel kasutatud R skripti leiab lisast 2.

Mudelite omavahelisel võrdlemisel kasutatakse Akaike informatsiooni kriteeriumit. Peaaegu võrdsete AIC väärtuste korral eelistatakse väiksema parameetrite arvuga mudelit. Eelnevalt kirjeldatud mudelite hulgast on sobivaimad ARIMA$(1,0,1) \times (2,0,1)_7$, ARIMA$(1,0,1) \times (1,0,2)_7$ ning ARIMA$(3,0,0) \times (2,0,1)_7$. Akaike informatsiooni kriteeriumite väärtused on vastavalt 4945,91, 4946,21 ja 4947,27. Kuna esimesel kahel mudelil on võrdne arv parameetreid, siis valitakse parimaks väiksema AIC väärtusega ARIMA$(1,0,1) \times (2,0,1)_7$ mudel.

Joonis 2: Mudeli ARIMA$(1,0,1) \times (2,0,1)_7$ Ljung-Box testi p-väärtused

Ka joonise 2 põhjal mudel ARIMA$(1,0,1) \times (2,0,1)_7$ sobib, sest vaadeldavas piirkonnas on kõik Ljung-Box testi p-väärtused suuremad kui 0,05.

Mudeli kuju on välja toodud tabelis 2.

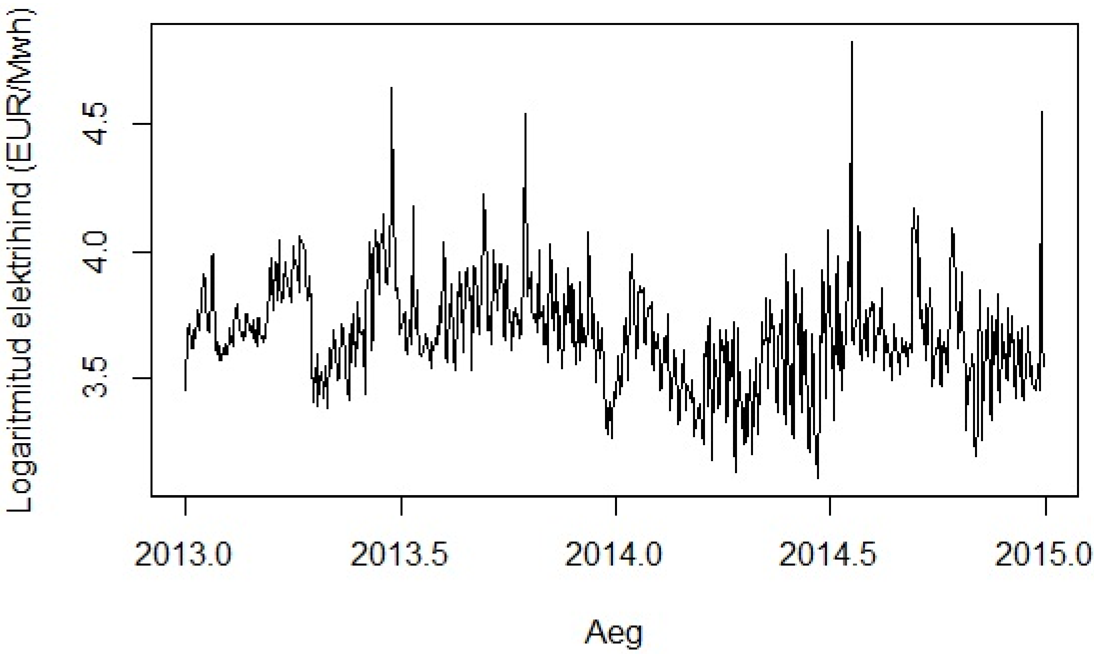
	ar1	**ma1**	**sar1**	**sar2**	**sma1**	**vabaliige**
kordajad	0,7881	-0,3596	1,1505	-0,1548	-0,9713	40,3418
s.e	0,0437	0,0659	0,0442	0,0426	0,0209	2,6055

Tabel 2: Mudeli ARIMA$(1,0,1) \times (2,0,1)_7$ kuju

Kui mudeli kordajate vastavad kahekordsed standardhälbed on väiksemad kui mudeli kordajate absoluutväärtused, siis eeldades vigade normaaljaotust saab need lugeda oluliseks olulisusnivool 0,05. Antud juhul on kõik kordajad koos vabaliikmega olulised, seega loetakse mudel ARIMA$(1, 0, 1) \times (2, 0, 1)_7$ sobivaks.

5.2 Eesti elektrienergia logaritmitud hinna aegrea analüüs

Kuna elektrihinna varieeruvus läbi kahe aasta on küllaltki suur ja esineb mitmeid suuremaid hinna kõikumisi (vt joonis 1), siis uuritakse lisaks tavalisele Eesti elektrienergia hinna aegreale logaritmitud rida. Logaritmimisega püütakse vähendada väärtuste vahelisi suuri muutuseid ja seeläbi leida väiksema veaga mudel.

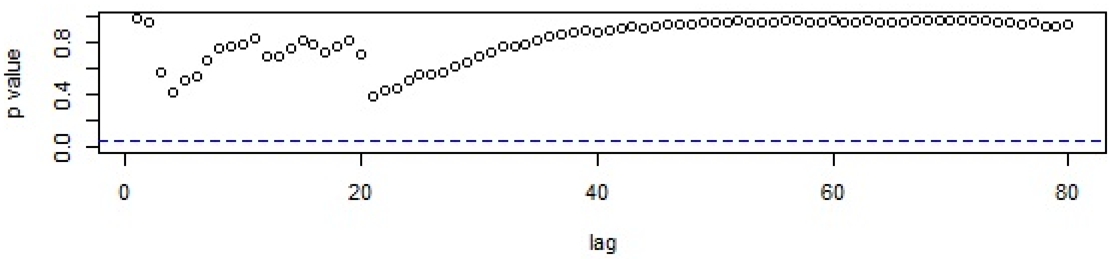


Joonis 3: Eesti päevased elektrienergia logaritmitud hinnad (EUR/MWh)

Logaritmitud real ei paista olevat trendi ega ka vastuolusid statsionaarsuse eeldustega. Phillips-Perroni testile tuginedes ei ole vaja rida diferentseerida.

Mudelite otsimine käib samamoodi nagu eelnevalt. Joonise pealt vaadatakse, kui palju jääb auto- või osaautokorrelatsioone piiridest välja ning otsustatakse kas AR, MA, ARMA või ARIMA mudeli kasuks. Antud rea autokorrelatsioonide ja osaautokorrelatsioonide graafik on leitav lisast 1 (vt joonis 15) ning mudelite sobitamisel kasutatud R skript lisast 2. Kui sobilikke mudeleid on mitu, valitakse väikseima AIC väärtusega mudel.

Kolm kõige sobivamat mudelit on ARIMA$(3,0,0) \times (2,0,1)_7$, ARIMA$(1,0,2) \times (2,0,1)_7$ ning ARIMA$(2,0,1) \times (2,0,1)_7$, AIC väärtustega vastavalt -773,73, -774,93 ja -775,17. Kuna kõikides mudelites on sama palju parameetreid, siis parimaks valitakse väikseima AIC väärtusega ARIMA$(2,0,1) \times (2,0,1)_7$ mudel. Kuna joonisel 4 on kõik Ljung-Box testi p-väärtused üleval pool 0,05 piiri, siis saab lugeda mudeli sobivaks.



Joonis 4: Mudeli $ARIMA(2,0,1) \times (2,0,1)_{7\,\log}$ Ljung-Box testi p-väärtused

Selles peatükis leitud parima mudeli AIC väärtus erineb oluliselt eelnevas peatükis leitud parimast mudelist. See on ka loomulik, sest kasutatud on erinevaid andmeid, st antud juhul logaritmitud rida ning eelnevalt originaalandmete rida. Seega ei ole mõistlik AIC väärtuseid omavahel võrrelda.

Parima, logaritmitud reale leitud, mudeli kuju näeb välja järgmiselt:

	ar1	**ar2**	**ma1**	**sar1**	**sar2**	**sma1**	**vabaliige**
kordajad	1,2396	-0,3031	-0,7166	1,1463	-0,1505	-0,9626	3,6678
s.e	0,1182	0,0898	0,1045	0,0438	0,0426	0,0204	0,0987

Tabel 3: Mudeli $ARIMA(2,0,1) \times (2,0,1)_{7\,\log}$ kuju

Selline mudel avaldub kujul

$$(1-1,2396B+0,3031B^2)(1-1,1463B^7+0,1505B^8)(W_t-3,6678)=$$
$$=(1-0,7166B)(1-0,9626B^7)A_t,$$

kus $W_t=\ln(Z_t)$.

Peale sulgude avamist

$$W_t-1,2396W_{t-1}+0,3031W_{t-2}-1,1463W_{t-7}+1,57145W_{t-8}-0,5341W_{t-9}+$$
$$+0,04562W_{t-10}-0,00098=A_t-0,7166A_{t-1}-0,9626A_{t-7}+0,6898A_{t-8}.$$

Asendades eelnevas võrrandis $W_t=\ln(Z_t)$ ja avaldades $\ln(Z_t)$, omandab mudel kuju

$$\ln(Z_t)=1,2396\ln(Z_{t-1})-0,3031\ln(Z_{t-2})+1,1463\ln(Z_{t-7})-1,57145\ln(Z_{t-8})+$$
$$+0,5341\ln(Z_{t-9})-0,04562\ln(Z_{t-10})+0,00098+A_t-0,7166A_{t-1}-$$
$$-0,9626A_{t-7}+0,6898A_{t-8}.$$

Võttes mõlemad pooled e astmesse, avaldub mudeli ümardatud lõplik kuju järgmiselt

$$Z_t=e^{1,24\ln(Z_{t-1})-0,303\ln(Z_{t-2})+1,146\ln(Z_{t-7})-1,571\ln(Z_{t-8})+0,534\ln(Z_{t-9})-0,046\ln(Z_{t-10})}*$$
$$*e^{0,001+A_t-0,717A_{t-1}-0,963A_{t-7}+0,69A_{t-8}}.$$

5.3 Eesti elektrienergia hinna originaal- ja logaritmitud aegreale leitud parimate mudelite võrdlus ühepäevaste prognooside baasil

Eesmärgiks on uurida järgmise päeva elektrihinna ennustust, seega võrreldakse treeningandmestiku abil sobitatud mudelite ühesammulisi prognoose testandmestiku väärtustega. Kuna treeningandmestik on kuni 31. detsember 2014, siis ühesammulised prognoosid tehakse perioodi jaoks 1. jaanuar kuni 31. märts 2015.

Ennustused tehakse vaid ühe sammu ehk antud juhul ühe päeva jaoks. Ühesammuline prognoosimine tähendab seda, et iga uue ennustuse tegemiseks kasutatakse varem sobitatud mudeli korral kogu infot, mis on ennustuse tegemise momendiks olemas. Pikki prognoose pole mõtet teha, sest statsionaarse mudeli abil arvutatud prognoosid koonduvad keskväärtuse hinnanguks. Nagu ka Eesti Energia endine juhatuse esimees Sandor Liive on öelnud, et elekrihinna ennustamine on tänamatu tegevus. Kui oleks vaja pikemat perioodi ennustada, saaks leitud mudeleid kasutada, kuid hinna prognoosimine on küllalt raske ülesanne ka juba järgmise päeva jaoks, seega keskendutakse antud töös vaid ühesammulistele ennustustele.

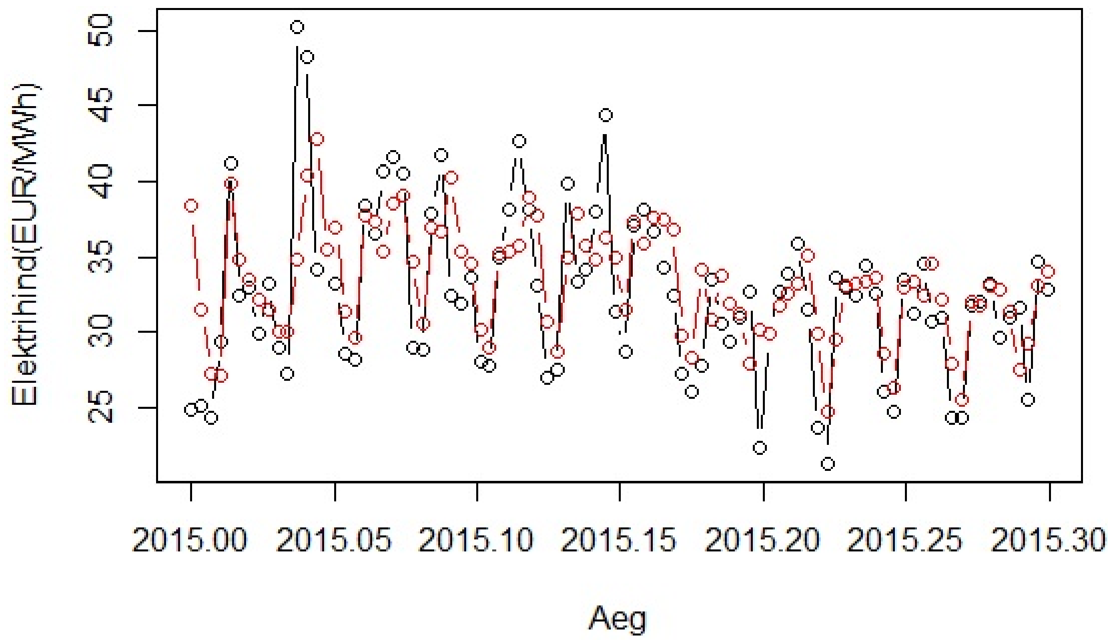
Prognooside leidmisel on kasutatatud peatükkides 5.1 ja 5.2 leitud parimaid mudeleid, milleks on vastavalt ARIMA$(1,0,1) \times (2,0,1)_7$ originaalrea jaoks ja ARIMA $(2,0,1) \times (2,0,1)_7$ logaritmitud rea korral. Selles peatükis enam uuesti mudelit ei sobitata, vaid fikseeritakse treeningandmestiku peal leitud parima mudeli koefitsiendid ning kasutatakse samasuguseid parameetreid kogu aegrea jaoks. Prognoosivigade mõõdikud leitakse mudeli poolt testperioodi jaoks tehtud ühesammuliste prognooside ja tegelike väärtuste erinevuste põhjal.

	MAD	**RMSE**	**MAPE**
ARIMA$(1,0,1) \times (2,0,1)_7$	3,50	4,47	11,35%
ARIMA$(2,0,1) \times (2,0,1)_{7\,\log}$	2,99	4,03	9,43%

Tabel 4: Ühesammuliste prognoosivigade mõõdikud

Tabelist 4 selgub, et logaritmitud rea korral on kõik prognoosivigade mõõdikud väiksemad. Keskmine suhteline viga on peaaegu 2% väiksem kui originaalse hinna aegrea korral. Seega annab logaritmimine paremaid tulemusi.

Kuna AIC väärtuseid ei saa üks-ühele võrrelda, siis prognoosivigade mõõdikute korral saab väita, et parem mudel on ARIMA$(2,0,1) \times (2,0,1)_7$, mis on leitud Eesti elektrienergia logaritmitud hinna reale. Et saada rohkem aimu parima mudeli prognoositud ja tegelike väärtuste erinevusest, uuritakse neid väärtuseid koos ühel graafikul.



Joonis 5: Ühesammulised ennustused (katkendlik punane joon) mudeliga $ARIMA(2,0,1) \times (2,0,1)_{7\,\log}$ *ja testaegrea (must joon) väärtused*

Jooniselt 5 on näha, et prognoosid nö ei püüa kinni väga kõrgeid elektrihindasid. Seega on antud mudeliga võimatu ekstreemseid väärtuseid prognoosida. Tundub, et madalamate hindade ennustamisega saab mudel enamiku ajast paremini hakkama. 2015. aasta 12. jaanuaril on tegelik elektrihind olnud küllalt kõrge, kõikudes 50 EUR/MWh ümbruses, aga ennustus on ligikaudu 35 EUR/MWh. Joonise 5 põhjal tundub ka, et veebruari keskpaigast alates on prognoosid tegelike väärtustega palju sarnasemad, kui perioodi esimeses pooles. Kolme kuu põhjal ei ole hea üldistusi teha, kuid võib tõepoolest olla, et uue aasta alguses on hinna käitumist raskem ennustada. Kui võrrelda mudeli ühesammuliste prognooside jääke, siis 12. jaanuaril on see lausa -15,39, mis on kogu vaadeldava perioodi suurim. Väikseim mudeli prognoosijääk on -0,01, mis realiseerus 15. jaanuaril, kui tegelik elektrihind oli 35,49 EUR/MWh ning ennustus oli 35,48 EUR/MWh.

Leitud mudeli põhjal ei ole siiski võimalik järgmise päeva elektrihinda kuigi täpselt ennustada, sest prognoosimisel tekkinud vead on liiga suured. Kuna antud tulemuse saamiseks on kasutatud vaid ajaloolisi hindasid, siis uuritakse lisaks faktoreid, mis võiksid mõjutada Eesti elektrienergia hinda.

6 Eesti elektrienergia hinna aegrea analüüs mitmemõõtmeliste mudelitega

Eelnevalt kasutati Eesti elektrihinna ennustamiseks vaid minevikuväärtuseid ning prognoosid erinesid tegelikkusest küllaltki palju. Selleks, et prognoose täpsustada, saab kasutada mõne teise juhusliku suuruse minevikuväärtuseid. Alapeatükis 2.2 oli välja toodud, et elektrihinna kujunemist mõjutavad õhutemperatuur ja kliima. Selline sõltuvus on iseenesest väga loomulik, sest talveajal kasutatakse elektrienergiat märgatavalt rohkem kui suveajal. Samamoodi on talvel ka küttekulud suuremad. Suvel võivad omakorda olla kasutusel konditsioneerid, mis mõjutavad tarbimist ja seeläbi ka elektrihinda. Seega võib elektritarbimine olla aastaajati väga erinev.

Töös uuritakse, kas Eesti päevaste õhutemperatuuride aegrida aitab prognoose täpsemaks muuta. Temperatuuride rida on saadud Eesti Meteoroloogia ja Hüdroloogia Instituudist. Treeningaegrida on perioodil 1. jaanuar 2013 – 31. detsember 2014 ning testaegrida 1. jaanuar – 31. märts 2015. Andmed on korrektsed, st aegreas ei leidu puuduvaid väärtuseid. Temperatuuri kasutamisel regressorina eeldatakse töös, et järgmise ööpäeva temperatuuriinfo on täpselt ette teada.

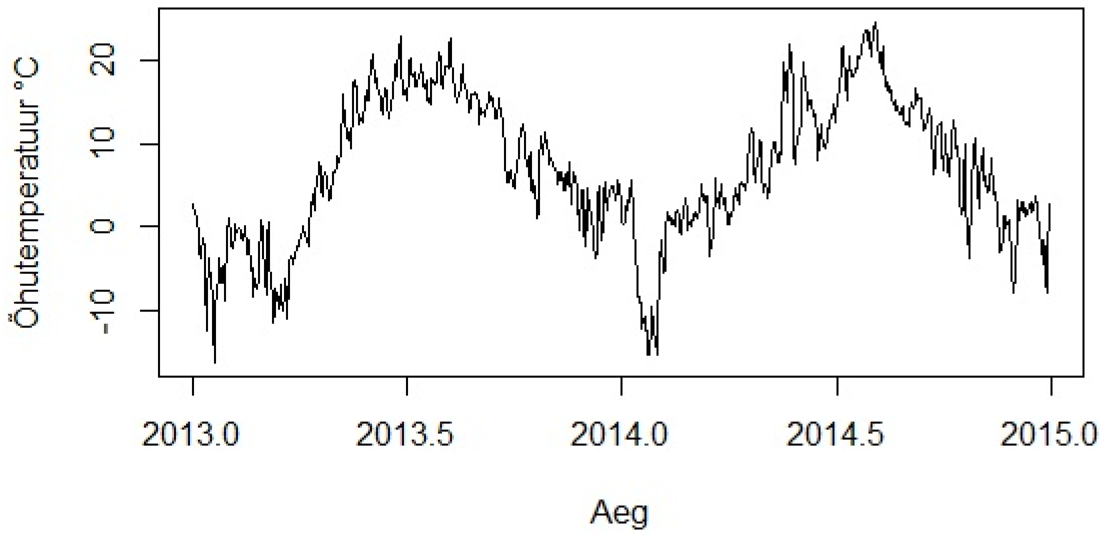
Kui kasutatakse kliima andmeid, võib tekkida küsimus, kas ka sademete hulk mõjutab elektrihinda. Eestis võiks sademete hulk mõjutada hüdroelektri tootmist, mille järel avalduks mõju elektrihinnale. Kuna aga Eestis on ligikaudu 600 000 majapidamist ja hüdroelektrijaamade poolt toodetud elektriga oleks võimalik varustada umbes 3500 majapidamist, mis teeb 0,6% kõikidest varustatavatest majapidamistest, siis olulist mõju ei saa siin olla.(Eesti Energia koduleheküljg [1], 2015; Eesti Statistikaameti andmebaaside kodulehekülg [21], 2015)

Kuna elektrihind peegeldab osaliselt ka tarbija käitumist, siis on alust arvata, et töö- ja puhkepäevade elektrihinnad käituvad erinevalt. Selleks kasutatakse aegrida, kus puhkepäevadele ning riigipühadele vastab 1 ja tööpäevadele 0. Treening- ja testaegrea pikkused on samad ehk vastavalt kaks aastat ja kolm kuud. Kuna järgmise päeva kohta on alati teada, kas tegemist on töö- või puhkepäevaga, siis sellise regressori kasutamine ei sisalda endas juhuslikkust.

Viimasena uuritakse õhutemperatuuride ning töö- ja puhkepäevi eristavaid aegridasid regressorina koos.

6.1 Eesti elektrihinna analüüs kasutades õhutemperatuure

Esmalt uuritakse Eesti õhutemperatuuride graafikut läbi kahe aasta.



Joonis 6: Eesti päevased õhutemperatuurid (°C)

Rida ei ole statsionaarne, kindlat keskmist ei paista olevat ja esineb perioodiline kõikumine. Perioodiks on üks aasta – suvel saavutab temperatuur maksimumi ning aasta alguses ja lõpus langeb temperatuur miinimumi lähedale.

Reaalses elus tõstavad elektrihinda nii kõrged kui ka väga madalad temperatuurid. Kui kasutada temperatuuride aegrida regressorina, siis ei ole lootust saada head tulemust, kuna regressorina kasutatud rida käitub minimaalsete väärtuste korral vastupidiselt maksimaalsetele väärtustele. Seega mõjutaksid kõrged õhutemperatuurid elektrihinda täpselt vastupidi kui madalad õhutemperatuurid, aga reaalses elus on nende mõjutus käitumuslikult sama.

Seega võiks temperatuuride aegrea puhul kasutada tükiti lineaarseid funktsioone ehk lineaarseid splaine. Näiteks tunnuse x korral tekitada tunnused

$$\begin{cases} x_1 = \max(a - x, 0) \\ x_2 = \max(x - a, 0) \\ x_3 = \max(x - b, 0), \end{cases} \tag{6.1}$$

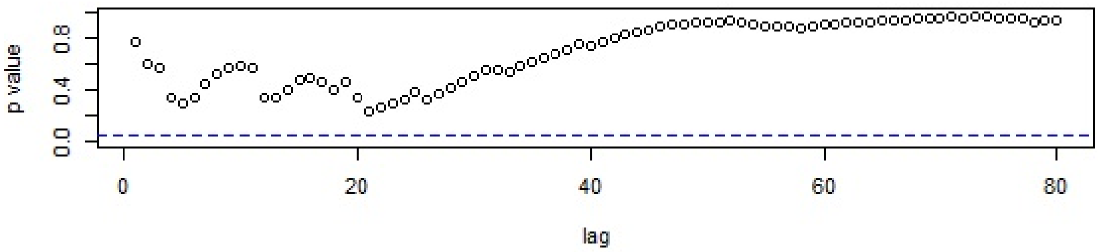
kus $b > a$ ja kasutada neid regressoritena. Sellisel juhul kasutatakse regressorina pidevaid ja tükiti lineaarseid funktsioone x-st, millel on murdepunktid kohtadel a ja b. Suurema arvu murdepunktide kasutamisel saab kirjeldada keerulisemaid mittelineaarseid sõltuvusi, kuid see toob kaasa ka vajaduse hinnata suuremat arvu parameetreid.

Tunnuseks x on mingi uuritavat aegrida mittelineaarselt mõjutav tunnus ning muutujad a ja b valitakse ennustatava- ja regressorrea kooskäitumist arvestades. Kuna regressoriks on temperatuuri aegrida, siis a ja b valimiseks on vaja hinnata temperatuure, mille korral elektritarbimine suureneb. Enamasti tekib selline olukord siis, kui kasutusele võetakse lisaseadmed, näiteks konditsioneerid ja kütteseadmed. Keskmine temperatuur suvel, kui hakatakse kasutama konditsioneere, võiks olla +15°C. Tuleb arvestada ka sellega, et päeval on temperatuurid kõrgemad ja öösel madalamad, aga parameetrite valikul lähtutakse ööpäeva keskmisest õhutemperatuurist. Keskmine temperatuur talvel, kui hakatakse kasutama kütteseadmeid, võiks olla -5°C. Seega $a = -5$, $b = 15$ ning x on Eesti päevaste õhutemperatuuride aegrida, regressorina kasutatakse funktsioone

$$\begin{cases} x_1 = \max(-5 - x, 0) \\ x_2 = \max(x + 5, 0) \\ x_3 = \max(x - 15, 0). \end{cases}$$

Sobiva mudeli otsimine käib sarnaselt eelnevaga, kuid nüüd lisatakse ka regressorread. Kuna temperatuuri mõju kirjeldav funktsioon pannakse kokku kolmest reast, siis regressoriks on kolmeveeruline maatriks. Antud rea autokorrelatsioonide ja osaautokorrelatsioonide graafik on leitav lisast 1 (vt joonis 16) ning kogu mudelite sobitamisel kasutatud R skript lisast 2. Eelistatakse mudelit, mille Akaike informatsiooni kriteerium on väikseim. Kui mõistliku parameetrite arvuga mudel on leitud, siis oluliselt suurema parameetrite arvuga mudeleid enam ei otsita.

Kõige paremaid tulemusi annavad ARIMA$(1,0,1) \times (1,0,2)_7$, ARIMA$(2,0,1) \times (2,0,1)_7$ ning ARIMA$(3,0,0)\times(2,0,1)_7$ mudelid, AIC väärtustega vastavalt 4941,83, 4940,33 ja 4942,9. Kuna kõikidel mudelitel on samasugune arv parameetreid, siis parimaks valitakse väikseima AIC väärtusega ARIMA$(2,0,1) \times (2,0,1)_7$ mudel. Sellist mudelit tähistatakse edaspidi ARIMAX$_1$.



Joonis 7: Mudeli ARIMAX$_1$ Ljung-Box testi p-väärtused

ARIMAX$_1$ mudeli kõik Ljung-Box testi p-väärtused on suuremad kui 0,05, seega saab lugeda mudeli sobivaks.

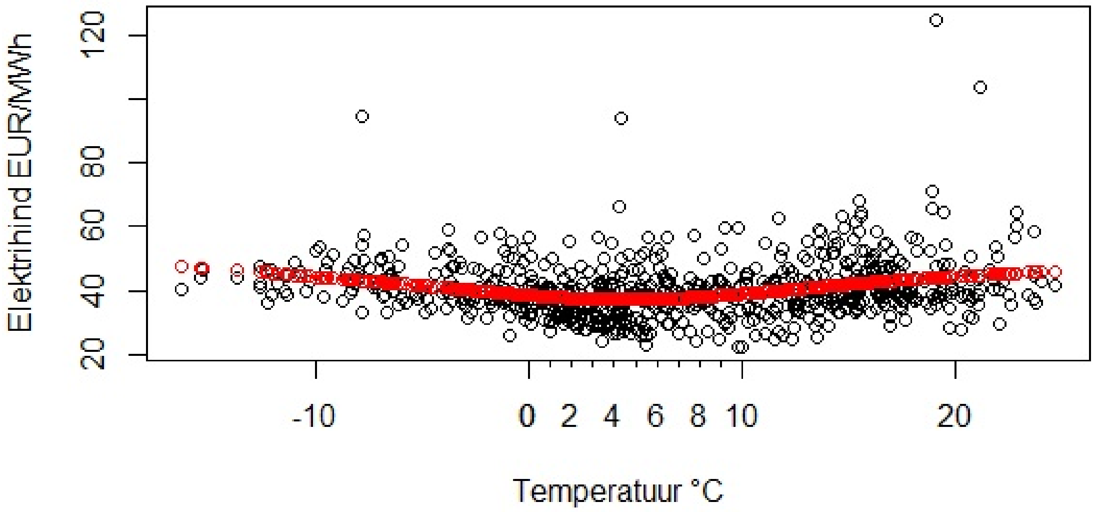
Mudel avaldub kujul:

	ar1	ar2	ma1	sar1	sar2	sma1	vabaliige	x1	x2	x3
kordajad	1,1867	-0,2415	-0,7517	1,1171	-0,1194	-0,9799	42,6262	0,6757	-0,2640	0,8009
s.e	0,1421	0,1025	0,1257	0,0439	0,0432	0,0191	4,1290	0,3342	0,1591	0,3577

Tabel 5: Mudeli ARIMAX$_1$ kuju

Tabelist 5 on näha, et tunnuse x_2 kordaja on ebaoluline, kuid kuna selle eemaldamine ei paranda AIC väärtust, siis jäetakse see mudelisse sisse.

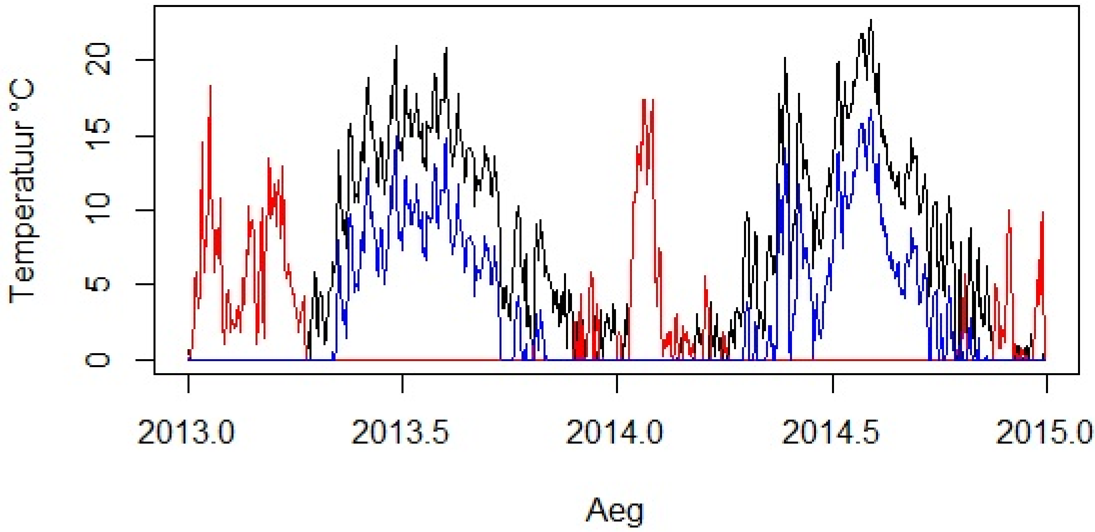
Eelnevalt leiti temperatuuri mõju kirjeldava tükiti lineaarse funktsiooni murdekohad ehk funktsioonide 6.1 parameetrid a ja b loogilise arutelu põhjal, kuid järgnevalt uuritakse, kas elektrihinna ja temperatuuri aegrea koos vaatlemine annaks alust teistsuguste parameetrite valikuks.



Joonis 8: Elektrihinna (EUR/MWh) ja temperatuuri (°C) koosmõju

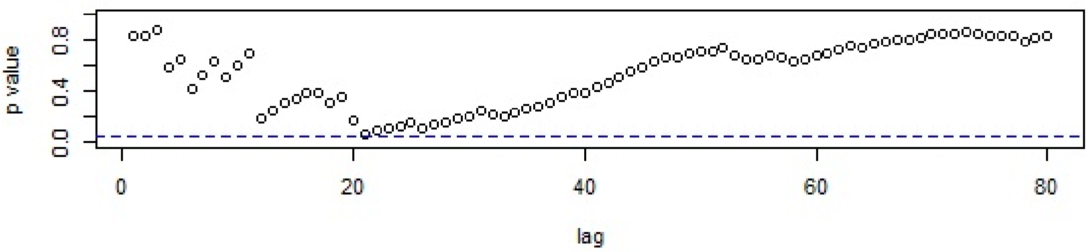
Joonise 8 põhjal tundub, et murdepunktid võiksid olla kohtadel 2°C ja 8°C. See tähendab seda, et vahemikus 2°C kuni 8°C on funktsioon konstantne ning edasi hakkab jälle kasvama.

Seega võiks regressoritena kasutada funktsioonide 6.1, kus $a = 2$ ja $b = 8$, abil temperaatuuride aegreast x moodustatud kolme aegrida. Vastavate regressorridade x_1, x_2 ja x_3 käitumine on kujutatud joonisel 9.



Joonis 9: Regressorread Eesti päevastest õhutemperatuuridest, kus funktsiooni x_1 tähistab punane joon, x_2 tähistab must joon ning funktsioon x_3 on tähistatud sinise joonega

Antud rea autokorrelatsioonide ja osaautokorrelatsioonide graafik on leitav lisast 1 (vt joonis 17) ning kogu sobitamine käib samamoodi nagu eelnevalt (vt lisa 2). Selliste regressorridade korral on kolm parimat mudelit ARIMA$(1,0,1)\times(1,0,2)_7$, ARIMA$(1,0,1)\times(2,0,1)_7$ ja ARIMA$(2,0,1)\times(1,0,2)_7$, AIC väärtustega vastavalt 4933,22, 4932,62 ning 4932,4. Kuna mudelitel on küllaltki sarnased AIC väärtused ja esimesel kahel mudelil on vähem kordajaid, siis parimaks mudeliks valitakse ARIMA$(1,0,1)\times(2,0,1)_7$ ning sellist mudelit tähistatakse edaspidi ARIMAX$_2$.



Joonis 10: Mudeli ARIMAX$_2$ Ljung-Box testi p-väärtused

Ka ARIMAX$_2$ mudeli kõik Ljung-Box testi p-väärtused on suuremad kui 0,05, seega saab lugeda mudeli sobivaks. Kuna ARIMAX$_1$ korral oli AIC väärtus suurem, siis valitakse antud peatükis parimaks mudeliks ARIMAX$_2$.

ARIMAX$_2$ mudel avaldub kujul:

	ar1	ma1	sar1	sar2	sma1	vabaliige	x1	x3
kordajad	0,7720	-0,3580	1,1501	-0,1531	-0,9766	37,9694	0,6430	0,3869
s.e	0,0467	0,0671	0,0433	0,0421	0,0203	2,5233	0,1693	0,1562

Tabel 6: Mudeli ARIMAX$_2$ kuju

Kuna ARIMAX$_2$ mudeli korral on kordaja x_2 ebaoluline, siis leitakse mudel ka ilma sellise regressorita. Kuna AIC väärtus paranes, siis jäetakse regressorite hulgast rida x_2 välja.

Eeldades vigade normaaljaotust on tabelis 6 kõik kordajad olulised olulisnivool 0,05, st kordajatele vastavad kahekordsed standardhälbed on väiksemad kui nende kordajate absoluutväärtused.

Järgnevalt tehakse kogu analüüs sarnaselt eelnevaga uuesti läbi, kuid nüüd kasutatakse logaritmitud elektrihinna aegrida. Eesti päevased elektrienergia logaritmitud hinnad on graafiliselt kujutatud joonisel 3. Kuna sellise rea puhul on sesoonses osas palju autokorrelatsioone usalduspiirde vahelt väljas, siis väga väikeste kordajatega mudelit ei õnnestu leida. Antud rea autokorrelatsioonide ja osaautokorrelatsioonide graafik on leitav lisast 1 (vt joonis 18) ning mudelite sobitamisel kasutatud R skript lisast 2. Parimaid tulemusi annavad mudelid ARIMA$(1,0,1) \times (1,0,4)_7$, ARIMA$(3,0,0) \times (1,0,4)_7$ ning ARIMA$(1,0,2) \times (1,0,4)_7$, AIC väärtustega vastavalt -792,15, -794,24 ja -795,5. Parimaks mudeliks on seega ARIMA$(1,0,2) \times (1,0,4)_7$, mida tähistatakse edaspidi ARIMAX$_{2\,\mathrm{log}}$.

	ar1	ma1	ma2	sar1	sma1	sma2	sma3	sma4
kordajad	0,8827	-0,3793	-0,1317	0,9979	-0,8168	-0,1198	0,0798	-0,1143
s.e	0,0435	0,0584	0,0549	0,0035	0,0444	0,0498	0,0504	0,0386

	vabaliige	x1	x2	x3
kordajad	3,6691	0,0131	-0,0131	0,0214
s.e	0,0982	0,0037	0,0061	0,0078

Tabel 7: Mudeli $ARIMAX_{2\,\log}$ kuju

Kuna ühemõõtmelisel juhul oli logaritmitud rea korral AIC väärtus -775,17, siis Akaike informatsiooni kriteeriumi alusel on ühemõõtmelisest mudelist parem mudel leitud, seega võib eeldada, et ka prognoosimisel tekkinud vead on väiksemad.

Võrdlemaks antud peatükis leitud mudeleid, uuritakse prognoosivigade mõõdikuid.

	MAD	RMSE	MAPE
ARIMAX_2	3,39	4,30	10,95%
$\mathrm{ARIMAX}_{2\,\log}$	3,13	4,03	9,87%

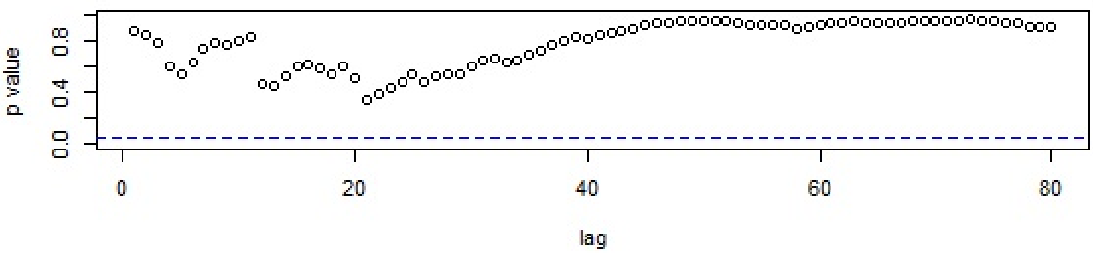
Tabel 8: Ühesammuliste prognoosivigade mõõdikud kasutades regressoritena õhutemperatuuride funktsioone

Taaskord annab logaritmitud rida paremaid tulemusi kui originaalne elektrihinna aegrida. Huvitav on aga see, et keskmine suhteline viga on ARIMAX mudeli logaritmitud rea korral suurem kui ühemõõtmelisel juhul (vt tabel 4). Kui AIC väärtuse põhjal sai kinnitada, et ühemõõtmelistest mudelitest on paremad mudelid leitud, siis prognoosivigade mõõdikud logaritmitud rea korral seda ei kinnita. Samas on erinevus niivõrd väike, et lõpptulemuses võib siiski temperatuurifunktsioonide kaasamine anda paremaid tulemusi. Siinjuures on oluline meeles pidada, et mudeli kasutamine nõuab heade prognooside olemasolu järgmise päeva temperatuuri jaoks. Seega, kui täpne järgmise päeva temperatuur ei ole teada, aitab ühemõõtmelise mudeli kasutamine kindlasti lisavigasid vältida.

Kuna aga viga ei õnnestunud väiksemaks saada, siis tasuks regressorite uurimist jätkata. Enne mitmemõõtmelise mudeli lõplikku valimist ja kirja panemist uuritaksegi, kas leidub veel ridasid, mille abil õnnestub järgmise päeva elektrihinda paremini ennustada.

6.2 Elektrihinna analüüs kasutades regressorina töö- ja puhkepäevi eristavat aegrida

Töö- ja puhkepäevi eristavat aegrida kasutatakse regressorina sarnaselt eelmisele peatükile, kuid nüüd ei ole tegemist enam maatriksiga. Kuna selline rida koosneb ainult 0-dest ja 1-dest ning moodustab ühe regressorrea, siis lisatakse see mudelisse vektorina. Treeningaegreaks on vahemikule 1. jaanuar 2013 kuni 31. detsember 2014 vastav kirjeldatud päevade vektor ning testaegrea moodustavad väärtused vahemikus 1. jaanuar kuni 31. märts 2015. Antud rea autokorrelatsioonide ja osaautokorrelatsioonide graafik on leitav lisast 1 (vt joonis 19) ning mudeli sobitamisel kasutatud R skript lisast 2. Kolm enim sobinud mudelit on ARIMA$(2,0,1)\times(1,0,0)_7$, ARIMA$(1,0,1)\times(0,0,3)_7$ ning ARIMA$(1,0,2)\times(1,0,0)_7$, vastavalt AIC väärtustega 4919,36, 4918,66 ja 4920,09. Kuna kõigi kolme parima mudeli AIC väärtused on küllalt sarnased, siis valitakse parimaks mudel, millel on üks parameeter vähem ehk siis ARIMA$(2,0,1)\times(1,0,0)_7$. Sellist mudelit tähistatakse edaspidi ARIMAX$_3$.



Joonis 11: Mudeli ARIMAX$_3$ Ljung-Box testi p-väärtused kasutades regressorrina töö-ja puhkepäevi eristavat aegrida

Kõik Ljung-Box testi p-väärtused on suuremad kui 0,05, seega saab lugeda mudeli sobivaks. Sellisel mudelil on oluline eelis, sest töö- ja puhkepäevi eristava aegrea järgmise päeva väärtus on alati teada. Kuna ühemõõtmelise mudeli korral AIC=4945,91 ja ARIMAX$_2$ korral AIC=4940,33, siis selles peatükis leitud mudel on Akaike informatsiooni kriteeriumi põhjal otsustades parim.

Mudel avaldub kujul:

	ar1	**ar2**	**ma1**	**sar1**	**vabaliige**	**nadalap**
kordajad	1,1342	-0,2112	-0,6956	0,1191	41,8285	-4,6973
s.e	0,1402	0,0994	0,1277	0,0432	1,1591	0,6033

Tabel 9: Mudeli $ARIMAX_3$ kuju

Järgnevalt uuritakse logaritmitud elektrihinna rida, kus regressoriks on töö- ja puhkepäevi eristav aegrida. Antud rea autokorrelatsioonide ja osaautokorrelatsioonide graafik on leitav lisast 1 (vt joonis 20). Kolm kõige paremini sobivat mudelit on ARIMA$(3,0,0) \times (0,0,3)_7$, ARIMA$(1,0,2) \times (0,0,3)_7$ ning ARIMA$(2,0,1) \times (0,0,3)_7$, vastavalt AIC väärtustega -807,76, -808,74 ja -808,3. Kuna kõikides mudelites on võrdne arv parameetreid, siis parimaks valitakse, nii nagu ka varasemalt, väikseima AIC väärtusega ARIMA $(1,0,2) \times (0,0,3)_7$ mudel. Sellist mudelit tähistatakse edaspidi $\mathrm{ARIMAX}_{3\,\mathrm{log}}$.

Mudel avaldub kujul:

	ar1	**ma1**	**ma2**	**sma1**	**sma2**	**sma3**	**vabaliige**	**nadalap**
kordajad	0,8811	-0,3698	-0,1246	0,1676	0,0581	0,1246	3,7142	-0,1263
s.e	0,0350	0,0525	0,0488	0,0409	0,0402	0,0380	0,0291	0,0137

Tabel 10: Mudeli $ARIMAX_{3\,\mathrm{log}}$ kuju

Võrreldes varasemate logaritmitud ridade mudelite AIC väärtustega on antud peatükis leitud AIC väärtus väikseim. Nagu eelnevalt näha, siis parem AIC väärtus ei tähenda erinevate mudelite puhul tingimata ka paremaid prognoosivigade mõõdikuid.

Et saada rohkem aimu mudelite sobivusest, uuritakse selles peatükis leitud parimate mudelite prognoosivigade mõõdikuid.

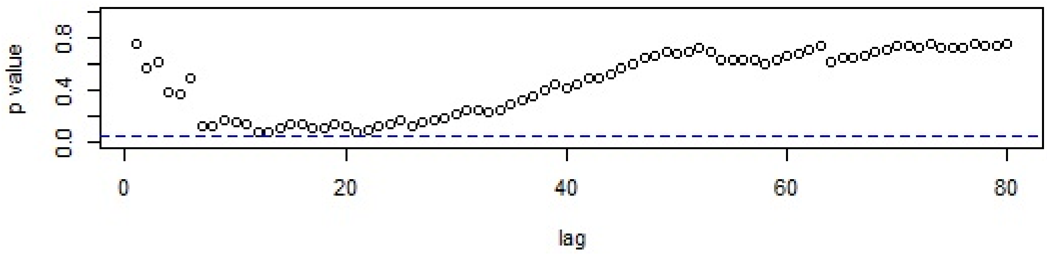
	MAD	RMSE	MAPE
ARIMAX_3	3,60	4,51	11,75%
$\text{ARIMAX}_{3\,\log}$	3,35	4,26	10,63%

Tabel 11: Ühesammuliste prognoosivigade mõõdikud kasutates regressorina töö- ja puhkepäevi eristavat aegrida

Olenemata sellest, et Akaike informatsiooni kriteeriumi väärtused olid vaadeldud mudelite korral oluliselt paremad kui eelnevalt, on testandmete prognoosimisel leitud vigade mõõdikud suuremad. Sellisel mudelil on aga suur eelis võrreldes temperatuuri kasutavate regressoritega, sest kasutatava regressorrea tulevikuväärtused on alati teada. Kuna sellise mudeliga saadud tulemused ei ole võrreldes eelnevatega oluliselt kehvemad, siis päris kõrvale ei tasu neid mudeleid jätta.

6.3 Elektrihinna analüüs kasutades regressorina funktsioone õhutemperatuuride aegreast ning töö- ja puhkepäevi eristavat aegrida

Kogu analüüs tehakse sarnaselt eelnevatele peatükkidele. Kõigepealt leitakse sobiv mudel ning regressoriks on nüüd maatriks, mis koosneb neljast veerust. Kolm veergu tulevad temperatuuri mõju kirjeldamiseks kasutatavatest funktsioonidest 6.1, kus $a = 2$ ning $b = 8$ ja viimase veeru moodustab töö- ja puhkepäevi eristav aegrida. Siinkohal pole regressorridade järjekord oluline. Antud rea autokorrelatsioonide ja osaautokorrelatsioonide graafik on leitav lisast 1 (vt joonis 21) ja mudelite sobitamisel kasutatud R skript lisast 2. Parimad mudelid antud regressorite korral on ARMA(2,0,1), $\text{ARIMA}(1,0,1) \times (0,0,3)_7$ ja $\text{ARIMA}(1,0,2) \times (0,0,3)_7$, vastavalt AIC väärtustega 4909,59, 4918,66 ning 4919,09. Selliste regressorite korral on mõneti üllatuslikult parim mudel ARMA(2,0,1), mida tähistatakse edaspidi ARIMAX_4.



Joonis 12: Mudeli ARIMAX$_4$ Ljung-Box testi p-väärtused

Kõik Ljung-Box testi p-väärtused on suuremad kui 0,05, mis kinnitab, et antud mudeli saab lugeda sobivaks.

Mudel avaldub kujul:

	ar1	ar2	ma1	vabaliige	x1	x2	x3	nadalap
kordajad	1,2378	-0,2696	-0,8112	41,9079	0,5867	-0,6852	1,0778	-4,7516
s.e	0,0809	0,0662	0,0640	1,8908	0,1727	0,3185	0,3918	0,5435

Tabel 12: ARIMAX$_4$ mudeli kuju

ARIMAX$_4$ mudeli korral on kordaja x_2 ebaoluline, sest selle kordaja absoluutväärtus on väiksem kui vastav kahekordne standardhälve, aga kuna ilma x_2-ta on mudeli AIC väärtus kehvem, siis jäetakse see alles.

Nii nagu ka eelnevates peatükkides uuritakse järgnevalt samasugust regressorite kombinatsiooni, aga logaritmitud elektrihinna rea korral. Antud rea autokorrelatsioonide ja osaautokorrelatsioonide graafik on leitav lisast 1 (vt joonis 22). Parimateks mudeliteks on ARIMA$(3,0,0) \times (1,0,0)_7$, ARIMA$(1,0,2) \times (1,0,0)_7$ ja ARIMA$(2,0,1) \times (1,0,0)_7$, AIC väärtustega vastavalt -816,57, -817,48 ning -824,87. Kuna kõikidel mudelitel on sama arv parameetreid, siis valitakse parimaks ARIMA$(2,0,1) \times (1,0,0)_7$, mida tähistatakse edaspidi kui ARIMAX$_{4\,\log}$.

Selline mudel avaldub kujul:

	ar1	ar2	ma1	sar1	vabaliige	x1	x2	x3	nadalap
kordajad	1,3676	-0,3873	-0,8563	0,1549	3,7327	0,0138	-0,0194	0,0264	-0,1268
s.e	0,0852	0,0732	0,0658	0,0400	0,0480	0,0037	0,0064	0,0080	0,0124

Tabel 13: ARIMAX$_{4\,\mathrm{log}}$ mudeli kuju

Kuna erinevatele ridadele sobitatud mudelite AIC väärtuseid ei saa üks-ühele võrrelda, siis uuritakse järgnevalt antud peatüki parimate mudelite prognoosivigade mõõdikuid.

	MAD	**RMSE**	**MAPE**
ARIMAX$_4$	3,40	4,33	11,07%
ARIMAX$_{4\,\mathrm{log}}$	3,16	4,00	9,97%

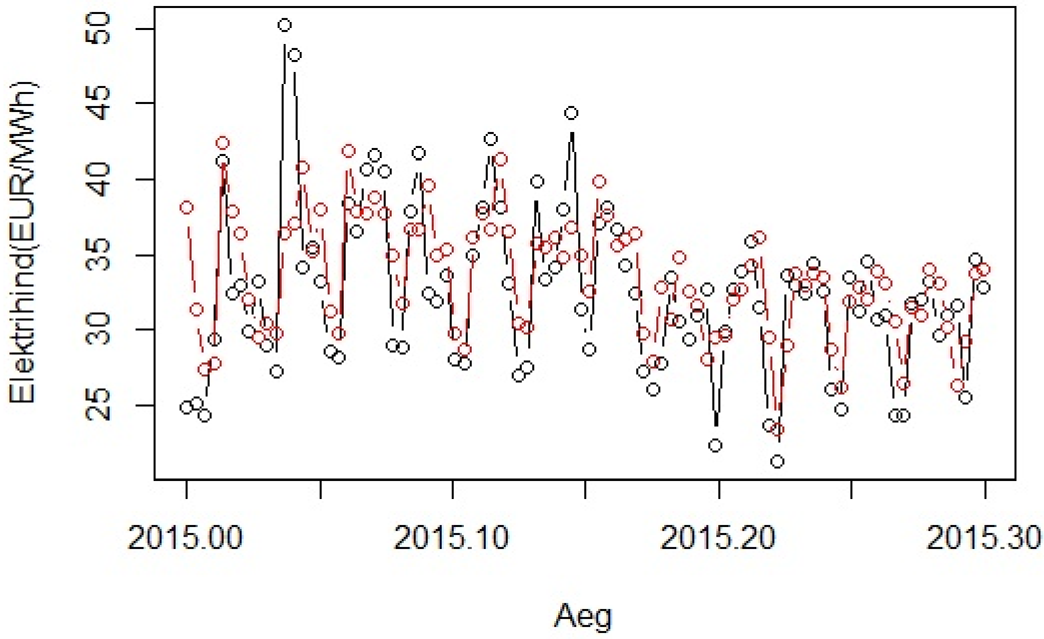
Tabel 14: Ühesammuliste prognoosivigade mõõdikud kasutates regressorina õhutemperatuuride funktsiooni ning töö- ja puhkepäevi eristavat aegrida

Tabeli 14 põhjal on taaskord logaritmitud rida andnud paremaid tulemusi. Kui võrrelda tabelit 11 tabeliga 14, siis tulemused on paranenud. Kui aga võrrelda tabelit 14 tabeliga 8, siis logaritmitud rea korral on tulemused halvemad, samas kui tavalise rea korral on mõõdikud natukene väiksemad.

Kui valida sobiv mudel Akaika informatsiooni kriteeriumite põhjal, siis samm-sammu haaval on iga mudeliga AIC väärtused järjest kahanenud ning viimase regressori kombinatsiooniga on saadud ka väikseima AIC väärtusega mudel ja viimase ARIMAX$_{4\,\mathrm{log}}$ võiks lugeda parimaks mudeliks. Kui aga sobiv mudel valida prognoosimõõdikute põhjal, siis nende väärtused käituvad hoopis teistmoodi. Kõige paremaid tulemusi andis hoopis ühemõõtmeline mudel, kuid kui uurida mitmemõõtmeliste seast, siis tundub, et sobivaim on ARIMAX$_{2\,\mathrm{log}}$. Kui otsus langetada ainult ruutkeskmise vea põhjal, siis osutuks parimaks samuti ARIMAX$_{4\,\mathrm{log}}$. Seega on erinevaid otsustuskriteeriumeid palju ning igal mudelil on omad plussid ja miinused. Kuna antud töö eesmärgiks oli leida võimalikult hea mudel järgmise päeva elektrihinna ennustamiseks, siis valitakse parimaks mitmemõõtmeliseks mudeliks prognoosimõõdikute põhjal ARIMAX$_{2\,\mathrm{log}}$.

6.4 Elektrihinna ennustamine üks päev ette kasutades parimat leitud mitmemõõtmelist mudelit

Antud peatükis kasutatakse mudelit ARIMA$(1,0,2) \times (1,0,4)_7$, mis on leitud logaritmitud elektrihinna reale, kus regressoriteks kasutatakse temperatuuri funktsioonide 6.1, kus $a = 2$ ja $b = 8$, põhjal leitud aegridu. Ühesammulised prognoosid leitakse kasutades treeningandmestikku ning võrreldakse neid testandmestiku väärtustega. Mudelit enam uuesti ei sobitada, vaid fikseeritakse ARIMAX$_{2\,\mathrm{log}}$ koefitsiendid ning kasutatakse samu parameetreid kogu aegrea jaoks. Kuna töö eesmärgiks on leida parim mudel ja katsetada meetodit, siis täiendavate prognoosivigade vältimiseks on kasutatud juba teadaolevat järgmise päeva temperatuuri selle päeva Eesti elektrihinna ennustamiseks.



Joonis 13: Ühesammulised ennustused ARIMAX$_{2\,\mathrm{log}}$ mudeliga (punane katkendlik joon) ja testaegrea väärtused (must joon)

Võrreldes joonise 13 ühesammulisi prognoose joonise 5 prognoosidega, tundub, et mitmemõõtmeline mudel püüab kõrgemaid hindasid isegi natukene paremini kui ühemõõtmeline. Madalate hindade korral olulist erinevust jooniste põhjal ei paista. Mudeli suurim ühesammuline prognoosijääk on 13,87, mis realiseerus 12. jaanuaril 2015. aastal ning väikseim jääk on 0,202, kui 23. märtsil elektrihind oli tegelikkuses 31,75 EUR/MWh ning ennustuse kohaselt 31,548 EUR/MWh.

Tabeli 7 abil avaldub parim, $\text{ARIMAX}_{2\,\log}$, mitmemõõtmeline mudel kujul:

$$\begin{gathered}\big(1-0,8827B\big)\big(1-0,9979B^7\big)\big(\ln(Z_t)-3,6691-0,0131X_1(t)+0,0131X_2(t)-\\ -0,0214X_3(t)\big)=\big(1-0,3793B-0,1317B^2\big)\big(1-0,8168B^7-0,1198B^8+\\ +0,0798B^9-0,1143B^{10}\big)A_t.\end{gathered}$$

Suuruse $\ln(Z_t)$, vabaliikme ja regressorridade kordajate vahe tähistatakse suurusega Wt:

$$W_t := \ln(Z_t)-3,6691-0,0131X_1(t)+0,0131X_2(t)-0,0214X_3(t).$$

Peale sulgude avamist:

$$\begin{gathered}W_t-0,88272W_{t-1}-0,9979W_{t-7}+0,8808W_{t-8}=A_t-0,3793A_{t-1}-0,1317A_{t-2}-\\ -0,8168A_{t-7}+0,19A_{t-8}+0,2328A_{t-9}-0,1288A_{t-10}+0,0329A_{t-11}+0,0151A_{t-12}.\end{gathered}$$

Peale ümardamist ja W_t avaldamist näeb mudeli lõplik kuju välja järgmiselt:

$$\begin{gathered}W_t=0,883W_{t-1}+0,998W_{t-7}-0,881W_{t-8}+A_t-0,379A_{t-1}-0,132A_{t-2}-\\ -0,817A_{t-7}+0,19A_{t-8}+0,233A_{t-9}-0,129A_{t-10}+0,033A_{t-11}+0,015A_{t-12},\end{gathered}$$

kus $\mathrm{W}_t=\ln(Z_t)-3,6691-0,0131X_1(t)+0,0131X_2(t)-0,0214X_3(t)$.

7 Võrdlus automaatse aegridade analüüsi koodiga

Järgnevalt uuritakse kas nö käsitsi uurimisega on võimalik lüüa automaatset aegridade analüüsimise koodi, mis vaatab läbi palju erinevaid mudeleid, süvenemata temaatikasse. Eesti Energias on koostatud kood, mis käib üle 20 erineva automaatse R-i sisseehitatud mudeliklassi, leiab nende seast parima ning annab ennustused järgnevateks perioodideks.

Automaatses koodis kasutatud ühemõõtmelised mudelid:

- Automaatne ARIMA (vt (Kangro [11], 2015))
- Osaline ARIMA mudel - ARFIMA (vt Wikipedia)
- Multiplikatiivne Holt-Winters (vt (Kangro [11], 2015))
- Aditiivne Holt-Winters (vt (Kangro [11], 2015))
- Struktuurne aegridade mudel (vt programmi R abi *StructTS*)
- Sesoonne kohandamine *loess* meetodil (vt (Kangro [11], 2015))
- Sesoonne kohandamine STL meetodil (vt programmi R abi *stlf*)
- TBATS mudel (vt programmi R abi *tbats*)
- BATS mudel (vt programmi R abi *bats*)
- Eksponentsiaalse silumise olekuruumi mudel (vt (Kangro [11], 2015))
- Närvivõrkude mudel ARIMA vigadega (vt programmi R abi *nnetar*)
- Theta meetod (vt programmi R abi *thetaf*)
- Silumine kuupsplainidega (vt programmi R abi *splinef*)
- Juhuslik ekslemine (vt programmi R abi *rwf*)

- Iid mudel (vt programmi R abi *meanf*)

Automaatses koodis kasutatud mitmemõõtmelised mudelid:

- VAR mudel (vt programmi R abi *VAR*)
- Mitmemõõtmeline lineaarne regressioon (vt Wikipedia)
- MARSS mudel (vt programmi R abi *MARSS*)
- ARIMAX mudel (vt (Kangro [11], 2015))
- VARMA (vt programmi R abi *VARMA*)
- Mitmemõõtmeline Cholesky volatiilsuse mudel (vt programmi R abi *MCholV*)

Automaatne kood tähendab seda, et koodi saab kasutada ükskõik millise aegrea analüüsimiseks. Miinuseks on see, et rea statsionaarsuse eeldus ei pruugi olla täidetud, sest seda ei kontrollita. Samuti ei vaadelda aegrea autokorrelatsioonide ega osaautokorrelatsioonide graafikuid, kust võib saada vajalikku infot. Kood saab ette uuritava aegrea ning vajadusel ka regressorread ning hakkab kohe mudeleid sobitama. Mudelite sobitamine on mitme etapiline:

1. Andmed jaotatakse treening- ja testandmestikuks. Et võrrelda automaatset koodi eelnevalt leitud parima mudeliga on treening- ja testandmestik jaotatud vastavalt kaks aastat ja kolm kuud.
2. Leitakse parim mudel iga klassi sees. Kuna programmis R on sisse ehitatud mudeliklassidel veel alammudelid, siis iga klassi sees valitakse omakorda parim mudel. Näiteks kui kasutatakse funktsiooni *auto.arima*, siis vaikimisi on maksimaalseteks p ja q väärtusteks 5, seega proovitakse läbi kõikvõimalikud kombinatsioonid, kus p ja q saavad väärtuseid nullist viieni. Lisaks p-le ja q-le on muidugi ka teisi parameetreid, näiteks tavaline ja sesoonne diferentseerimine. Parima mudeli iga klassi sees leiab programm R ise sisseehitatud

meetoditega. Näiteks funktsiooni *auto.arima* korral leitakse parim mudel kasutatades AIC (3.8), AICc(3.9) või BIC väärtuseid.

3. Kõikide mudeliklasside parimate mudelitega tehakse ühesammulised prognoosid kasutades treeningandmestikku.

4. Võrreldakse prognoose tegelikult realiseerunud tulemustega ehk siis testandmestiku väärtustega. Võrdlemine käib kasutatedes suuruseid MAD, RMSE ja MAPE. Need suurused normeeritakse, et väärtused jääksid vahemikku 0-1 ning antakse kaalud, millise tähtsusega need on. Antud juhul on kõik mõõdikud võrdse tähtsusega.

5. Leitakse igale mudelile lõplik skoor ehk liidetakse statistikute väärtused kokku ning väikseima skooriga mudel valitakse parimaks mudeliks.

Kolme parima ühemõõtmelise mudeli mõõdikud, mis on leitud kasutades Eesti Energia automaatset koodi, on välja toodud järgnevas tabelis.

	ARIMA	**ARFIMA**	**Juhuslik ekslemine**
MAD	4,27	3,93	4,34
RMSE	5,4	4,93	5,96
MAPE	13,91%	12,9%	13,38%

Tabel 15: Ühesammuliste prognoosivigade mõõdikud ühemõõtmeliste mudelite korral, leitud automaatse koodi abil

Ühemõõtmeliste mudelite seast valis programm parimaks ARIMA mudelite klassi üldistuse ARFIMA, mille korral on lisaks tavalistele diferentsidele võimalik leida ka nn murruline diferents. Kuna aga leitud mudeli prognoosivead on oluliselt halvemad kui autori leitud mudelitel, ei hakata siin töös mainitud mudeli definitsiooni ja sellega seotud mõisteid lähemalt kirjeldama. Parim mudel on ühe autoregressiivse parameetriga, mille väärtus on 0,1054716 ning d=0,3462161.

Kolme parima mitmemõõtmelise mudeli mõõdikud on välja toodud tabelis 16.

	VAR	ARIMAX	Cholesky
MAD	4,38	4,14	5,37
RMSE	5,55	5,12	6,41
MAPE	14,42%	13,17%	15,86%

Tabel 16: Ühesammuliste prognoosivigade mõõdikud mitmemõõtmeliste mudelite korral, leitud automaatse koodi abil

Mitmemõõtmeliste mudelite seast osutus parimaks ühe korra diferentseeritud nelja MA liikmega automaatne *ARIMAX* mudel. See on hea näide sellest, et automaatne kood ei anna paremaid tulemusi kui käsitsi sobitamine. Parima ARIMAX mudeli korral on rida diferentsitud, kuid nagu selgus peatükist 5.1, ei ole diferentseerimine Eesti elektrihinna rea puhul otstarbekas ning liigselt diferentseeritud rida on ebatäpsem, kui õige arv diferentseeritud või üldse diferentseerimata rida. Kui võis oletada, et mitmemõõtmelised mudelid peaksid andma parema tulemuse, kui ühemõõtmelised, siis ka selline väide automaatse koodi puhul ei pea paika. Automaatse koodi abil leitud parim mudel on ühemõõtmeline ning keskmine suhteline viga on 12,9%. Kuna käsitsi sobitamisega õnnestus leida väiksema suhtelise veaga mudel, saab väita, et selline automaatne kood ei suuda Eesti elektrihinna aegrea korral paremaid ühesammulisi ennustusi leida.

Kokkuvõte

Käesolevas magistritöös vaadeldi järgmise päeva elektrihinna prognoosimist erinevate meetoditega. Analüüs teostati nii ühemõõtmeliste kui ka mitmemõõtmeliste ARIMA mudelitega. Lisaks tavalisele elektrihinna aegreale uuriti ka logaritmitud rida. Kuna alates 2013. aastast kujuneb Eesti elektrihind vabalt börsil ja hinnakõikumised on olnud väga suured, siis püüti logaritmisega vähendada väärtuste vahelisi suuri muutuseid.

Analüüs teostati Eesti päevaste elektrihindade aegreal alates 1. jaanuar 2013 kuni 31. detsember 2014 ning rida alates 1. jaanuar 2015 kuni 31. märts 2015 kasutati selleks, et kontrollida, kui hästi suudab leitud mudel andmeid prognoosida. Mudeleid võrreldi Akaike informatsioonikriteeriumi põhjal ning mudelite prognoosivõimet hinnati mõõdikute MAD, RMSE ja MAPE abil.

Esmalt otsiti mudeleid, mis kasutasid ainult ajaloolisi elektrihindasid. Sellisel juhul andis parimaid tulemusi logaritmitud elektrhihinna reale sobitatud ARIMA$(2,0,1)$ $\times(2,0,1)$ mudel perioodiga 7. Keskmine suhteline viga oli sellise mudeli korral 9,43%.

Järgnevalt teostati analüüs mitmemõõtmeliste mudelitega. Prognoose püüti täpsustada regressorite kaasamisega. Selleks kasutati Eesti õhutemperatuuride ning töö- ja puhkepäevi eristavat aegrida. Parima tulemuse andis logaritmitud elektrihinna aegreale leitud ARIMA$(1,0,2) \times (1,0,4)_7$ mudel, kus regressoriteks kasutati temperatuuri funktsioonide põhjal leitud aegridu. Sellise mudeli korral oli keskmine suhteline viga 9,87%. Seega osutus ühemõõtmeline mudel mitmemõõtmelisest mudelist paremaks.

Viimaseks võrreldi töös leitud parimat mudelit automaatse aegridade analüüsi koodi abil leitud parima mudeliga. Saadi kinnitust, et eelnev põhjalik analüüs oli ennast ära tasunud. Võrreldes automaatse koodiga väheneb keskmine suhteline viga peaaegu 4% ning arvestades vea suurusjärku, on selline vea parandus väga hea.

Kuna antud töös vaadeldi vaid kahte tegurit, mis võiksid mõjutada elektrihinda, siis edasiseks uurimiseks tasuks vaadelda ka lisaks muid faktoreid, näiteks riikidevahelisi ülekandevõimsuseid või analüüsida rikete/avariide tekkimise võimalust. Kuna prognoosimisel tekkinud vead jäid ka parima mudeli korral küllalt suureks, siis tasuks jätkata uurimist keerukamate ja teist tüüpi mudelitega.

Kirjandus

[1] Eesti Energia kodulehekülg. Kasutatud 04.03.2015 `http://www.energia.ee/`.

[2] Elering AS-i kodulehekülg. Kasutatud 23.02.2015 `http://elering.ee/`.

[3] Eleringi toimetised nr2/2012(4). *Elektrituru käsiraamat (lk 22-23, 25-26).* Tallinn. `http://www.elering.ee/public/Infokeskus/Elektrituru_kasiraamat.pdf`.

[4] Konkurentsiamet (2013). *Aruanne elektri- ja gaasiturust Eestis (lk 5, 7).* Tallinn. `http://www.konkurentsiamet.ee/file.php?24368`.

[5] Estlink 2 kodulehekülg. Kasutatud 09.03.2015 `http://estlink2.elering.ee/`.

[6] Energiatalgute kodulehekülg. Kasutatud 18.02.2015 `http://www.energiatalgud.ee/index.php?title=Elektriturg`.

[7] Elektribörsi Nord Pool Spot'i kodulehekülg. Kasutatud 02.02.2015 `http://www.nordpoolspot.com/historical-market-data/`.

[8] Avatud elektrituru kodulehekülg. Kasutatud 10.03.2015 `http://avatud2013.ee/`.

[9] Imatra Elekter AS-i kodulehekülg. Kasutatud 03.03.2015 `http://imatraelekter.ee/elektrimuuk/nord-pool-spot/`.

[10] Eleringi toimetised nr2/2014(7). *Eesti elektrisüsteemi varustuskindluse aruanne 2014 (lk 47).* Tallinn. `http://elering.ee/public/Infokeskus/Aruanded/Elering_varustuskindluse_aruanne_2014_1.pdf`.

[11] Kangro, R.(2015). *Aegridade analüüs.* Konspekt. Tartu Ülikool, matemaatika- ja informaatikateaduskond.

[12] Shumway, R. H., Stoffer, D. S.(2011). *Time Series Analysis and Its Applications (p 22).* Third Edition. New York: Springer.

[13] Box, G. E. P., Jenkins, G. M., Reinsel, G. C.(2008) *Time Series Analysis. Forecasting and Control (p 25).* Fourth Edition. New Jersey: Wiley.

[14] Wikipedia kodulehekülg.
`http://en.wikipedia.org/wiki/Phillips%E2%80%93Perron_test`

[15] Wikipedia kodulehekülg.
`http://en.wikipedia.org/wiki/Ljung%E2%80%93Box_test`

[16] Wikipedia kodulehekülg. Kasutatud 17.02.2015
`http://en.wikipedia.org/wiki/Akaike_information_criterion.`

[17] Koppel, K.(2013). *Eesti Energia: ühest rekordpäevast ei tasu teha liiga suuri järeldusi.* Eesti Rahvusringhääling, 25.juuni. `http://uudised.err.ee/v/majandus/93c62485-1737-4ea1-b4ae-4acb0ed3d04d`

[18] Kallas, R.(2013). *Hando Sutter: börsil ei saa hind olla anomaalselt kõrge.* Ärileht, 16.oktoober.
`http://m.arileht.delfi.ee/majandus/article.php?id=66920478`

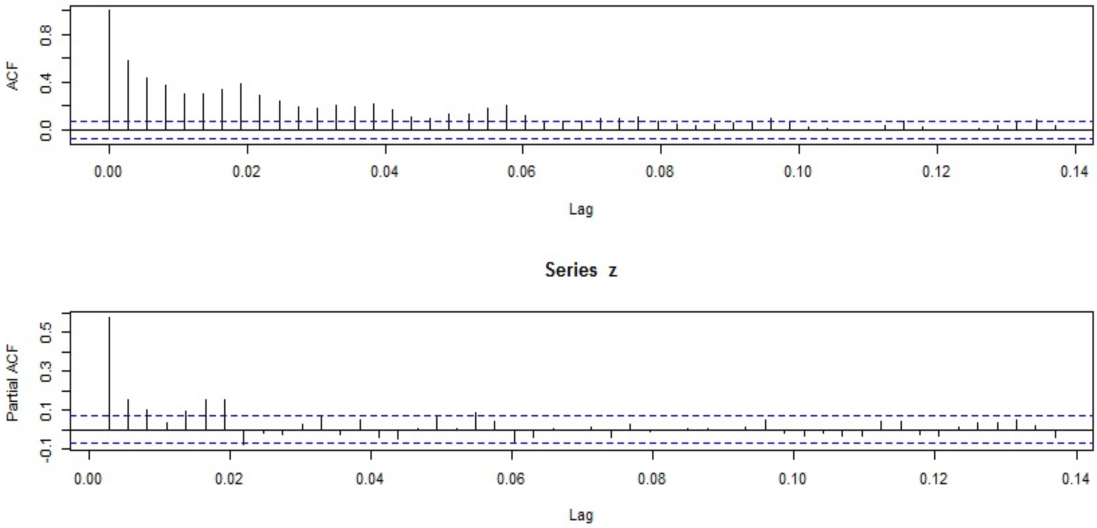
[19] Ärileht. (2014). *Hinnašokk Eesti moodi: elektri megavatt-tund maksab 11 tunni vältel üle 203 euro.* 19.juuli. `http://m.arileht.delfi.ee/majandus/article.php?id=69385199`

[20] Õepa, A. (2014). *Elektri hinnarekordis on süüdi külm ilm ja Soome-Rootsi kaabel* Postimees, 29.detsember. `http://majandus24.postimees.ee/3038753/elektri-hinnarekordis-on-suudi-kulm-ilm-ja-soome-rootsi-kaabel`

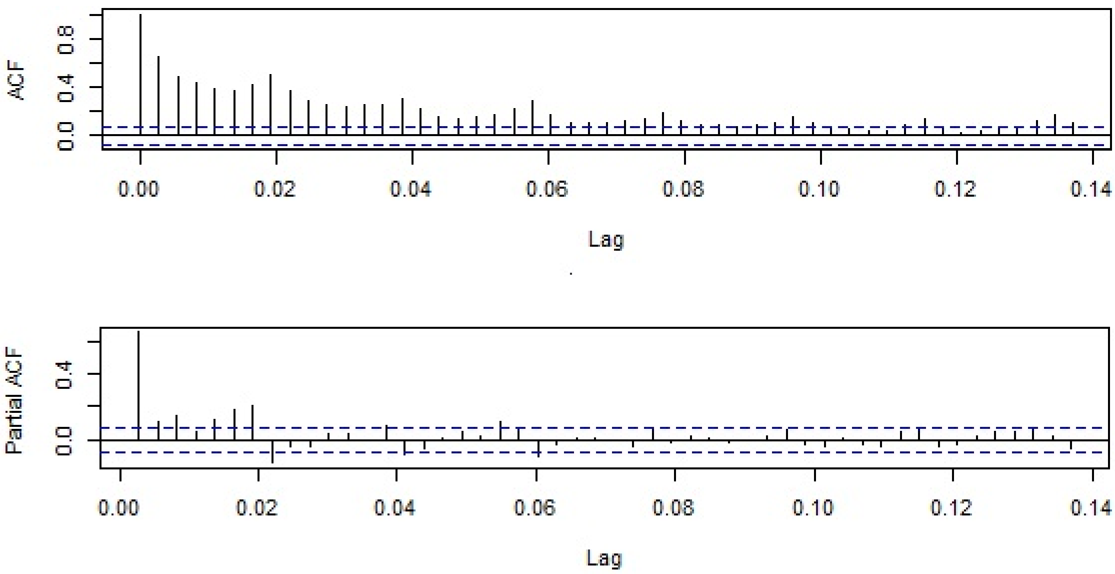
[21] Eesti Statistikaameti andmebaaside kodulehekülg. Kasutatud 03.03.2015
`http://pub.stat.ee/px-web.2001/dialog/varval.asp?ma=RL718&ti=ASUSTATUD+TAVAELURUUMID%2C+LEIBKONNAD+JA+ELANIKUD+ELURUUMI+`

ASUKOHA+JA+T%DC%DCBI+J%C4RGI&path=../database/Rahvaloendus/rel2000/03eluruumid/&search=LEIBKOND&lang=2

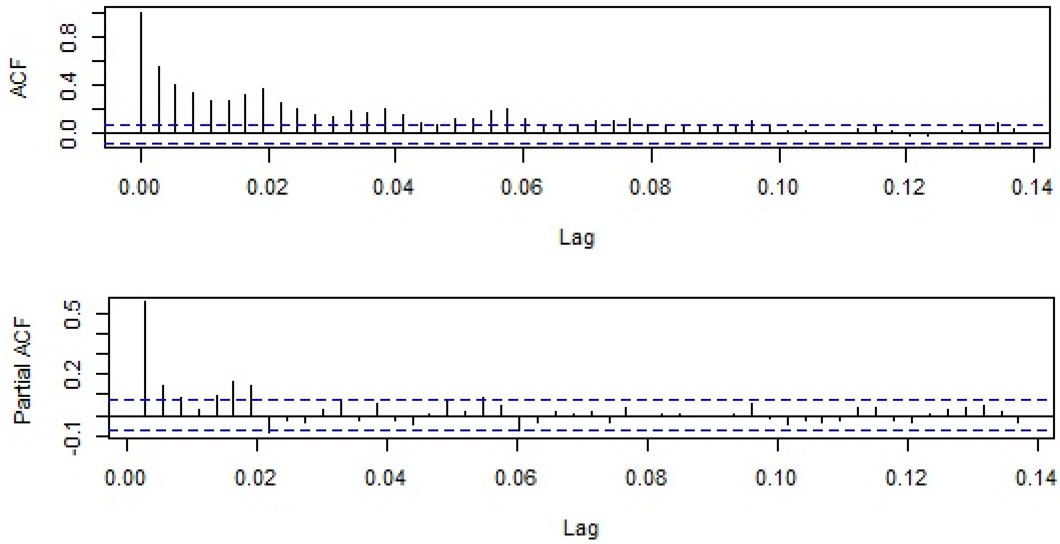
Lisa 1



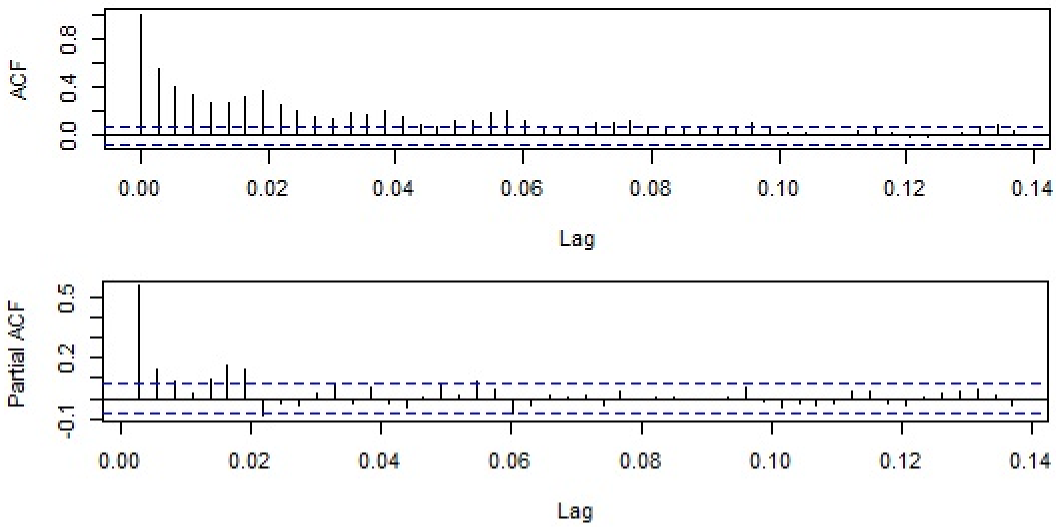
Joonis 14: Eesti elektrienergia hindade autokorrelatsioonid ja osaautokorrelatsioonid



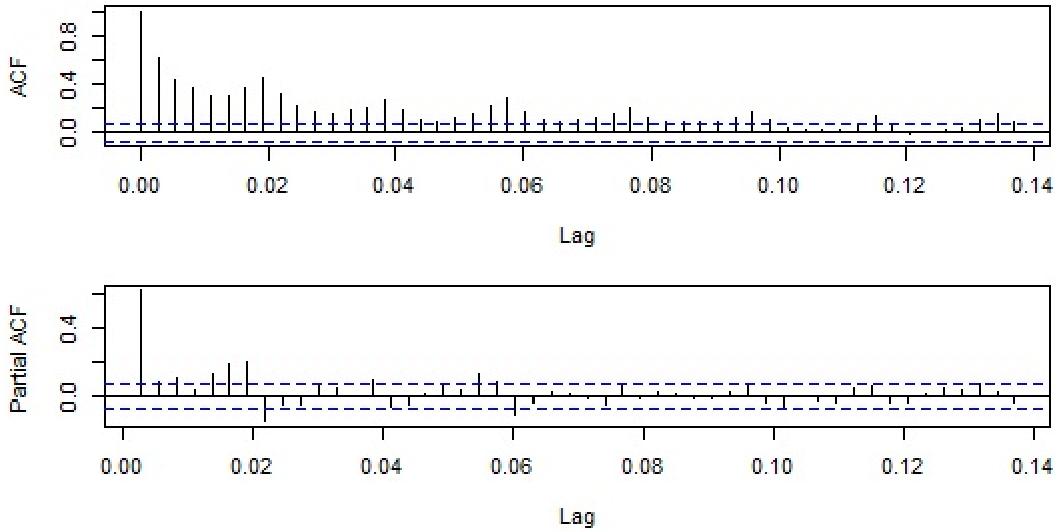
Joonis 15: Eesti elektrienergia logaritmitud hindade autokorrelatsioonid ja osaautokorrelatsioonid



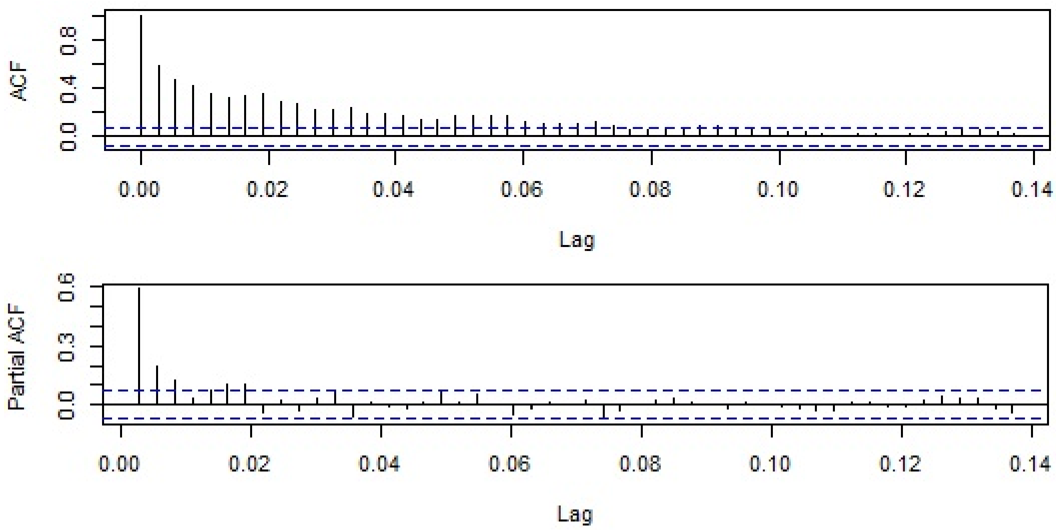
Joonis 16: Eesti elektrienergia hindade autokorrelatsioonid ja osaautokorrelatsioonid, kus regressoriteks on kasutatud temperatuuri funktsioone 6.1, kus $a = -5$ *ja* $b = 15$



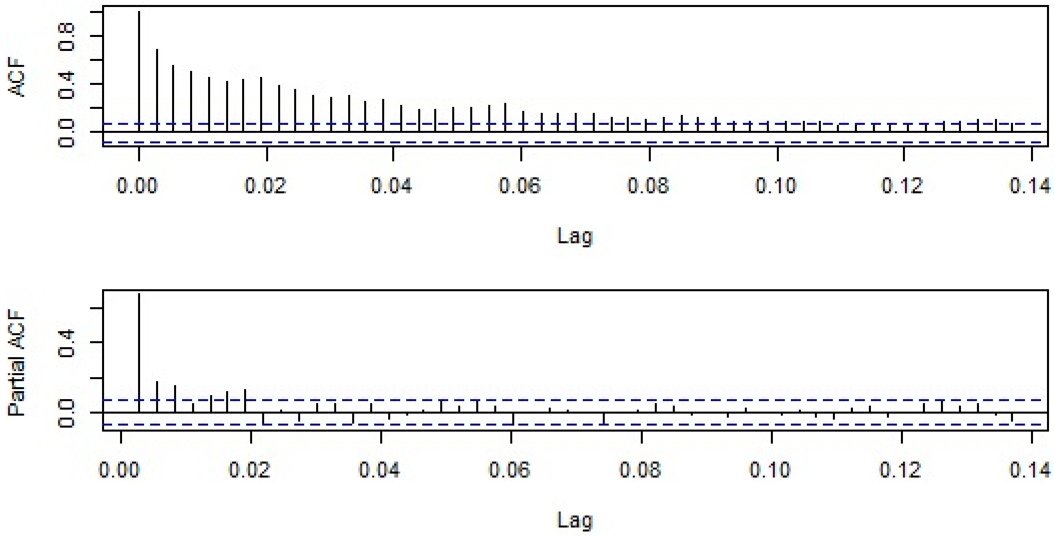
Joonis 17: Eesti elektrienergia hindade autokorrelatsioonid ja osaautokorrelatsioonid, kus regressoriteks on kasutatud temperatuuri funktsioone 6.1, kus $a = 2$ *ja* $b = 8$



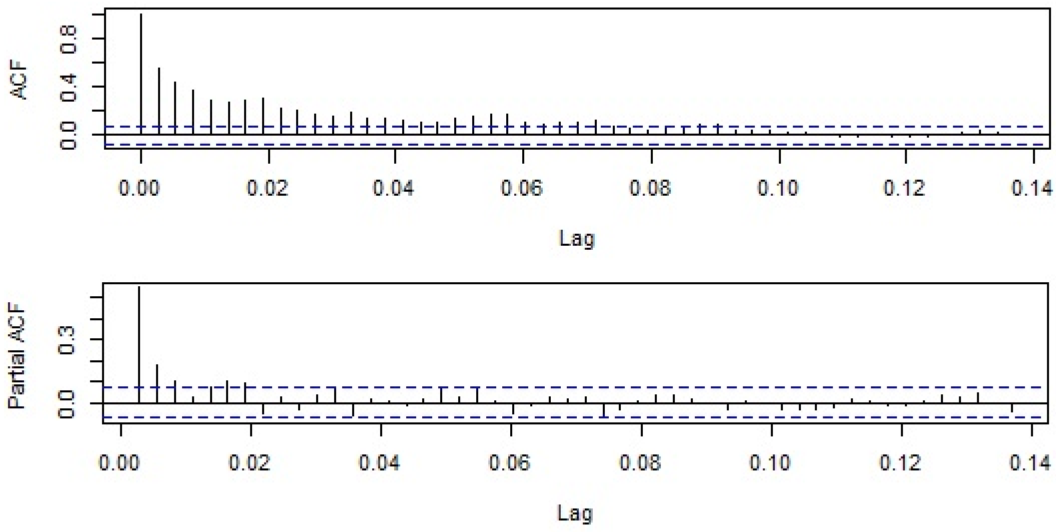
Joonis 18: Eesti elektrienergia logaritmitud hindade autokorrelatsioonid ja osaautokorrelatsioonid, kus regressoriteks on kasutatud temperatuuri funktsioone 6.1, kus $a = 2$ *ja* $b = 8$



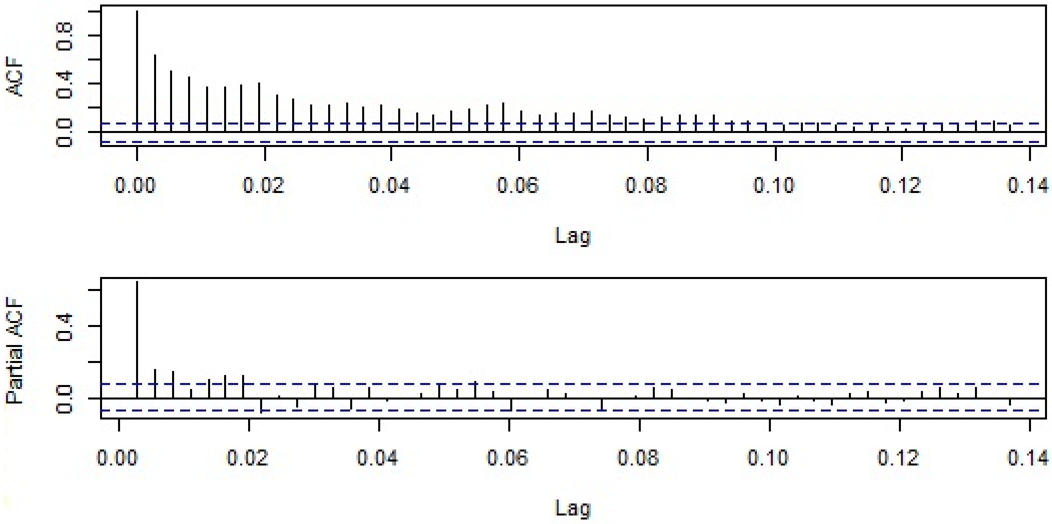
Joonis 19: Eesti elektrienergia hindade autokorrelatsioonid ja osaautokorrelatsioonid, kus regressoriks on kasutatud töö- ja puhkepäevi eristavat aegrida



Joonis 20: Eesti elektrienergia logaritmitud hindade autokorrelatsioonid ja osaautokorrelatsioonid, kus regressoriks on kasutatud töö- ja puhkepäevi eristavat aegrida



Joonis 21: Eesti elektrienergia hindade autokorrelatsioonid ja osaautokorrelatsioonid, kus regressoriteks on kasutatud temperatuuri funktsioone $(a = 2, b = 8)$ *ning töö- ja puhkepäevi eristavat aegrida*



Joonis 22: Eesti elektrienergia logaritmitud hindade autokorrelatsioonid ja osaautokorrelatsioonid, kus regressoriteks on kasutatud temperatuuri funktsioone ($a = 2, b = 8$) *ning töö- ja puhkepäevi eristavat aegrida*

Lisa 2

```
library(forecast)

# Treeningandmestik
andmed1=read.table("C:\\Users\\kartpal\\paevased_2013_2014.txt",
sep="",skip=3)
elekt_hind1=andmed1$V2
hind_treen=ts(elekt_hind1,start=2013,frequency=365)
sum(is.na(hind_treen)) #ei ole puuduvaid andmeid

plot(hind_treen,type="l",ylab="Elektrihind EUR/Mwh",xlab="Aeg")
PP.test(hind_treen)

#millised on 4 suurimat vaartust
max_vaartused = order(hind_treen,decreasing = T)[1:4]
hind_treen[max_vaartused]
#hind_treen[566],[176],[728],[289]

graafikud=function(z,mitu=30){
  layout(1:3)
  plot(z,type="l")
  acf(z,mitu)
  pacf(z,mitu)
  layout(1)}
graafikud(hind_treen,50)

#joonise pohjal voiks sobida  AR(1),AR(3),ARMA(1,0,1),ARMA(1,0,2,),
#(2,0,1); sesoonses osas ARMA(0,0,3), ARMA(1,0,2), ARMA(2,0,1),
#ARMA(1,0,1),AR(1)
m1=arima(hind_treen,order=c(1,0,0))#ei sobi
m2=arima(hind_treen,order=c(1,0,1))#ei sobi
m3=arima(hind_treen,order=c(3,0,0))#ei sobi
m4=arima(hind_treen,order=c(1,0,2))#ei sobi
m5=arima(hind_treen,order=c(2,0,1))#ei sobi
m6=arima(hind_treen, order=c(1,0,0),seasonal=list(order=c(1,0,0),
period=7))#ei sobi
m7=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,0),
period=7))#ei sobi
m8=arima(hind_treen, order=c(1,0,2),seasonal=list(order=c(1,0,0),
period=7))#ma2 ebaoluline
m9=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(1,0,0),
period=7))#AIC=4969.69
m10=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(1,0,0),
period=7))#AIC=4970.32
m11=arima(hind_treen, order=c(1,0,0),seasonal=list(order=c(1,0,1),
period=7))#ei sobi
m12=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,1),
period=7))#AIC=4954.6
m13=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(1,0,1),
period=7))#AIC=4952.26
m14=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(1,0,1),
period=7))#AIC=4955.72
m15=arima(hind_treen, order=c(1,0,0),seasonal=list(order=c(0,0,3),
```

```
period=7))#ei sobi
m16=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(0,0,3),
period=7))#AIC=4962.87
m17=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(0,0,3),
period=7))#ar2 ebaoluline
m18=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(0,0,3),
period=7)) #AIC=4964.01
m19=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,2),
period=7)) #AIC=4946.21
m20=arima(hind_treen, order=c(1,0,0),seasonal=list(order=c(1,0,2),
period=7))#ei sobi
m21=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(1,0,2),
period=7))#ei sobi
m22=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(1,0,2),
period=7))#ei sobi
m23=arima(hind_treen, order=c(1,0,0),seasonal=list(order=c(2,0,1),
period=7))#ei sobi
m24=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(2,0,1),
period=7))#AIC=4945.91, parim
m25=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(2,0,1),
period=7))#ar2 ebaoluline
m26=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(2,0,1),
period=7))#AIC=4947.27
tsdiag(m24,80)
m24 #parima mudeli kuju

#logaritmitud rida
log_hind_treen=log(hind_treen)
log_hind_test=log(hind_test)
plot(log_hind_treen,type="l",xlab="Aeg",ylab="Logaritmitud
elektrihind (EUR/Mwh)")
graafikud(log_hind_treen,50)

#joonise pohjal AR(1),AR(3),(1,0,1),(1,0,2,),(2,0,1);
#sesoonses osas vahemalt 6 autokorrelatsiooni piiridest valjas,
#seega otsitakse seal AR voi segamudeleid,voiks sobida AR(3), sest
#sellest edasi on piirides, tasuks proovida ka ARMA(2,0,1)
m1_log=arima(log_hind_treen,order=c(1,0,1),seasonal=list(order=c(2,0,1),
period=7))#AIC=-770.97
m2_log=arima(log_hind_treen,order=c(1,0,2),seasonal=list(order=c(2,0,1),
period=7))#AIC=-774.93
m3_log=arima(log_hind_treen,order=c(2,0,1),seasonal=list(order=c(2,0,1),
period=7))#AIC=-775.17, parim
m4_log=arima(log_hind_treen,order=c(3,0,0),seasonal=list(order=c(2,0,1),
period=7))#AIC=-773.73
m5_log=arima(log_hind_treen,order=c(3,0,0),seasonal=list(order=c(3,0,0),
period=7))#AIC=-755.62
m6_log=arima(log_hind_treen,order=c(1,0,1),seasonal=list(order=c(3,0,0),
period=7))#AIC=-752.43
m7_log=arima(log_hind_treen,order=c(1,0,2),seasonal=list(order=c(3,0,0),
period=7))#AIC=-754.98
m8_log=arima(log_hind_treen,order=c(2,0,1),seasonal=list(order=c(3,0,0),
period=7))#AIC=-754.07
tsdiag(m3_log,80)
```

```
m3_log

#Testandmestik
andmed2=read.table("C:\\Users\\kartpal\\hind_2015.txt",sep="",skip=1)
elekt_hind2=andmed2$V2
hind_test=ts(elekt_hind2,start=2015,frequency=365)

#moodikud tavalise rea korral
moodikud=function(mudel,test_rida){
  vead=residuals(mudel)[731:820]
  MAD=mean(abs(vead))
  MSE=mean(vead**2)
  MAPE=mean(abs(vead/test_rida))
  RMSE=sqrt(MSE)
  return(list(MAD=MAD,RMSE=RMSE,MAPE=MAPE))}

#ennustamine tavaline
m24_test=arima(c(hind_treen,hind_test), order=c(1,0,1),
                seasonal=list(order=c(2,0,1),period=7),fixed=m24$coef)
moodikud(m24_test,hind_test)

#moodikud logaritmitud rea korral
moodikud_log=function(mudel,rida){ #sisendiks logaritmimata rida
  prognoosid_log=log(rida)-residuals(mudel)[731:820]
  prognoosid=exp(prognoosid_log)
  vead=rida-prognoosid
  MAD=mean(abs(vead))
  MSE=mean(vead**2)
  RMSE=sqrt(MSE)
  MAPE=mean(abs((vead)/rida))
  return(list(MAD=MAD,RMSE=RMSE,MAPE=MAPE))}

#ennustamine logaritmitud
m3_log_test=arima(c(log_hind_treen,log_hind_test), order=c(2,0,1),
seasonal=list(order=c(2,0,1),period=7),fixed=m3_log$coef)
moodikud_log(m3_log_test,hind_test)

#prognooside graafik
prog_log=log_hind_test-residuals(m3_log_test)[731:820]
progn_log=exp(prog_log)
prog_log_ts=ts(progn_log,start=2015,frequency=365)
plot(hind_test,type="b",xlab="Aeg", ylab="Elektrihind(EUR/MWh)")
lines(prog_log_ts, type="b",col="red")
vead_log=hind_test-progn_log
max(abs(vead_log))
min(abs(vead_log))

#temperatuur treening
temp1=read.table("C:\\Users\\kartpal\\temp.txt",sep="",skip=1)
temperatuur1=temp1$V2
temp_treen=ts(temperatuur1,start=2013,frequency=365)
sum(is.na(temp_treen)) #ei ole puuduvaid

#temperatuur test
temp2=read.table("C:\\Users\\kartpal\\temp_rida_2015.txt",sep="")
```

```
temp_test=ts(temp2,start=2015,frequency=365)
sum(is.na(temp_test)) #ei ole puuduvaid

plot(temp_treen,type="l",
     ylab="Ohutemperatuur C",xlab="Aeg")

#regressorread 1
x1=pmax(-5-temp_treen,0)
x2=pmax(temp_treen+5,0)
x3=pmax(temp_treen-15,0)

#vajalik sobiva mudeli leidmiseks
m=arima(hind_treen,xreg=cbind(x1,x2,x3))
v=residuals(m)
graafikud(v,50)

#ei uurita AR, MA ega ARMA mudeleid,need ei andnud varasemalt
#tulemusi; joonise pohjal voiks sobida AR(3),ARMA(1,0,1),(1,0,2,),
#(2,0,1); sesoonses osas 3 autokorrelatsiooni ja 1 osaautokorrelatsioon
#kindlalt valjas
mm1=arima(hind_treen, order=c(1,0,2),seasonal=list(order=c(1,0,0),
period=7),xreg=cbind(x1,x2,x3))#ma2 ebaoluline
#seega pole motet mittesesoonses osas (1,0,2) mudeleid edasi uurida
mm2=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(1,0,0),
period=7),xreg=cbind(x1,x2,x3))#AIC=4964.81
mm3=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(1,0,0),
period=7),xreg=cbind(x1,x2,x3))#ei sobi
mm4=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(0,0,3),
period=7),xreg=cbind(x1,x2,x3))#AIC=4958.92
mm5=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(0,0,3),
period=7),xreg=cbind(x1,x2,x3))#AIC=4959.17
mm6=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(0,0,3),
period=7),xreg=cbind(x1,x2,x3))#AIC=4960.16
mm7=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,2),
period=7),xreg=cbind(x1,x2,x3))#AIC=4941.83
mm8=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(1,0,2),
period=7),xreg=cbind(x1,x2,x3))#ei sobi
mm9=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(1,0,2),
period=7),xreg=cbind(x1,x2,x3))#AIC=4943.56
mm10=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(2,0,1),
period=7),xreg=cbind(x1,x2,x3))#ei sobi
mm11=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(2,0,1),
period=7),xreg=cbind(x1,x2,x3))#AIC=4940.33, parim
tsdiag(mm11,80)
mm11
mm12=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(2,0,1),
period=7),xreg=cbind(x1,x2,x3))#AIC=4942.9


# temperatuur ja elektrihind koos, graafik
m=loess(hind_treen~temp_treen)
plot(temp_treen,hind_treen,ylab="Elektrihind EUR/MWh",
xlab="Temperatuur C")
axis(side = 1, at = seq(0, 10, by = 1), tcl = -0.2)
points(temp_treen,predict(m),col="red")
```

```
#regressorread 2
x11=pmax(2-temp_treen,0)
x22=pmax(temp_treen-2,0)
x33=pmax(temp_treen-8,0)

#temperatuuri funktsioonid mitmemootmelise jaoks
x11_reg=pmax(2-c(temp_treen,temp_test),0)
x22_reg=pmax(c(temp_treen,temp_test)-2,0)
x33_reg=pmax(c(temp_treen,temp_test)-8,0)

plot(x22,type="l",ylab="Temperatuur C",xlab="Aeg")
lines(x11,type="l",col=2)
lines(x33,type="l",col=4)

#vajalik sobiva mudeli leidmiseks
m1=arima(hind_treen,xreg=cbind(x1,x2,x3))
v1=residuals(m1)
graafikud(v1,50)

#joonise pohjal voiks sobida AR(1),AR(3),ARMA(1,0,1),ARMA(1,0,2,),
#(2,0,1); mittesesoonses AR(1),AR(3),ARMA(1,0,1),ARMA(1,0,2,),(2,0,1)
#sesoonses osas ARMA(0,0,3), ARMA(1,0,2), ARMA(2,0,1), voiks ka AR(1)
mm1_2=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(0,0,3),
period=7),xreg=cbind(x11,x22,x33))#AIC=4948.91
mm2_2=arima(hind_treen, order=c(1,0,2),seasonal=list(order=c(0,0,3),
period=7),xreg=cbind(x11,x22,x33))#AIC=4950.52
mm3_2=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(0,0,3),
period=7),xreg=cbind(x11,x22,x33))#AIC=4950.27
mm4_2=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,2),
period=7),xreg=cbind(x11,x22,x33))#AIC=4933.22
mm5_2=arima(hind_treen, order=c(1,0,2),seasonal=list(order=c(1,0,2),
period=7),xreg=cbind(x11,x22,x33))#AIC=4934.21
mm6_2=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(1,0,2),
period=7),xreg=cbind(x11,x22,x33))#AIC=4932.4
mm7_2=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(1,0,2),
period=7),xreg=cbind(x11,x22,x33))#AIC=4934.8
mm8_2=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(2,0,1),
period=7),xreg=cbind(x11,x22,x33))#AIC=4932.85
#ilma x2-ta
mm9_2=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(2,0,1),
period=7),xreg=cbind(x11,x33))#AIC=4932.62,parim
tsdiag(mm9_2,80)
mm9_2
mm10_2=arima(hind_treen, order=c(1,0,2),seasonal=list(order=c(2,0,1),
period=7),xreg=cbind(x11,x22,x33))#AIC=4933.97
mm11_2=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(2,0,1),
period=7),xreg=cbind(x11,x22,x33))#ei sobi

#logaritmitud elekt_hind,regressoriks temperatuur
m2=arima(log_hind_treen,xreg=cbind(x11,x22,x33))
v2=residuals(m2)
graafikud(v2,50)

#joonise pohjal voiks sobida AR(1),AR(3),ARMA(1,0,1),ARMA(1,0,2,),(2,0,1);
```

```
#mittesesoonses AR(1),AR(3),ARMA(1,0,1),ARMA(1,0,2,),(2,0,1)
#sesoonses osas on 5 autokorrelatsiooni kindlalt valjas
#sesoonses osas AR(3),ARMA(0,0,4)/ARMA(0,0,5), ARMA(1,0,3)/ARMA(1,0,4)
#esmalt proovitakse vaiksema arvuga, kui ei sobi, siis suurendatakse
mm1_log=arima(log_hind_treen, order=c(1,0,0),seasonal=list(order=c(3,0,0),
period=7),xreg=cbind(x11,x22,x33))#ei sobi
mm2_log=arima(log_hind_treen, order=c(1,0,1),seasonal=list(order=c(3,0,0),
period=7),xreg=cbind(x11,x22,x33))#AIC=-772.55
mm3_log=arima(log_hind_treen, order=c(1,0,2),seasonal=list(order=c(3,0,0),
period=7),xreg=cbind(x11,x22,x33))#AIC=-774.14
mm4_log=arima(log_hind_treen, order=c(2,0,1),seasonal=list(order=c(3,0,0),
period=7),xreg=cbind(x11,x22,x33))#AIC=-773.6
mm5_log=arima(log_hind_treen, order=c(3,0,0),seasonal=list(order=c(3,0,0),
period=7),xreg=cbind(x11,x22,x33))#AIC=-775.06
mm6_log=arima(log_hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,3),
period=7),xreg=cbind(x11,x22,x33))#sma3 ebaoluline
#sesoonses osas (1,0,3) ei sobi, sest sma3 ebaoluline, seega vaja suurendada
#parameetrite arvu
mm7_log=arima(log_hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,4),
period=7),xreg=cbind(x11,x22,x33))#AIC=-792.15
mm8_log=arima(log_hind_treen, order=c(1,0,2),seasonal=list(order=c(1,0,4),
period=7),xreg=cbind(x11,x22,x33))#AIC=-795.5, parim
tsdiag(mm8_log,80)
mm8_log
mm9_log=arima(log_hind_treen, order=c(2,0,1),seasonal=list(order=c(1,0,4),
period=7),xreg=cbind(x11,x22,x33))#ei sobi
mm10_log=arima(log_hind_treen, order=c(3,0,0),seasonal=list(order=c(1,0,4),
period=7),xreg=cbind(x11,x22,x33))#AIC=-794.24
mm11_log=arima(log_hind_treen, order=c(1,0,1),seasonal=list(order=c(0,0,4),
period=7),xreg=cbind(x11,x22,x33))#AIC=-755.04
mm12_log=arima(log_hind_treen, order=c(3,0,0),seasonal=list(order=c(0,0,4),
period=7),xreg=cbind(x11,x22,x33))#AIC=-758.29
mm13_log=arima(log_hind_treen, order=c(1,0,2),seasonal=list(order=c(0,0,4),
period=7),xreg=cbind(x11,x22,x33))#AIC=-758.06
mm14_log=arima(log_hind_treen, order=c(2,0,1),seasonal=list(order=c(0,0,4),
period=7),xreg=cbind(x11,x22,x33))#AIC=-757.88
mm15_log=arima(log_hind_treen, order=c(1,0,1),seasonal=list(order=c(0,0,5),
period=7),xreg=cbind(x11,x22,x33))#AIC=-757
mm16_log=arima(log_hind_treen, order=c(3,0,0),seasonal=list(order=c(0,0,5),
period=7),xreg=cbind(x11,x22,x33))#AIC=-760.21
mm17_log=arima(log_hind_treen, order=c(1,0,2),seasonal=list(order=c(0,0,5),
period=7),xreg=cbind(x11,x22,x33))#AIC=-760.2
mm18_log=arima(log_hind_treen, order=c(2,0,1),seasonal=list(order=c(0,0,5),
period=7),xreg=cbind(x11,x22,x33))#AIC=-760.09

#moodikud arimax_2 jaoks
arimax_2_test=arima(c(hind_treen,hind_test),order=c(1,0,1),
seasonal=list(order=c(2,0,1),period=7),xreg=cbind(x11_reg,x33_reg),
fixed=mm9_2$coef)
moodikud(arimax_2_test,hind_test)

#moodikud arimax_2log jaoks
arimax2log_test=arima(c(log_hind_treen,log_hind_test), order=c(1,0,2),
seasonal=list(order=c(1,0,4),period=7),xreg=cbind(x11_reg,x22_reg,x33_reg),
```

```
fixed=mm8_log$coef)
moodikud_log(arimax2log_test,hind_test)

#Treening nadalapaev
nadalapaevad1=read.table("C:\\Users\\kartpal\\tp_nv_1_0_tabel2.txt",sep="")
np_treen=ts(nadalapaevad1,start=2013,frequency=365)

#Test nadalapaev
nadalapaevad2=read.table("C:\\Users\\kartpal\\nadalapaevad_2015.txt",sep="")
np_test=ts(nadalapaevad2,start=2015,frequency=365)

m3=arima(hind_treen,xreg=np_treen)
v3=residuals(m3)
graafikud(v3,50)

#joonise pohjal voiks sobida AR(1),AR(3),ARMA(1,0,1),ARMA(1,0,3),(2,0,1);
#joonise jargi sesoonses osas AR(1) voi MA(3)/MA(4)
mm1_3=arima(hind_treen,order=c(1,0,2),xreg=np_treen)#AIC=4927.13
mm2_3=arima(hind_treen,order=c(2,0,1),xreg=np_treen)#AIC=4925.45
mm3_3=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,0),
period=7),xreg=np_treen)#AIC=4920.22
mm4_3=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(1,0,0),
period=7),xreg=np_treen)#AIC=4921.33
mm5_3=arima(hind_treen, order=c(1,0,2),seasonal=list(order=c(1,0,0),
period=7),xreg=np_treen)#AIC=4920.09
mm6_3=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(1,0,0),
period=7),xreg=np_treen)#AIC=4919.36, parim
tsdiag(mm6_3,80)
mm6_3
mm7_3=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(0,0,4),
period=7),xreg=np_treen)#sma4 ebaoluline, seega piisab vaadata (0,0,3)
mm8_3=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(0,0,3),
period=7),xreg=np_treen)#AIC=4918.66
mm9_3=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(0,0,3),
period=7),xreg=np_treen)#AIC=4920.1
mm10_3=arima(hind_treen, order=c(1,0,2),seasonal=list(order=c(0,0,3),
period=7),xreg=np_treen)#ma2 ebaoluline
mm11_3=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(0,0,3),
period=7),xreg=np_treen)#ar2 ebaoluline
mm12_3=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,3),
period=7),xreg=np_treen)#sar1 ebaoluline, ei ole motet edasi uurida
mm13_3=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,2),
period=7),xreg=np_treen)#sma2 ebaoluline, ei ole motet edasi uurida

#logaritmitud elektrihind, regressoriks nadalapaevad
m4=arima(log_hind_treen,xreg=np_treen)
v4=residuals(m4)
graafikud(v4,50)

#sesoonses osas 4/5 autokorrelatsiooni ja 1 osaautokorrelatsioon valjas
#mittesesoonses osas AR(3),ARMA(1,0,1),ARMA(1,0,3),(2,0,1)
mm1_2log=arima(log_hind_treen, order=c(3,0,0),seasonal=list(order=c(1,0,0),
period=7),xreg=np_treen)#AIC=-801
mm2_2log=arima(log_hind_treen, order=c(1,0,2),seasonal=list(order=c(1,0,0),
period=7),xreg=np_treen)#AIC=-802.78
```

```
mm3_2log=arima(log_hind_treen, order=c(2,0,1),seasonal=list(order=c(1,0,0),
period=7),xreg=np_treen)#AIC=-802.93
mm4_2log=arima(log_hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,3),
period=7),xreg=np_treen)#sar1 ebaoluline
#pole motet (1,0,3) sesoonses osas edasi otsida
mm5_2log=arima(log_hind_treen, order=c(1,0,1),seasonal=list(order=c(0,0,4),
period=7),xreg=np_treen)#sma4 ebaoluline
mm6_2log=arima(log_hind_treen, order=c(1,0,1),seasonal=list(order=c(0,0,5),
period=7),xreg=np_treen)#ei sobi
#sma4 ebaoluline,sma5 ebaoluline, seega piisab (0,0,3) uurimisest
mm7_2log=arima(log_hind_treen, order=c(1,0,1),seasonal=list(order=c(0,0,3),
period=7),xreg=np_treen)#AIC=-804.71
mm8_2log=arima(log_hind_treen, order=c(3,0,0),seasonal=list(order=c(0,0,3),
period=7),xreg=np_treen)#AIC=-807.76
mm9_2log=arima(log_hind_treen, order=c(1,0,2),seasonal=list(order=c(0,0,3),
period=7),xreg=np_treen)#AIC=-808.74,parim
tdiag(mm9_2log,80)
mm9_2log
mm10_2log=arima(log_hind_treen, order=c(2,0,1),seasonal=list(order=c(0,0,3),
period=7),xreg=np_treen)#AIC=-808.3

#moodikud arimax_3 jaoks
arimax_3_test=arima(c(hind_treen,hind_test), order=c(2,0,1),
seasonal=list(order=c(1,0,0),period=7),xreg=c(np_treen,np_test),
fixed=mm6_3$coef)
moodikud(arimax_3_test,hind_test)

#moodikud arimax_3log jaoks
arimax3log_test=arima(c(log_hind_treen,log_hind_test), order=c(1,0,2),
seasonal=list(order=c(0,0,3),period=7),xreg=c(np_treen,np_test),
fixed=mm9_2log$coef)
moodikud_log(arimax3log_test,hind_test)


#regressorid kokku
m5=arima(hind_treen,xreg=cbind(x11,x22,x33,np_treen))
v5=residuals(m5)
graafikud(v5,50)
#sesoonses osas 1 osaautokorrelatsioon ja 4autokorrelatsiooni valjas
#sesoonses osas tasuks proovida MA(4), voib-olla sobib ka juba MA(3)
#AR(3),ARMA(1,0,1),ARMA(1,0,3),(2,0,1)
mm1_4=arima(hind_treen,order=c(2,0,1),xreg=cbind(x11,x22,x33,np_treen))
#AIC=4909.59, parim
#ilma x2-ta
mm2_4=arima(hind_treen,order=c(2,0,1),xreg=cbind(x11,x33,np_treen))
#AIC=4912.39
mm3_4=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,0),
period=7),xreg=np_treen)#AIC=4920.22
mm4_4=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(1,0,0),
period=7),xreg=np_treen)#AIC=4921.33
mm5_4=arima(hind_treen, order=c(1,0,2),seasonal=list(order=c(1,0,0),
period=7),xreg=np_treen)#AIC=4920.09
mm6_4=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(1,0,0),
period=7),xreg=np_treen)#AIC=4919.36
```

```
mm7_4=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(0,0,4),
period=7),xreg=np_treen)#sma4 ebaoluline, seega piisab vaadata (0,0,3)
mm8_4=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(0,0,3),
period=7),xreg=np_treen)#AIC=4918.66
mm9_4=arima(hind_treen, order=c(3,0,0),seasonal=list(order=c(0,0,3),
period=7),xreg=np_treen)#AIC=4920.1
mm10_4=arima(hind_treen, order=c(1,0,2),seasonal=list(order=c(0,0,3),
period=7),xreg=np_treen)#AIC=4919.09
mm11_4=arima(hind_treen, order=c(2,0,1),seasonal=list(order=c(0,0,3),
period=7),xreg=np_treen)#ar2 ebaoluline
mm12_4=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,3),
period=7),xreg=np_treen)
#sar1 on ebaoluline, seega sesoonsesse osasse ei sobi (1,0,3) mudelid
mm13_4=arima(hind_treen, order=c(1,0,1),seasonal=list(order=c(1,0,2),
period=7),xreg=np_treen)#sma2 ebaoluline

#logaritmitud regressorid kokku
m6=arima(log_hind_treen,xreg=cbind(x11,x22,x33,np_treen))
v6=residuals(m6)
graafikud(v6,50)
#sesoonses osas 1 osaautokorrelatsioon ja 5 autokorrelatsiooni valjas
#AR(3),ARMA(1,0,1),ARMA(1,0,3),(2,0,1)
mm1_3log=arima(log_hind_treen, order=c(3,0,0),
seasonal=list(order=c(1,0,0),period=7),xreg=cbind(x11,x33,np_treen))
#AIC=-816.57
mm2_3log=arima(log_hind_treen, order=c(1,0,2),
seasonal=list(order=c(1,0,0),period=7),xreg=cbind(x11,x33,np_treen))
#AIC=-817.48
mm3_3log=arima(log_hind_treen, order=c(2,0,1),
seasonal=list(order=c(1,0,0),period=7),xreg=cbind(x11,x22,x33,np_treen))
#AIC=-824.87,parim
mm4_3log=arima(log_hind_treen, order=c(1,0,1),
seasonal=list(order=c(0,0,5),period=7),xreg=np_treen)#sma5 ebaoluline
mm5_3log=arima(log_hind_treen, order=c(1,0,1),
seasonal=list(order=c(0,0,4),period=7),xreg=np_treen)#sma4 ebaoluline
mm6_3log=arima(log_hind_treen, order=c(1,0,1),
seasonal=list(order=c(0,0,3),period=7),xreg=np_treen)#AIC=-804.71
mm7_3log=arima(log_hind_treen, order=c(3,0,0),
seasonal=list(order=c(0,0,3),period=7),xreg=np_treen)#AIC=-807.76
mm8_3log=arima(log_hind_treen, order=c(1,0,2),
seasonal=list(order=c(0,0,3),period=7),xreg=np_treen)#AIC=-808.74
mm9_3log=arima(log_hind_treen, order=c(2,0,1),
seasonal=list(order=c(0,0,3),period=7),xreg=np_treen)#AIC=-808.3
mm10_3log=arima(log_hind_treen, order=c(1,0,1),
seasonal=list(order=c(1,0,4),period=7),xreg=np_treen)#sar1,sma4 ebaoluline
mm11_3log=arima(log_hind_treen, order=c(1,0,1),
seasonal=list(order=c(1,0,3),period=7),xreg=np_treen)#sar1 on ebaoluline
#sesoonses osas ei sobi (1,0,2), ka (2,0,1) ei sobi

#moodikud arimax4
arimax4_test=arima(c(hind_treen,hind_test),order=c(2,0,1),
xreg=cbind(x11_reg,x22_reg,x33_reg,c(np_treen,np_test)),fixed=mm1_4$coef)
moodikud(arimax4_test,hind_test)
```

```
#moodikud arimax4_log
arimax4log_test=arima(c(log_hind_treen,log_hind_test), order=c(2,0,1),
seasonal=list(order=c(1,0,0),period=7),xreg=cbind(x11_reg,x22_reg,x33_reg,
c(np_treen,np_test)),fixed=mm3_3log$coef)
moodikud_log(arimax4log_test,hind_test)

#prognoosid parim mitmemootmeline

prog_log=log_hind_test-residuals(arimax2log_test)[731:820]
progn_log=exp(prog_log)
prog_log_ts=ts(progn_log,start=2015,frequency=297)
plot(hind_test,type="b",xlab="Aeg", ylab="Elektrihind(EUR/MWh)")
lines(prog_log_ts, type="b",col="red")
vead_log=hind_test-progn_log
max(abs(vead_log))
min(abs(vead_log))
```

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks.

Mina, Kärt Päll (sünnikuupäev: 17.01.1991)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

   Eesti elektrienergia hinna analüüs ja ühesammuline prognoosimine ARIMA tüüpi mudelitega,

   mille juhendaja on Raul Kangro,

   1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace'is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

   1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartu, 12. mai 2015. a.