

## XIV PRAKTIKUM

**Teemad:** ARIMA modelleerimine. Parima mudeli valik Akaike informatsioonikriteeriumi põhjal. Staatile ja dünaamiline prognoosimine.

**Failid** *eestiTHI.gdt; eestiSKP.gdt; ukhp.gdt; aktsiad\_BLT\_HAE.gdt; hae.inp; uk\_tarbimine.gdt*

## Ülesanne 1. Eesti THI modelleerimine

Fail *eestiTHI.gdt*

THI Eesti tarbijahinnaindeks (1997=100), kuud 1998 jaan – 2013 dets.

- 1) Tutvuda aegrea diagrammiga. Veenduda, et esineb trend.
- 2) Trendi eemaldamiseks võtta aegreast 1. järku diferentsid. Uurida diagrammilt, kas trend õnnestus eemaldada.
- 3) ARMA modelleerimiseks tuleb aegrea mudel identifitseerida, selleks uurida korrelogramme. Millist tüüpi mudeliga võiks diferentside aegrida korrelogrammide põhjal modelleerida?
- 4) Hinnata vastava mudeli parameetrid ja kontrollida, kas need on statistiliselt olulised.
- 5) Kontrollida, kas jäägid moodustavad valge müra.
- 6) Kontrollida, kas jäägid alluvad normaaljaotusele.
- 7) Aktuaalsete väärtuste ja mudelväärtuste graafikute vaatamiseks valida mudeli aknas *Graphs-> Fitted, actual plot ->Against time*
- 8) Kasutada hinnatud mudelit 2014. a väärtuste prognoosimiseks *Analysis->Forecasts*. Kuna andmebaasis on väärtused ainult 2013. detsembrini, tuleb lisada vaatlusi ja selleks avaneb aken *Add observations*. Lisada 12 väärtust, st valime prognoosi pikkuseks 12 perioodi (1 aasta). Prognoositud väärtusi saab vaadata nii graafikul kui tabelis. Tabelis on lisaks antud ka prognoosi standardviga ja usalduspiirid (viimased veerud).
- 9) Eelmises punktis leiti prognoos 1. järku diferentside jaoks, sest modelleeriti just seda aegrida. Enamasti ei soovi me prognoosida mitte diferentse, vaid aegrea taset, antud juhul siis THI-d. Kui 1. järku diferentside jaoks on ARMA(p,q) mudel, siis aegrea tasemele vastab mudel ARIMA(p,1,q). Järelikult tuleb läbi viia ARIMA modelleerimine. Selleks valida sõltuvaks tunnuseks vastava aegrea tase (THI) ja mudeli spetsifikatsiooni aknas märkida 1. järku diferentside mudeli vastavad ARMA järgud ja lisaks modelleeritavate diferentside järk. Antud juhul, kuna 1. järku diferentsid olid AR(2) tüüpi, siis

*Non-seasonal*

*AR order:* 2

*Difference:* 1

*MA order:* 0

Tulemuseks saadakse sama mudel, mis punktis 4). Erinevus on protseduuris: punktis 4) me kasutasime eelnevalt leitud 1. järku diferentse, aga nüüd leitakse 1. järku diferentsid programmi poolt automaatselt enne mudeli hindamist.

- 10) Hinnata mudeli prognoosimisvõimet valimi sees. Selleks mudeli aknas *Analysis-> Display actual, fitted residual*. Kuvatakse tabel tegelike väärtuste, mudelväärtuste ja jääkidega. Tabeli lõpus on prognoosimisvõime näitajad. Leida sealt

a) Keskmise viga ME .....

b) Juuritud keskmine ruutviga RMSE .....

- c) Keskmise suhteline absoluutviga MAPE (protsenti) .....  
Heaks loetakse prognoose, mille korral MAPE ei ületa 20%. Kas seda prognoosi saab pidada heaks?
- d) Theil'i U .....  
Kas prognoos on parem või halvem kui naiivne prognoos?

11) Viia läbi THI 2014. a. väärtuste prognoosimine. Võrrelda 2014. aasta prognoositud väärtusi all tabelis esitatud tegelike väärtusega. Kas tegelikud väärtused mahuvad prognoosi usaldusvahemiku piiridesse?

2014:1	2014:2	2014:3	2014:4	2014:5	2014:6	2014:7	2014:8	2014:9	2014:10
195,95	196,19	196,77	196,72	196,90	196,97	196,67	196,07	195,73	195,72

**Ülesanne 2. Eesti SKP modelleerimine**

Fail *eestiSKP.gdt*

SKP SKP sesoonselt ja tööpäevade arvuga korrigeerimata, jooksevhindades (mln eurot)  
1995:1 – 2007:4

- 1) Tutvuda aegrea diagrammiga. Veenduda, et esineb trend ja sesoonsus.
- 2) Sesoonsuse eemaldamiseks võtta sesoonsed diferentsid. Uurida nende aegrea diagrammi. Kas trend õnnestus ka eemaldada?
- 3) Trendi eemaldamiseks võtta sesoonsetest diferentsidest veel 1. järku diferentsid ja uurida nende aegrea diagrammi.
- 4) Viimasel teisendusel saadud aegrea (sesoonsete diferentside 1. järku diferentsid) modelleerimiseks tuleb kõigepealt kontrollida, kas autokorrelatsioon esineb ja kui esineb, siis identifitseerida aegrida (st leida sobiv ARMA tüüp). Selleks uurida korrelogramme. Kas autokorrelatsioon esineb? Kas korrelogrammide abil on võimalik aegrida identifitseerida?
- 5) Kui korrelogrammide abil aegrida identifitseerida ei õnnestu, tuleb proovida erinevaid ARMA mudeleid ja valida välja see, mille korral Akaike informatsioonikriteerium on kõige väiksem. Hinnata järgmisi ARMA mudeleid: AR(1), AR(2), MA(1), MA(2), ARMA(1,1). Iga mudeli korral salvestada Akaike informatsioonikriteerium.  
*Save -> Akaike Information Criterion*  
Nimi võiks viidata mudeli kujule, näiteks `aic_ar1`, `aic_ar2`, ....
- 6) Salvestatud Akaike informatsioonikriteeriumite võrdlemiseks avada ikoonide aknas ikoon *Scalars*. Parim mudel on see, mille korral Akaike informatsioonikriteerium on väikseim.
- 7) Parimaks osutunud mudelis on konstantne liige statistiliselt mitteoluline. Seepärast viia läbi mudeli hindamine ilma konstandita, mis tähendab, et aegrea keskvärtus on 0. Kuidas muutus Akaike informatsioonikriteerium?
- 8) Viia läbi väljavalitud mudeli diagnostika: kas parameetrid on statistiliselt olulised, kas jäägid moodustavad valge müra (autokorrelatsioon puudub), kas jäägid alluvad normaaljaotusele.
- 9) Kui mudel on korrektne (mudeli diagnostika läbitud), võib seda kasutada prognoosimiseks. Prognoos viia läbi SKP jaoks, mitte selle diferentside jaoks. Selleks hinnata vastavat mudelit, kus on märgitud diferentside järk ( mittesessoonne 1 ja sessoonne 1). Kontrollida, kas tuleb sama mudel, mis saadi sesoonsete diferentside 1. järku diferentside modelleerimisel.

- 10) Võrrelda 2008. a 1. kvartali prognoositud väärtust SKP tegeliku väärtusega, mis oli 3810,236 mln eurot. Kas tegelik väärtus jääb prognoosi usalduspiiridesse?

### Ülesanne 3. Majade keskmine hind Suurbritannias

Fail *ukhp.gdt*

HP majade keskmine hind perioodil 1996. a jaanuar kuni 2007. a mai, kuised andmed, naelades.

- 1) Tutvuda aegrea diagrammiga ja veenduda, et aegrida on mittestatsionaarne ning trend on mittelineaarne.
- 2) Mittelineaarsuse eemaldamiseks logaritmid aegrida.
- 3) Trendi eemaldamiseks leida logaritmitud aegrea 1. järku diferentsid  $d_1$  HP.
- 4) Uurida diferentsitud logaritmade aegrea korrelogrammi. Kas korrelogrammi põhjal on võimalik aegrida korralikult identifitseerida, st määrata ARMA tüüpi?
- 5) Viia läbi AR(2) mudeli hindamine. Kas kõik parameetrid on statistiliselt olulised?
- 6) Uurida jääkliikmete korrelogrammi (*Graphs*-> *Residual correlogramm*). Kontrollida, kas jääkliikmed moodustavad valge müra. Korrelogrammi tabelist on näha, et autokorrelatsioon puudub kuni viitajani 11. Alates viitajast 12 on autokorrelatsioon olemas. See on seletatav sellega, et tegemist on kuiste andmetega.
- 7) AR(2) mudelisse tuleks lisada 12. järku viitajaga autoregressiivne liige. Selleks tuleb viia läbi uus ARMA mudeli hindamine ning mudeli spetsifitseerimise aknas märkida AR järgu juures „*specific lags*“ ning vastavasse kasti kirjutada 1,2,12. Sellise mudeli tähistus on AR([1,2,12]).
- 8) Mudeli diagnostika
  - a) Kas parameetrid on statistiliselt olulised?
  - b) Kas jäägid moodustavad valge müra?
  - c) Kas jäägid alluvad normaaljaotusele?
- 9) Hinna logaritmi prognoosimise jaoks viia läbi mudeli hindamine, kus sõltuvaks tunnuseks on hinna logaritmi ja mudeli spetsifikatsioonis määratakse AR järkudele lisaks ka diferentside järk. Kontrollida, kas tuleb sama mudel, mis punktis 7).
- 10) Viia läbi prognoosimine perioodiks 2007:06 kuni 2007:11 ja leida hinna logaritmi prognoos 2007. aasta juunis.
- 11) Logaritmi prognoosi alusel arvutada majade keskmise hinna prognoos 2007. aasta juunis.

### Ülesanne 4. Baltika ja Harju Elekter aktsiate tulumäärad.


Fail *aktsiad\_BLT\_HAE.gdt* ja skriptifail *hae.inp*

Andmebaasis on kahe Tallinna börsil noteeritud aktsia tulumäär päevas ajavahemikul 1.12.2009 – 30.11.2010 (protsentides).

HAE1T	Harju elekter
BLT1T	Baltika

- 1) Uurida mõlema aktsia tulumäära korrelogrammi. Kas autokorrelatsioon esineb? Kas ühe või mõlema aktsia tulumäära aegrida vastab valgele mürale?
- 2) Aktivate tulumäärade käitumist analüüsivad uurimised on näidanud, et enamasti järgivad aktivad peaaegu normaaljaotust. Kontrolli, kas nende aktsiate tulumäärade korral peab see paika.
- 3) Kumb aktsia on volatiilsem (tulumäära standardhälve suurem)? Kontrollida oma järeldust vastava *F*-testi abil (dispersioonide erinevuse testimine).
- 4) HAE1T aktsia ARIMA modelleerimine.

Proovida läbi järgmised mudelid AR(1), MA(1), ARMA(1,1). Iga mudeli korral kontrollida jääkliikmete vastavust valgele mürale ning märkida üles või salvestada Akaike informatsioonikriteerium AIC. Milline neist kolmest mudelist on AIC põhjal kõige parem?

- 5) Selline üksikmuutujate mudelite koostamine võtab palju aega. Mugavam on kasutada programmeerimist ja koostada Gretli skript, mis järjest viib läbi erinevate mudelite hindamise ning salvestab iga mudeli korral AIC väärtuse vastavasse tabelisse. Pärast võib tabelist määrata, millise mudeli korral on AIC kõige väiksem, hinnata seda mudelit ja viia läbi jääkliikmete testimine.
  - a) Gretlis avada skripti fail *hae.inp* (*File->Script files->User file*). Tutvuda skriptiga ja seejärel käivitada see ikoonist *Run* . Skript viib läbi ARMA mudelite hindamise AR ja MA järkudega 0 kuni 5 (kõikvõimalikud kombinatsioonid) ja salvestab iga mudeli korral Akaike informatsioonikriteeriumi AIC väärtuse.
  - b) Peale skripti töö lõppemist avada ikoonide aknas maatriks A, mis sisaldab erinevate mudelite AIC väärtusi. NB! Kuna maatriksi reanumber ega veerunumber ei saa olla null, siis näiteks mudelile ARMA(1,0) vastav AIC väärtus asub kohal (2,1), st 2. rea 1. veerg.
  - c) Leida maatriksist, millise ARMA mudeli korral on AIC väärtus kõige väiksem.
- 6) Viia läbi vastava ARMA mudeli hindamine ning selle jääkliikmete diagnostika.

### Ülesanne 5. Majapidamiste tarbimiskulud Suurbritannias

Fail *uk\_tarbimine.gdt*

SKP arvutamisel kulumeetodil on üheks oluliseks komponendiks majapidamiste tarbimiskulutused. SKP prognoosimisel on vajalik järelikult nende kulutuste prognoosimine. Majapidamiste tarbimiskulud kokku koosnevad kuludest erinevatele teenustele ja kaupadele.

Andmebaasis on Suurbritannia majapidamiste lõpptarbimise kulud toidule, riieale ja transpordile<sup>1</sup>. Tegemist on kvartaalsete andmetega, 1985 I kv kuni 2016 II kv. Eesmärgiks on kõikide aegridade jaoks leida sobiv ARIMA mudel ja selle alusel teha prognoos järgnevas neljaks kvarataliks 2016 III kv kuni 2017 II kv.

toit	Majapidamiste kulud toidule, mln GBP
riided	Majapidamiste kulud riieale, mln GBP
transport	Majapidamiste kulud transpordile, mln GBP

- 1) Tutvuda aegridade diagrammidega ja veenduda, et esineb kvartaalne sesoonsus.
- 2) Kõikide aegridade korral leida sesoonsed diferentsid. Aegrea mudelid leida sesoonsete diferentside jaoks.
- 3) Uurida sesoonsete diferentside korrelogramme. Kas nende abil on võimalik üheselt määrata aegrea tüüp?
- 4) Proovida kõikide sesoonsete diferentside jaoks mudeleid AR(1), AR(2), ARMA(1,1). Välistada need mudelid, mille kordajad pole statistiliselt olulised nivool 0,05 ja mille korral jäägid ei moodusta valget müra viitajani 21 (testimiseks kasutada jääkliikmete korrelogramme).
- 5) Nende aegridade korral, kus sõelale jäi rohkem kui üks mudel, valida sõelale jäänud mudelite hulgast parim välja Akaike kriteeriumi alusel.
- 6) Kasutades sesoonsete diferentside jaoks leitud mudeleid, modelleerida aegrea tasemeid vastavate ARIMA mudelitega (sesoone diferentsi järk 1).
- 7) Võrrelda mudelite prognoosimisvõimet, kasutades keskmist suhtelist absoluutviga MAPE. Millist aegrida on võimalik õige paremini prognoosida?
- 8) Prognoosida kulutusi toidule, riieale ja transpordile neli kvartalit edasi, so perioodiks 2016 III kv – 2017 II kv.

<sup>1</sup> UK Office of National Statistics <https://www.ons.gov.uk>

VASTUSED

Ül 1. 3). AR(2) 4) Mudel on  $y = 0,26 + 0,307y_{t-1} + 0,142y_{t-2} + u_t$ , kus  $y$  on THI esimest järku diferentsid. Aegrea keskväärtus on 0,468. 5) Jäägid moodustavad valge müra, Box-Ljungu Q-statistiku olulisuse tõenäosus viitajani 12 on  $0,3515 > 0,05$ . 6) Jäägid ei allu normaaljaotusele, vastava testi (Doornik-Hansen) olulisuse tõenäosus  $p = 0,00012 < 0,05$ . 8) THI 1. järku diferentsi prognoositud väärtus 2014. aasta jaanuaris on 0,09, veebr 0,25. 10 a) ME = -0,0017, b) RMSE = 0,628, c) MAPE = 0,327%, väga hea prognoos. d) Theili U = 0,765 < 1, prognoos on parem kui naiivne. 11) Kõik tabelis esitatud 2014. aasta tegelikud väärtused jäävad prognoosi usalduspiiridesse.

Ül.2. 6) AR(1) AIC=514,52, AR(2) AIC=516,33, MA(1) AIC=517,29, MA(2) AIC=508,78, ARMA(1,1) AIC=519,45. Nendest mudelitest on Akaike informatsioonikriteeriumi järgi parim mudel MA(2). 7) Vähenes, nüüd AIC=508,2. 8) Jäägid moodustavad valge müra: autokorrelatsioon puudub viitajani 4 (Q statistiku olulisuse tõenäosus  $p = 0,1931 > 0,05$ ). Jäägid alluvad normaaljaotusele  $p = 0,0669 > 0,05$ . 10) 2008. a I kvartali SKP prognoositud väärtus 4298,72 mln eurot, usalduspiiridega 4183,18066 - 4414,26198 mln eurot. Tegelik väärtus ei jää prognoosi usalduspiiridesse.

Ül. 3.5)

AR2: ARMA, using observations 1996:01-2007:05 (T = 137)

Dependent variable: d\_l\_HP

Standard errors based on Hessian

	<i>Coefficient</i>	<i>Std. Error</i>	<i>z</i>	<i>p-value</i>	
const	0,00923825	0,00134361	6,8757	<0,0001	***
phi_1	0,233586	0,0841477	2,7759	0,0055	***
phi_2	0,21075	0,0843135	2,4996	0,0124	**

7)

AR1,2,12: ARMA, using observations 1996:01-2007:05 (T = 137)

Dependent variable: d\_l\_HP

Standard errors based on Hessian

	<i>Coefficient</i>	<i>Std. Error</i>	<i>z</i>	<i>p-value</i>	
const	0,00906481	0,00184697	4,9079	<0,0001	***
phi_1	0,176745	0,0839906	2,1043	0,0353	**
phi_2	0,228694	0,0817722	2,7967	0,0052	***
phi_12	0,221368	0,0798409	2,7726	0,0056	***

8) a) Kõik parameetrid on statistiliselt olulised. b) Jääkliikmed moodustavad valge müra (Box-Ljungu Q-statistiku olulisuse tõenäosus  $p = 0,1189$ ). c) Jääkliikmed alluvad normaaljaotusele ( $p = 0,489$ ). 10) 2007. a juuni prognoositud väärtus keskmise hinna logaritmi jaoks on 12,1197. 11) 2007. a juuni prognoositud väärtus keskmise hinna jaoks ca 183,5 tuhat naela.

Ül. 4. 1) HAE1T autokorrelatsioon esineb, 1. viitajale vastava Q-statistiku olulisuse tõenäosus 0,008. BLT1T autokorrelatsioon puudub viitajani 24, Q-statistiku olulisuse tõenäosus viitaja 24 korral 0,878, järelikult BLT1T aktsia tulumäära aegrida on valge müra. 2) Kumbki ei allu normaaljaotusele: Jarque-Bera testi olulisuse tõenäosus HAE1T korral  $2,55 \cdot 10^{-17}$ , BLT1T korral  $1,76 \cdot 10^{-13}$ . St BLT1T ei ole Gaussi valge müra. 3) BLT1T on volatiilsem: BLT1T standardhälve 3,14, HAE1T korral 2,24. F-testi olulisuse tõenäosus (ühepoolne test)  $5,038 \cdot 10^{-8}$ , järelikult Baltika aktsia tulumäära dispersioon on suurem. 4) AR(1) mudelil AIC= 1123,403, MA(1): AIC=1122,042, ARMA(1,1): AIC=1123,608. Kõikide mudelite jääkliikmed moodustavad valge müra (viitajani 24), Q-statistiku olulisuse tõenäosus > 0,05. Esineb aga üksikuid autokorrelatsioonikordajate väärtusi, mis väljuvad usalduspiiridest, st on

oluliselt erinevad nullist. Neist kolmest mudelist on AIC põhjal parim mudel MA(1). **5)** Kõige väiksem on AIC väärtus mudeli ARMA(3,2) korral: 1121,1767. **6)** Selle mudeli korral kõik parameetrid, välja arvatud konstant, on statistiliselt olulised. Jäägid moodustavad valge müra viitajani 5 (Ljung-Box statistiku olulisuse tõenäosus  $p=0,7407$ ). Jäägid ei allu normaaljaotusele.

**Ül. 5. 3)** Toit ja riided võiksid olla AR(1) või AR(2). Transport: raske määrata. **4)** Toit: järele jäävad AR(2) ja ARMA(1,1). Riided: järele jäävad AR(1), AR(2) ja ARMA(1,1). Transport: järele jääb AR(2). ARMA(1,1) korral MA kordaja ei ole statistiliselt oluline nivool 0,05, on nivool 0,1. **5)** Toit: AR(2) AIC=1753,46; ARMA(1,1) AIC=1754,77; parem on AR(2). Riided: AR(1) AIC=1634,31; AR(2) AIC=1626,39; ARMA(1,1) AIC=1631,59. Parim on ARMA(1,1). Lõplik valik sesoonsete diferentside jaoks: toit AR(2), riided ARMA(1,1), transport AR(2). **7)** MAPE toit 1,71%, riided 1,83%, transport 2,28%. Kõige paremini saab prognoosida kulutusi toidule. **8)** Prognoosid 2017. a II kvartaliks: kulutused toidule 20 904 mln naela, kulutused riidele 14 052 mln naela, kulutused transpordile 42 485 mln naela.