

Aikasarjatutkimuksia Valkeakosken kaupunki-ilman
hajurikkipitoisuuksista

Tampereen yliopisto
Informaatiotieteiden tiedekunta

VÄISÄNEN, JAANI

Pro gradu -tutkielma
Tilastotiede
Lokakuu 2004

TAMPEREEN YLIOPISTO

Informaatiotieteiden tiedekunta/Tilastotieteen laitos

VÄISÄNEN, JAANI. Aikasarjatutkimuksia Vakekosken kaupunki-
ilman hajurikkipitoisuuksista

Tilastotieteen pro gradu –tutkielma, 122 s. + 9 liites.

Tilastotiede

Lokakuu 2004

Avainsanat: *saastepitoisuus, autokorrelaatio, interventioanalyysi,*
vektoriautoregressiiviset mallit

Kaupunkien saasteongelmat ovat akuutteja Suomessakin. Asutuskeskuksia ja taajamia lähellä olevat tehtaat vähentävät asukkaiden viihtyvyyttä ja voivat luoda vakavia terveysriskejä iäkkäille ja huonokuntoisille. Tämän tutkielman tarkoituksena on tutkia ilman hajurikkipitoisuuksia Valkeakosken keskustassa vuosilta 1990-2001. Tutkimuksessa myöskin käytetään hyväksi tietoja ilman rikkidioksidipitoisuuksista. Em. pitoisuuksien lisäksi saatavilla ovat myöskin tiedot vallitsevista sääolosuhteista kyseiseltä ajalta.

Tutkimus voidaan jakaa pääpiirteissään komeen osaan. Ensimmäisessä osassa(luku 4) käytetään regressioanalyysin teoriaa hyväksi osoitettaessa hajurikkipitoisuuksien vaihtelua vuorokaudenajan mukaan. Toinen osa perustuu pääasiassa aikasarjojen käyttöön ja tarkastelee hajurikkipitoisuuksien ja MILOS –säähavaintoasemalta saatujen lämpötilamittausten ristikorrelaatiofunktioita. Kolmas osa(luku 13) osoittaa edellä tehtyjen ympäristönsuojeluinvestointien vaikutukse interventioanalyysin avulla. Neljännessä osassa(luvut 15 ja 17) hajurikkiiyhdisteille ja rikkidioksidille muodostetaan vektoriautoregressiiviset mallit ja osoitetaan molempien sarjojen suhde toisiinsa. Samalla myös todetaan, että vaikka sarjojen välillä näyttääkin vallitsevan tietty relaatio, eivät sarjat ole yhteisintegroituneita.

1. JOHDANTO	5
2. AINEISTO	5
2.1. RIKKIVETY	6
2.2. RIKKIDIOKSIDI	7
2.3. AINEISTON TAUSTAA	7
2.4. TECO –MITTAUS	8
2.5. OPSIS –MITTAUS	8
2.6. MILOS –SÄÄHAVAINTOASEMA	9
3. AIKASARJA-ANALYYSIN PERUSKÄSITTEITÄ.....	9
3.1. AIKASARJA	9
3.2. STATIONAARISUUS.....	10
3.3. AUTOKORRELAATIO JA OSITTAISAUTOKORRELAATIO	11
3.4. VALKOINEN KOHINA	14
3.5. VIIVEPERAATTORIT.....	15
4. VUOROKAUDENAJAN VAIKUTUS HAJURIKKIPITOISUUKSIIN.....	16
4.1. GRAAFISET TARKASTELUT	16
4.2. VUOROKAUDENAJAN TARKASTELUA DUMMYMUUTTUJIEN AVULLA.....	18
5. ARMA(P,Q) –MALLIN ESITTELY	24
5.1. AR(P) -PROSESSI	24
5.2. AR(P) –PROSESSIN STATIONAARISUUSEHDOT	26
5.3. MA(Q) –PROSESSI.....	27
5.4. MA(Q) –PROSESSIN KÄÄNNETTÄVYYSEHDOT	28
5.5. ARMA(P,Q) –PROSESSI	29
6. ARMA(P,Q) –MALLIEN OMINAISUUDET	30
6.1. AR(P) –MALLIN ACF	30
6.2. AR(P) –PROSESSIN PACF	31
6.3. MA(Q) –PROSESSIN ACF.....	32
6.4. MA(Q) –PROSESSIN PACF	33
6.5. ARMA(P,Q) –PROSESSIN AFC:STÄ JA PAFC:STÄ	34
7. EPÄSTATIONAARISET AIKASARJAT	35
7.1. ARIMA –MALLIT	35
7.2. KAUSIMALLIT	36
8. BOXIN JA JENKINSIN MENETELMÄ.....	37
8.1. IDENTIFIOINNISTA.....	37
8.1.1. Otosautokorrelaatiofunktio.....	38
8.1.2. Portmanteau –testi	38
8.1.3. Käännepestetesti	39
8.1.4. Merkkitesti	39
8.1.5. Järjestystesti	40
8.2. SARMA –MALLIN IDENTIFIOINNISTA	41
8.3. ESTIMOINNISTA.....	42
8.4. MALLIN DIAGNOSTISISTA TARKASTELUISTA.....	42
9. YKSIKÖJUURIEN TESTAUS.....	43
9.1. DICKEY-FULLER –TESTAUS	44
9.2. LAAJENNETTU DICKEY-FULLER –TESTAUS	45
10. HETEROSKEDASTISUUDEN SALLIVAT MALLIT.....	46
10.1. ARCH –PROSESSI	47
10.2. GARCH –PROSESSI	49

10.3. HETEROSKEDASTISUUDEN IDENTIFIOINNISTA	49
11. MONILOTTEISET AIKASARJAT	51
11.1 STATIONAARISUUS	51
11.3. MONILOTTEINEN AIKASARJA JA VALKOINEN KOHINA	54
11.4. 2 -ULOTTEISEN AIKASARJAN RIIPPUMATTOMUUDEN TESTAUS	54
11.4.1. Ristikorrelaatioestimaattien keskivirheet	54
11.4.2. Esivalkaisu	56
11.4.3. Testaus.....	56
12. LÄMPÖTILAN VAIKUTUS TRS:ÄÄN.....	57
12.1. ELOKUU 1992.....	58
12.2.1. Lämpö, elokuu 1992.....	59
12.2.2. TRS elokuu 1992.....	64
12.3. LOPUT KUUKAUDET	68
13. INTERVENTIOANALYYSI.....	69
13.1. INTERVENTIOANALYYSIN TEORIA	70
13.2. INTERVENTIOIDEN TUTKIMINEN TRS –SARJASSA	71
13.2.1. Säterin suunnalta tulleet päästöt	72
13.2.2. UPM:n suunnalta tulleet päästöt.....	76
14. VEKTORIAUTOREGRESSIIVISET MALLIT.....	79
14.1. VAR –MALLIYHTÄLÖT	79
14.2. ESTIMOINNISTA.....	82
14.3. IDENTIFIOINNISTA	83
14.4. IMPULSSIVASTEFUNKTIO	85
14.5. VARIANSSIN DEKOMPOINTI	87
14.6. GRANGER –KAUSAALISUUS.....	90
14.7. VIIVEITTEN LUKUMÄÄRÄ	91
15. SO2:N JA TRS:N YHTEISVAIKUTUKSET	92
15.1. SÄTERIN SUUNNALTA TULEVAT PÄÄSTÖT	92
15.1.1. Viivepituuden määrittäminen	93
15.1.2. Innovaatioiden laskenta.....	98
15.1.3. Granger –kausaalisuus.....	101
15.2. UPM:N SUUNNALTA TULEVAT PÄÄSTÖT.....	101
15.2.1. Viivepituuden määrittäminen.....	102
15.2.2. Innovaatioiden laskenta.....	106
15.2.3. Granger –kausaalisuus.....	109
16. YHTEISINTEGROITUVUUS.....	110
16.1. YHTEISINTEGROITUVUUDEN MÄÄRITELMÄ	110
16.2. VIRHEENKORJAUS	111
16.3. YHTEISINTEGROITUVUUDEN TESTAUS	113
17. TRS:N JA SO2:N YHTEISINTEGROITUVUUS.....	115
17.1. SÄTERIN TUULISEKTORI.....	116
17.2. UPM:N TUULISEKTORI.....	117
18. LOPPUSANAT.....	119
19. LÄHDELUETTELO.....	121
LIITE 1. TIETOKONEKOODIT	123

1. Johdanto

Tiedot Valkeakosken keskustan saastepäästöistä on kerätty pistekohtaisesti TECO-mittarilla(hajurikkiyhdisteet) ja optisesti OPSIS –laitteella(rikkidioksidi). Säähavaintoasema MILOS on kerännyt tiedot sääolosuhteista. Valkeakosken tekee mielenkiintoiseksi tutkimuspaikaksi se, että siellä on suhteellisen lähellä keskustaa päästöjä tuottavia tehtaita, jotka sijaitsevat eri puolilla mittausasemia. Tämän vuoksi päästöjä tutkittaessa on kiinnitettävä erityistä huomiota tuulen suuntaan, eli minkä tehtaan vaikutuspiiristä tulevia päästöjä mitataan. Päästöjen lähteitä ovat Kemira Fibersin tehdas, UPM-Kymmenen tehtaas sekä Säterin kaatopaikka.

Tämä on ensimmäinen kerta, kun kyseessä olevasta aineistosta tehdään laajamittaista tutkimusta. Valkeakoskelta ja sen ympäristöstä on aiemmin kerätty saastepitoisuuksia, mutta menetelmät ja saastekomponentit ovat olleet erilaiset(Herrmann, R. ja Hübner, D. 1984), sekä tutkimuksen jälkeen UPM - Kymmenen tehtailla on tapahtunut joitain merkittäviä ilmansuojeluinvestointeja. Merkittävimpinä väkevien hajukaasujen poltto on aloitettu vuoden 1990 pääsiäisen jälkeen, vuonna 1995 laimeimmista kaasuseoksista poistettiin väkevimmät ja lisättiin polttoon mukaan ja vuonna 2000 väkevien hajukaasujakeiden poltto siirrettiin soodakattilaan asennettuun erilliseen polttimeen.

2. Aineisto

Tutkimuksessa tullaan käsittelemään rikkivetyä(H_2S), joka muodostaa 70-80% pelkistetyistä, haisevista, rikkiyhdisteistä, joihin kuuluvat lisäksi metyylimerkaptani (CH_3-SH), dimetyylisulfidi ($CH_3 - S - CH_3$),

dimetyylidisulfidi ((CH₃)₂ - S₂), rikkihiili (CS₂) ja karbonyylisulfidi (COS). Lisäksi tutkitaan rikkidioksidin(SO₂) ja rikkivedyn välisiä suhteita.

2.1. Rikkivety

Rikkivety on värtön kaasu, joka liukenee moniin nesteisiin, mm. veteen ja alkoholiin. Vaikka suurin osa ilmakehän rikkivedystä on luonnollista alkuperää, aiheuttaa teollistuminen vahvasti päästöjä urbaaneilla alueilla.

Rikkivedyn pääasiallinen reitti kehoon kulkee hengityselimien kautta. Vaikka rikkivedyn tarkkaa imeytymisprosenttia keuhkoissa ei tiedetäkään, on se asiantuntijoiden mielestä luultavasti hyvinkin suuri. Ensimmäisiä merkkejä altistumisesta rikkivedylle on sen epämiellyttävä tuoksu. On kuitenkin huomattava, että pelkät hajuhaitat eivät ole vielä terveydelle vaarallisia. Vaikka tieteellisin menetelmin ei voida määrittää tarkkaa pitoisuutta, jossa haju alkaa tuntua, on tutkittu, että puolen tunnin annostus, joka ylittää 7 µg/m³, on aiheuttanut reaktioita altistuneiden joukossa Taulukossa **2.1.** on WHO:n tarjoamat annostus-vaikutus –suhteet rikkivedylle.

Taulukko 2.1.

H ₂ S –pitoisuus(µg/m ³)	Vaikutus
15-30	Silmien ärsyntyminen
70-140	Vakava silmävamma
210-350	Hajuaistin menetys
450-750	Keuhkojen turvotus, hengenvaara
750-1400	Hengityksen kiihtyminen, hengityshalvaus
1400-2800	Välitön pyörtyminen, hengitysteiden halvaus

Informaatiota pitkäaikaisesta altistumisesta rikkivedylle on niukalti. 81 suomalaista puunjalostustehtaan työntekijää, jotka olivat altistuneet rikkivedylle(<30 µg/m³) verrattiin 81 ei –altistuneeseen. Tuloksista kävi ilmi, että altistuneet koehenkilöt osoittivat alentunutta keskittymiskykyä, toistuvaa

päänsärkyä ja levottomuutta kuin ei-altistuneet koehenkilöt. Tulokset eivät kuitenkaan olleet tilastollisesti merkittävät.(WHO Air Quality Guidelines 2000.)

2.2. Rikkidioksidi

Kuten rikkivety, myös rikkidioksidi on väritön ja helposti veteen liukeneva kaasu. Rikkidioksidia esiintyy luonnossa runsaasti esim. tulivuorten läheisyydessä. Teollistuneilla alueilla kuitenkin suurimmat päästöt ovat ihmisen aiheuttamia. Eritoten kiinteistöjen lämmitys hiilen avulla aiheuttaa runsaita rikkidioksidipitoisuuksia ilmassa.

Samoin kuin rikkivedyn tapauksessa, kulkeutuu rikkidioksidi ihmisen elimistöön lähes ainoastaan hengitysilman kautta. Rikkidioksidin pitkäaikaisia vaikutuksia ei ole tutkittu muuten kuin koe-eläimillä, jolloin on havaittu, että yli $28.6 \mu\text{g}/\text{m}^3$:n pitkäaikainen annostus vahingoittaa ilmasteitten epiteeliä. Ihmisillä samoja oireita havaitaan kroonista keuhkoputkentulehdusta sairastavilla henkilöillä. Lyhytaikaisia vaikutuksia on testattu ihmisillä laboratorioolosuhteissa. Näissä kokeissa on havaittu, että jotkut yksilöt(erityisesti astmaatikot) ovat huomattavasti enemmän resistenssejä rikkidioksidille kuin toiset. Kuitenkin myös heillä keuhkojen toiminta heikkenee 10 minuutin altistumisen aikana noin $11000 \mu\text{g}/\text{m}^3$ suuruisilla päästöillä. Fyysisen stressin on todettu voivan alentaa tätä kynnyksiä. Oireet rikkidioksidille altistuneelle tulevat hyvin nopeasti. Näitä voivat lyhyellä altistumisella(<24 h) olla erilaiset hengitysteitten ongelmat kuten hengenahdistus ja vinkuna hengitettäessä.(WHO Air Quality Guidelines 2000.)

2.3. Aineiston taustaa

Aineiston kerääjänä on toiminut Alpo Päällysaho. Hän on mitannut ilman saastepitoisuuksia Valkeakoskella vuosina 1990-2001. Saastepitoisuuksia on mitattu kahdella eri menetelmällä: pistekohtaisesti TECO- mittarilla Valkeakosken terveyskeskuksen katolta noin kahdeksan metriä katutasosta, sekä OPSIS-laitteella, joka perustuu ns. DOAS- menetelmään(differentiaalinen optinen absorptiospektroskopia).

2.4. TECO –mittaus

Valkeakosken TECO –mittari sijaitsee Valkeakosken terveyskeskuksen katolla noin kahdeksan metrin korkeudella katutasosta. Mittaus on pistekohtainen, eli siinä ei mitata saastepitoisuuksia miltään tietyltä alueelta, vaan vain yhdestä pisteestä, mikä täytyy ottaa huomioon johtopäätöksiä tehtäessä.

TECO-mittarin tulokset perustuvat UV-fluoresenssiin. Näyteilmasta poistetaan rikkidioksidi ja hajurikkiyhdisteet (pelkistetyt rikkiyhdisteet) poltetaan konverterissa, jolloin syntyy rikkioksidia, jonka määrä mitataan. Tämä rikkidioksidi muunnetaan ekvivalentiksi määräksi rikkivetyä, merkitään trs.

2.5. OPSIS –mittaus

DOAS-menetelmä on hieman monimutkaisempi prosessi kuin edellinen TECO –mittaus. OPSIS –laite koostuu kolmesta eri osasta: valonlähteestä, vastaanottimesta ja analyysiyksiköstä.

Valonlähteenä DOAS –systeemeissä toimii 75, 150 tai 300 watin ksenonlamppu. Tästä valonlähteestä keskitetään parabolisen peilin avulla kapea valonsäde, joka matkaa ilman halki muutamasta sadasta metrillä muutamaan kilometriin.

Vastaanottimessa valonsäde kaapataan jälleen peilien avulla ja syötetään valokaapelia pitkin analyysiyksikköön. Analyysiyksikkö pystyy nyt Beer-Lambertin lain mukaan valon eri aallonpituuksia hyväksikäyttäen laskemaan tiettyjen kaasujen määrän valonsäteen reitiltä. Valkeakosken OPSIS –laite mittaa seuraavat yhdisteet: rikkidioksidi(SO₂), rikkihiili(CS₂), karbonyylisulfidi(COS), typpidioksidi(NO₂), typpioksidi(NO), otsooni(O₃), bentseeni(C₆H₆) ja tolueeni(C₇H₈). Laitteella on Valkeakoskella kaksi linjaa, joista Linja 1 (vesitorni-terveyskeskus) on 1387m. ja Linja 2 (terveyskeskus-tallinmäki) on 303m. Molemmilta linjoilta kukin yhdiste mitataan n. 12 kertaa tunnissa puolen minuutin ajan. (Opsis User's Manual.)

2.6. MILOS –säähavaintoasema

Kolmas tutkimuksessa käytettävä mittausasema on MILOS-säähavaintoasema, jonka avulla mitataan tuulen suunta, nopeus, lämpötila ja ilmanpaine. Kaikkien mittauslaitteiden vuorokauden- ja kellonajat synkronoidaan keskenään, joten tulokset, joita saadaan tunnin välein, ovat yhtäpitäviä toistensa kanssa.

3. Aikasarja-analyysin peruskäsitteitä

Seuraavassa kappaleessa käsitellään pääpiirteittäin joitain aikasarja-analyysin peruskäsitteitä lähtien siitä, mitä aikasarjalla oikeastaan tarkoitetaan, ja käydään läpi joitain peruslaskutoimituksia. Tarkoitus on perehdyttää lukija niihin aikasarja-analyysin peruskäsitteisiin, joihin tullaan myöhemmin viittaamaan, kun aihepiiriä laajennetaan.

3.1. Aikasarja

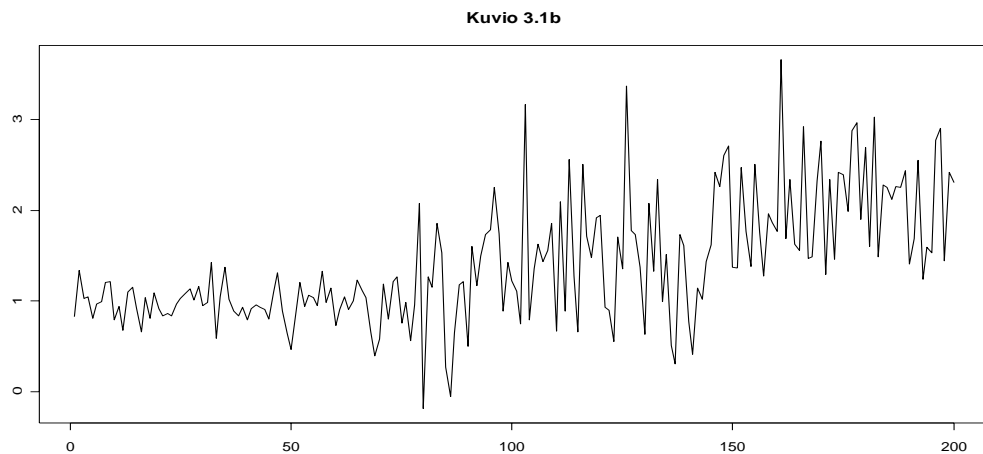
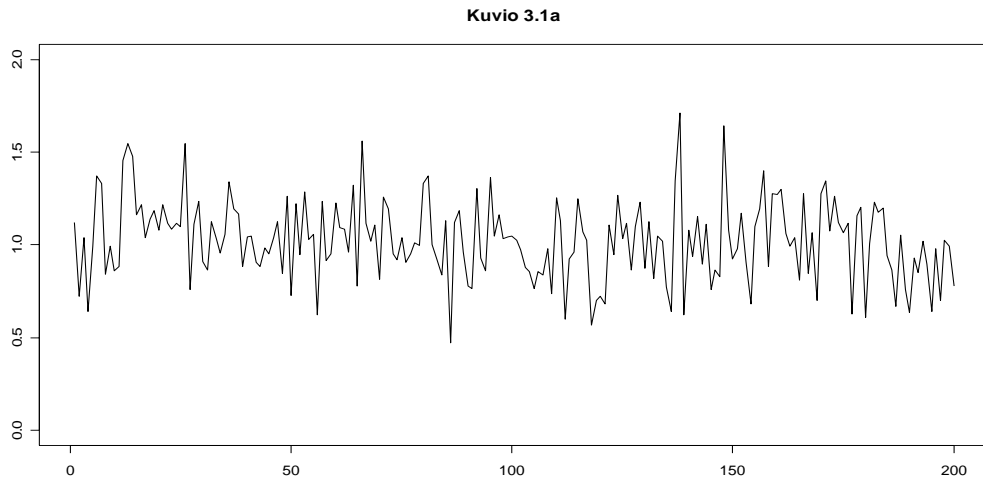
Aikasarja on joukko tiettyjä peräkkäisiä havaintoja y_t , joista jokainen on havaittu jonakin tiettyinä ajanhetkenä t . Aikasarja voi olla joko diskreetti tai jatkuva. Diskreetissä aikasarjassa, joita tässä tutkielmassa käsitellään, ajanhetket t muodostavat diskreetin sarjan, useimmiten tietyillä aikaväleillä. Jatkuvassa aikasarjassa taas havainnot ovat saatu jatkuvasti tietyllä aikavälillä $T_0=[t_1, t_2]$.

Aikasarjamalli havainnoille y_t on tapa ilmaista sarja satunnaismuuttujia Y_t , jonka yksi realisaatio y_t on. Aikasarjan siis ajatellaan syntyneen tietyn satunnaisprosessin Y_t tuloksena. On huomioitava, että aikasarjan jokainen havainto on satunnaismuuttuja, joten jokaisella havainnolla on oma todennäköisyysjakaumansa. Kun havaitulle aikasarjalle on estimoitu tyydyttävä aikasarjamalli, voidaan tätä mallia käyttää joko pelkästään sarjan kompaktiin kuvaamiseen tai tulevien havaintojen ennustamiseen. (Brockwell & Davis s. 1-2, 6.)

Tapahtumia voidaan ennustaa aikasarjan sisäistä riippuvuutta hyväksi käyttäen. Jos esimerkiksi mitataan valtion väkilukua tietyllä aikavälillä, on selvää, että edellisten vuosien väkiluku vaikuttaa lähitulevaisuuteen sijoittuviin ennusteisiin.

3.2. Stationaarisuus

Aikasarjaa, jonka keskimääräinen taso ja varianssi pysyvät samana, sanotaan stationaariseksi. Stationaarisuus on joissain tapauksissa helppo havaita aikasarjan kuvaajasta. Kuviossa **3.1a** ja **3.1b** ovat kuvaajat sekä stationaarisesta että epästationaarista aikasarjasta.



Kuvioista on helppo huomata, kuinka jälkimmäisen aikasarjan varianssi kasvaa, kun $t < 75$, samalla, kun sarjan taso nousee.

3.3. Autokorrelaatio ja osittaisautokorrelaatio

Karkeasti ottaen aikasarjan $\{y_t\}$ sanotaan olevan stationaarinen, jos sen tilastolliset ominaisuudet ovat samat kuin sarjan $\{y_{t+h}\}$, $h \neq 0$. Stationaarisuudessa keskitytään kuitenkin vain $\{y_t\}$:n ensimmäisiin ja toisiin momentteihin. Näistä saadaan seuraavat määritelmät aikasarjan keskiarvo- ja kovarianssifunktioille. (Brockwell & Davis, 14-15.)

Määritelmä 3.1.

Olkoon $\{y_t\}$ aikasarja, jolle $E(y_t)^2 < \infty$

Tällöin $\{y_t\}$:n keskiarvofunktio on

$$\mu_y(t) = E(y_t)$$

ja $\{y_t\}$:n kovarianssifunktio on

$$\gamma_y(r, s) = \text{Cov}(y_r, y_s) = E[(y_r - \mu_y(r))(y_s - \mu_y(s))]$$

Kaikille kokonaisluvuille r ja s .

Stationaarisuuden määritelmä **3.2.** seuraa edellä mainittuja määritelmiä käyttämällä seuraavasti (Kendall & Ord, 51):

Määritelmä 3.2.

$\{y_t\}$ on (heikosti) stationaarinen jos

$$\mu_y(t) \text{ on riippumaton } t\text{:stä}$$

ja

$$\gamma_y(t+h, t) \text{ on riippumaton } t\text{:stä kaikilla } h.$$

Aikasarja on vahvasti stationaarinen, kun havaintojen $y_t \dots y_{t+h}$ yhteisjakauma on riippumaton t :stä kaikilla $n \geq 0$. Kovarianssifunktiolla tarkoitetaan yleensä autokovarianssifunktiota, eli aikasarjan $\{y_t\}$ kovarianssia aikasiirrettyyn aikasarjaan $\{y_{t+h}\}$. Autokovarianssifunktiosta saadaan määritelmän **3.3.** mukaan autokorrelaatiofunktio. (Kendall & Ord, 52.)

Määritelmä 3.3.

Olkoon $\{y_t\}$ stationaarinen aikasarja. Tällöin $\{y_t\}$:n autokovarianssifunktio viiveellä h on

$$\gamma_y(h) = \text{Cov}(y_{t+h}, y_t)$$

Ja $\{Y_t\}$: autokorrelaatiofunktio viiveellä h on

$$\rho_y(h) = \text{Cor}(y_{t+h}, y_t) = \frac{\gamma_y(h)}{\gamma_y(0)}$$

Autokorrelaation lisäksi sarjalle voidaan määrittää myös osittaisautokorrelaatiofunktio. Osittaisautokorrelaatiofunktio eliminoi arvojen y_t ja y_{t-h} välissä olevien havaintojen vaikutuksen. Osittaisautokorrelaatiofunktio määritellään seuraavasti (Enders, 82):

Määritelmä 3.4.

Olkoon $\{y_t\}$ stationaarinen aikasarja. Tällöin $\{y_t\}$:n osittaisautokorrelaatiofunktio viiveellä h on

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\phi_{hh} = \frac{\rho_h - \sum_{i=1}^{h-1} \phi_{h-1,i} \rho_{h-i}}{1 - \sum_{i=1}^{h-1} \phi_{h-1,i} \rho_i} \quad h=3,4,5,\dots$$

$$\text{,missä } \phi_{hi} = \phi_{h-1,i} - \phi_{hh} \phi_{h-1,h-i} \quad i=1,2,3,\dots$$

Otoksesta lasketut estimaatit keskiarvolle, autokovariansille, autokorrelaatiolle sekä osittaisautokorrelaatiolle lasketaan seuraavasti(Wei, 17-18, 21-23):

Määritelmä 3.5a.

Olkoon $\{y_t\}$ stationaarinen aikasarja. Tällöin otoksesta laskettu estimaatti $\{y_t\}$:n keskiarvolle on

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Määritelmä 3.5b.

Olkoon $\{y_t\}$ stationaarinen aikasarja. Tällöin otoksesta laskettu estimaatti $\{y_t\}$:n autokovariansille viiveellä h on

$$\hat{\gamma}_h = \frac{1}{n} \sum_{i=1}^{n-h} (y_i - \bar{y})(y_{i+h} - \bar{y})$$

Määritelmä 3.5c.

Olkoon $\{y_t\}$ stationaarinen aikasarja. Tällöin otoksesta laskettu estimaatti $\{y_t\}$:n autokorrelaatiolle viiveellä h on

$$\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0} = \frac{\sum_{i=1}^{n-h} (y_i - \bar{y})(y_{i+h} - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Määritelmä 3.5d.

Olkoon $\{y_t\}$ stationaarinen aikasarja. Tällöin otoksesta laskettu estimaatti $\{y_t\}$:n osittaisautokorrelaatiolle viiveellä h on

$$\hat{\phi}_{11} = \hat{\rho}_1$$

$$\hat{\phi}_{h+1,h+1} = \frac{\hat{\rho}_{h+1} - \sum_{i=1}^h \hat{\phi}_{hi} \hat{\rho}_{h+1-i}}{1 - \sum_{i=1}^h \hat{\phi}_{hi} \hat{\rho}_i}$$

$$\hat{\phi}_{h+1,j} = \hat{\phi}_{hj} - \hat{\phi}_{h+1,h+1} \hat{\phi}_{h,h+1-i} \quad j=1, \dots, h$$

Auto- ja osittaisautokorrelaatiofunktioita kutsutaan yleisesti nimillä ACF(AutoCorrelation Function) ja PACF(Partial AutoCorrelation Function).

3.4. Valkoinen kohina

Eniten käytetty stationaarinen prosessi on Valkoinen Kohina(White Noise). Siinä $\{y_t\}$ on sarja korreloimattomia satunnaismuuttujia, joiden keskiarvo on nolla ja varianssi σ^2 , jotka kumpikaan eivät riipu ajanhetkestä t , joten on helppo todeta valkoisen kohinan olevan vähintään heikosti stationaarinen. (HUOMAUTUS.

Vahvaa stationaarisuutta tarvitaan aikasarjoja mallinnettaessa hyvin vähän. Tässä tutkielmassa stationaarisuudella tarkoitetaan vastedes heikkoa stationaarisuutta, ellei toisin mainita). Valkoista kohinaa merkitään seuraavasti(Brockwell & Davis s. 15):

$$\{y_t\} \sim \text{WN}(0, \sigma^2)$$

Toinen yleisesti käytössä oleva stationaarinen prosessi on IID(Independent and Identically Distributed) –kohina. Siinä $\{y_t\}$ on joukko riippumattomia ja samoin jakautuneita satunnaislukuja, joilla on keskiarvo nolla. Koska arvot ovat samoin jakautuneita, on niillä äärellinen ajasta riippumaton varianssi. Näiden ominaisuuksien perusteella IID –kohina on vahvasti stationaarinen prosessi, jota merkitään(Brockwell & Davis s. 15-16)

$$\{y_t\} \sim \text{IID}(0, \sigma^2)$$

3.5. Viiveoperaattorit

Eräs eniten käytetty muunnos aikasarjalle on viiveoperaattori B (Backshift operator, joskus myös L (Lag)). Viiveoperaattori ilmaisee havainnon olevan viivästetty, eli jos tarkastellaan havaintoa ajanhetkellä t , on yhdellä viivästetty havainto aikasarjan realisaatio ajanhetkellä $t-1$.

B –operaattorin avulla saadaan toinen erittäin käyttökelpoinen operaattori, jota kutsutaan differenssioperaattoriksi ∇ . Differenssioperaattori tarkoittaa, että tämänhetkisestä havainnosta vähennetään viivästetty havainto. Seuraavassa muodolliset määritelmät edellä esitetyille(Brockwell & Davis, 28 Hamilton, 26).

Määritelmä 3.6.

Olkoon $\{y_t\}$ aikasarja, jolloin viiveellä

1 differoitu aikasarja on

$$\nabla y_t = y_t - y_{t-1} = (1 - B)y_t,$$

missä

$$By_t = y_{t-1}$$

Operaattorien ∇ ja B potenssit määritellään normaalisti, eli $B^j y_t = y_{t-j}$ ja $\nabla^j y_t = \nabla(\nabla^{j-1} y_t)$, missä $j \geq 1$ ja $\nabla^0 y_t = y_t$. Esimerkiksi

$$\nabla^2 y_t =$$

$$\begin{aligned} \nabla(\nabla y_t) &= (1 - B)(1 - B)y_t = (1 - B^2)y_t = (1 - 2B + B^2)y_t = y_t - 2By_t + B^2 y_t \\ &= y_t - 2y_{t-1} + y_{t-2} \end{aligned}$$

Edellä määritettyä differointia käytetään myöhemmin Box-Jenkins -mallien yhteydessä yksikköjuurien ja kausikomponenttien poistoon.

4. Vuorokaudenajan vaikutus hajurikkipitoisuuksiin

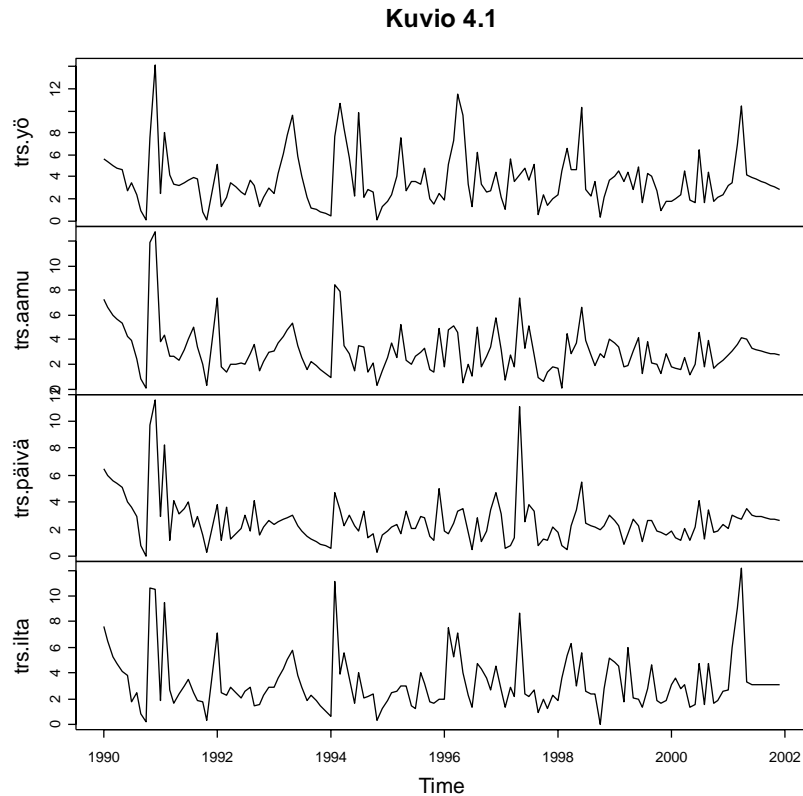
Seuraavassa kappaleessa tarkastellaan eri vuorokaudenaikoja, ja niiden mahdollista vaikutusta ilman haiseviin rikkiyhdisteisiin. Tarkoituksena on tehdä perustellut analyysit hypoteesille, jonka mukaan ilman hajurikkipitoisuus todellakin vaihtelee eri vuorokaudenaikoina.

4.1. Graafiset tarkastelut

Jotta saadaan jonkinlainen yleiskäsitys siitä, kuinka hajurikki- eli TRS - pitoisuudet (Total Redusable Sulfur, tutkimuksissa aikasarja $\{trs\}$) käyttäytyvät suhteessa vuorokaudenaikaan, aloitetaan tutkiminen graafisilla tarkasteluilla.

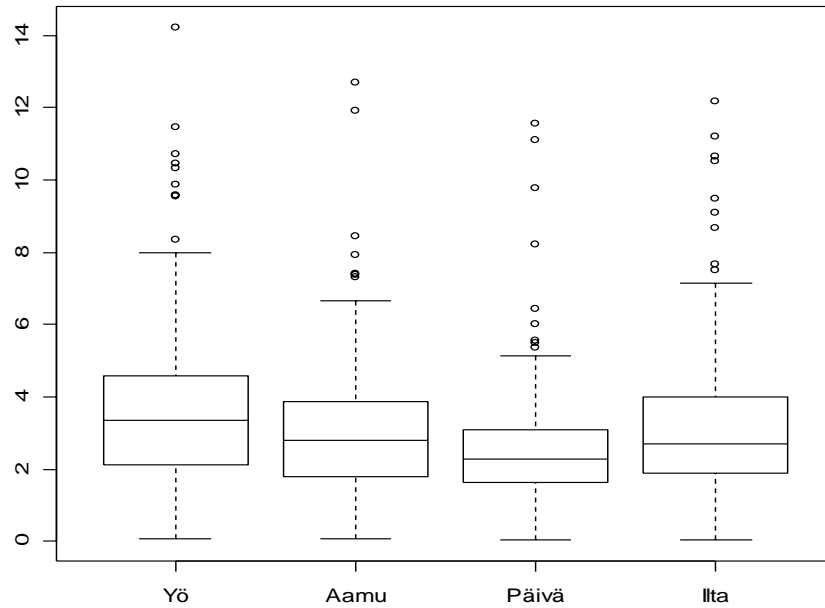
Vuorokaudenaikaa tutkitaan siten, että vuorokausi jaetaan neljään kahdeksan tunnin mittaiseen osaan seuraavassa kuvatulla tavalla. Yö: klo 00:00-07:00, Aamu: 08:00-11:00, Päivä: 12:00-17:00, Ilta: 18:00-23:00. Havainnot ovat tuntiaineistosta laskettuja kuukausikeskiarvoja.

Tarkastellaan ensin vuorokaudenaikoja ja trs-pitoisuuksia. Kuviossa **4.1** ovat aikasarjat kaikilta vuorokaudenajoilta koko tarkastelujakson ajalta.



Toistaiseksi kuviosta on vaikea tehdä valideja päätelmiä, sillä erot ovat selkeästi pienet. Tutkitaan edelleen graafisesti viiksilaatikkokuvion avulla (kuvio **4.2**).

Kuvio 4.2



Viiksilaatikkokuviosta tarkemmin mm. Iversen & Gergen s.82. Kuvio 4.2. on tehty R:n `boxplot()` -funktiolla, jossa 'viiksien' pituus on maksimissaan 1,5 kertaa 'laatikon' pituus. Näin äärimmäiset arvot saadaan paremmin esille.

Nyt saadaan hieman parempi käsitys datan muodosta. Kuvioista näyttäisi, että päästöt olisivat suurimmillaan yöllä sekä illalla ja alhaisimmillaan päivällä. Jokaisena vuorokaudenaikana havaitut äärimmäiset arvot hankaloittavat hieman tulkintaa.

Vaikka nyt ollaan saatu hieman suuntaa siihen, kuinka vuorokaudenaika vaikuttaa `trs` -päästöihin, ei niihin tilastollisesti voi sanoa vielä mitään. Tähän palataan seuraavassa kappaleessa, jossa tarkastellaan pitoisuuksien muuttumista tilastollisin menetelmin.

4.2. Vuorokaudenajan tarkastelua dummymuuttujien avulla

Graafisten tarkastelujen valossa oletettiin siis, että `trs` -pitoisuudet olisivat suurimmillaan yöllä ja alhaisimmillaan päivällä. Seuraavassa tutkitaan asiaa tarkemmin regressioanalyysillä käyttäen apuna dummymuuttujia.

Käytetään hyväksi lineaarista regressiota, jossa selitettävänä muuttujana on ilman trs –pitoisuus. Normaali lineaarisen regression yhtälömuoto on

$$Y = X\beta + \varepsilon$$

, missä Y on vektori, jossa ovat selitettävän muuttujan arvot,
 β on vektori, jossa ovat regressiokertoimet,
 X on matriisi, jossa ovat selitettävien muuttujien arvot
 ja ε on vektori, jonka alkiot ovat satunnaisvirheitä

Regressioanalyysin kertoimet saadaan normaalisti matriisiyhtälöstä

$$\beta = (X^T X)^{-1} * X^T Y \quad (4.1)$$

, missä β on vektori, jossa ovat regressiokertoimet β_0, \dots, β_n , X on havaintomatriisi, jossa on yksikkövektori ja selittävät muuttujat ja Y on vektori, jossa on selitettävät havainnot. Tällöin regressioyhtälö on muotoa: (Wetherill, 8-9.)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_n x_n$$

Hatut parametrien päällä tarkoittavat niiden olevan teoreettisten parametrien estimaatteja, jotka ovat harhattomia (Puntanen, 146-147). Jos halutaan, ettei yhtälössä ole vakioerointa β_0 , jätetään X :stä yksikkövektori pois (Pedhazur, 68-70). Jäännöstermin ε oletetaan olevan siis korreloimaton, nollakeskiarvoinen ja vakiovarianssinen satunnaismuuttuja. Todellisuudessa ε edustaa joukkoa 'latenteja' muuttujia, joita ei regressiomalliin ole otettu mukaan (Tiao(ed), 157). Tässä tapauksessa kuitenkin edellinen huomio jätetään huomiotta, ja kohdellaan ε :tä, kuten määritelmä sanoo.

Kun dummymuuttujia käytetään pelkästään ryhmien välisten erojen selvittämiseen, ei välttämättä tarvita erillisiä selittäviä muuttujia. (4.1):stä huomataan, että jos selitettävää muuttujaa selitetään pelkällä vakioilla (eli vektori $X = \mathbf{1}$), on regressiokertoimen arvo selitettävän keskiarvo. Dummymuuttujia käytettäessä muodostuu matriisi X jälleen yksikkövektorista, mutta nyt selittävinä

muuttujina ovat lisäksi tiettyyn ryhmään kuulumisen siten että muuttuja saa arvon 1, jos se kuuluu ryhmään ja arvon 0, jos näin ei ole. Matriisissa X on siis ryhmien määrä= n saraketta. Kun sarakkeet ovat siten, että $\sum (X_2+\dots+X_n)=1$, saadaan kuitenkin, että matriisi $(X'X)$ on singulaarinen, eikä siten käännettävissä.

Ratkaisuna tähän löydetään, että annetaan X :n yksikkövektori jollekin selittävistä muuttujista, jolloin X :stä häviää yksi sarake ja $(X'X)$ on käännettävissä (Pedhazur, 276-277). Tässä nimenomaisessa ongelmassa dummymuuttujia käytetään saamaan tietoa eri selittävien muuttujakategorioiden (vuorokaudenaika) vaikutusta selitettävään muuttujaan (trs). Luokitellaan siis eri vuorokaudenajoille muuttujat seuraavasti: $Dy=yö$, $Da=aamu$, $Dp=päivä$ ja $Di=ilta$. Muuttujille annetaan arvot seuraavasti: $Dy=1$ aina, $Da=1$, kun mittaus on tapahtunut aamulla, muulloin 0, Dp ja Di samoin.

Tarkoituksena on, että pidetään yöllä mitattuja arvoja eräänlaisena referenssipisteenä, joihin muita arvoja verrataan. Eli kun mittaukset ovat tapahtuneet yöllä, regressioyhtälönä on $\hat{y} = \hat{\beta}_0$, kun taas mittaukset ovat tapahtuneet aamulla, regressioyhtälönä on $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1$, jossa $\hat{\beta}_0$ on yöllä mitattujen arvojen vaikutus ja $\hat{\beta}_1$ aamulla mitattujen arvojen vaikutus. Päivällä regressioyhtälö näyttäisi seuraavalta: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_2$. Aineisto siis näyttää taulukossa

4.1 seuraavalta:

Taulukko 4.1

N	trs	Dy	Da	Dp	Di
1	5.660	1	0	0	0
2	5.321	1	0	0	0
3	4.983	1	0	0	0
...
144	2.883	1	0	0	0
145	7.320	1	1	0	0
146	6.646	1	1	0	0
...
288	2.731	1	1	0	0
289	6.432	1	0	1	0
290	6.001	1	0	1	0
...
432	2.645	1	0	1	0
433	7.656	1	0	0	1
434	6.482	1	0	0	1
...
574	3.056	1	0	0	1
575	3.058	1	0	0	1
576	3.060	1	0	0	1

Näin nähdään, kuinka regressiomalliksi tulee

$$trs = Dy + Da + Dp + Di + \varepsilon$$

Selvästi, kun $n=(1,144)$, ei $\{trs\}$:ään kohdistu muiden vuorokaudeaikojen vaikutuksia kuin yöllä mitatut arvot. Kun taas $n=(145,288)$, $\{trs\}$:ään kohdistuu yöllä mitattujen arvojen lisäksi aamulla mitatut arvot. Seuraavassa taulukossa **4.2** kyseisen regressiomallin tulostukset SAS –ohjelmalla laskettuna (Bowerman s. 502). Taulukossa on myös mukana parametrit $P1$, $P2$ ja $P3$. Nämä kuvaavat päästöjen eroja silloin, kun refrenssiajankohtina käytettyjä yöhavaintoja ei ole verrattavissa (esim. aamun ja päivän ero).

Taulukko 4.2

Parametri	Estimaatti	Keskivirhe	T-arvo	P-arvo
Vakio	3.732	0.176	21.19	<0.0001
Dy	0.000	.	.	.
Da	-0.612	0.249	-2.46	0.0142
Dp	-1.110	0.249	-4.46	<0.0001
Di	-0.444	0.249	-1.78	0.0753
$P1$	-0.498	0.249	-2.00	0.0460
$P2$	0.169	0.249	0.68	0.4990
$P3$	0.667	0.249	2.68	0.0077

Parametrejä luetaan seuraavasti:

Vakio: Yön keskiarvo = $\mu_{yö}$

Da : $\mu_{aamu} - \mu_{yö}$

Dp : $\mu_{päivä} - \mu_{yö}$

Di : $\mu_{ilta} - \mu_{yö}$

$P1$: $\mu_{päivä} - \mu_{aamu}$

$P2$: $\mu_{ilta} - \mu_{aamu}$

$P3$: $\mu_{ilta} - \mu_{päivä}$

Koska $Dy=1$, ei sille ole laskettu estimaattia tulostukseen. Tämän voi jättää huomiotta. SAS:in tulostus ei laske luottamusvälejä parametreille, mutta ne ovat laskettavissa 95:lle prosentille seuraavalla kaavalla(Bowerman, 494):

$$\hat{\beta} \pm t_{0,025}(n - np) * s.e.(\hat{\beta})$$

, missä n =havaintojen kukumäärä

, np =parametrien määrä

ja $s.e.(\hat{\beta})$ = $\hat{\beta}$:n keskivirhe

Parametriestimaatteja voidaan tutkia seuraavasti:

Vakio=3.732=yöllä mitattujen päästöjen keskiarvo. P –arvo $H_0: \beta=0$:lle, on sen verran pieni, että parametri on merkitsevä 99% merkitsevyystasolla. Parametrin 95% luottamusväli on [3.386315, 4.0776850].

$D_a=-0.612$ =aamun ja yön päästöjen erotus. Estimaatti on negatiivinen, joten voidaan päätellä, että aamulla on noin 0.612 yksikköä pienemmät päästöt kuin yöllä. P –arvo on 0.0142, joten H_0 :n voi hylätä 95% riskitasolla, joten parametri on merkitsevä. Parametrin 95% luottamusväli on [-1.101066, -0.1229341].

$D_p=-1.110$ =päivän ja yön päästöjen erotus. Estimaatti on negatiivinen, joten voidaan päätellä, että päivällä on noin 1.110 yksikköä pienemmät päästöt kuin yöllä. P –arvo on <0.0001 , joten H_0 :n voi hylätä 99% riskitasolla, joten parametri on merkitsevä. Parametrin 95% luottamusväli on [-1.599066, -0.6209341].

$D_i=-0.444$ =illan ja yön päästöjen erotus. Estimaatti on negatiivinen, joten voidaan päätellä, että illalla on noin 0.444 yksikköä pienemmät päästöt kuin yöllä. P –arvo on 0.0753, joten H_0 :n voi hylätä 90% riskitasolla, joten parametri on merkitsevä. Parametrin 95% luottamusväli on [-0.933066, 0.0450659].

Jo näistä luvuista saadaan käsitys, että päästöt ovat todella suurimmillaan öiseen aikaan, ja että illalla päästöt ovat toiseksi suurimpia. Itse asiassa on huomattava, että illan 95% luottamusväli kipuaa positiiviselle puolelle, joten

tähän luottamusväliin kuuluu myös se mahdollisuus, että päästöt olisivat illalla kovempia kuin yöllä. Tarkastellaan kuitenkin estimaatteja $P1$, $P2$ ja $P3$, josko ne toisivat lisävalaistusta/varmuutta päättelyihin.

$P1 = -0.498$ = päivän ja aamun päästöjen erotus. Estimaatti on negatiivinen, joten voidaan päätellä, että päivällä on noin 0.498 yksikköä pienemmät päästöt kuin aamulla. P –arvo on 0.0460, joten H_0 :n voi hylätä 95% riskitasolla, joten parametri on merkitsevä. Parametrin 95% luottamusväli on $[-0.987066, -0.0089341]$.

$P2 = 0.169$ = illan ja aamun päästöjen erotus. Estimaatti on positiivinen, joten voidaan päätellä, että illalla on noin 0.169 yksikköä pienemmät päästöt kuin aamulla. P –arvo on kuitenkin 0.4990, joten H_0 jää tässä tapauksessa voimaan.

$P3 = 0.667$ = illan ja päivän päästöjen erotus. Estimaatti on positiivinen, joten voidaan päätellä, että illalla on noin 0.667 yksikköä suuremmat päästöt kuin päivällä. P –arvo on 0.0077, joten H_0 :n voi hylätä 99% riskitasolla. Parametrin 95% luottamusväli on $[0.177934, 1.1560660]$.

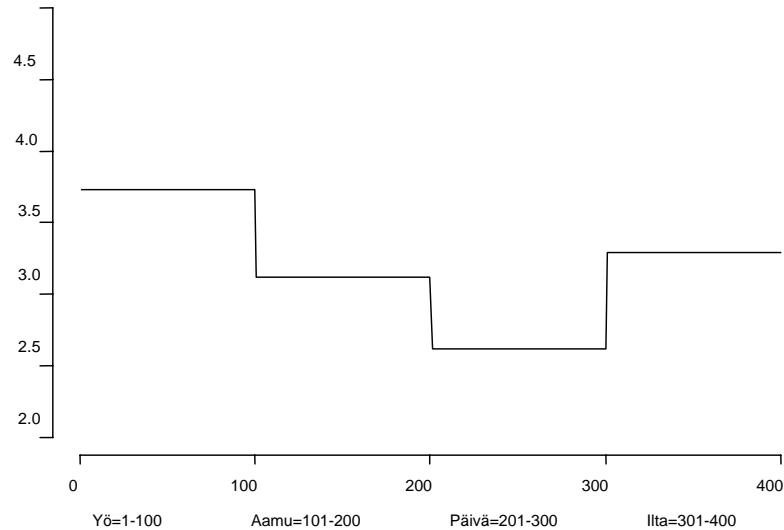
Edellä pääteltiin, että yöllä olisi kaikkein suurimmat päästöt. Jos tarkastellaan neljän ensimmäisen estimaatin suuruuksia, voidaan päätellä seuraavanlainen järjestys suurimmasta pienimpään: Yö-Ilta-Aamu-Päivä. Taulukosta 4.2 saadaan estimaatit eri vuorokaudenaikojen keskiarvoille seuraavasti.

Taulukko 4.2

$\mu_{y\ddot{o}} = 3.732$
$\mu_{aamu} = D_a + \mu_{y\ddot{o}} = -0.612 + 3.732 = 3.12$
$\mu_{p\ddot{a}iv\ddot{a}} = D_p + \mu_{y\ddot{o}} = -1.110 + 3.732 = 2.62$
$\mu_{ilta} = D_i + \mu_{y\ddot{o}} = -0.444 + 3.732 = 3.29$

Kuviossa 4.3 on kuvattu graafinen esitys trs –pitoisuuksien keskiarvoille vuorokaudenaikojen mukaan.

Kuvio 4.3



5. ARMA(p,q) –mallin esittely

Tässä kappaleessa tutustutaan hieman tarkemmin aikasarja-analyysin peruspilareihin, ARMA(p,q) –malleihin, joita usein kutsutaan myös Box-Jenkins -malleiksi. Ensin tarkastellaan pelkkää autoregressiivitä AR(p) –mallia, minkä jälkeen tutustutaan liukuvan keskiarvon MA(q) –malleihin ja lopuksi yhdistetään nämä mallit. Tarkoituksena on käydä läpi mallien muodostuminen ja välttämättömät käännettävyys- ja stationaarisuusehdot mallien toteutumiseksi.

5.1. AR(p) -prosessi

Oletetaan, että meillä on stationaarinen aikasarja $\{y_t\}$. Autoregressiiviset prosessit ja niiden ennustamiskyky perustuu ns. aikasarjan pitkään muistiin, eli prosessia mallinnetaan sen menneillä havainnoilla. Kirjoitetaan $\{y_t\}$ autoregressiivisessä muodossa, missä $\{y_t\}$:tä kuvataan sen menneillä arvoilla ja satunnaisvirheellä:

$$y_t = \pi_1 y_{(t-1)} + \pi_2 y_{(t-2)} + \dots + u_t$$

, missä $u_t \sim WN(0, \sigma^2)$

ja $1 + \sum |\pi_i| < \infty$

Box ja Jenkins sanovat prosessia, jonka voi kirjoittaa edellisessä muodossa, käännettäväksi (invertible). Heidän mielestään prosessi, joka ei ole käännettävä, on merkityksetön ennustamisen kannalta (Box & Jenkins, 51). Kun mallia rajoitetaan siten, että parametreja on äärellinen määrä p , saadaan p :n asteen autoregressiivinen prosessi (Wei, 42.)

Määritelmä 5.1.

$\{Y_t\}$:tä sanotaan p :n asteen autoregressiiviseksi prosessiksi, AR(p), jos

$$y_t = \phi_0 + \phi_1 y_{(t-1)} + \dots + \phi_p y_{(t-p)} + u_t \quad (5.1)$$

, missä $u_t \sim WN(0, \sigma^2)$ ja ϕ_0, \dots, ϕ_p ovat vakioita.

AR(p) – prosessi voidaan esittää myös muodoissa (mukailtu Box & Jenkins, 9):

$$(1 - \phi_1 B - \dots - \phi_p B^p) y_t = \phi_0 + u_t$$

ja

$$\phi(B) y_t = \phi_0 + u_t$$

missä B on kappaleessa 3 määritelty viiveoperaattori ja $(1 - \phi_1 B - \dots - \phi_p B^p) = \phi(B)$ mallin karakteristinen polynomi. Mikäli sarja on keskistetty, on $\phi_0 = 0$, ja (5.1) voidaan kirjoittaa seuraavasti (Box & Jenkins, 9, Hamilton, 53):

$$y_t = \phi_1 y_{(t-1)} + \dots + \phi_p y_{(t-p)} + u_t$$

$$(1 - \phi_1 B - \dots - \phi_p B^p) y_t = u_t \quad (5.3)$$

5.2. AR(p) –prosessin stationaarisuusehdot

Aivan kappaleen alussa totesimme, että prosessin oletetaan olevan stationaarinen. Tarkastellaan seuraavanlaista tapahtumaa. Olkoon (5.3)

$$(1 - \phi_1 B - \dots - \phi_p B^p) y_t = u_t, \text{ josta}$$

$$y_t = \frac{u_t}{(1 - \phi_1 B - \dots - \phi_p B^p)} \quad (5.4)$$

Edellä mainittu on mielekäs vain, jos yhtälöllä $1 - \phi_1 z - \dots - \phi_p z^p = 0$ ei ole ratkaisuja yksikköympyrällä kompleksitasossa. Kun merkitään $1 - \phi_1 B - \dots - \phi_p B^p = \phi(B)$, edellisestä saadaan $\phi(1) \neq 0$.

Yksinkertaisimmassa AR(1) –tapauksessa ehto olisi siis, että $\phi_1 \neq 1$. Tilanteen näin ollessa $y_1 = y_0 + u_1$, $y_2 = y_1 + u_2 = y_0 + u_1 + u_2$... $y_t = y_0 + (u_1 + \dots + u_t)$. Koska $u_t \sim \text{WN}(0, \sigma^2)$, on

$$E(y_t) = E(y_0) + (0 + \dots + 0) = E(y_0) \quad (= \text{vakio}).$$

$$\text{Var}(y_t) = \text{Var}(y_0) + (\sigma^2 + \dots + \sigma^2) = t\sigma^2$$

joka riippuu ajan hetkestä t , joten edellä määritelty stationaarisuusehto ei enää päde.

Edellisen kappaleen mukaan AR(p) –prosessi on käännettävä. Juuri kuitenkin nähtiin, että käännettävät prosessit eivät välttämättä ole stationaarisia. Jotta AR(p) –prosessi olisi stationaarinen, täytyy se Woldin mukaan olla kirjoitettavissa seuraavassa MA(∞) –muodossa:

$$y_t = \frac{1}{\phi(B)} u_t = \theta(B)$$

$$, \text{ missä } \sum_{i=0}^{\infty} \theta_i^2 < \infty$$

Edellisen toteutumisen ehtona on, että mallin karakteristisen yhtälön $\phi(B) = 0$ juurien (kompleksisten tai reaalisten) tulisi löytyä yksikköympyrän ulkopuolelta, jotta AR(p) –prosessi olisi stationaarinen. (Wei, 26, 32-33, Box & Jenkins, 53-54.)

5.3. MA(q) –prosessi

Kun AR(p) –prosessin sanottiin noudattavan aikasarjan ns. pitkää muistia, MA(q) –prosessi puolestaan noudattaa lyhyttä muistia. Kun AR(p) –mallissa uusia havaintoja mallinnetaan menneillä havainnoilla, MA(q) –mallissa niitä mallinnetaan menneillä satunnaisvirheillä seuraavasti:

$$y_t = \psi_0 + u_t + \psi_1 u_{(t-1)} + \psi_2 u_{(t-2)} + \dots$$

$$, \text{ missä } u_t \sim WN(0, \sigma^2)$$

$$\text{ja } \sum \psi_i^2 < \infty$$

Nyt rajoitetaan taas mallia siten, että mallissa on q parametria, jolloin aikasarjan $\{Y_t\}$ sanotaan noudattavan q:nneen asteen liukuvan keskiarvon, MA(q), mallia. Mallia voidaan rajoittaa pelkän q: rajallisen arvon (Hamilton, 50) lisäksi siten että $\psi_i = -\theta_i$ (Wei, 46). Tässä käytetään jälkimmäistä tapaa, jolloin q:nneen asteen liukuvan keskiarvon prosessille saadaan määritelmä (Kendall & Ord, 63):

Määritelmä 5.2

$\{Y_t\}$:tä sanotaan q :n asteen liukuvan keskiarvon prosessiksi, MA(q), jos

$$y_t = \theta_0 - u_t - \theta_1 u_{(t-1)} - \dots - \theta_q u_{(t-q)} \quad (5.5)$$

missä $u_t \sim WN(0, \sigma^2)$ ja $\theta_0 \dots \theta_q$ ovat vakioita

Nyt nähdään heti, että koska y_t on valkoisen kohinan lineaarikombinaatio, on MA(q) –prosessi aina stationaarinen (Wei s. 24). Kuten autoregressiivisessä tapauksessa, jos sarja on keskistetty, on $\theta_0 = 0$, ja (5.5) voidaan kirjoittaa seuraavasti (Box & Jenkins s. 10):

$$y_t = u_t - \theta_1 u_{(t-1)} - \dots - \theta_q u_{(t-q)} \quad (5.6)$$

5.4. MA(q) –prosessin käännettävyysehdot

Edellä mainittiin, että Boxin ja Jenkinsin mielestä sarjan pitäisi olla käännettävä, jotta sillä olisi ennustamisen kannalta hyödyllisiä ominaisuuksia. Oletetaan, että $\theta_0 = 0$, jolloin MA(q) –prosessin voi edelliseen tapaan kirjoittaa myös muodossa

$$y_t = (1 - \theta_1 B - \dots - \theta_q B^q) u_t = \theta(B) u_t$$

Nyt samoin kuin edellä, saadaan ehto, että $\theta(B)$:n juurien pitää löytyä yksikköympyrän ulkopuolelta. Tätä kutsutaan käännettävyysehdoksi (invertibility). (Wei, 47, Box & Jenkins, 67.)

5.5. ARMA(p,q) –prosessi

Usein pelkkä autoregressiivinen tai liukuvan keskiarvon malli ei ole käytännöllinen kuvaamaan stationaarista ja käännettävää prosessia siitä yksinkertaisesta syystä, että estimoitavien parametrien määrä voi kasvaa liikaa. Tämän vuoksi AR(p) –ja MA(q) –mallit voidaan yhdistää autoregressiivisen liukuvan keskiarvon prosessiksi, ARMA(p,q). Nämä prosessit saadaan seuraavasta määritelmästä(Wei s. 56).

Määritelmä 5.3

$\{y_t\}$:tä sanotaan p:nneen asteen autoregressiiviseksi ja q:nneen asteen liukuvan keskiarvon prosessiksi, ARMA(p,q), jos

$$y_t - \phi_1 y_{(t-1)} - \dots - \phi_p y_{(t-p)} = u_t - \theta_1 u_{(t-1)} - \dots - \theta_q u_{(t-q)} \quad (5.7)$$

, missä $u_t \sim WN(0, \sigma^2)$

Prosessi voidaan myös edelliseen tapaan kirjoittaa lyhyemmin seuraavasti

$$\phi(B)y_t = \theta(B)u_t$$

Jotta prosessi olisi sekä käännettävä että stationaarinen, on molempien karakterististen polynomien $\phi(B)=0$ ja $\theta(B)=0$ juuret löydyttävä yksikköympyrän ulkopuolelta(Wei s. 56). Jos juuria ei ole eksplisiittisesti saatavilla, voidaan käyttää seuraavia testejä, kun $B = \pm 1$ (Kendall & Ord 112):

$$|\phi_p| < 1, |\theta_q| < 1, \phi(\pm 1) > 0, \theta(\pm 1) > 0$$

On huomattava, että nämä ehdot ovat välttämättömiä, mutta eivät riittäviä.

6. ARMA(p,q) –mallien ominaisuudet

Tässä kappaleessa käsitellään ARMA(p,q) –prosessien ominaisuuksia, eritoten niiden autokorrelaatio- ja osittaisautokorrelaatiofunktioita. Nämä ominaisuudet auttavat meitä myöhemmin tunnistamaan ja estimoimaan aikasarjalle sopivan mallin.

6.1. AR(p) –mallin ACF

Oletetaan, että meillä on p:nneen asteen autoregressiivinen prosessi

$$y_t = \phi_1 y_{(t-1)} + \dots + \phi_p y_{(t-p)} + u_t \quad (6.1)$$

Nyt kerrotaan (6.1) $y_{(t-h)}$: lla, jotta saadaan

$$y_t y_{(t-h)} = \phi_1 y_{(t-1)} y_{(t-h)} + \dots + \phi_p y_{(t-p)} y_{(t-h)} + u_t y_{(t-h)} \quad (6.2)$$

Kun (6.2):sta otetaan odotusarvot, määritelmän 3.3 mukaan saadaan

$$\gamma_h = \phi_1 \gamma_{(h-1)} + \dots + \phi_p \gamma_{(h-p)} \quad (6.3)$$

On huomattava, että $E(u_t y_{(t-h)})$ häviää, kun $h > 0$. Kun (6.3) jaetaan γ_0 :lla, saadaan määritelmä 3.3:n mukaan

$$\rho_k = \phi_1 \rho_{(k-1)} + \dots + \phi_p \rho_{(k-p)} \quad (6.4)$$

mikä voidaan kirjoittaa seuraavasti

$$\phi(B)\rho_h = 0$$

, missä $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, mikä on sama, mikä saatiin edellä, paitsi että B viittaa nyt h :hon eikä t :hen. Nyt stationaarisuusehtojen mukaan $\phi(B)$:n juurien pitäisi löytyä yksikköympyrän ulkopuolelta. Kun karakteristisen yhtälön

$$(1 - \phi_1 B - \dots - \phi_p B^p) = \phi(B)$$

juuret $1/G_1 \dots 1/G_p$ ovat erilliset, ykeinen ratkaisu on

$$A_1 G_1^h + \dots + A_p G_p^h$$

Ratkaisussa joko a) G_i on reaalinen tai b) juuripari (G_i, G_j) on kompleksinen

Tapauksessa a) $1/G_i^h$ kasvaessa äärettömyyteen, G_i^h :n on lähestyttävä nollaa, joten autokorrelaatio lähestyy nollaa h :n kasvaessa. Tapauksessa b) autokorrelaatiofunktio saa termin $d^h \sin(2\pi fh + F)$, josta seuraa nollaan vaimeneva siniaalto.

Yleisesti siis AR(p) -prosessin autokorrelaatio joko hiipuu eksponentiaalisesti tai sinikäyränä nollaan. (Wei, 44-46, Box & Jenkins, 54-55.)

6.2. AR(p) -prosessin PACF

Edellisessä kappaleessa opittiin tunnistamaan AR(p) -prosessi sen autokorrelaatiofunktion avulla. Osittaisautokorrelaatiofunktio auttaa tunnistamaan AR(p) -prosessin asteen p.

Olkoon ϕ_{kj} j :s parametri k :nnen asteen autoregressiivisessä prosessissa.

Nyt (6.3):sta saadaan

$$\rho_j = \phi_{k1} \rho_{(j-1)} + \dots + \phi_{(k,k-1)} \rho_{(j-k+1)} + \phi_{kk} \rho_{(j-k)} \quad (6.5)$$

$, j=1 \dots k$

, josta saadaan seuraavat Yule-Walker -yhtälöt:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}$$

eli

$$P_k \phi_k = \rho_k$$

Näistä yhtälöistä voidaan ratkaista

$$\begin{aligned} \phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_1 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \end{aligned}$$

Yleisesti ϕ_{kk} :n osoittajassa oleva determinantti on sama kuin nimittäjässä, paitsi että sen viimeinen sarake on korvattu ρ_k lla. Tätä ϕ_{kk} :ta, joka on k:n funktio, sanotaan osittaisautokorrelaatiofunktioksi.

P:nen asteen autoregressiivisessä prosessissa ϕ_{kk} on nolla kaikilla $k > p$, eli AR(p) –prosessin tunnistaa osittaisautokorrelaatiofunktioista katkoksesta viiveellä p.(Box & Jenkins,64-65.)

6.3. MA(q) –prosessin ACF

Oletetaan, että meillä on q:nnen asteen liukuvan keskiarvon prosessi

$$y_t = \theta_0 - u_t - \theta_1 u_{(t-1)} - \dots - \theta_q u_{(t-q)} \quad (6.6)$$

jolle saadaan määriteltyä autokovarianssi seuraavasti

$$\gamma_h = E[(u_t - \theta_1 u_{(t-1)} - \dots - \theta_q u_{(t-q)})(u_{(t-h)} - \theta_1 u_{(t-h-1)} - \dots - \theta_q u_{(t-h-q)})] \quad (6.7)$$

, josta saadaan prosessin varianssi

$$\gamma_0 = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2$$

ja autokovarianssi viiveellä h

$$\gamma_h = \begin{cases} (-\theta_h + \theta_1 \theta_{(h+1)} + \dots + \theta_{(q-h)} \theta_q) \sigma^2 & k = 1, \dots, q \\ 0 & k > q \end{cases} \quad (6.8)$$

Tästä seuraa, että autokorrelaatiofunktio MA(q) –mallille on

$$\rho_h = \begin{cases} \frac{-\theta_h + \theta_1 \theta_{(h+1)} + \dots + \theta_{(q-h)} \theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} & k = 1, \dots, q \\ 0 & k > q \end{cases}$$

Näin havaitaan, että MA(q) –prosessin autokorrelaatiofunktio katkeaa viiveen q jälkeen. (Box & Jenkins, 68.)

6.4. MA(q) –prosessin PACF

Aloitetaan MA(1) –prosessista. Edellä mainitun perusteella siis

$$\rho_1 = \frac{-\theta_1}{1 + \theta_1^2}$$

Nyt sijoitetaan MA(1) –prosessin ρ_i s.e. $\rho_i = 0$, kaikille $i > 1$, edellä mainittuihin Yule-Walker –yhtälöihin ja saadaan

$$\phi_{kk} = \frac{-\theta_1^k (1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}}$$

Näin ollen $|\phi_{kk}| < \theta_1^k$ ja nähdään, että kun k kasvaa, osittaisautokorrelaatio vaimenee eksponentiaalisesti. Kyseessä on siis päinvastainen ilmiö, mitä AR(p) – prosessin auto- ja osittaisautokorrelaatiofunktioissa. Sama ilmiö toteutuu kaikille MA(q) –prosesseille. (Box & Jenkins, 70.)

6.5. ARMA(p,q) –prosessin AFC:stä ja PAFC:stä

Yhdistettyjen AR(p) – ja MA(q) –prosessien autokorrelaatiofunktio voidaan johtaa samalla tavoin kuin edellä AR(p) –prosessin yhteydessä. Kerrotaan ARMA(p,q) – yhtälöön

$$y_t - \phi_1 y_{(t-1)} - \dots - \phi_p y_{(t-p)} = u_t - \theta_1 u_{(t-1)} - \dots - \theta_q u_{(t-q)}$$

termit $y_{(t-h)}$:lla ja otetaan odotusarvot, saadaan

$$\gamma_h = \phi_1 \gamma_{(h-1)} + \dots + \phi_p \gamma_{(h-p)} + \gamma_{yu}(h) - \theta_1 \gamma_{yu}(h-1) - \dots - \theta_q \gamma_{yu}(h-q) \quad (6.9)$$

missä $\gamma_{yu}(h)$ on ristikorrelaatio y :n ja u :n välillä viiveellä h . Ristikorrelaatio on määritelty seuraavasti: $\gamma_{yu}(h) = E(y_{(t-h)}, u(t))$ ja siihen palataan tarkemmin kappaleessa 11. Nyt kuitenkin riittää tieto, että $y_{(t-h)}$ on riippuva vain $t-h$:hon saakka, joten $\gamma_{yu}(h) = 0$ kaikilla $h > 0$.

Edelläolevasta seuraa, että

$$\gamma_h = \phi_1 \gamma_{(h-1)} + \dots + \phi_p \gamma_{(h-p)}$$

ja

$$\rho_h = \phi_1 \rho_{(h-1)} + \dots + \phi_p \rho_{(h-p)} \quad \text{kun } h \geq q+1$$

Nyt, jos $q < p$, autokorrelaatiofunktio hiipuu AR(p) –prosessin tapaan eksponentiaalisesti ja/tai siniaallon tavoin. Jos puolestaan $q \geq p$, löytyy autokorrelaatiofunktioista $q-p+1$ arvoa, jotka eivät hiivu, edellä mainitulla tavalla. (Box & Jenkins, 75.)

ARMA(p,q) –prosessin osittaisautokorrelaatiofunktio puolestaan käyttäytyy liukuvan keskiarvon prosessin autokorrelaation tavoin.

7. Epästationaariset aikasarjat

Kappaleessa 7 laajennetaan kappaleessa 6 esitettyjä stationaarisen aikasarjan prosesseja käsittelemään sarjoja, jotka ovat epästationaarisia. Ensin esitellään integroituneille sarjoille tarkoitetut ARIMA –mallit, minkä jälkeen katsastetaan kausivaihteleville sarjoille tarkoitetut SARIMA –mallit.

7.1. ARIMA –mallit

Edellä käsiteltiin ARMA –mallien ominaisuuksia, joista keskeisimpänä oli niiden stationaarisuus. Kappaleessa 3 differoinnin yhteydessä todettiin, että differointi on käyttökelpoinen tapa puhdistaa aikasarja trendistä tai kausikomponentista. Differointi pätee yleensäkin melko hyvin aikasarjan saattamiseen stationaariseksi. Tästä johtuen eräs ARMA –mallien yleistyksistä ovat ARIMA –mallit. ARIMA –mallit ovat yksinkertaisuudessaan ARMA –malleja, jotka äärellisellä määrällä d differointeja palaavat ARMA –malleiksi

Tästä seuraa suora määritelmä ARIMA –malleille yhdistämällä ARMA –mallit ja d kertaa differoidun aikasarjan y_t seuraavasti:

$$\phi(B)\nabla^d y_t = \theta(B)u_t$$

$$\phi(B)(1-B)^d y_t = \theta(B)u_t$$

, missä $\phi(B)$ ja $\theta(B)$ ovat edellä määritellyn mukaisia polynomeja. (Brockwell & Davis, 178, Hamilton, 437.)

Käytännössä ARIMA –malleja käytetään kuten ARMA –mallejakin, eli kun on saatu aikasarja d kertaa differoimalla stationaariseksi, sovelletaan siihen normaalia ARMA –mallia, ja lisätään integroinnin aste d lopulliseen malliin. Useimmat tietokoneohjelmat laskevat mallin myös suoraan käyttäjän antamien arvojen mukaan, jolloin on oltava tarkkana oikeasta integraation asteesta d .

7.2. Kausimallit

Useat luonnossa esiintyvät ilmiöt vaihtelevat kausittain, vaikka pitkän ajan keskiarvo voi säilyäkin samana. Joka kesä on lämpimämpää kuin talvella, mutta vuoden keskilämpötilat eivät hirveästi muutu vuosien saatossa. Tällaisia tilanteita varten tavallinen ARMA –malli ei ole riittävä työkalu aikasarjan dynamiikan kuvaamiseksi, vaan tarvitaan kausittaista (seasonal) ARMA –mallia, SARMA –mallia.

Kausittaiset aikasarjat on yleensä helppo tunnistaa jo niiden kuvaajasta. Myös autokorrelaatiofunktio on kätevä työkalu, kun tutkitaan, onko sarjassa kausivaihtelua. Korkeat ACF:n arvot ja jaksottaiset huiput ovat lähes aina merkkejä siitä, että aikasarjassa on kausivaihtelua. Jaksottaiset huiput merkitsevät, että eri jaksojen samaa kautta vastaavat arvot ovat voimakkaasti korreloituneita. Esimerkiksi viime vuoden tammikuun kylmä lämpötila (muihin kuukausiin nähden) kertoi, että tämänkin vuoden tammikusta tulisi kylmempi kuin muista kuukausista.

Koska tällaista sarjaa on mahdotonta mallintaa sellaisenaan, turvaudutaan jälleen differointiin. Nyt vain ei differoida d kertaa, vaan D kertaa viiveellä s . Edellä käytyä lämpötilaesimerkkiä differoitaisiin nyt viiveellä 12, jolloin sarjan kausivaihtelu jää vielä näkyviin ACF:ään, mutta aikasarja muuttuu (mahdollisesti) stationaariseksi. SARIMA(p,d,q)*(P,D,Q) s –malli merkitään seuraavasti:

$$\phi(B)\Phi(B^s)x_t = \theta(B)\Theta(B^s)u_t$$

, missä $\phi(B)$, $\theta(B)$ ja u_t ovat kuten normaalissa ARMA –mallissa. $x_t = (1-B)^d(1-B^s)^D y_t$, eli x_t on y_t , kun se on differoitu normaalisti d kertaa ja vielä D kertaa viiveellä s , joka on jakson pituus. $\Phi(B^s)$ ja $\Theta(B^s)$ puolestaan

mallintavat kausivaihtelua samoin kuin $\phi(B)$ ja $\theta(B)$ kauden sisäistä.(Brockwell & Davis, 201-203.) Kappaleessa 8 käsitellään parametrien s, P, D ja Q määrittelyä helpottavia tekniikoita.

8. Boxin ja Jenkinsin menetelmä

Box ja Jenkins(1970) suosittelevat kolmivaiheista menettelyä yksiulotteisen aikasarjan estimoimiseksi ja ennustamiseksi.

1)Mallin identifiointi

2)Mallin estimointi

3)Mallin diagnostiset tarkastelut

Tässä kappaleessa käydään jokainen kolme edellä esitettyä vaihetta läpi. Koska tämän tutkielman tarkoituksena ei ole varsinaisesti tutkia erilaisia ARMA – malleja, ja niiden ominaisuuksia, vaan tutkia ilman hajurikkipitoisuuksia, tämän osion tehtävänä on antaa lukijalle riittävät tiedot, jotka auttavat myöhemmin tulevien analyysien seurannassa.

8.1. Identifioinnista

Mallin identifiointin tarkoituksena on löytää sopivat p ja q(sekä mahdollisesti P ja Q), jotka sopivat (S)ARMA –malliin, joka on aiemmin määritelty. Ensimmäinen tehtävä on kuitenkin piirtää aikasarjan kuvaaja ja tehdä sille sopivat muunnokset. Harvat luonnossa tai taloudessa esiintyvät aikasarjat ovat luonnostaan sellaisia, että niitä voitaisiin alkaa mallintaman saman tien. Aikasarjan pitää olla stationaarinen, jotta sitä voitaisiin mallintaa ARMA –malleilla. Useiten käytetyt muunnokset ovat varianssin vakauttamiseksi tehtävät muunnokset, kuten logaritointi ja neliöjuuren otto, sekä differointi. Kappaleessa 9 käsitellään testejä, joiden avulla tutkitaan yksikköjuuren olemassaoloa sarjassa. Näiden testien ja sopivien differointien ja muunnosten avulla pyritään saamaan sarja stationaariseksi.

Kun aikasarja on saatu stationaariseksi, seuraava askel on tarkistella uutta sarjaa, joita tässä kutsutaan residuaaleiksi. Jos residuaalien välillä ei ole minkäänlaista riippuvuutta, voidaan niitä pitää riippumattomina satunnaismuuttujina, eikä niitä ole syytä mallintaa enää pidemmälle(Brockwell &

Davis, 34). Tarkoitus on poistaa yksikköjuurista aiheutuvat riippuvuudet ja jättää muu riippuvuus tarkasteltavaksi ja sitä kautta mallinnettavaksi. Seuraavassa on muutamia yleisimmin käytettyjä testejä riippumattomuuden testaamiseksi.

8.1.1. Otosautokorrelaatiofunktio

Kun otoskoko on suuri, IID –sarjan otosautokorrelaatiot ovat likimain riippumattomia ja noudattavat $N(0, \frac{1}{n})$ –jakaumaa. Tästä seuraa, että jos residuaalit olisivat jakautuneet, kuten IID –sarja, noin 95% otosautokorrelaatioista pitäisi jäädä rajojen $\pm 1.96/\sqrt{n}$ sisään. Eli jos lasketaan otosautokorrelaatiot aina viiveeseen sata asti, ja löydetään huomattavasti enemmän kuin viisi rajojen ulkopuolelle jäävää korrelaatiota, hypoteesi riippumattomuudesta joudutaan hylkäämään.

8.1.2. Portmanteau –testi

Otosautokorrelaatioita voidaan tutkia myös yhden tunnusluvun avulla, jolloin hypoteesi ja tarkastelun kohteena oleva ilmiö ovat samat kuin edellisessä. Tarkastellaan tunnuslukua

$$Q = n \sum \hat{\rho}^2$$

Nyt, jos residuaalien havainnot ovat IID –jakautuneet, kohdan 4.2. perusteella voidaan sanoa, että Q on jakautunut $N(0,1)$ –muuttujien neliösummana, eli X^2 –jakautunut vapausasteilla h. Näin ollen liian suuret Q:n arvot johtavat H_0 :n hylkäämiseen, tarkemmin H_0 hylätään riskitasolla alpha, jos $Q > X^2_{(1-\alpha, h)}$.

Useimmissa tietokoneohjelmissa käytetään testin hienostuneempia versioita. Ljung ja Box(1978) ovat osoittaneet, että tunnusluku

$$Q_{LB} = n(n+2) \sum_{i=1}^h \hat{\rho}_i^2 / (n-i)$$

noudattaa X^2 –jakaumaa.(Brockwell & Davis, 34-35, Wei, 149-150.)

8.1.3. Käännepestetesti

Oletetaan, että saatavilla on havainnot $y_1 \dots y_n$, joista haluttaisiin tutkia edellä mainittua nollahypoteesia. Määritellään termi käännepestete seuraavasti: Käännepestete hetkellä i , $1 < i < n$, ilmenee, jos $y_{(i-1)} < y_i$ ja $y_i > y_{(i+1)}$ tai $y_{(i-1)} > y_i$ ja $y_i < y_{(i+1)}$, eli havainto y_i on 'piikin huippu'.

IID –sarjassa pisteessä i käännepesteteen todennäköisyys on siis $2/3$. Jos T on käännepesteteiden määrä n –mittaisessa IID –sarjassa, niin T :n odotusarvoksi tulee

$$E(T) = 2(n-2)/3$$

Voidaan myös näyttää, että kyseisessä sarjassa varianssi on

$$Var(T) = (16n-29)/90$$

Suuri $T-E(T)$:n arvo ilmaisee, että sarja vaihtelee odotettua nopeammin, kun taas selvästi nolaa pienempi $T-E(T)$:n arvo ilmaisee, että sarjan vierekkäisillä havainnoilla on positiivista korrelaatiota.

Jos siis havainnot todella ovat IID –sarjasta, voidaan niiden sanoa noudattavan $N(\mu_T, \sigma_T^2)$ –jakaumaa, jota voidaan testata. (Brockwell & Davis, 35.)

8.1.4. Merkkitetesti

Oletetaan jälleen, että käytössä on havainnot $y_1 \dots y_n$. Lasketaan luku S , joka on kaikkien niiden i määrä, kun $y_i > y_{(i-1)}$. Jos havainnot ovat peräisin IID –sarjasta, on selvää, että S :n odotusarvo on

$$E(S) = 1/2(n-1)$$

ja varianssi

$$Var(S) = (n+1)/12$$

Nyt, kuten edellä, S noudattaa suurilla n :n arvoilla $N(\mu_S, \sigma_S^2)$ – jakaumaa, jota voidaan testata. (Brockwell & Davis, 35-36.)

8.1.5. Järjestystesti

Olkoon käytössä jälleen havainnot $y_1 \dots y_n$. Olkoon P parien (i, j) lukumäärä siten että $y_j > y_i$ ja $j > i$. Parien suurin mahdollinen lukumäärä (tällöin jokainen havainto olisi edellistä suurempi) on $\binom{n}{2} = \frac{1}{2} n(n-1)$. Jos sarja on IID, jokaisella tapahtumalla $y_j > y_i$ on todennäköisyys $\frac{1}{2}$. Näin ollen P :n odotusarvo on

$$E(P) = \frac{1}{4} n(n-1)$$

ja varianssi

$$Var(P) = \frac{n(n-1)(2n+5)}{72},$$

, jolloin P approksimoi $N(\mu_P, \sigma_P^2)$ – jakaumaa. (Brockwell & Davis, 36.)

Sääntönä on, ettei johtopäätöksi ikinä pitäisi tehdä vain yhden testin perusteella, vaan tehdä ainakin kolme. Testien kasvaessa on kuitenkin syytä muistaa, että samalla kasvaa todennäköisyys hylätä oikea hypoteesi.

Kun ollaan saatu tyydyttävä sarja, pyritään identifioimaan mallin asteet. Tärkeimmät apuvälineet mallin identifioinnissa ovat (otos)ACF ja (otos)PACF. Asteet p ja q identifioidaan vertaamalla saatua ACF:ää ja PACF:ää teoreettisiin vastaaviin. Kappaleessa 6 käsiteltiin AR – ja MA – prosessien ACF:n ja PACF:n yleistä luonnetta. Taulukossa **8.1** on vielä lyhyesti esiteltyinä eri vaihtoehdot (McLeod, 11-47 – 11-50):

Taulukko 8.1

Prosessi	ACF	PACF
AR(p)	Hiipuu pois eksponentiaalisesti tai siniaallon tavoin	Katkeaa p:n viiveen jälkeen
MA(q)	Katkeaa q:n viiveen jälkeen	Hiipuu pois eksponentiaalisesti tai siniaallon tavoin
ARMA(p,q)	Hiipuu (q-p):n viiveen jälkeen	Hiipuu (p-q):n viiveen jälkeen

ARMA –prosessin astetta tutkittaessa kannattaa muistaa, että p ja q ovat harvoin suurempia kuin kaksi, ja koska malleja voi rakentaa helposti ja vaivatta tietokoneella, ei ole huono käytäntö tutkia samalla useampia malleja, kuin vain sitä, mikä ensin identifiointivaiheessa saadaan. Jos aikasarjaa on jouduttu differoimaan, on syytä muistaa lisätä integroinnin aste malliin (ARMA –mallista tulee ARIMA –malli)

8.2. SARMA –mallin identifioinnista

Identifiointi kausimalleille tapahtuu samanlaisia periaatteita noudattaen kuin normaaleillekin malleille. Nyt on syytä huomata, että tässä vaiheessa on oltava selvillä jakson pituus s. Sarjaa on mahdollisesti jo differoitu d kertaa yhdellä viiveellä ja D kertaa viiveellä s.

Kun tarkastellaan sekä sarjan auto- että osittaisautokorrelaatiofunktioita, on kiinnitettävä ensin huomio kausittaisiin korrelaatioihin, eli jos jakson pituus on 24, tarkastellaan arvoja, jotka ovat 24:n kerrannaisia. Näistä arvoista etsitään kuvioita, jotka sopivat teoreettisiin arvoihin. Teoreettiset arvot kausikerrannaisille ovat samat kuin normaaleillekin viiveille, eli jos aikasarja osoittaa SAR(2) – taipumuksia, vaimenevat autokorrelaatiot $n*s$ ($n=1, \dots, N$) joko eksponentiaalisesti tai siniaallon tavoin ja osittaisautokorrelaatiot $n*s$ katkeavat $2*s:n$ jälkeen. Taulukkoa **8.1** voidaan käyttää referenssinä (mukailtu Bowerman & O’Connell s. 102).

Kun on saatu sopivat P ja Q, tarkastellaan vielä normaalisti autokorrelaatioita kappaleen **8.1** mukaan, jotta saadaan sopivat p ja q.

8.3. Estimoinnista

ARMA –mallien parametrit estimoidaan käytännössä aina tietokoneella. Eri ohjelmat saattavat laskea estimaatit hieman eri tavalla, joten kannattaa aina tarkistaa tapa, jolla estimaatteja ollaan laskemassa.

Kun kyseessä on puhdas autoregressiivinen malli, voidaan käyttää lineaarista pienimmän neliösumman menetelmää, sillä AR –malleja pidetään ehdollisina regressiomalleina(Pandit, 153). Näin saadut estimaatit eivät kuitenkaan ole harhattomia, kuten tavallisessa lineaarisessa regressiossa(Hamilton s. 215). Suurilla otoskoilla ei kuitenkaan ole käytännössä eroa, onko estimaatit saatu ML- vai regressiomenetelmällä(Harvey, 122).

Kun malliin otetaan mukaan liukuvan keskiarvon parametreja, ei regressio olekaan enää lineaarinen, joten tarvitaan epälineaarista pienimmän neliösumman menetelmää. Tästä enemmän esimerkiksi Boxin ja Jenkinsin kirjassa ”Time Series Analysis Forecasting And Control”. Kun käytetään iteratiivista estimointialgoritmia, on syytä kiinnittää huomiota myös parametriestimaattien alkuarvoihin, jotka on saatu teoreettista autokorrelaatiofunktioita tarkastelemalla. Tämä siksi, että jos alkuarvo on mennyt pahasti pieleen, on mahdollista, että algoritmi lähtee menemään ’väärään suuntaan’, tai antaa muuten vain epätarkkoja estimaatteja. Tätä ongelmaa ei kuitenkaan tarvitse pohtia tietokoneohjelmien kanssa.

Tässä tutkimuksessa ARMA –mallien estimointiin käytetty R –ohjelma laskee estimaatit käyttäen tila-avaruus –representaatiota ARMA –prosessista(R –manuaali v. 1.9).

8.4. Mallin diagnostisista tarkasteluista

Kun malli on identifioitu ja siihen liittyvät parametrit estimoitu, on jäljellä mallin tarkastelu. Tässä vaiheessa päätetään, onko malli hyväksyttävä kuvaus aikasarjan luoneesta prosessista vai onko mallissa parannettavaa. ARMA –mallien perusolettamus on, että u_t :t ovat korreloimattomia satunnaisvirheitä, joiden keskiarvo on nolla ja varianssi vakio. Mille tahansa mallille estimoidut

jäännökset \hat{u}_t ovat estimaatteja näistä virheistä. Näin ollen diagnostiset tarkastelut suoritetaan estimoidulle residuaalisarjalle \hat{u}_t . (Wei, 149.)

Kun residuaalisarjan on oltava keskenään korreloimattomia, nollakeskiarvoisia ja vakiovarianssisia satunnaisvirheitä (siis valkoista kohinaa), voidaan käyttää esimerkiksi kappaleiden 8.1.1-8.1.5 testejä. Jos sarja on testien mukaan valkoista kohinaa, voidaan olettaa, että malli on siltä osin kunnossa.

Jokaiselle estimoidulle parametrille saadaan myös keskivirhe. Tätä keskivirhettä voidaan käyttää hyväksi testattaessa, ovatko mallin parametrit tilastollisesti merkittäviä. Muodollisemmin testataan nollahypoteesia $\beta=0$, missä β on estimoitu parametri. Testaukseen liittyvän testisuureen, t -arvon, kaava on seuraava (Bowerman & O'Connell s. 138-139):

$$t(\beta) = \frac{\hat{\beta}}{\sigma_{\hat{\beta}}}$$

, missä $\sigma_{\hat{\beta}}$ on estimoitavan parametrin keskivirhe. Testisuureen arvo noudattaa likimain t -jakaumaa vapausastein $n-np$, missä n on havaintojen määrä ja np estimoitavien parametrien määrä.

9. Yksikköjuurien testaus

ARMA -maillien esittelyn yhteydessä määriteltiin stationaarisuus- ja käännettävyyshdot, joihin liittyy olennaisena osana yksikköjuuri, joita siis ei saisi aikasarjassa olla, ennen kuin sitä aletaan mallintamaan. Autokorrelaatiofunktion tarkastelu voi antaa viitteitä yksikköjuuren olemassaolosta. Esimerkiksi hitaasti vähenevä ACF voi olla merkki, että sarjassa on yksikköjuuri. Toisaalta hitaasti vähenevä ACF voi olla myös merkki trendistä sarjassa. Kummassakaan tapauksissa aikasarja ei ole stationaarinen, jolloin yksikköjuuri on poistettava esim. differoimalla. Yksikköjuuren olemassaolosta ei siis voida tehdä pitäviä päätelmiä pelkän autokorrelaatiofunktion perusteella, vaikka siitä mahdollisesti saataisiinkin viitteitä siihen.

Epästationaarisella aikasarjalla on merkittäviä eroavuuksia stationaariseen aikasarjaan nähden(Enders s.212):

- 1)Epästationaarisella sarjalla ei ole pitkän aikavälin keskiarvoa, mihin sarja palautuisi
- 2)Epästationaarisen sarjan varianssilla ei ole äärellistä raja-arvoa, kun t kasvaa äärettömyyteen.
- 3)Otosautokorrelaatiot eivät vaimene kovinkaan nopeasti.

9.1. Dickey-Fuller –testaus

Kappaleessa 5.5 esitetyn stationaarisuusehdon mukaan aikasarja on stationaarinen, jos siihen liittyvällä autoregressiivisellä polynomilla ei ole juuria yksikköympyrällä(Enders, 76-77). Tämä pätee niin puhtaisiin autoregressiivisiin prosesseihin kuin autoregressiivisiin liukuvan keskiarvon prosesseihin. Dickey ja Fuller(1979, 1981) ovat kehittäneet testin, jonka avulla yksikköjuurta voidaan testata. Tarkastellaan AR(1) –mallia ja sitä, onko mallissa $\phi_1 = 1$ (prosessissa on tällöin yksikköjuuri). Vähennetään AR(1) –yhtälöstä molemmilta puolilta $y_{(t-1)}$, jolloin saadaan

$$y_t - y_{(t-1)} = \phi_1 y_{(t-1)} + u_t - y_{(t-1)}$$

$$\nabla y_t = \gamma y_{(t-1)} + u_t$$

,missä $\gamma = \phi_1 - 1$. Nyt $H_0: \phi_1 = 1$ on $\gamma = 0$. Testi on yksisuuntainen, jolloin $H_1: \gamma < 0$. Nyt on helppo estimoida malli pienimmän neliösumman menetelmällä. Toisin kuin tavallisessa PNS –estimoinnissa, γ :n estimaatti ei ole t –jakautunut, vaan se noudattaa ns. Dickey-Fuller –jakaumaa.

Dickey ja Fuller ovat myös esittäneet seuraavat kaksi muuta regressioyhtälöä, joiden avulla yksikköjuuren olemassaoloa voidaan tutkia:

$$\nabla y_t = a_0 + \gamma y_{(t-1)} + u_t$$

$$\nabla y_t = a_0 + \gamma y_{(t-1)} + a_1 t + u_t$$

Ensimmäinen näistä vaihtoehtoisista yhtälöistä lisää vakion(driftin) malliin ja jälkimmäinen lisää vielä lineaarisen trendin. Vaihtoehtoisten testien avulla suoritettavia hypoteeseja suoritetaan samalla tavalla kuin edellä on mainittu. On kuitenkin huomattava, että t –arvojen kriittiset arvot vaihtelevat sen mukaan, onko mallissa drifttiä ja/tai trendiä mukana(Enders, 221).

9.2. Laajennettu Dickey-Fuller –testaus

Kaikkia aikasarjoja ei voida riittävästi kuvata ensimmäisen asteen autoregressiivisellä prosessilla. Dickey ja Fuller ovat laajentaneet testiään koskemaan p :nnen asteen autoregressiivisiä prosesseja. Kirjoitetaan p :nnen asteen autoregressiivinen prosessi seuraavasti:

$$y_t = \phi_0 + \phi_1 y_{(t-1)} + \dots + \phi_{(p-1)} y_{(t-p+1)} + \phi_p y_{(t-p)} + u_t$$

Nyt lisätään ja vähennetään molemmilta puolilta $\phi_p y_{(t-p+1)}$, jotta saadaan

$$y_t = \phi_0 + \phi_1 y_{(t-1)} + \dots + (\phi_{(p-1)} + \phi_p) y_{(t-p+1)} - \phi_p \nabla y_{(t-p+1)} + u_t$$

Lisätään ja vähennetään seuraavaksi $(\phi_{(p-1)} + \phi_p) y_{(t-p+2)}$ ja saadaan

$$y_t = \phi_0 + \phi_1 y_{(t-1)} + \dots - (\phi_{(p-1)} + \phi_p) \nabla y_{(t-p+2)} - \phi_p \nabla y_{(t-p+1)} + u_t$$

Kun jatketaan samaan tahtiin, saadaan lopulta y :n differoidulle arvolle

$$\nabla y_t = \phi_0 + \gamma y_{(t-1)} + \sum_{i=2}^p \beta_i \nabla y_{(t-i+1)} + u_t$$

$$, \text{ missä } \gamma = -\left(1 - \sum_{i=1}^p \phi_i\right)$$

$$\text{ja } \beta_i = \sum_{j=i}^p \phi_j$$

Testissä tutkitaan jälleen parametria γ samoilla Dickey-Fuller –taulukon arvoilla kuin AR(1) –prosessin tapauksessakin. Kuten edellä, voidaan myös p:n viiveen tapauksessa tutkia sarjaa, jossa on drifti ja/tai trendi. Yhtälöt ovat silloin seuraavat (Enders, 222, 225):

$$\nabla y_t = a_0 + \gamma y_{(t-1)} + \sum_{i=2}^p \beta_i \nabla y_{(t-i+1)} + u_t$$

$$\nabla y_t = a_0 + \gamma y_{(t-1)} + a_2 t + \sum_{i=2}^p \beta_i \nabla y_{(t-i+1)} + u_t$$

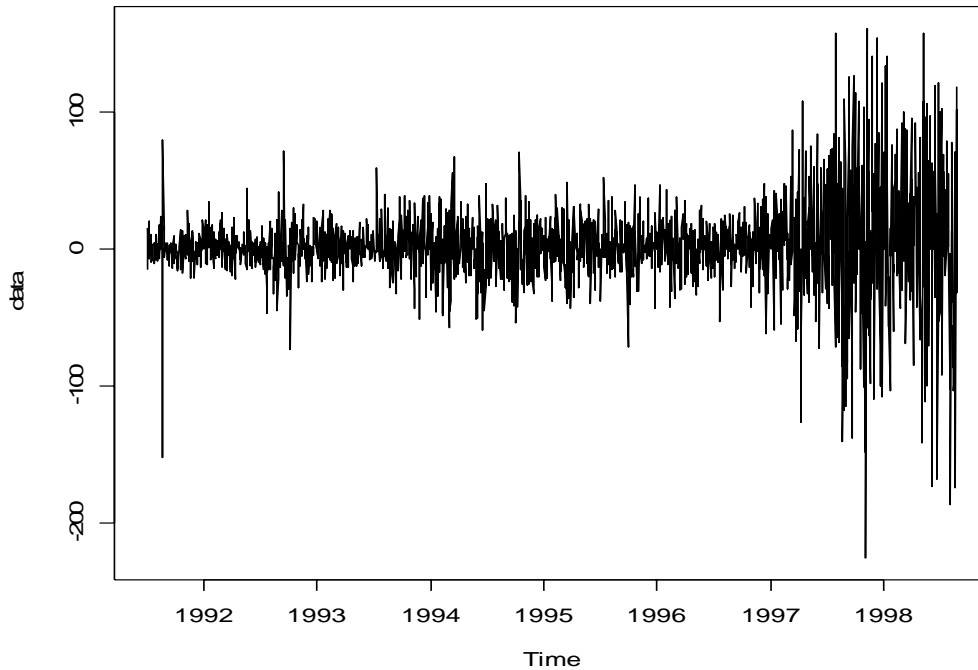
Dickey-Fuller –testeistä on monia laajennuksia. Niillä voidaan tutkia mm. useampien kuin yhden yksikköjuuren olemassaoloa ja kausittaisia yksikköjuuria. Myös usein käytetty Phillips-Perron –testi on yleistys Dickey-Fuller –testistä.

Dickey-Fuller –testin on kuitenkin todettu johtavan usein vääriin johtopäätöksiin varsinkin ekonometristen aikasarjojen kohdalla. Tähän Kwiatkowski, Phillips, Schmidt ja Shin ovat kehittäneet oman testisuurensa sekä trendi- että tasostationaarisuuden testaamiseen. Testissä aikasarja on ilmaistu deterministisen trendin, satunnaiskävelyn ja satunnaisvirheen summana. Testisuure testaa nollihypoteesia, jonka mukaan satunnaiskävelyn varianssi olisi nolla. Testaus suoritetaan tutkimuksessa R –ohjelmalla. Testistä tarkemmin Kwiatkowski et al, 159.

10. Heteroskedastisuuden sallivat mallit

Kappaleen 5 mukaan rakennetuissa Box-Jenkins –malleissa eräs oletus oli, että sarjan varianssin on pysyttävä vakiona. Sarjan tällaista ominaisuutta kutsutaan homoskedastisuudeksi. Kuitenkin varsinkaan ekonometrisissä aikasarjoissa tilanne ei usein ole niin yksinkertainen. Esimerkiksi osakekurssit voivat olla hyvin pitkän aikaa 'rauhallisessa' tilassa, jolloin niiden varianssi on hyvin pientä. Toisaalta jokin tapahtuma voi järkyttää kurssia siten että ajanhetkellä $\{t, t+h\}$ varianssin vaihtelu on hyvinkin suurta. Tällaista muuttujaa sanotaan heteroskedastiseksi. Kuviossa **10.1** on esimerkki tällaisesta sarjasta.

Kuvio 10.1



Tässä kappaleessa tutustutaan heteroskedastisuuden salliviin malleihin, joiden avulla sarjan ehdollista varianssia voidaan mallintaa. Samaten käydään läpi menetelmiä heteroskedastisuuden havaitsemiseen aikasarjassa. On osoitettu (Andersen & Bollerslev, 1), että autoregressiiviset heteroskedastisuuden sallivat mallit tuottavat yllättävän hyviä volatilitteettiestimaatteja siitä huolimatta, että näitä malleja on joskus kritisoitu riittämättöminä käytäntöön.

10.1. ARCH –prosessi

Oletetaan, että käytössä on edellä mainitun kaltainen aikasarja, jossa on pitkiä rauhallisia jaksoja, joita seuraavat jaksot, jolloin vaihtelu on suurta. Tällöin voidaan sanoa, että aikasarjalla on ajankohdasta t riippuva ehdollinen varianssi.

Tästä ehdollisen varianssin vaihtelusta huolimatta aikasarjan pitkän aikavälin varianssi voi olla vakio. Tällainen tilanne syntyy, jos sarjan ehdollinen varianssi noudattaa vaikkapa AR(1) –prosessia. Engle(1982) on osoittanut, että on mahdollista estimoida samanaikaisesti tällaisen sarjan sekä keskiarvo että varianssi. Tarkastellaan yksinkertaista AR(1) –mallia

$$y_t = \phi_0 + \phi_1 y_{(t-1)} + u_t$$

Nyt merkitään y_t :n ehdollista varianssia h_t :llä seuraavasti

$$h_t = \text{Var}(y_t | y_{(t-1)}, y_{(t-2)}, \dots) = E_{t-1}[(y_t - \phi_0 - \phi_1 y_{(t-1)})^2] = E_{t-1}(u_t^2)$$

Box-Jenkins –metodologialla ehdollinen varianssi h_t on vakio(ei-ehdollinen varianssi σ^2). Nyt kuitenkin oletetaan, että näin ei ole, vaan h_t :n oletetaan noudattavan seuraavaan kaavaa:

$$h_t = \alpha_0 + \alpha_1 u_{(t-1)}^2 + \dots + \alpha_s u_{(t-s)}^2$$

Nyt voidaan laskea varianssiyhtälölle parametrien estimaatit AR(s) –mallin avulla:

$$u_t^2 = \alpha_0 + \alpha_1 u_{(t-1)}^2 + \dots + \alpha_s u_{(t-s)}^2 + \eta_t$$

, missä η_t on valkoisen kohinan prosessi. Jos kertoimet $\alpha_1, \dots, \alpha_s$ ovat kaikkia nollia, on ehdollinen varianssi yksinkertaisesti α_0 , mikä merkitsee, että heteroskedastisuutta ei ole.

Tästä saadaan estimaatit \hat{h}_t , joiden avulla lasketaan tasomalli uudelleen GLS –tekniikalla(kts. esim. Freund & Minton, 99-100) käyttäen painoina $\frac{1}{\hat{h}_t}$:stä. Näitten residuaalien neliöistä saadaan uudet ennusteet ehdolliselle varianssille. Menettelyä toistetaan niin kauan, kunnes estimaatit konvergoivat(yleensä kaksikin kierrosta on tarpeeksi).

Jos $s=1$, saadaan $h_t = \alpha_0 + \alpha_1 u_{(t-1)}^2$, jonka ei-ehdollinen varianssi lasketaan seuraavasti:

$$E(h_t) = \alpha_0 + \alpha_1 E(u_{(t-1)}^2) = \alpha_0 + \alpha_1 \sigma^2$$

$$, \text{ josta puolestaan saadaan } \sigma^2 = \frac{\alpha_0}{1 - \alpha_1}$$

Jotta σ^2 olisi positiivinen, vaaditaan, että $\alpha_0 > 0$ ja $0 < \alpha_1 < 1$.

Edellä saatua mallia kutsutaan ARCH(s) –malliksi (Autoregressive Conditional Heteroskedasticity). S määriytyy autoregressiivisten viiveiden määräästä varianssiyhtälössä, jonka avulla ehdollinen varianssi on ennustettavissa seuraavasti: (Enders s. 142, Kanto & Kahra s. 25.)

$$\hat{h}_{(t+1)} = E_t(\hat{u}_{(t+1)}^2) = \alpha_0 + \alpha_1 \hat{u}_t^2 + \alpha_2 \hat{u}_{(t-1)}^2 + \dots + \alpha_{(s+1)} \hat{u}_{(t-s+1)}^2$$

10.2. Garch –prosessi

Bollerslev(1986) on laajentanut Englen ARCH –mallia siten, että nyt ehdollinen varianssi voi noudattaa pelkän AR –prosessin sijaan ARMA –prosessia. Nyt virheprosessin voi määrittellä seuraavasti (Enders, 146-147):

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i u_{(t-i)}^2 + \sum_{i=1}^p \beta_i h_{(t-i)} \quad (10.1)$$

Edellä esitettyssä yleistetyssä (Generalized) ARCH(p,q), eli GARCH(p,q) –mallissa on siis sekä autoregressiivisiä että liukuvan keskiarvon parametrejä. GARCH –mallin etuna ARCH –malliin on se, että sitä käytettäessä päästään usein entistä vähäparametrisempaan malliin, mihin päästäisiin pelkästään ARCH –mallia käyttämällä. Huomattavaa on, että jos asetetaan $p=0$, saadaan GARCH(0,q) – eli ARCH(q) -malli.

10.3. Heteroskedastisuuden identifioinnista

Eräs GARCH –mallin tärkeimpiä ominaisuuksia on, että kun virhetermien ajatellaan noudattavan ARMA –prosessia, voidaan odottaa, että myös estimoidun mallin residuaaleilla on sama ARMA –ominaisuus. Toisin sanoen mallin residuaalien pitäisi käyttäytyä valkoisen kohinan tavoin, mikäli malli on riittävän hyvä. Koska ehdollinen varianssi on $E_{t-1}u_t$, voidaan yhtälö (10.1) kirjoittaa muodossa

$$E_{t-1}u_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i u_{(t-i)}^2 + \sum_{i=1}^p \beta_i h_{(t-i)}$$

, joka muistuttaa u_t^2 :n ARMA –prosessia. Siis jos sarjassa on ehdollista heteroskedastisuutta, saadaan se seuraavasti selville:

1) Estimoidaan sarja y_t ARMA- tai regressiomallilla. Talletetaan residuaalien neliöt u_t^2 ja lasketaan residuaalien varianssi

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^T \hat{u}_i^2}{T}$$

, jossa T on residuaalien lukumäärä.

2) Lasketaan neliöityjen residuaalien otosautokorrelaatiot

$$\rho_h = \frac{\sum_{i=h+1}^T (\hat{u}_i^2 - \hat{\sigma}^2)(\hat{u}_{i-h}^2 - \hat{\sigma}^2)}{\sum_{i=1}^T (\hat{u}_i^2 - \hat{\sigma}^2)^2}$$

3) Kun otoskoko on riittävän suuri, ρ_h :n keskihajonta on noin $T^{-\frac{1}{2}}$. Yksittäiset ρ_h :n arvot, jotka ovat tilastollisesti nolasta poikkeavia, ovat merkinä GARCH –virheistä. Ljung-Box –testiä voidaan myös soveltaa seuraavasti:

$$Q = T(T+2) \frac{\sum_{i=1}^n \rho_i}{T-i}$$

Q noudattaa χ^2 -jakaumaa n :llä vapausasteella, kun virheiden neliöt ovat korreloimattomia, eli jos hylätään H_0 , jonka mukaan \hat{u}_t^2 ovat korreloimattomia, hylätään samalla homoskedastisuusoletus. (Enders, 147-148.)

Engle(1982) on kehittänyt ARCH –ominaisuuden testaamiseen ns. Lagrangen kerrointestin, jossa on kaksi vaihetta.

1) Estimoidaan y_t :lle PNS –menetelmällä mahdollisimman hyvä AR(n) – tai regressiomalli

$$y_t = \phi_0 + \phi_1 y_{(t-1)} + \dots + \phi_n y_{(t-n)} + u_t$$

2) Muodostetaan virheistä sarja $\{\hat{u}_t^2\}$ ja regressioyhtälö

$$\hat{u}_t^2 = \alpha_0 + \alpha_1 \hat{u}_{(t-1)}^2 + \alpha_2 \hat{u}_{(t-2)}^2 + \dots + \alpha_q \hat{u}_{(t-q)}^2$$

Jos sarjassa ei ole ARCH- tai GARCH –ominaisuutta, kaikkien estimaattien $\alpha_1, \dots, \alpha_q$ pitäisi olla nollia, eli regression selitysteorian R^2 :n pitäisi olla alhainen. Tällä residuaalilla H_0 :aa (ei ARCH –virheitä) voidaan testata tunnusluvulla TR^2 , joka konvergoi χ^2 -jakaumaan q :lla vapausasteella (Enders, 148-149).

11. Moniulotteiset aikasarjat

Kappaleessa 11 käydään läpi N -ulotteisen aikasarjan peruskäsitteitä, kuten stationaarisuus ja valkoinen kohina. Kappaleen pääasiallinen tehtävä on kuitenkin tutustuttaa lukija jo edellä mainittuun ristikorrelaation käsitteeseen ja kahden aikasarjan riippumattomuuden testaamiseen. Tätä tekniikkaa käytetään yhdessä edellä käsiteltyjen Box-Jenkins –mallien ja heteroskedastisuuden sallivien mallien kanssa seuraavassa kappaleessa, jossa tutkitaan, ovatko Valkeakosken keskustan hajurikkipäästöt ja lämpötila riippuvia keskenään.

11.1 Stationaarisuus

Käytännössä monia aikasarjoja on paras mallintaa ei vain pelkästään sarjan sisäisten havaintojen riippuvuudella, vaan myös riippuvuudella kokonaan toisesta aikasarjasta. Näin aikasarjaa voidaan käsitellä n –ulotteisena vektorina, jonka kaikki n elementtiä ovat aikasarjoja.

Stationaarisuus moniulotteisille aikasarjoille määritellään samoin, kuin yksiulotteisillekin. Ensin on määriteltävä n –ulotteinen aikasarja X_t seuraavasti: $X_t = (y_{t1} \dots y_{tn})'$, jossa y_{ti} , $i=1 \dots n$, ovat aikasarjoja. Määritellään X_t :n keskiarvovektori ja kovarianssimatriisi seuraavasti (Brockwell & Davis, 218, 223):

Määritelmä 11.1.

Olkkoon X_t aikasarjavektori s.e. $X_t = (y_{t1} \dots y_{tn})'$, niin sen keskiarvovektori on

$$\mu_t = EX_t = \begin{bmatrix} \mu_{t1} \\ \vdots \\ \mu_{tn} \end{bmatrix}$$

ja kovarianssimatriisi viiveellä h on

$$\Gamma_{(t+h,t)} = \begin{bmatrix} \gamma_{11}(t+h,t) & \dots & \gamma_{1n}(t+h,t) \\ \vdots & \ddots & \vdots \\ \gamma_{n1}(t+h,t) & \dots & \gamma_{nn}(t+h,t) \end{bmatrix}$$

$\Gamma_{(h+t,t)} = \Gamma_h$:n diagonaali-alkiot ovat yksiulotteisten aikasarjojen $y_1 \dots y_n$ autokovariansseja viiveellä h . Diagonaalin ala- ja yläpuolella olevat alkiot ovat kovariansseja $y_{(t+h,i)}$:n ja $y_{t,j}$:n, $i \neq j$, välillä. On huomattava, että $\gamma_{ij}(h) = \gamma_{ji}(-h)$. Seuraavassa yksiulotteisesta aikasarjasta saatu määritelmä moniulotteisen aikasarjan stationaarisuudelle.

Määritelmä 11.2.

X_t on (heikosti)stationaarinen, jos

μ_t on riippumaton t :stä ja

$\Gamma_{(t+h,t)}$ on riippumaton t :stä kaikilla h

Estimaatit μ_t :lle ja Γ_h :lle saadaan seuraavasti (Brockwell & Davis, 227, 229):

$$\bar{X}_t = \frac{1}{n} \sum_{t=1}^n X_t$$

, missä n on X_t :n aikasarjan pituus

$$\hat{\Gamma}_h = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X})(X_t - \bar{X})' & , 0 \leq h \leq n-1 \\ \hat{\Gamma}_{-h} & , -n+1 \leq h \leq 0 \end{cases}$$

Analogisesti yksiulotteisen tapauksen kanssa sarjojen välinen korrelaatio viiveellä h , $\rho_{ij}(h)$, määritellään seuraavasti (Brockwell & Davis s. 229):

$$\rho_{ij}(h) = \hat{\gamma}_{ij}(h) (\hat{\gamma}_{ii}(0) \hat{\gamma}_{jj}(0))^{-\frac{1}{2}}$$

Edellä määritellyn mukaan ristikorrelaation käsite siis ilmaisee korrelaation sarjan i t :nnen havainnon ja sarjan j $t+h$:nnen havainnon välillä. Näin voidaan myöhemmin testata kahden aikasarjan välistä riippuvuutta.

11.3. Moniulotteinen aikasarja ja valkoinen kohina

Kuten yksiulotteisessa, on moniulotteisessakin tapauksessa yksinkertaisin aikasarjan muoto valkoinen kohina, joka määritellään seuraavasti (Brockwell & Davis, 225):

Määritelmä 11.3.

n -ulotteisen aikasarjavektorin X_t sanotaan noudattavan valkoisen kohinanprosessia keskiarvovektorilla $\mathbf{0}$ ja kovarianssimatriisilla Σ , merk.

$$X_t \sim WN(\mathbf{0}, \Sigma)$$

jos ja vain jos X_t on stationaarinen, sillä on nollavektori keskiarvona ja kovarianssimatriisin funktiona

$$\begin{cases} \Gamma_h = \Sigma & , \text{jos } h = 0 \\ 0 & , \text{muuten} \end{cases}$$

11.4. 2 -ulotteisen aikasarjan riippumattomuuden testaus

Kahden sarjan välinen korrelaatio, ristikorrelaatio, on käytännöllinen apuväline kahden sarjan riippuvuuden testaamiseen. Ristikorrelaatiota ei kuitenkaan voida tulkita sellaisenaan, vaan sarjoille on tehtävä erinäisiä esivalmisteluita.

11.4.1. Ristikorrelaatioestimaattien keskivirheet

Käypä tapa testata, ovatko ristikorrelaatiofunktion arvot nollia, on verrata niitä niiden keskivirheisiin. Bartlett(1978) on johtanut seuraavat kaavat

ristikorrelaatiofunktioestimaattien $\hat{\rho}_{12}(h)$ ja $\hat{\rho}_{12}(h+k)$ kovarianssille (Box & Jenkins, 376, Brockwell & Davis, 232):

$$\begin{aligned} \text{Cov}(\hat{\rho}_{12}(h), \hat{\rho}_{12}(h+k)) &= \frac{1}{n-k} \sum_{i=-\infty}^{\infty} [\rho_{11}(i)\rho_{22}(i+k) + \rho_{12}(-i)\rho_{12}(i+2h+k) \\ &\quad + \rho_{12}(h)\rho_{12}(h+k) \{ \rho_{12}^2(i) + \frac{1}{2}\rho_{11}^2(i) + \frac{1}{2}\rho_{22}^2(i) \} \\ &\quad - \rho_{12}(h) \{ \rho_{11}(i)\rho_{12}(i+h+k) + \rho_{12}(-i)\rho_{22}(i+h+k) \} \\ &\quad - \rho_{12}(h+k) \{ \rho_{11}(i)\rho_{12}(i+h) + \rho_{12}(-i)\rho_{22}(i+h) \}] \quad (11.1) \end{aligned}$$

Erityisesti, jos $k=0$, saadaan varianssiksi

$$\begin{aligned} \text{Var}(\hat{\rho}_{12}(h)) &= \frac{1}{n-h} \sum_{i=-\infty}^{\infty} [\rho_{11}(i)\rho_{22}(i) + \rho_{12}(h+i)\rho_{12}(h-i) \\ &\quad + \rho_{12}^2(h) \{ \rho_{11}(i)\rho_{12}(i+h) + \frac{1}{2}\rho_{11}^2(i) + \frac{1}{2}\rho_{22}^2(i) \} \\ &\quad - 2\rho_{12}(h) \{ \rho_{11}(i)\rho_{12}(i+h) + \rho_{12}(-i)\rho_{22}(i+h) \}] \end{aligned}$$

Usein käy niin, että kaksi prosessia ovat korreloituneet vain tietyllä välillä. Oletetaan, että $\rho_{12}(i) \neq 0$ vain tietyllä välillä $Q_1 \leq i \leq Q_2$. Tällöin jos ei h , $h+k$ eikä $h+0.5k$ ole tällä välillä, kaavasta (11.1) tippuvat loppuosan termit pois, ja saadaan

$$\text{Cov}(\rho_{12}(h), \rho_{12}(h+k)) = \frac{1}{n-h} \sum_{i=-\infty}^{\infty} \rho_{11}(i)\rho_{22}(i+k) \quad (11.2)$$

Jos taas h ei ole tällä välillä, saadaan varianssin kaavasta yksinkertaisempi muoto

$$\text{Var}(\rho_{12}(h)) = \frac{1}{n-h} \sum_{i=-\infty}^{\infty} \rho_{11}(i)\rho_{22}(i) \quad (11.3)$$

Nollahypoteesin, jonka mukaan sarjat eivät ole korreloituneet, vallitessa kaavat (11.2) ja (11.3) pätevät kaikille viiveille h . (Box & Jenkins, 377.).

11.4.2. Esivalkaisu

Valaistaan esivalkaisun käsitettä Brockwellin esimerkillä. Oletetaan, että y_1 ja y_2 ovat kaksi riippumatonta AR(1) –prosessia, joilla on $\rho_{11}(h) = \rho_{22}(h) = 0.8^{|h|}$. Tällöin edellisen kappaleen mukaan suurilla n saadaan $\rho_{12}(h)$:n varianssiksi päättymättömän geometrisen sarjan mukaan $\frac{1}{n} \left(1 + 2 \sum_{i=1}^{\infty} 0.64^i \right) = 4.56 \frac{1}{n}$. Tällöin ei olisi yllättävää saada niinkin suuria arvoja kuin $3 \frac{1}{\sqrt{n}} \rho_{12}(h)$:lle. Näin ollen siis normaalia 95%:n approksimointia ei voida käyttää. (Brockwell & Davis, 230.)

Sen sijaan, että laskettaisiin joka sarjalle varianssit erikseen, käytetään ns. esivalkaisua. Ideana on se, että varianssin laskeminen yksinkertaistuu, jos sarjat ovat valkoista kohinaa. Jos y_1 ja y_2 ovat käännettäviä ARMA(p,q) – prosesseja, voidaan esivalkaisu suorittaa seuraavien filttereiden avulla. (Box & Jenkins, 380)

$$z_{ii} = \frac{\phi(B_i)}{\theta(B_i)} y_{ii}$$

,missä $\phi(B)$ ja $\theta(B)$ ovat prosessin autoregressiiviset ja liukuvan keskiarvon karakteristiset polynomit ja $i=1,2$.

Koska käytännössä todellinen malli on lähes aina tuntematon, koska dataa ajalta $t \leq 0$ ei ole saatavilla, käytetään z_{ii} :n estimaatteina residuaaleja malleista, joita on sovitettu y_1 :een ja y_2 :een (Brockwell & Davis, 230). Jos sijoitetut mallit olisivat todellisia malleja, residuaalisarjahan olisi silloin valkoista kohinaa.

11.4.3. Testaus

Testattaessa H_0 :aa, jonka mukaan y_1 ja y_2 ovat riippumattomia, tarkkaillaan hypoteesia, jonka mukaan myös esivalkaistut sarjat z_1 ja z_2 ovat riippumattomia. Tämä on nyt helposti testattavissa, sillä teoreettiset autokorrelaatiot $\rho_{12}(h)$ ja $\rho_{12}(k)$, $h \neq k$, ovat riippumattomia ja normaalisti jakautuneita keskiarvolla 0 ja

varianssilla $\frac{1}{n}$ suurilla n . Näin ollen testaus voidaan siis suorittaa vertailemalla havaittuja $\rho_{12}(h)$:n arvoja $\pm 1.96/n^{0.5}$:een.(Brockwell & Davis, 231)

Edellä esivalkaistiin molemmat sarjat y_1 ja y_2 . Näin on tapana tehdä, jos epäillään, että sarjassa saattaisi olla palautetta(y_1 vaikuttaa y_2 :een ja päinvastoin)(Wei, 299).

12. Lämpötilan vaikutus trs:ään

Seuraavaksi tarkastellaan, kuinka ilman lämpötila vaikuttaa keskustan hajurikkipitoisuuksiin. Aineistot, joita tarkastellaan, ovat kuukausiaikasarjoja ja niiden mittausintervalli on yksi tunti. Ensin tutkitaan aikasarjojen ristikorrelaatioita, jotka edellisen kappaleen mukaisesti esivalkaistaan ja tutkitaan riippuvuutta. Jos riippuvuutta on havaittavisa, voidaan aikasarjoille rakentaa esim. siirtofunktio- tai vektoriautoregressiivinen malli. Esivalkaisun teoria on käsitelty edellä, joten tässä kappaleessa käydään läpi vain tulokset.

Koko aineistossa on 144 kappaletta näitä kuukausiaineistoja, joten ei ole järkevää alkaa heti mallintamaan kaikkia. Sen sijaan ollaan arvottu kahdeksan kuukautta ositetulla satunnaisotannalla, jossa jokaisessa ositteessa olevalla alkiolla on yhtä suuri todennäköisyys tulla valituksi otokseen(tarkemmin kts. esim. Hansen et al 180,181). Aineisto on ositettu neljään osaan siten että talvikuukausia ovat kuukaudet joulukuusta helmikuuhun, kevätkuukausia maaliskuusta toukokuuhun, kesäkuukausia kesäkuusta elokuuhun ja syyskuukausia syyskuusta marraskuuhun. Jokaisesta ositteesta on siis valittu kaksi kuukautta. Näin tarkasteltaviksi aineistoiksi tulivat seuraavat kuukaudet:

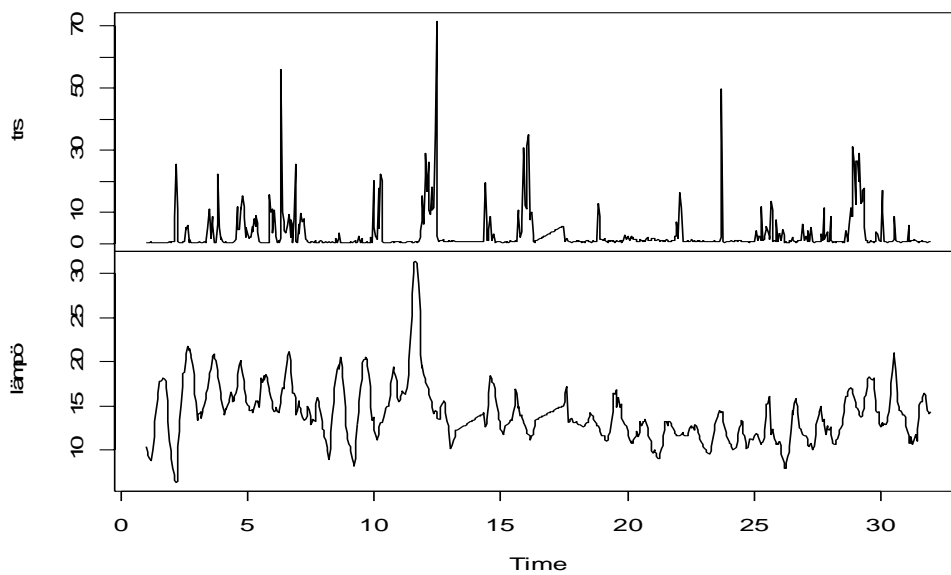
- 1)Elokuu 1992
- 2)Elokuu 1993
- 3)Lokakuu 1994
- 4)Maaliskuu 1995
- 5)Syyskuu 1997
- 6)Joulukuu 1998
- 7)Tammikuu 2000
- 8)Toukokuu 2000

Jos näistä aineistoista saadaan mallintamisen arvoisia tuloksia, voidaan ottaa lisää kuukausia tarkasteluun.

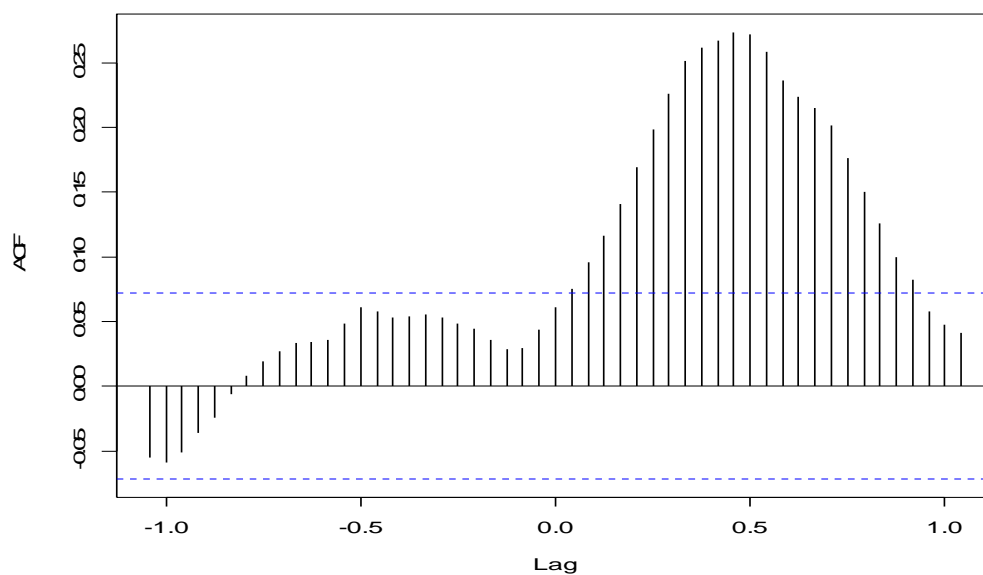
12.1. Elokuu 1992

Tarkastellaan lämpötilan ja trs:n suhdetta vuoden 1992 elokuussa. Kuviossa **12.1.** näkyy molemmat aikasarjat plotattuina, kun taas kuviossa **12.2** on sarjojen välinen ristikorrelaatio.

Kuvio 12.1



Kuvio 12.2

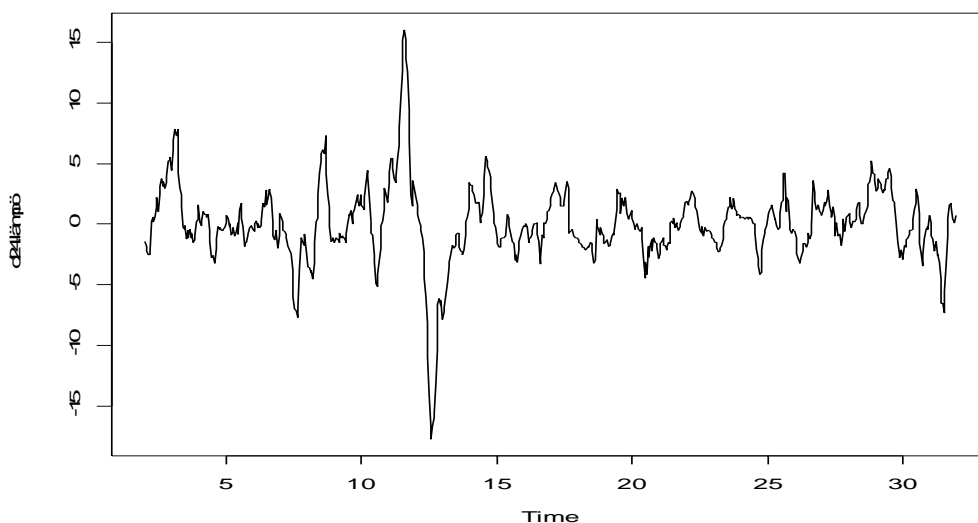


Kuviosta **12.1** nähdään selvästi sekä lämpötilan kausittainen vaihtelu että hajurikkipitoisuuksien vaihteleva varianssi. Kuvio **12.2** näyttää suuria ristikorrelaation arvoja kahden sarjan välillä. Edellä kuitenkin todettiin, että olisi virheellistä mennä harkitsemattomasti tulkitsemaan näitä aikasarjoja. Kuviosta näyttäisi siltä, että sarjoilla olisi voimakaskin keskinäinen riippuvuus, mutta koska lämpötila on vahvasti autokorreloitunut, on mahdollista, että tuo havaittu riippuvuus johtuu juuri siitä.

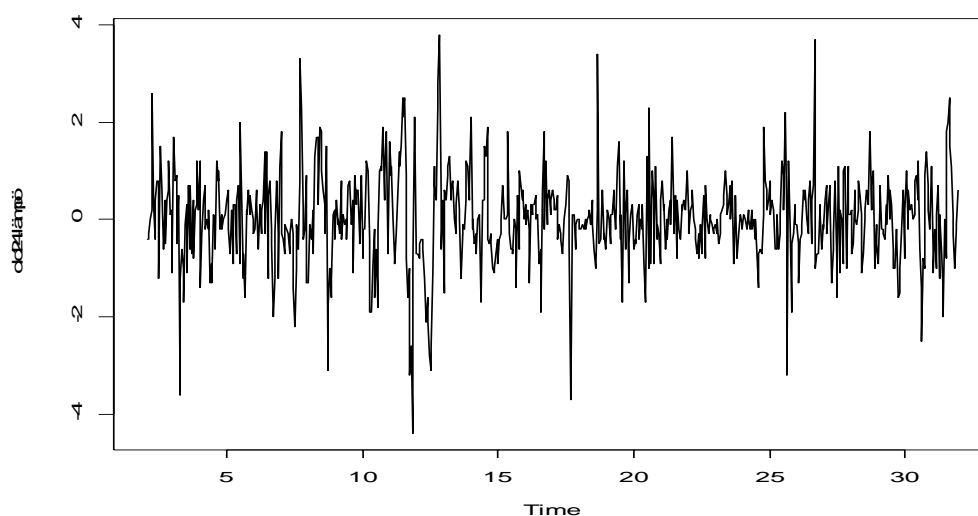
12.2.1. Lämpö, elokuu 1992

Edellisen kappaleen mukaisesti pyritään valkaisemaan molemmat sarjat siten, että sovitetaan ensin niihin sopivat mallit, ja tarkastellaan residuaalien ristikorrelaatioita. Jos aloitetaan mallintaminen lämpötilasarjasta $\{lämpö\}$, nähdään heti luonnollisesti, että sarjassa on selvää kausivaihtelua ja että kauden pituus on 24 tuntia. Yritetään ensin poistaa kausivaihtelu differoimalla sarja kauden pituuden mittaisella viiveellä (Kuvio **12.3**). Sarja ei vielä vaikuta silmämääräisesti kovin stationariselta, mutta kausivaihtelu on hävinnyt hyvin. Kun sarja differoidaan vielä kerran viiveellä yksi, saadaan sarja $\{d12.d.lämpö\}$, joka näyttäisi silmämääräisesti sen verran hyvältä (kuvio **12.4**), että voitaisiin alkaa soveltamaan yksikköjuuritestejä stationarisuuden toteamiseksi.

Kuvio 12.3



Kuvio 12.4



Yksikköjuuren olemassaoloa tarkastellaan R -ohjelman tseries -paketista löytyvillä laajennetulla Dickey Fuller -testauksella, Phillips-Perron -testauksella ja Kwiatkowski-Phillips-Schmidt-Shin(level-)testauksella. ADF- ja PP -testien H_0 :na on, että sarjasta löytyy yksikköjuuri kun taas KPSS -testin H_0 :na on stationaarisuus. Taulukossa 12.1 ovat p -arvot em. testeille.

Taulukko 12.1.

Testi	P -arvo
ADF	<0.01
PP	0.01
KPSS	>0.1

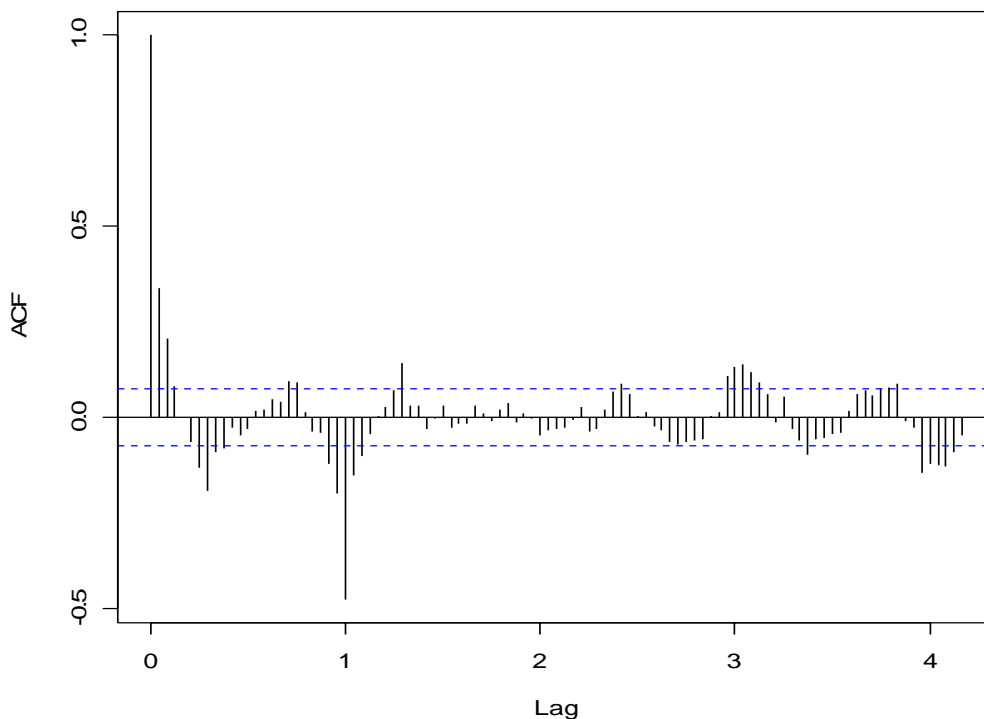
Kun otetaan nollahypoteesit huomioon, voidaan päätellä sarjan $\{d.d24.lämpö\}$ olevan stationaarinen. Seuraavaksi varmistetaan, ettei sarja ole valkoista kohinaa, missä tapauksessa sitä ei olisi mielekästä mallintaa. Sarjaa testataan R:stä löytyvien Box-Pierce-, Ljung-Box- ja Jarque-Bera -testien avulla. Lisäksi käytetään Arto Luoman kodaamaa funktiota, joka testaa aikasarjaa käännepestetestillä. Tulokset näkyvät taulukossa 12.2.

Taulukko 12.2.

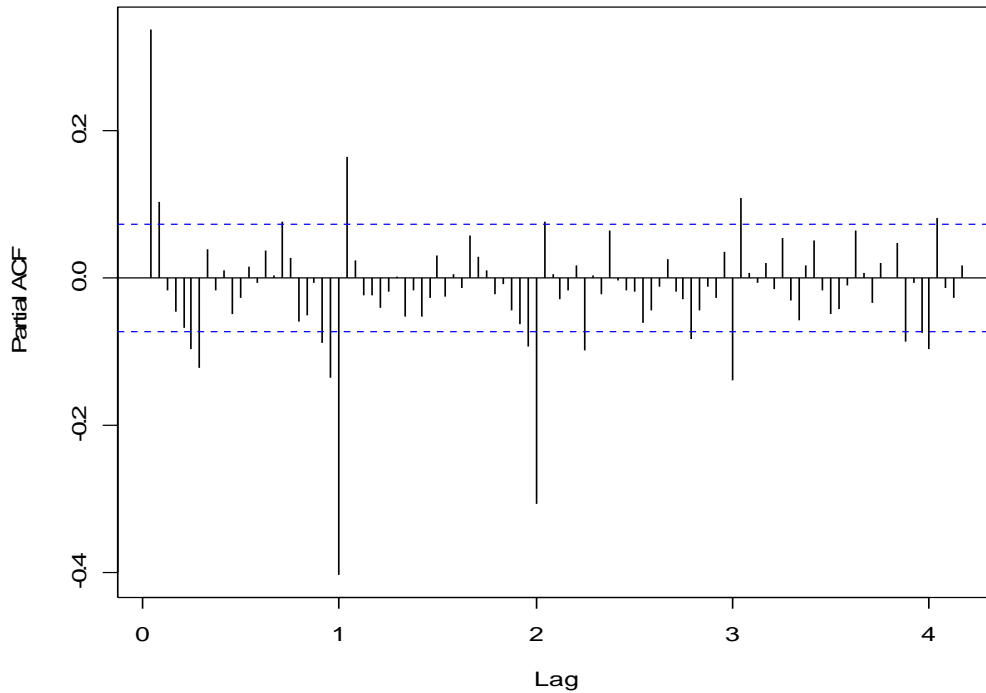
Testi	P -arvo
Box-Ljung	< 2.2e-16
Box-Pierce	< 2.2e-16
Jarque-Bera	< 2.2e-16
Käännepiste	0.0002823

Koska kaikissa testeissä nollahypoteesina on riippumattomuus, voidaan olettaa, että sarjassa on olemassa riippuvuutta, jota voidaan mallintaa. Ensimmäinen tehtävä on tutkia otoksesta saatuja autokorrelaatioita sekä osittaisautokorrelaatioita, jotta saataisiin selville sopivat mallin asteet. Koska voidaan hyvällä syyllä olettaa, että kyseessä olisi kausimalli, aloitetaan päättelystä kausimallin autoregressiivisten ja liukuvan keskiarvon parametrien määrää. Tätä varten kuviossa **12.5a** ja **12.5b** on $\{d.d24.lämpö\}$:n auto- ja osittaisautokorrelaatiofunktioiden kuvaajat siten, että viiveitä on reilusti enemmän kuin normaalia autokorrelaatiota tutkittaessa.

Kuvio 12.5a



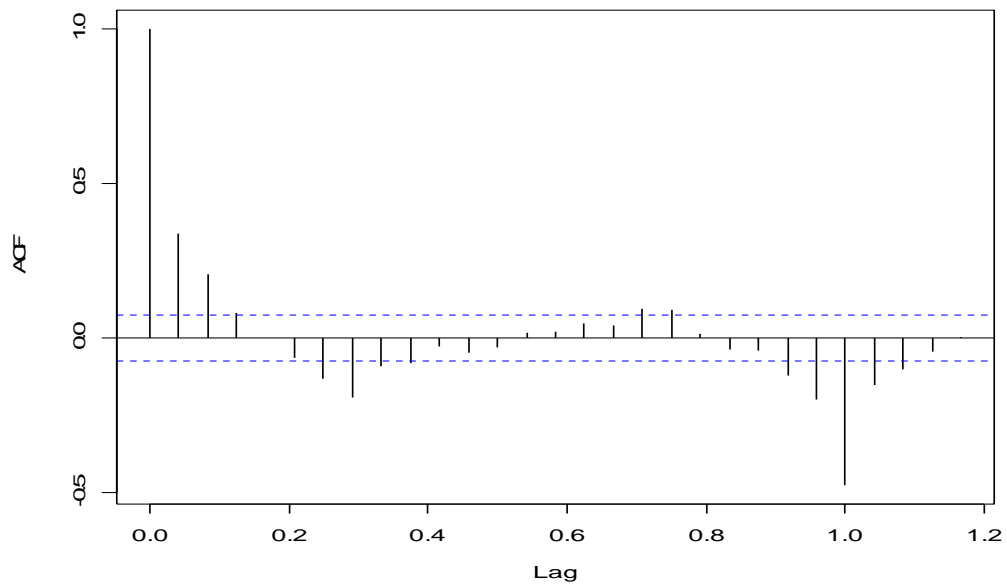
Kuvio 12.5b



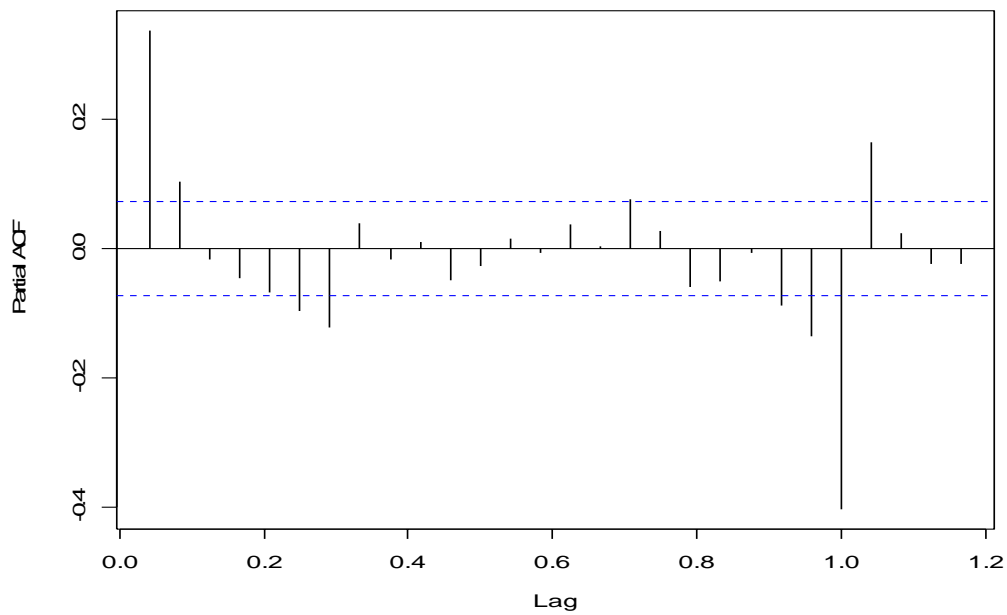
Autokorrelaatiofunktiossa ainoa selvä piikki kausivaihteluissa on ensimmäisen kauden(eli 24:n) kohdalla, mikä viittaisi SMA(1) -malliin. Tämä malli saa vahvistusta osittaisautokorrelaatiofunktion kuvaajasta, joka näyttää, kuinka kausivaihtelujen autokorrelaatiot hiipuvat eksponentiaalisesti.

Seuraavaksi etsitään kauden sisäisten autoregressiivisten ja liukuvan keskiarvon parametrien määrät jälleen auto- ja osittaisautokorrelaatiofunktioiden avulla. Nyt kuvioissa **12.6a** ja **12.6b** viiveet loppuvat ensimmäisiin kausivaihteluihin.

Kuvio 12.6a



Kuvio 12.6b



Autokorrelaatiofunktio näyttäisi hiipuvan eksponentiaalisesti, joten se voisi merkitä AR -prosessia. Kun osittaisautokorrelaatiofunktiossa taas ainoa selvä piikki on ykkösen kohdalla, voidaan olettaa SARIMA(1,1,0)*(0,1,1)*24 -mallin olevan hyvä yleistyks aikasarjalle.

Edelliseen tapaan parametrisoidulle mallille saadaan seuraava malli:

$$(1-0.3523B)\{d.d24.lämpö\}=(1-0.8921B^{24})\{u_t\} \quad (12.1)$$

, missä $\{d.d24.lämpö\}=(1-B)(1-B^{12})\{lämpö\}$
ja $\{u_t\}\sim WN(0, 0.4358)$

Näin ollaan saatu malli (12.1). Seuraavana toimenpiteenä ovat diagnostiset tarkastelut, jotta voidaan todeta mallin olevan pätevä. Jotta voitaisiin sanoa aikasarjan olevan mallinnettu loppuun saakka, on jäännössarjan oltava riippumatonta valkoista kohinaa. Taulukossa **12.3** ovat tulokset edellä käytetyistä riippumattomuustesteistä. Nyt vain riippumattomuushypoteesien täytyy pitää mallin jäännöksien.

Taulukko 12.3

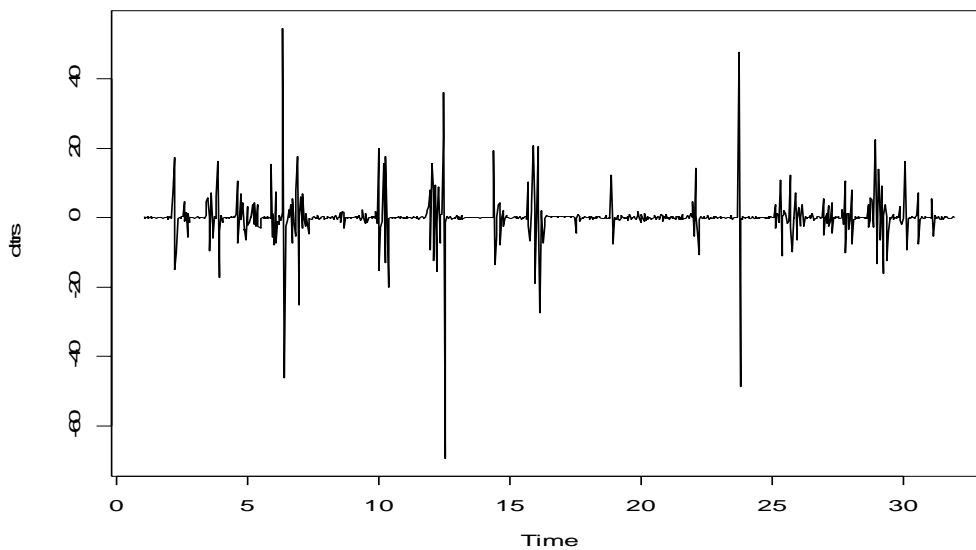
Testi	P –arvo
Box-Ljung	0.4512
Box-Pierce	0.4521
Jarque-Bera	< 2.2e-16
Käännepiste	0.706

Ainoastaan Jarque-Bera –testin kohdalla nollihypoteesi voidaan hylätä. Tämä ei ole kovinkaan vakavaa, sillä Jarque-Bera –testi testaa tarkkaan ottaen sarjan normaaliutta, ei riippumattomuutta. Keskiarvot parametreille ovat 0.0352 ja 0.0236. Kun vapausasteita on $744-2=742$, saadaan t –jakaumasta, että molemmat estimoidut parametrit ovat merkitseviä. Näin ollen edellä saadun mallin residuaaleja voidaan käyttää esivalkaisussa.

12.2.2. TRS elokuu 1992

Seuraavaksi mallinnetaan hajurikkupitoisuuksia $\{trs\}$. Kuviosta **12.1** nähdään, että sarjassa on selkeitä piikkejä epäsäännöllisin väliajoin, joten kausivaihtelusta ei voi olla kyse. Kun sarja differoidaan, merk $\{d.trs\}$ (kuvio **12.7**), nähdään vielä selvemmin, että sarjan taso ei ole ajasta riippuvainen mutta vaihtelu sitä vastoin on.

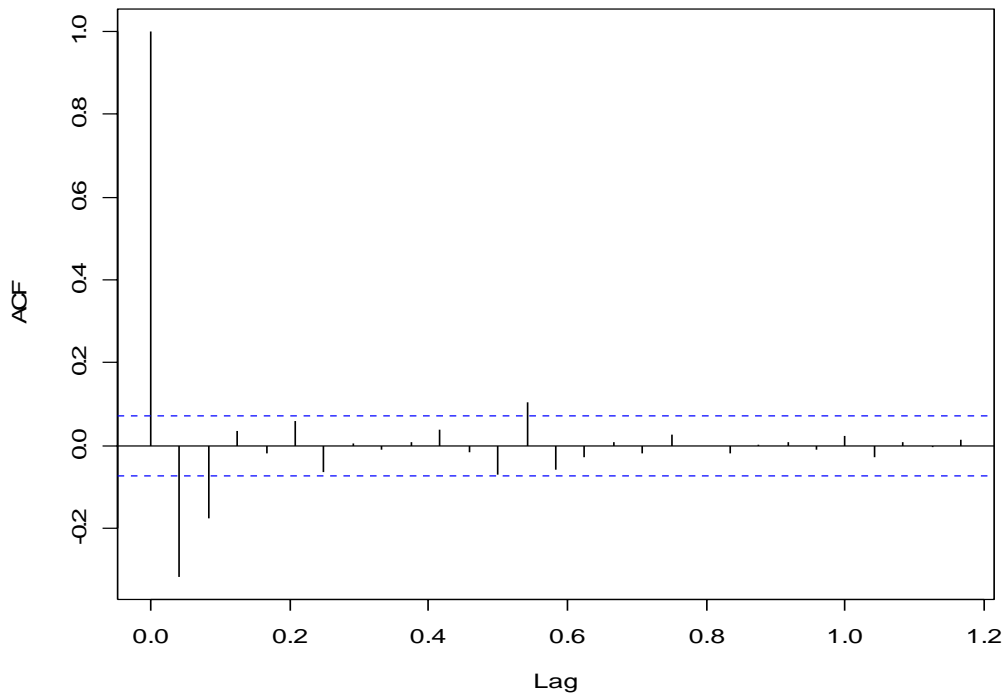
Kuvio 12.7



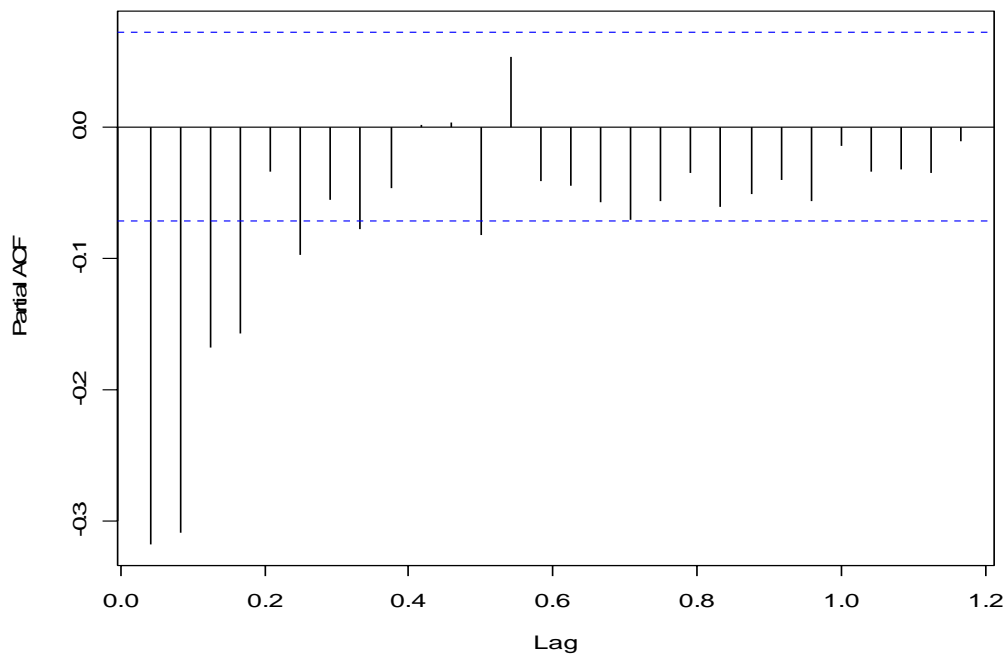
Kuten edellä ollaan todettu, ARMA –mallit toimivat vain stationaarisille aikasarjoille, ja stationaarisuuden määrittelyssä ilmaistaan selvästi, että sarjan varianssin pitää olla vakio ajan suhteen. Tämä jättää ARMA –mallit pois $\{d.trs\}$:n mallintamisesta. Sarjan vaihteleva varianssi antaa kuitenkin vahvan viittauksen siihen, että kappaleessa 10 esitetyt heteroskedastisuuden sallivat mallit voisivat toimia.

Ensin täytyy todeta (G)ARCH –ominaisuus sarjassa. Tämä tehdään luvussa 10.3 ensin mainitulla tavalla. Sovitetaan ensin aineistoon mahdollisimman hyvin sopiva ARMA -malli. Tarkastellaan kuvioissa **12.8a** ja **12.8b** olevia otoksesta laskettuja auto- ja osittaisautokorrelaatiofunktioita.

Kuvio 12.8a



Kuvio 12.8b



Autokorrelaatiofunktiossa on vain kaksi selvää piikkiä, jotka sattuvat viiveiden yksi ja kaksi kohdalle. Tämä antaisi viitettä MA(2) -mallista. Kun vielä osittaisautokorrelaatiofunktio osoittaa hiipumisen merkkejä, voidaan olettaa, että MA(2) -malli kuvaa systeemiä niin hyvin kuin sitä ARMA -malleilla voidaan kuvata. Kun MA(2) -mallin residuaalit talletetaan ja korotetaan toiseen

potenssiin, saadaan Box-Ljung –arvoksi 68.1773, jota vastaava p –arvo on $< 2.2e-16$. Nollahypoteesin voi siis hylätä ja kappaleen 10.3 mukaisesti samalla hylätään myös olettaus, että sarjassa ei olisi heteroskedastisuutta.

GARCH(0,2) –malli antaa hyvät tulokset niin estimaateille kuin residuaaleillekin mutta GARCH(0,1) malli antaa suurinpiirtein samat tulokset, mutta koska siinä on vähemmän estimoitavia parametrejä, otetaan se tarkasteltavaksi malliksi. GARCH(0,1) –mallissa ehdolliseksi varianssiksi saadaan

$$h_t = 5.75778 + 13.99532 u_{(t-1)}^2$$

Taulukossa 12.4 ovat edellä saadun GARCH(0,1) –mallin residuaalien diagnostiset tarkastelut Box-Ljung -testeillä.

Taulukko 12.4

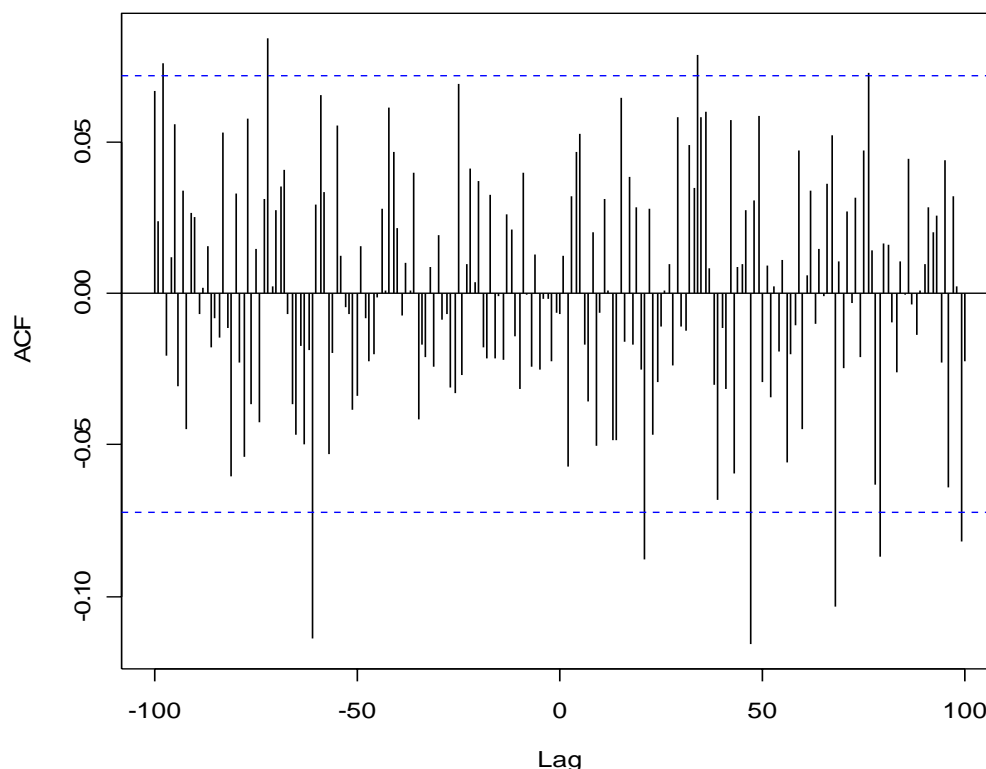
Box-Ljung -testi	P –arvo
Residuaaleille	0.3855
Neliöidylle residuaaleille	0.5861

Molemmissa tapauksissa nollahypoteesi jää voimaan. Residuaalien tapauksessa se tarkoittaa sitä, että residuaalit todellakin ovat riippumattomia, ja siten niitä voi käyttää esivalkaisussa. Neliöityjen residuaalien kohdalla se taas tarkoittaa sitä, että jäännöksissä ei enää ole GARCH –ominaisuutta jäljellä.

12.2.3. Ristikorrelaation tarkastelu

Nyt kun ollaan saatu sekä $\{\text{lämpö}\}$:lle että $\{\text{trs}\}$:lle jäännössarjat, voidaan lopulta tutkia näiden sarjojen välistä ristikorrelaatiota, jotka ovat kuviossa 12.9.

Kuvio 12.9



Kuviossa näkyy viisi merkittävää piikkiä positiiviseen viiveeseen sata mennessä, mikä on normaalia, kun kyseessä ovat 95%:n luottamusvälit. Negatiivisilla viiveillä piikkejä on vieläkin vähemmän.

Edellä mainitun perusteella voidaan olettaa, etteivät kaksi aikasarjaa ole korreloituneet keskenään. Näin voidaan sanoa, ettei hajurikkipitoisuuksien määrä riippunut ilman lämpötilasta ainakaan 1992 elokuussa.

12.3. Loput kuukaudet

Koska sarjat ovat luonteeltaan hyvin samanlaisia, tekniikka muitten kuukausien ristikorrelaatioiden tutkimiseen on käytännössä sama. Taulukossa **12.5** on kirjattuna saadut mallit lopulle seitsemälle sarjalle.

Taulukko 12.5

	Lämpö	Trs
Elokuu 1993	SARIMA(110)(011)24	GARCH(0,1)
Lokakuu 1994	SARIMA(310)(011)24	GARCH(0,1)

Maaliskuu 1995	SARIMA(210)(011)24	GARCH(0,1)
Syyskuu 1997	SARIMA(110)(011)24	GARCH(0,1)
Joulukuu 1998	SARIMA(121)(011)24	GARCH(0,1)
Tammikuu 2000	SARIMA(121)(011)24	GARCH(0,1)
Toukokuu 2000	SARIMA(121)(011)24	GARCH(0,1)

Kuten taulukosta näkyy, ovat kaikki sarjat mallinnettavissa lähes samalla tavoin. Jokainen lämpösarja on luonnollisesti kausisarja, jonka lisäksi jokaisen mallinnuksessa käytetään autoregressiivisiä parametreja. Kahdesti differoidut sarjat vaativat lisäksi yhden liukuvan keskiarvon parametrin mukaan. Hajurikkisarjoja voi kaikkia mallintaa GARCH(0,1) –mallilla, sillä sarjat ovat luonteeltaan vielä enemmän toistensa kaltaisia kuin lämpösarjat. Valitut mallit ovat kaikkein yksinkertaisimpia kaikista mahdollisista kriteerit täyttävistä malleista.

Jäännössarjojen ristikorrelaatioita tarkastellessa yhdessäkin tapauksessa ei ollut havaittavissa riippuvuutta sarjojen välillä. Koska sarjojen mallinnus näissä kahdeksassa tapauksessa noudattaa pitkälti samaa kaavaa, ei ole mitään syytä olettaa, että tilanne olisi toinen muita kuukausia tarkisteltaessa. Näin voidaan päätellä, että hajurikkipitoisuudet eivät todellakaan riipu lämpötilasta, eikä siten ole mielekäästä alkaa tehdä ennustemalleja näiden havaintojen pohjalta.

13. Interventioanalyysi

N-ulotteisten aikasarjojen paras ominaisuus on se, että ne mallintavat sarjojen välisiä vaihteluita. Esimerkiksi tiettyjä pörssikursseja on mielekkäämpi mallintaa niiden suhteilla toisiin kurssiin, kuin samojen kurssien aikaisempiin havaintoihin. Näin ollen toista aikasarjamuuttujaa voidaan pitää 'riippuvana' muuttujana ja toista 'riippumattomana' tai 'eksogeenisenä' muuttujana. Näin, jos ollaan varmoja, että syy ja seuraus sarjojen välillä kulkevat vain toiseen suuntaan.

Boxin ja Tiaon(1975) kehittämän interventioanalyysin tarkoituksena on estimoida aikasarjassa tapahtuvia tason muutoksia tiettyjen interventio pisteiden suhteen. Esimerkkinä muutaman vuoden päästä voitaisiin tutkia, kuinka Suomessa alkoholiveronalennus on vaikuttanut alkoholijuomien kulutukseen. Silloin

interventiopisteenä olisi päivä, jolloin alkoholivero pieneni ja alkoholijuomien hinnat halpenivat. Eräs tapa tehdä tämä olisi mitata keskiarvo interventiota edeltävältä ajalta ja verrata sitä intervention jälkeiseen keskiarvoon. Tällainen lähestymistapa ei kuitenkaan ole aikasarja-analyysissä riittävä, sillä intervention vaikutus ei välttämättä ole välitön, tai se voi jopa hiipua kokonaan pois.

13.1. Interventioanalyysin teoria

Interventioanalyysin avulla voidaan muodollisesti testata aikasarjan keskiarvon muuttumista. Eräs malli, jonka Enders, Sandler ja Cauley(1990) ovat kehittäneet, tutkii metallinpaljastimien asentamisen vaikutusta USA:n lentokentille vuonna 1973 lentokonekaappausten määrään. Seuraavassa on mallin periaate. Malli on AR(1) –malli, jossa on ns. interventiotermi mukana.

$$y_t = \phi_0 + \phi_1 y_{(t-1)} + c_0 z_t + u_t$$

, missä z_t on interventio(tai dummy) muuttuja, joka saa arvon 0 interventiota edeltävänä ajanjaksona, ja arvon 1 sen jälkeisenä aikana. Oletetaan, että intervention ajankohta on h .

Kun $t < h$, $z_t = 0$, jolloin mallin vakiotermin on ϕ_0 ja pitkän välin keskiarvo $\frac{\phi_0}{1 - \phi_1}$. Kun saavutaan pisteeseen h , interventiomuuttuja alkaa saada arvon 1, joten vakiotermin arvoksi muuttuu $\phi_0 + c_0$. Näin ollen uusi, intervention jälkeinen, pitkän ajan keskiarvo on $\frac{\phi_0 + c_0}{1 - \phi_1}$. Intervention välitöntä vaikutusta kuvaa termi c_0 , jota voidaan testata tavallisella t-testillä. Jos c_0 eroaa tilastollisesti nolasta, ja on negatiivinen, voidaan päätellä myös intervention vaikutuksen olevan negatiivinen.(Enders, 270-275.)

Mallin ei tietenkään tarvitse välttämättä olla edellä mainittu AR(1), vaan yleinen Box-Jenkins –malli interventioanalyysille on(Enders, 272):

$$y_t = \phi_0 + \phi(B)y_{(t-1)} + c_0 z_t + \theta(B)u_t$$

, missä vakiotermin erottamiseksi $\phi(B)$ on nyt $(\phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p)$
ja $\theta(B)$ kuten kappaleessa 5 on määritelty

Endersin et al esimerkissä oletettiin, että metallinpaljastimien vaikutus on välitön ja pysyvä. Näin ei kuitenkaan aina käytännössä ole, vaan intervention vaikutus voi viivästyä s :n aikayksikön verran. Intervention täysi vaikutus ei myöskään aina välttämättä näy heti, vaan asteittain, tai se voi jopa häipyä pois. Tällöin tutkijan on valittava aineistoon parhaiten sopiva interventiofunktio. Tällä tarkoitetaan z_t :n sopivien arvojen etsimistä. Oletetaan, että interventio tapahtuu ajanhetkellä h , mutta sen vaikutukset alkavat näkyä hitaasti kolmen aikayksikön aikana vasta s yksikköä interventiosta. Tällöin z_t :n voisi määritellä seuraavasti:

$$z_t = 0, \text{ kun } t < h + s$$

$$z_t = 0.5, \text{ kun } t = h + s$$

$$z_t = 1, \text{ kun } t > h + s$$

Edellinen oli vain esimerkki eräästä interventiofunktioista. Teoreettisesti interventiofunktioita on ääretön määrä. Käytännössä on todettu useimpien ilmiöiden noudattavan jotakin seuraavista interventiofunktioista. 1) Impulssi: tapahtuma on yllätyksellinen luonteeltaan ja vaikutukset hiipuvat ajan myötä pois. 2) Askel: Intervention vaikutus on pysyvä ainakin tarkasteltavalla aikavälillä. 3) Ramppi: Intervention vaikutus muuttuu tasaista vauhtia. (McLeod, 13-8 – 13-11.) Tutkijan on syytä tutkia käsillä olevaa ilmiötä ja tehdä omat päätöksensä intervention luonteesta.

13.2. Interventioiden tutkiminen trs –sarjassa

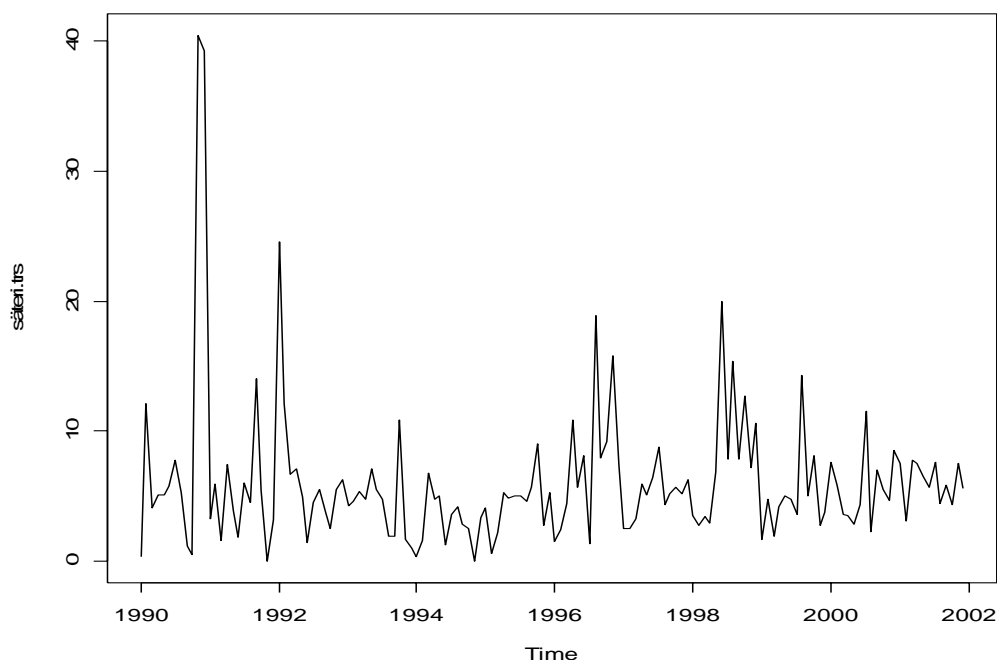
Koska interventiolla yleensä on pysyvä, tai ainakin pitkäaikainen vaikutus, käytetään näissä analyyseissä kuukausiaineistoa. Aineistot ovat jaoteltu kahtia siten että toisessa aineistossa ovat saastemittaukset tuulisektorista 17-54, jolloin saadaan Säterin kaatopaikan suunnasta tulevat päästöt. Toisessa aineistossa ovat saastemittaukset tuulisektorilta 225-251, jolloin kyseessä ovat UPM:n tehtailta tulevat päästöt.

13.2.1. Säterin suunnalta tulleet päästöt

Säterin suunnalla ei ole tehty varsinaisia ilmansuojeluinvestointeja. Säterin kaatopaikalle lopetettiin jäteköyden vienti vuoden 1998 lopussa, mikä on ainoa edes jollakin tavalla ilmastonsuojeluun liittyvä muutos. Näin tuo vuoden 1998 loppu on luonnollinen valinta interventiopisteeksi ja nimenomaan sellaiseksi, että hajurikkipitoisuuksien voisi luulla laskevan.

Kuviosta **13.1** kuitenkin nähdään, ettei tuohon aikaan ole ainakaan silmämääräisesti tapahtunut trs -pitoisuuksien laskemista.

Kuvio 13.1

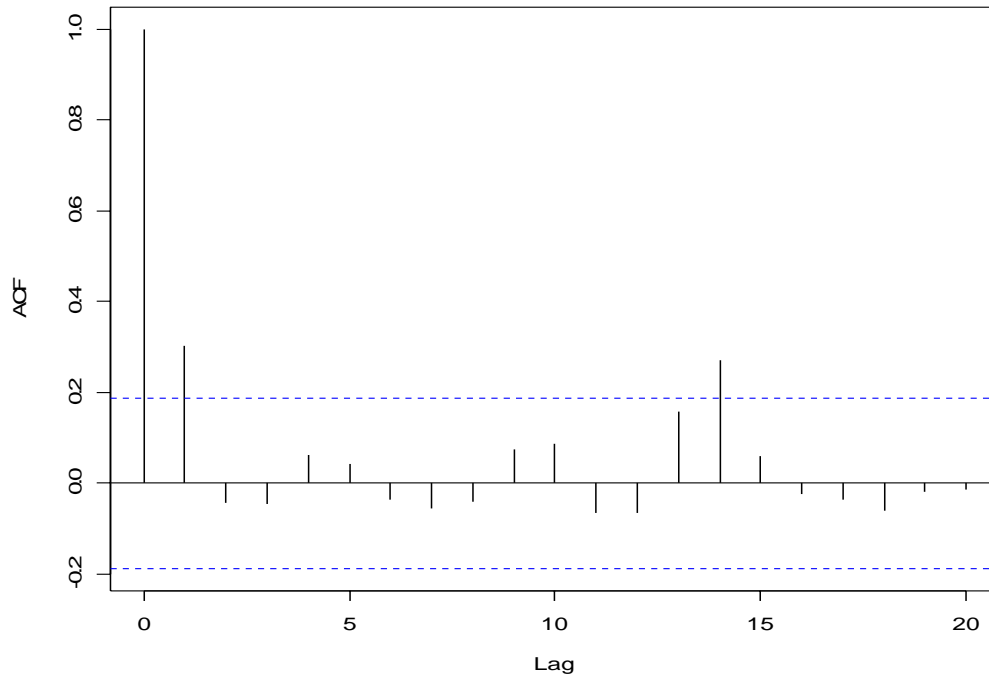


Pelkän silmämääräisyyden perusteella ei kuitenkaan voi luottaa siihen, etteikö todellisuudessa jonkinasteista interventioita olisi tapahtunut.

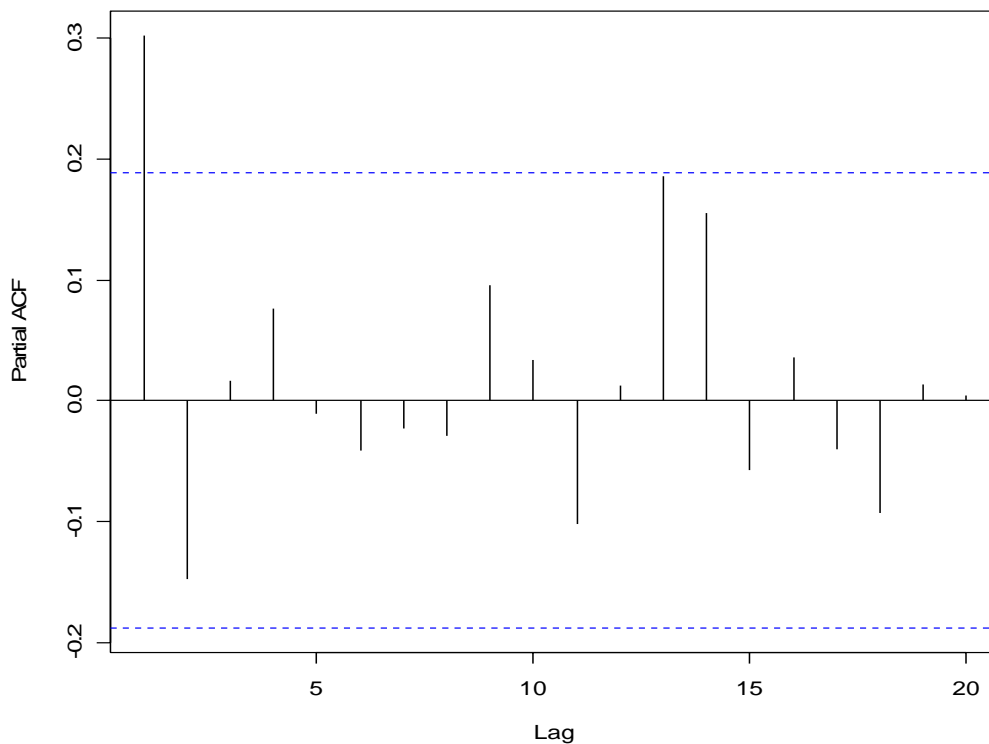
Valitaan interventiopisteeksi 108(joulukuu 1998), ja koska havaintojen välillä on kuukausi, oletetaan, että intervention vaikutus oli pysyvä ja välitön, eli $z_t=0$, kun $t \leq 108$ ja $z_t=1$, kun $t > 108$. Analyysissä hankaluuksia tuottaa lyhyt 'häntä', eli intervention jälkeen käytettävissä on vain $144-108=36$ havaintoa. Tämän vuoksi luonnollisesti on järkevintä valita mallinnettavaksi alueeksi interventiota ennen tapahtuneet havainnot. Merkataan tätä sarjaa $\{\text{säteri.trs.alku}\}$

Adf-, PP- ja kpss –testit osoittivat, että interventiota edeltävästä sarjasta ei löydy yksikköjuurta, kun taas riippumattomuustestit osoitivat, että sarja ei ole valkoista kohinaa, joten muunnoksia ei tarvita. Kuvioissa **13.2a** ja **13.2b** ovat sarjan autokorrelaatio- sekä osittaisautokorrelaatiofunktiot.

Kuvio 13.2a



Kuvio 13.2b



Ainoat merkittävät piikit löytyvät autokorrelaatiofunktiossa ykkösen ja neljäntoista kohdalta, kun taas osittaisautokorrelaatiofunktiossa ainoa piikki on ykkösen kohdalla. Autokorrelaatioissa on havaittavissa pientä siniaaltomaista käyttäytymistä, mikä voi viitata AR(p) –prosessiin. Jos unohdetaan viive neljäntoista autokorrelaatioissa, voidaan olettaa, että mallin asteluvut eivät ole ykköstä korkeammat. Näin ollen tarkasteltavaksi jää kolme mallia, joista valitaan yksinkertaisuuden vuoksi AR(1) –malli(kaikki kolme mallia antoivat hyväksyttävät tulokset diagnostisissa tarkasteluissa). Seuraavaksi tehdään lineaarinen malli, joka on AR(1) –mallin muotoinen ja jossa on interventiotermi z_t mukana. Taulukossa **13.1** on tulostus mallista parametrien keskivirheineen ja merkitsevyytensä osineen.

Taulukko 13.1

Parametri	Estimaatti	Keskivirhe	T -arvo	P –arvo
ϕ_0	4.58265	0.71902	6.373	2.50e-09
ϕ_1	0.27236	0.08093	3.365	0.000986
c_0	-0.55285	1.03228	-0.536	0.593110

Mallin mukaan keskiarvo ennen interventiota on $\frac{\phi_0}{1-\phi_1}=4.58265/(1-0.27236)=6.297963$. Intervention välitön vaikutus on -0.55285 , josta uusi keskiarvo $\frac{\phi_0 + c_0}{1-\phi_1}=(4.58265-0.55285)/(1-0.27236)=5.538178$. Näin ollen intervention pitkän ajan vaikutus on $5.538178-6.297963=-0.759785$, eli jäteköyden viennin lopettamisella näyttäisi äkikseltään todellakin olleen ilman laatua parantava vaikutus. Näin ei kuitenkaan voida sanoa, sillä kuten taulukosta **13.1** nähdään, on interventiotermin t –arvo -0.536 , joka ei ole merkitsevä, minkä vuoksi termi on nolla, eikä interventiolla todellisuudessa ole merkitystä.

Seuraavassa vielä tulostus olettaen, että aikasarjan luonut prosessi on luonteeltaan valkoista kohinaa(mikä voi myös olla hyvin lähellä todellisuutta). Tällöin riittää laskea lineaarinen malli pelkällä vakio –ja interventiotermillä:

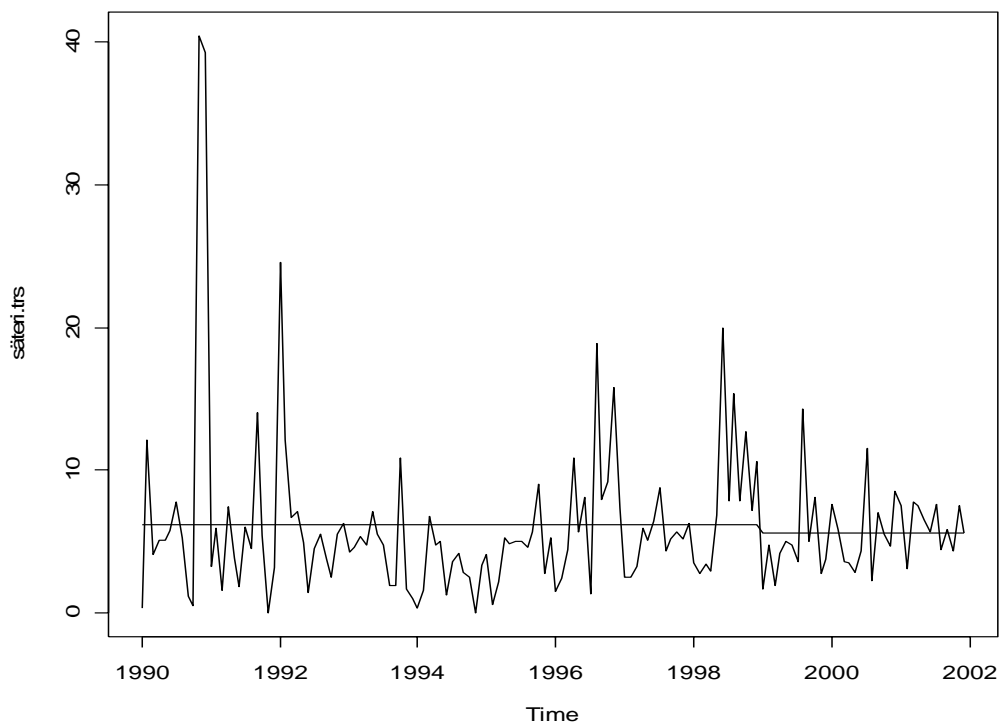
$$\{\text{säteri.trs}\} = a_0 + c_0 z_t + u_t$$

Taulukossa **13.2.** on kyseisen mallin kertoimet keskivirheineen ja t –arvoineen, sekä kuviossa **13.3.** intervention vaikutukset.

Taulukko 13.2.

Parametri	Estimaatti	Keskivirhe	T -arvo	P –arvo
a_0	6.2078	0.5340	11.626	<2e-16
c_0	-0.6172	1.0680	-0.578	0.564

Kuvio 13.3

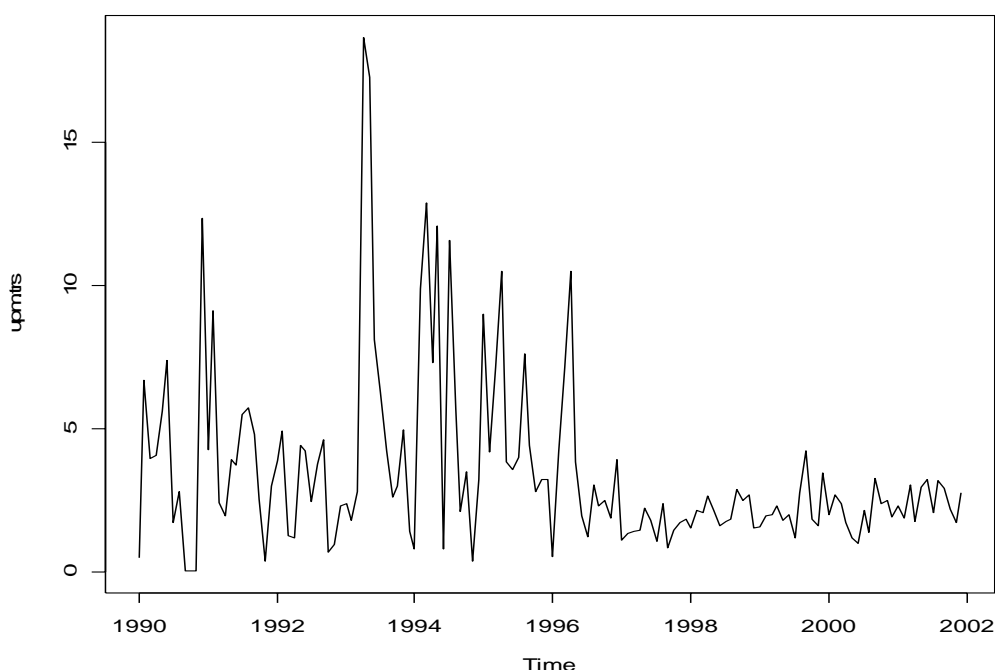


Taulukosta nähdään jälleen, kuinka interventiotermi ei ole tilastollisesti merkitsevä. Kuvioista puolestaan nähdään, kuinka pieni merkitys interventiotermillä on noin pienellä merkitsevyystasolla. Tulosten valossa voidaan huoletta todeta, ettei jäteköyden viemisen lopettamisella ollut vaikutuksia ympäristölle suuntaan eikä toiseen

13.2.2. UPM:n suunnalta tulleet päästöt

UPM:n tehtailla ollaan tehty enemmän muutoksia. Vuonna 1990 aloitettiin väkevien hajukaasujen poltto. Tämän intervention vaikutusta ei kuitenkaan voida käytössä olevilla havainnoilla tutkia. 1995 tiettyjä väkevimpä jakeita laimeista otettiin mukaan polttoon(Päällysaho). Kuviosta **13.4.** näkyikin selvä lasku sekä varianssin pieneneminen trs -pitoisuuksissa.

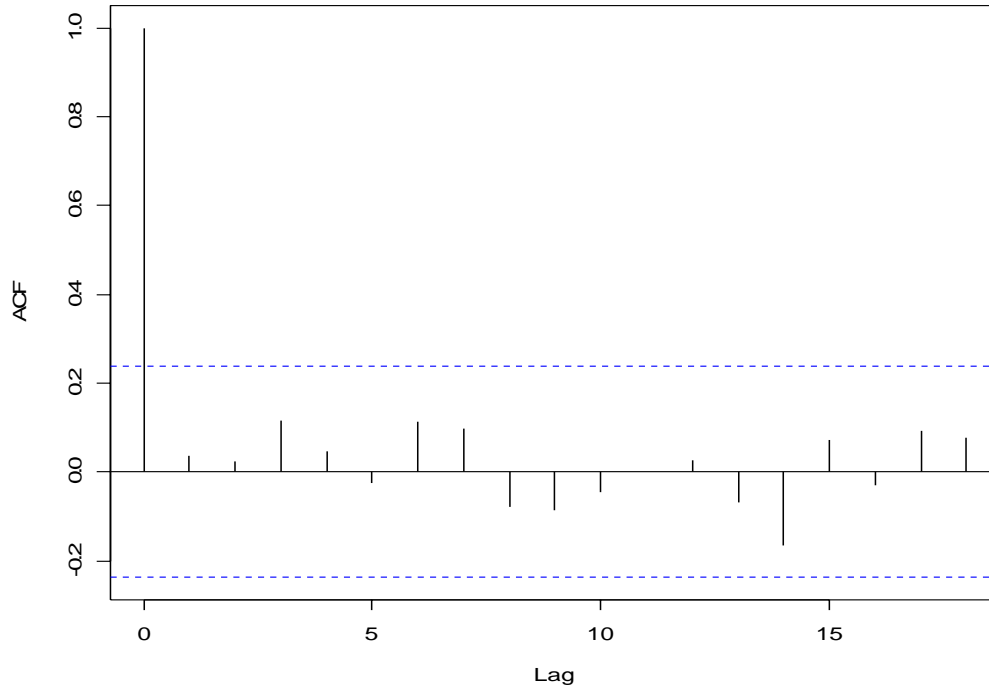
Kuvio 13.4



Tutkitaan ensin tämän interventiopisteen vaikutusta. Voitaneen olettaa, että vuonna 1995 tapahtuneen intervention vaikutus ei ole alkanut näkyä päästöissä vasta kuin vuonna 1996. Tällöin interventiofunktio olisi viivästettyä muotoa. Oletetaan kuitenkin, että intervention vaikutus tässäkin tapauksessa on välitön, eli interventiofunktio hyppää suoraan ykköseen. Kun valitaan interventiopisteeksi 76, saadaan pisteen molemmille puolille jäävien sarjojen pituuksiksi 76 ja 68, eli sarja menee suunnilleen 'keskeltä poikki'. Voidaan siis valita kumpi puoli tahansa mallinnettavaksi. Koska kuviosta **13.4.** nähdään, että varianssi pysyy paremmin aisoissa interventiopisteen oikealla puolella, valitaan sarjan loppupää mallinnettavaksi.

Adf-, PP- ja KPSS –testit osoittavat, ettei sarjasta löydy yksikköjuurta. Kuitenkin Box- ja Jarque-Bera –testit osoittavat, että sarjassa ei ole riippuvuutta, minkä voi hyvin ymmärtää, kun katsoo kuviossa **13.5.** olevaa autokorrelaatiofunktion kuvaajaa.

Kuvio 13.5



Sarjaa ei siis ole mielekästä mallintaa Box-Jenkins –menetelmillä, vaan edellisen kappaleen lopun tyyliin tavallisella lineaarisella mallilla, jossa ei vakiotermin lisäksi ole muuta kuin interventiotermi. Taulukossa **13.3.** ovat kertoimet keskivirheineen ja t –arvoineen tällaiselle mallille

Taulukko 13.3.

Parametri	Estimaatti	Keskivirhe	T -arvo	P –arvo
a_0	4.7058	0.3239	14.530	< 2e-16
c_0	-2.5475	0.4713	-5.405	2.67e-07

Taulukosta nähdään selvästi, kuinka kaikki kertoimet ovat merkitseviä, joten interventiolla on selvä vaikutus. Keskiarvo ennen interventiota on 4.7058, ja intervention välitön ja samalla pysyvä vaikutus on -2.5475, josta uudeksi keskiarvoksi saadaan $4.7058 - 2.5475 = 2.1583$.

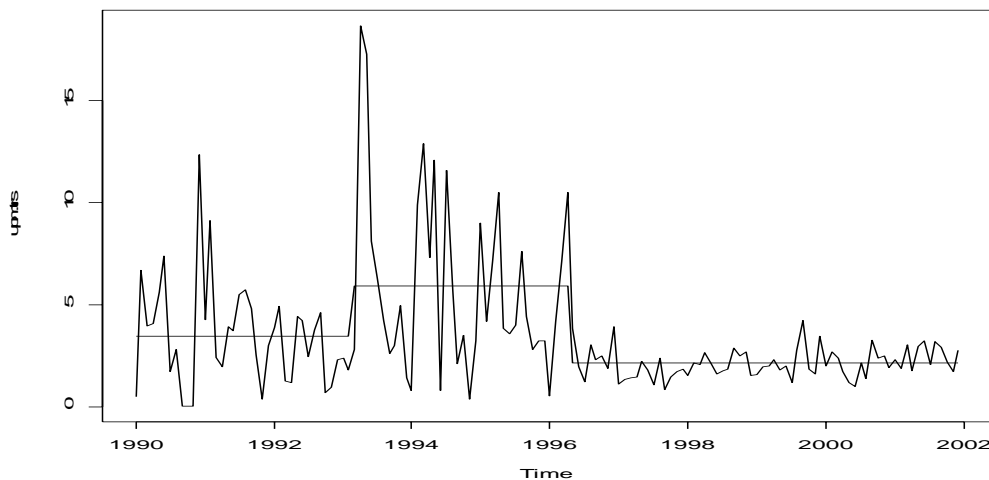
Aikasarjassa voi kuitenkin olla useampi kuin yksi interventio. Koska nyt tiedetään, että aikasarjan jälkimmäinen puolisko on valkoista kohinaa, voidaan helposti tutkia interventioita lineaarisen mallin avulla sarjan ensimmäiseltä puoliskolta. Vaikka mistään interventiosta ei olekaan varsinaista tietoa, voidaan sarjasta esim. silmämääräisesti ottaa jokin piste, asettaa se interventioasteeksi, ja tutkia sen vaikutusta. Nyt kuviosta voidaan nähdä, että pisteen $t=76$ lisäksi, on sarjan tasossa pieni muutos pisteessä $t=39$. Edelliseen tapaan selitetään trs – pitoisuuksia lineaarisen mallin avulla, jossa selittäjinä ovat vakio-termi ja kaksi interventio-termiä $z_1=0$, kun $t<39$, $z_1=1$, kun $t\geq 39$ ja $z_2=0$, kun $t<76$, $z_2=1$, kun $t\geq 76$. Taulukossa **13.4.** on tulostus kyseisestä mallista.

Taulukko 13.4.

Parametri	Estimaatti	Keskivirhe	T -arvo	P -arvo
a_0	3.4668	0.4353	7.964	5.03e-13
c_1	2.4779	0.6156	4.025	9.26e-05
c_2	-3.7864	0.5435	-6.967	1.14e-10

Nyt nähdään, että edellä laskettu interventio on vielä merkitsevä, mutta niin on myös uusi malliin mukaan otettu interventio. Keskiarvo ennen ensimmäistä interventiota on 3.4668, josta se muuttuu ensimmäisen intervention jälkeen $3.4668+2.4779=5.9447$:een. Tämä taas puolestaan putoaa toisen intervention jälkeen $5.9447-3.7864=2.1583$:een, mikä saatiin jo edellisestä mallista. Kuviossa **13.6.** on esitys interventioiden kulusta.

Kuvio 13.9



Kumpi malleista sitten on parempi? Molemmissa malleissa kaikki kertoimet olivat merkitseviä, mutta jälkimmäisen mallin selitysvaste oli selvästi edellistä korkeampi, joten jälkimmäistä mallia voidaan pitää siinä mielessä edellistä 'parempana'. Molemmat mallit ovat sikäli 'oikeita', että interventiokertoimet olivat merkitseviä, mutta jälkimmäinen malli antaa lisäinformaatiota, mitä edellisestä puuttuu. Ilman jälkimmäistä mallia ei olisi aivan päivänselvää, että hajurikkipitoisuudet olivat nousussa kolmen vuoden ajan ennen tiettyjen laimeitten jakeitten polttoon ottoa.

14. Vektoriautoregressiiviset mallit

Kappaleessa 11 käsiteltiin ristikorrelaatiota kahden aikasarjamuuttujan välillä, eli tutkittiin kahden aikasarjan mahdollista vaikutusta toisiinsa. Seuraavassa kappaleessa asia viedään pitemmälle, kun esitellään vektoriautoregressiiviset mallit (VAR –mallit). VAR –analyysi liittyy läheisesti aikasarjatutkimuksen nykyisen trendiin, jossa aikasarjoja tutkitaan moniulotteisina yhtälöjärjestelminä.

VAR –analyysiä vastaavaa tutkimusta voidaan tehdä myös siirtofunktioimalleilla, joka on yleistys edellä käsitellyistä interventioanalyysistä. Siirtofunktioimallien haittana on se, että malliin pitää aina asettaa 'selittäjä' ja 'selitettävä', eli sarjojen välistä vuorovaikutusta ei mallinneta. Tämä rajoitus poistuu VAR –analyysissä.

14.1. VAR –malliyhtälöt

Jos ei olla varmoja, ovatko muuttujat eksogeenisiä, voidaan niitä tutkia symmetrisesti. Kahden muuttujan y_t ja z_t tapauksessa voidaan siis olettaa, että y_t :n menneisyys vaikuttaa z_t :n menneisyyteen ja päinvastoin. Tarkastellaan seuraavaa kaksiulotteista järjestelmää:

$$y_t = b_{10} - b_{12} z_t + \gamma_{11} y_{(t-1)} + \gamma_{12} z_{(t-1)} + u_{yt} \quad (14.1)$$

$$z_t = b_{20} - b_{21} y_t + \gamma_{21} y_{(t-1)} + \gamma_{22} z_{(t-1)} + u_{zt} \quad (14.2)$$

, missä on oletettu, että 1) y_t ja z_t ovat stationaarisia, 2) u_{y_t} ja u_{z_t} ovat valkoisen kohinan virheprosesseja, joilla on keskihajonnat σ_y ja σ_z ja 3) u_{y_t} ja u_{z_t} ovat korreloimattomia.

Yhtälöt (14.1) ja (14.2) muodostavat ensimmäisen asteen vektoriautoregressiivisen mallin. Nimitys tulee siitä, että mallin pisin viivepituus on yksi. Kyseinen malli on yksinkertaisin mahdollinen VAR –malli, joten käytetään sitä kuvaamaan yleisempiä VAR –mallien ominaisuuksia. Yhtälöistä nähdään, kuinka mallissa esiintyy palautetta. Esimerkiksi b_{21} on y_t :n samanhetkinen vaikutus z_t :hen. Virhesarjoista on huomattava, että jos $b_{12} \neq 0$, on shokilla(tai innovaatiolla) u_{z_t} epäsuora vaikutus z_t :hen ja toisinpäin.

Yhtälöt (14.1) ja (14.2) eivät ole redusoidun muodon yhtälöitä, sillä y_t vaikuttaa z_t :hen ja päinvastoin. Yhtälöt voidaan kuitenkin muuttaa matriisilaskennan avulla käytännöllisempään muotoon seuraavasti:

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{(t-1)} \\ z_{(t-1)} \end{bmatrix} + \begin{bmatrix} u_{y_t} \\ u_{z_t} \end{bmatrix} \quad (14.3)$$

Tätä matriisiyhtäöä voidaan nyt merkitä seuraavasti:

$$\mathbf{B}\mathbf{x}_t = \mathbf{\Gamma}_0 + \mathbf{\Gamma}_1\mathbf{x}_{(t-1)} + \mathbf{u}_t$$

, missä

$$\mathbf{B} = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}, \quad \mathbf{x}_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix}, \quad \mathbf{\Gamma}_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix}$$

$$\mathbf{\Gamma}_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}, \quad \mathbf{u}_t = \begin{bmatrix} u_{y_t} \\ u_{z_t} \end{bmatrix}$$

Yhtälö (14.3) on VAR –mallin primitiivimuoto. Jos (14.3) kerrotaan vasemmalta \mathbf{B}^{-1} :llä, saadaan VAR –mallin standardimuoto:

$$\mathbf{x}_t = \mathbf{A}_0 + \mathbf{A}_1\mathbf{x}_{(t-1)} + \mathbf{e}_t \quad (14.4)$$

, missä $A_0 = \mathbf{B}^{-1}\Gamma_0$

$$A_1 = \mathbf{B}^{-1}\Gamma_1$$

$$e_t = \mathbf{B}^{-1}u_t$$

Nyt (14.4) voidaan kirjoittaa seuraavasti:

$$y_t = a_{10} + a_{11}y_{(t-1)} + a_{12}z_{(t-1)} + e_{1t} \quad (14.5)$$

$$z_t = a_{20} + a_{21}y_{(t-1)} + a_{22}z_{(t-1)} + e_{2t} \quad (14.6)$$

Yhtälöissä (14.5) ja (14.6) on huomattava, että virhetermit e_{1t} ja e_{2t} koostuvat molemmat nyt kahdesta sokista seuraavasti. Koska $e_t = \mathbf{B}^{-1}u_t$, saadaan

$$e_{1t} = \frac{u_{yt} - b_{12}u_{zt}}{1 - b_{12}b_{21}} \quad (14.7)$$

$$e_{2t} = \frac{u_{zt} - b_{21}u_{yt}}{1 - b_{12}b_{21}} \quad (14.8)$$

Koska edellä määriteltiin u_{yt} ja u_{zt} valkoisen kohinan prosessiksi, ovat niiden keskiarvot nollia, varianssit vakioita ja ne ovat erikseen autokorreloimattomia. Vaikka virhesarjoissa ei olekaan autokorrelaatiota, on syytä huomata, että ne ovat korreloituneita. Sarjojen kovarianssi voidaan laskea seuraavasti:

$$E(e_{1t}, e_{2t}) = \frac{E[(u_{yt} - b_{12}u_{zt})(u_{zt} - b_{21}u_{yt})]}{(1 - b_{12}b_{21})^2} = -\frac{b_{21}\sigma_y^2 + b_{12}\sigma_z^2}{(1 - b_{12}b_{21})}$$

Yleensä virhetermien e_{1t} ja e_{2t} kovarianssimatriisi määritellään seuraavasti:

$$\Sigma = \begin{bmatrix} \text{Var}(e_{1t}) & \text{Cov}(e_{1t}, e_{2t}) \\ \text{Cov}(e_{2t}, e_{1t}) & \text{Var}(e_{2t}) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

Kahden muuttujan ensimmäisen asteen VAR –yhtälöistä saadaan $n:n$ muuttujan p –asteiset standardimuodon yhtälöt analogisesti. Kasvatetaan ensin viivepituus yhdestä kahteen: (Enders, 294-296)

$$\mathbf{x}_t = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{x}_{(t-1)} + \mathbf{A}_2 \mathbf{x}_{(t-2)} + \mathbf{e}_t$$

, eli

$$y_t = a_{10} + a_{1.11}y_{(t-1)} + a_{1.12}z_{(t-1)} + a_{2.11}y_{(t-2)} + a_{2.12}z_{(t-2)} + e_{1t}$$

$$z_t = a_{20} + a_{1.21}y_{(t-1)} + a_{1.22}z_{(t-1)} + a_{2.21}y_{(t-2)} + a_{2.22}z_{(t-2)} + e_{2t}$$

Tästä voidaan lisätä viivepituutta aina p :hen saakka, jotta saadaan 2 –muuttujan p –asteinen VAR –malli:

$$\mathbf{x}_t = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{x}_{(t-1)} + \mathbf{A}_2 \mathbf{x}_{(t-2)} + \dots + \mathbf{A}_p \mathbf{x}_{(t-p)} + \mathbf{e}_t \quad (14.9)$$

Nyt jos (14.9):ään laitetaan \mathbf{x}_t –vektoriksi $(x_1, x_2, \dots, x_n)'$, saadaan $n:n$ muuttujan p –asteinen VAR –malli (Enders, 301).

14.2. Estimoinnista

Kun tarkastellaan mallia (14.9), huomataan, kuinka paljon parametreja tarvitaan estimoida mihinkin malliin. Vektorissa \mathbf{A}_0 on n vakiotermiä ja jokaisessa \mathbf{A}_i –matriisissa on n^2 parametriä. Kun \mathbf{A}_i –matriiseja on yhteensä p , saadaan lopullisten parametrien määräksi $n+pn^2$. Box-Jenkins –lähestymistapa yksiulotteisissa tapauksissa painottaa ’säästäväisyyteen’ parametrien valinnassa, eli yritetään poistaa mallista kaikki parametrit, joiden ei katsota antavan merkittävää lisäinformaatiota lyhyen aikavälin ennusteeseen.

Sims(1980) on kritisoinut tätä käytäntöä varsinkin moniulotteisissa tapauksissa, joissa esiintyy palautetta eri yhtälöitten välillä. Simsin metodologian mukaan VAR –malleissa riittää testata sopivaa viivepituutta, ja kun se ollaan löydetty, ei ole enää syytä yrittää karsia parametrien määrää. Sims painottaa, että VAR –mallin todellinen päämäärä on ennemminkin löytää yhtälöiden välillä

vaikuttavat suhteet, kuin tehdä lyhyen aikavälin ennusteita. Huolimattomasti asetetut rajoitukset voivat hukata tärkeää informaatiota. Lisäksi selittäjät ovat usein kolinearisia, joten yksittäisiin estimaatteihin kohdistuvat t –testit eivät välttämättä ole oikeassa (Enders, 301).

Edellä oletettiin, että muuttujat olisivat stationaarisia. Yksiulotteisessa tapauksessa yksikköjuuren kohdalla aikasarja saatiin stationaariseksi differoimalla. VAR –mallien kohdalla on kuitenkin vallalla käsitys, että muuttujia ei pitäisi differoida siinäkään tapauksessa, että niillä olisi yksikköjuuri. Perusteluina sanotaan, kuten edellä, että differointi voi heittää hukkaan informaatiota (Doan, 283), ja mallin päätarkoitus ei kuitenkaan ole lyhyen aikavälin ennusteiden tekeminen. Toistaiseksi käsitellään muuttujia siten, kuin ne olisivat stationaarisia. Aiheeseen palataan myöhemmin, kun käsitellään yhteisintegraatiota ja virheenkorjausta.

Koska (14.9):n oikea puoli sisältää vain muuttujien tunnettuja arvoja, virhetermeissä ei ole autokorrelaatiota ja niiden varianssit ovat vakiot, voidaan jokainen systeemin yhtälö estimoida pienimmän neliösumman menetelmällä. PNS –estimaatit ovat konsistentteja ja asympotoottisesti tehokkaita, eli mikäli kaikissa yhtälöissä on samat selittäjät ja samat viivepituudet, PNS –estimointi yhtälö kerrallaan tuottaa saman tuloksen, kuin jos yhtälöt estimoitaisiin samanaikaisesti (Enders, 301). Tästä syystä VAR –mallit kannattaa rakentaa juuri siten, että yhtälöiden oikeat puolet ovat samoja.

14.3. Identifioinnista

Palataan takaisin yhtälöihin (14.1) ja (14.2). Näitä primitiivimuodon yhtälöitä ei voida estimoida suoraan. Syynä on se, että y_t korreloi u_{zt} :n kanssa ja toisin päin. PNS –estimoinnin oletuksissa virhetermin pitää olla riippumaton selittäjistä. Standardimuodon yhtälöissä (14.5) ja (14.6) vastaavaa ongelmaa ei kuitenkaan ole. Standardimuodosta voidaan laskea estimaatit A_0 :n kahdelle elementille ja A_1 :n neljälle elementille. Sen lisäksi, kun saadaan yhtälöiden residuaalit e_{1t} ja e_{2t} , voidaan laskea niiden varianssit ja kovarianssi. Standardimuodosta saadaan siis laskettua yhteensä yhdeksän parametria.

Kun taas tarkastellaan (14.1):stä ja (14.2):sta, huomataan, että primitiivimuodosta saadaan parametreja kymmenen (b_{10} , b_{20} , b_{12} , b_{21} , γ_{11} , γ_{12} ,

γ_{21} , γ_{22} ja virhetermien keskihajonnat σ_y ja σ_z). Näin ollen kaikkea primitiivimuodon sisältämää informaatiota ei voida ilmaista standardimuodon avulla. Tämä rajoite voidaan kuitenkin kiertää rajoittamalla primitiivimuotoa sopivalla tavalla. Yleisin tällainen rajoite on Simsin(1980) esittämä rekursiivinen lähestymistapa. Asetetaan primitiivimuodon parametriksi b_{21} nolla(Enders s.302). Näin saadaan (14.1) ja (14.2) seuraavaan muotoon:

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{(t-1)} + \gamma_{12}z_{(t-1)} + u_{yt} \quad (14.10)$$

$$z_t = b_{20} + \gamma_{21}y_{(t-1)} + \gamma_{22}z_{(t-1)} + u_{zt} \quad (14.11)$$

Nyt \mathbf{B}^{-1} -matriisi on

$$\begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}$$

Nyt primitiivimuodon kertominen vasemmalta rajoitetulla \mathbf{B}^{-1} :llä antaa

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} - b_{12}b_{20} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} - b_{12}\gamma_{22} & \gamma_{12} - b_{12}\gamma_{22} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{(t-1)} \\ z_{(t-1)} \end{bmatrix} + \begin{bmatrix} u_{yt} - b_{12}u_{zt} \\ u_{zt} \end{bmatrix} \quad (14.12)$$

Nyt kun (14.12):n ja standardimuodon parametrit asetetaan yhtäsuuriksi, saadaan

$$a_{10} = b_{10} - b_{12}b_{20}$$

$$a_{11} = \gamma_{11} - b_{12}\gamma_{22}$$

$$a_{12} = \gamma_{12} - b_{12}\gamma_{22}$$

$$a_{20} = b_{20}$$

$$a_{21} = \gamma_{21}$$

$$a_{22} = \gamma_{22}$$

Koska $e_{1t} = u_{yt} - b_{12}u_{zt}$ ja $e_{2t} = u_{zt}$, saadaan kovarianssimatriisin parametrit

$$\sigma_1^2 = \sigma_y^2 + b_{12}^2 \sigma_z^2$$

$$\sigma_2^2 = \sigma_z^2$$

$$\sigma_{12} = -b_{12} \sigma_z^2$$

Näin ollaan saatu yhdeksän standardimuodon parametriestimaattia vastaamaan yhdeksää (yksi kymmenestä siis rajoitettiin nolaksi) primitiivimuodon parametria. Nyt myös residuaalien u_{yt} ja u_{zt} estimaatit voidaan ratkaista. z_t :n residuaalit ovat yksinkertaisesti sarja e_{2t} ja u_{yt} saadaan yhtälöstä $e_{1t} = u_{yt} - b_{12} u_{zt}$.

Kun tarkastellaan rajoitusta, jonka mukaan $b_{21}=0$, huomataan, ettei y_t :llä ole välitöntä vaikutusta z_t :hen. Sekä u_{yt} että u_{zt} –sokit vaikuttavat välittömästi y_t :hen, mutta vain u_{zt} –sokit vaikuttavat välittömästi z_t :hen. Tällaista residuaalien dekomponointia kutsutaan Choleskin hajotelmaksi. Yhtä hyvin Choleskin hajotelmana olisi voitu käyttää rajoitetta $b_{12}=0$. (Enders, 302,303.)

14.4. Impulssivastefunktio

Aivan kuten yksiulotteisessa tapauksessa, on myös vektoriarvoisella autoregressiivisellä prosessilla liukuvan keskiarvon esitysmuoto (VMA) (Enders s. 305). VMA –muoto ilmaisee muuttujat niiden nykyisten ja menneiden sokkien avulla. VMA –esitys on oleellisessa osassa Simsin (1980) metodologiassa, jossa tutkitaan sokkien vaikutusta muuttujien aikauriin. Käytetään jälleen kahden muuttujan ja yhden viiveen VAR –mallia (Kanto & Kahra s. 55, 56):

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{(t-1)} \\ z_{(t-1)} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

Jos μ_y ja μ_z ovat y_t :n ja z_t :n keskiarvot, saadaan iteroimalla prosessia taaksepäin samoin kuin yksiulotteisessa tapauksessa

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_z \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} e_{1(t-i)} \\ e_{2(t-i)} \end{bmatrix} \quad (14.13)$$

Nyt muuttujat ovat ilmaistu sokkien e_{1t} ja e_{2t} avulla. Käyttökelpoisempaa on ilmaista ne u_{yt} :n ja u_{zt} :n avulla. Käyttämällä (14.7) ja (14.8) saadaan (14.13):sta (Enders, 305, 306):

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_z \end{bmatrix} + \frac{1}{1 - b_{12}b_{21}} \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} u_{yt} \\ u_{zt} \end{bmatrix} \quad (14.14)$$

Selventämiseksi määritellään matriisi ϕ_i seuraavasti:

$$\phi_i = \frac{A_1^i}{1 - b_{12}b_{21}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix}$$

, jolloin (14.14):stä saadaan

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_z \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} u_{y(t-i)} \\ u_{z(t-i)} \end{bmatrix} \quad (14.15)$$

Liukuvan keskiarvon esitysmuoto on erittäin hyödyllinen työkalu, kun halutaan tutkia muuttujien välisiä vuorovaikutussuhteita. Matriisin ϕ_i parametreja voidaan käyttää sokkien u_{yt} ja u_{zt} vaikutusten selvittämiseen. ϕ_0 :n neljä elementtiä ovat vaikutuskertoimia, joista esimerkiksi $\phi_{21}(0)$ ja $\phi_{22}(0)$ ovat u_{yt} :n ja u_{zt} :n yhden yksikön suuruisten muutosten samanaikainen vaikutus z_t :hen. Vastaavasti $\phi_{21}(1)$ ja $\phi_{22}(1)$ ovat $u_{y(t-1)}$:n ja $u_{z(t-1)}$:n yhden yksikön suuruisten muutosten vaikutus z_t :hen. Yksikkömuutosten kumulatiiviset vaikutukset saadaan summaamalla sopivat vaikutuskertoimet. Neljää kerrointa $\phi_{11}(i)$, $\phi_{12}(i)$, $\phi_{21}(i)$ ja $\phi_{22}(i)$ sanotaan impulssivastefunktioksi. Piirtämällä impulssivastefunktioiden kuvaajat saadaan hyvä kuva sokkien vaikutuksista sarjojen aikauriin. (Enders, 309.)

Koska kaikkia rajoittamattoman primitiivimuodon parametreja ei voida laskea standardimuodosta, joudutaan jälleen rajoittaman primitiivimuotoa esimerkiksi edellä kuvatulla Choleskin hajotelmalla. Koska sokilla u_{zt} on

samanaikainen vaikutus sekä y_t :hen että z_t :hen, mutta u_{y_t} :llä samanaikainen vaikutus rajoittuu y_t :hen, tulee malliin epäsymmetriaa. Tämän takia Choleskyn hajotelman järjestys riippuu e_{1t} :n ja e_{2t} :n välisestä korrelaatiokertoimesta. Olkoon korrelaatiokerroin

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1\sigma_2} = \frac{\text{Cov}(e_{1t}, e_{2t})}{\sqrt{\text{Var}(e_{1t})}\sqrt{\text{Var}(e_{2t})}} = 0$$

Tällöin on oltava $\text{Cov}(e_{1t}, e_{2t}) = -b_{12}\sigma_z^2 = 0$. Näin ollen $e_{1t} = u_{y_t}$ ja $e_{2t} = u_{z_t}$, eikä järjestyksellä ole väliä. Jos taas $\rho_{12} = 1$, tulee systeemiin sokki, joka samanaikaisesti vaikuttaa molempiin muuttujiin, eli sokeista tulee identtiset. Näin ollen jos Choleskyn hajotelmaksi valitaan $b_{21}=0$, on $e_{1t} = e_{2t} = u_{z_t}$. Jos taas hajotelmaksi valitaan $b_{12}=0$, saadaan $e_{1t} = e_{2t} = u_{y_t}$. Yleensä käytetään nyrkkisääntöä, jos $|\rho_{12}| > 0.2$, korrelaation katsotaan olevan merkittävä. Jos näin on, suositellaan laskettavaksi impulssivastefunktiot kummallakin järjestyksellä. Jos tulokset poikkeavat dramaattisesti toisistaan, on syytä tarkistella muuttujien välisiä relaatioita syvällisemmin.

14.5. Varianssin dekomponointi

Kuten edellä on jo todettu, vektoriautoregressiiviset mallien pääasiallinen tarkoitus ei ole luoda lyhyen aikavälin ennusteita. VAR –malleilla pystytään kuitenkin tarkastelemaan ennustevirheiden ominaisuuksia, josta on hyötyä, kun pyritään selvittämään muuttujien välisiä suhteita.

Tarkastellaan jälleen ensimmäisen asteen VAR –mallia, josta halutaan laskea ehdolliset ennusteet $\mathbf{x}_{(t+i)}$, eli päivitetään standardimuodon yhtälöä (14.4) yhdellä:

$$\mathbf{x}_{(t+1)} = \mathbf{A}_0 + \mathbf{A}_1\mathbf{x}_t + \mathbf{e}_{(t+1)}$$

Nyt joudutaan ottamaan ehdollinen odotusarvo:

$$E(\mathbf{x}_{(t+1)}|\mathbf{x}_t) = \mathbf{A}_0 + \mathbf{A}_1\mathbf{x}_t$$

Näin yhden periodin ennustevirhe on $\mathbf{x}(t+1) - E(\mathbf{x}(t+1)|\mathbf{x}_t) = \mathbf{e}_{(t+1)}$. Kahden periodin ennuste saadaan, kun päivitetään standardimuotoa taas yhdellä:

$$\mathbf{x}_{(t+2)} = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{x}_{(t+1)} + \mathbf{e}_{(t+2)} = \mathbf{A}_0 + \mathbf{A}_1 (\mathbf{A}_0 + \mathbf{A}_1 \mathbf{x}_t + \mathbf{e}_{(t+1)}) + \mathbf{e}_{(t+2)}$$

Nyt otetaan ehdollinen odotusarvo ja saadaan:

$$E(\mathbf{x}_{(t+2)}|\mathbf{x}_t) = (\mathbf{I} + \mathbf{A}_1) \mathbf{A}_0 + \mathbf{A}_1^2 \mathbf{x}_t$$

Näin kahden periodin ennustevirhe on $\mathbf{e}_{(t+2)} + \mathbf{A}_1 \mathbf{e}_{(t+1)}$. Nyt saadaan helposti yleistys n :n periodin ennusteelle ja sen virheelle. Ennuste on:

$$E(\mathbf{x}_{(t+n)}|\mathbf{x}_t) = (\mathbf{I} + \mathbf{A}_1 + \mathbf{A}_1^2 + \dots + \mathbf{A}_1^{(n-1)}) \mathbf{A}_0 + \mathbf{A}_1^n \mathbf{x}_t$$

ja ennustevirhe on

$$\mathbf{e}_{(t+n)} + \mathbf{A}_1 \mathbf{e}_{(t+n-1)} + \mathbf{A}_1^2 \mathbf{e}_{(t+n-2)} + \dots + \mathbf{A}_1^{(n-1)} \mathbf{e}_{(t+1)}$$

On myös mahdollista ilmaista ennustevirheet liukuvan keskiravon muodossa (14.15), eli \mathbf{u}_t :n avulla. Tarkastellaan ehdollista n :n periodin ennustetta yhtälöön (14.15):

$$\mathbf{x}_{(t+n)} = \boldsymbol{\mu}_t + \sum_{i=0}^{\infty} \boldsymbol{\phi}_i \mathbf{u}_{(t+n-i)}$$

Tästä saadaan n :n periodin ennustevirhe, joka on

$$\mathbf{x}_{(t+n)} - E(\mathbf{x}_{(t+n)} | \mathbf{x}_t) = \sum_{i=0}^{n-1} \boldsymbol{\phi}_i \mathbf{u}_{(t+n-i)} \quad (14.16)$$

Keskitytään yhteen muuttajaan vektorissa \mathbf{x}_t , olkoon se y_t . y_t :n n :n periodin ennustevirhe (14.16) mukaan on

$$y_{(t+n)} - E(y_{(t+n)} | y_t) = \phi_{11}(0)u_{y(t+n)} + \phi_{11}(1)u_{y(t+n-1)} + \dots + \phi_{11}(n-1)u_{y(t+1)} \\ + \phi_{12}(0)u_{z(t+n)} + \phi_{12}(1)u_{z(t+n-1)} + \dots + \phi_{12}(n-1)u_{z(t+1)}$$

Näin saadaan n :n periodin ennustevarianssi y_t :lle:

$$\sigma_y(n)^2 = \sigma_y^2 [\phi_{11}(0)^2 + \phi_{11}(1)^2 + \dots + \phi_{11}(n-1)^2] \\ + \sigma_z^2 [\phi_{12}(0)^2 + \phi_{12}(1)^2 + \dots + \phi_{12}(n-1)^2]$$

Koska kaikki $\phi_{jk}(i)$:n arvot ovat ei-negatiivisia, ennustevirheen varianssi kasvaa, kun n kasvaa. Ennustevarianssi on kuitenkin mahdollista hajottaa siten että saadaan $\sigma_y(n)^2$:n suhde sokkeihin u_{yt} ja u_{zt} seuraavasti. (Enders, 310, 311.)

$$\frac{\sigma_y^2 [\phi_{11}(0)^2 + \phi_{11}(1)^2 + \dots + \phi_{11}(n-1)^2]}{\sigma_y(n)^2}$$

$$\frac{\sigma_z^2 [\phi_{12}(0)^2 + \phi_{12}(1)^2 + \dots + \phi_{12}(n-1)^2]}{\sigma_y(n)^2}$$

Tämä varianssin dekomponointi kertoo, missä määrin sarjan ennustevirheen varianssin vaihtelu johtuu sarjan omista sokeista ja missä määrin muiden sarjojen sokeista. Jos u_{zt} –sokit eivät pysty selittämään yhtään y_t :n ennusteen virhevariانسsia, sanotaan, että y_t on eksogeeninen muuttuja. Tällöin y_t –sarja kehittyy täysin riippumattomana z_t –sarjasta ja sen sokeista. Jos taas tilanne olisi toisinpäin, eli u_{zt} –sokit selittäisivät 100% y_t :n ennustevirheen varianssista, sanottaisiin y_t :n olevan kokonaan endogeeninen.

Varianssin dekomponointiin sisältyy sama ongelma kuin impulssivastefunktioon. Standardimuodosta ei voida saada parametrimatriisia \mathbf{B} rajoittamatta kaikkea primitiivimuodon informaatiota, jolloin joudutaan käyttämään Choleskyn hajotelmaa. Tämä ilmenee siinä, että jos käytetään rajoitusta $b_{21}=0$, selittää u_{zt} 100% z_t :n yhden periodin ennustevirheen varianssista. Tämä efekti kuitenkin hiipuu pois, kun n kasvaa. Suurilla n varianssien dekomponoinnin

pitäisi konvergoitua. Lisäksi, jos muuttujien välinen korrelaatiokerroin poikkeaa merkittävästi nolasta, tulisi dekomponoinnit tehdä käyttäen eri järjestystä.

14.6. Granger –kausaalisuus

Vektoriautoregressivisissä malleissa tulee usein esille kysymys, kuinka hyödyllinen jokin muuttuja z_t on ennustettaessa y_t :tä. Granger –kausaalisuuden käsite pyrkii osaltaan vastaamaan tähän kysymykseen. Mikäli muuttuja z_t ei voi auttaa muuttujan y_t ennustamisessa, niin silloin z_t ei ole Granger –kausaalinen muuttujan y_t suhteen. Muodollisemmin z_t ei ole Granger -kausaalinen y_t :n suhteen, jos kaikilla $s > 0$, ennusteen $y_{(t+s)}$ keskineliövirhe on sama, olivatpa ennusteet perustuneet pelkästään y_t :n menneisiin arvoihin tai sitten sekä y_t :n että z_t :n menneisiin arvoihin. (Hamilton, 303.)

Granger –kausaalisuutta voidaan testata erilaisilla menetelmillä, mutta Hamilton suosittelee seuraavaa kaksiulotteisessa tapauksessa: Muodostetaan lineaarinen malli 1 y_t :lle, jossa y_t :tä selitetään sen omilla sekä z_t :n viiveillä aina p :hen viiveeseen saakka.

$$y_t = c_1 + \alpha_1 y_{(t-1)} + \dots + \alpha_p y_{(t-p)} + \beta_1 z_{(t-1)} + \dots + \beta_p z_{(t-p)} + u_{1t}$$

Testattava hypoteesi on $\beta_1 = \beta_2 = \dots = \beta_p = 0$. Seuraavaksi muodostetaan malli 0, jossa y_t :tä selitetään pelkästään sen omilla viiveillä:

$$y_t = c_0 + \alpha_1 y_{(t-1)} + \dots + \alpha_p y_{(t-p)} + u_{0t}$$

Merkitään mallin 0 antamien residuaalien neliösummaa RSS_0 :lla ja mallin 1 antamien residuaalien neliösummaa RSS_1 :llä. Nyt voidaan testata testisuuretta

$$S_1 = \frac{(RSS_0 - RSS_1) / p}{RSS_1 / (T - 2p - 1)}$$

, joka noudattaa $F(p, T-2p-1)$ –jakaumaa. Nollahypoteesinahan oli, että z_t : lla ei ole Granger –kausaalista vaikutusta y_t :hen.

14.7. Viiveitten lukumäärä

Aivan kuten yksiulotteisessa tapauksessa, myös vektoriautoregressiivisessä tapauksessa oikean viivepituuden löytäminen on oleellinen osa mallin rakentamista. VAR –mallien tapauksessa on erityisesti syytä ottaa huomioon, että kun viivepituus (ja muuttujien määrä) kasvaa, kasvaa myös estimoitavien parametrien määrä kappaleessa 14.2 kuvaillun mukaisesti, jolloin menetetään vapausasteita. Näin ollen on tärkeää sisällyttää malliin vain ne viivepituudet, joiden katsotaan olevan merkittäviä mallin kannalta, jottei menetettäisi enää turhia vapausasteita.

Etsittäessä sopivaa viivepituutta aloitetaan pisimmästä 'järjellisestä' viivepituudesta ja verrataan sitä valittuun lyhyempään viivepituuteen. Tämä tapahtuu siten, että muodostetaan ensin rajoittamaton malli, jossa on pisin 'järjellinen' viivepituus, ja muodostetaan sen residuaalien kovarianssimatriisi Σ_u . Seuraavaksi muodostetaan uusi, rajoitettu, malli, jota vastaan rajoittamatonta mallia tarkastellaan. Merkitään tämän mallin residuaalien kovarianssimatriisia Σ_r . Huomioitavaa on, että rajoitettu malli on estimoitava käyttämällä samaa otosväliä kuin rajoittamatonkin. Nyt halutaan selvittää, onko rajoitettu malli mahdollisesti parempi kuin rajoittamaton. Tätä rajoitusta testataan LR – testillä siten, että testisuure on: (Doan, 287.)

$$(T - c)(\log|\Sigma_r| - \log|\Sigma_u|)$$

, missä T = käytettävien havaintojen lukumäärä

c = estimoitavien parametrien lukumäärä rajoittamattoman mallin yhtälössä.

Tämä testisuure noudattaa asymptoottisesti X^2 –jakaumaa vapausasteinaan rajoitteiden määrä mallissa. Testatessa nollahypoteesina on, että rajoitetun mallin viivepituudet käyvät aineistoon. Jos ensimmäisessä testissä H_0 hyväksytään, tehdään toinen testaus, jossa vähennetään viivepituutta edelleen, ja nyt äskeisestä rajoitetusta mallista tulee uusi rajoittamaton malli.

Enders kuitenkin kehottaa varovaisuuteen, kun viiveitä lyhennetään tähän malliin. Koska joka vähennysvaiheessa menetetään hieman selittävyttä, voi se kasvaa merkittäväksi, jolloin nollahypoteesia ei välttämättä hylätä, vaikka näin pitäisikin tapahtua. Käytännössä tämä voi ilmetä siten, että tarkasteltaessa H_0 : 12 viivettä vs. 8 viivettä hyväksytään samoin kuin H_0 : 8 viivettä vs. 6, mutta kuitenkin H_0 : 12 vs. 6 viivettä hylätään. (Enders, 314).

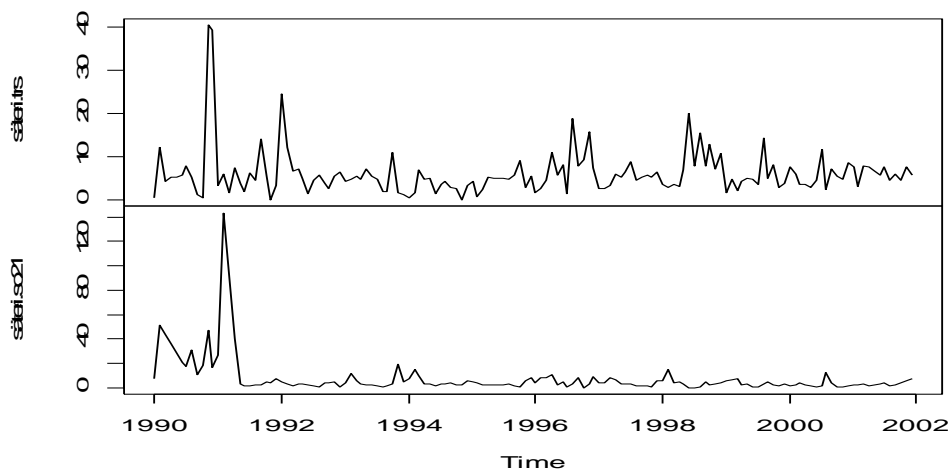
15. SO₂:n ja TRS:n yhteisvaikutukset

Tässä kappaleessa rakennetaan vektoriautoregressiivinen malli kahdelta eri tuulisektorilta tuleville trs –päästöille, josta on mukana rikkidioksidin vaikutus. Kuten kappaleessa 14 mainittiin, analyysin tehtävänä ei ole ennustaa tulevia päästöjä, vaan tutkia näiden kahden saastuttajan yhteisiä aikauria.

15.1. Säterin suunnalta tulevat päästöt

Kappaleen 13.2.1 tapaan tuulisektori tälle tutkimukselle on 17-54. Kaikki tämän kappaleen analyysit on tehty RATS –ohjelmalla, ellei toisin mainita. Aineistoa ei ole millään tavalla muutettu, eli mahdolliset yksikköjuuret ovat aineistossa mukana. Yksikköjuuria käsitellään myöhemmin, kun tutkitaan näiden kahden saastuttajan mahdollista yhteisintegraatiota. Kuviossa **15.1.** ovat sekä trs- että SO₂:n kuvaajat.

Kuvio 15.1



15.1.1. Viivepituuden määrittäminen

Ensin määritellään sopiva viivepituus p VAR –mallille, joka on määrätty seuraavasti:

$$\mathbf{x}_t = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{x}_{(t-1)} + \mathbf{A}_2 \mathbf{x}_{(t-2)} + \dots + \mathbf{A}_p \mathbf{x}_{(t-p)} + \mathbf{e}_t$$

, missä $\mathbf{x}_t = \{trs, so2l\}$ '

, $\mathbf{A}_0 = (1 \times 2)$ –vektori

ja $\mathbf{A}_1 \dots \mathbf{A}_p = (2 \times 2)$ –matriiseja, kuten edellä ollaan määrätty.

Tutkitaan ensin moniulotteisia informaatiokriteereitä AIC:tä ja BIC:tä. Nämä on laskettu, kuten Doan on esittänyt. (Doan, 288,289)

$$AIC = -0.5 * T * \log|\Sigma| - 2N$$

$$BIC = -0.5 * T * \log|\Sigma| - N * \log(T)$$

, missä $T =$ havaintojen lukumäärä ja $N = n(np+1) = n^2p + n$

Informaatiokriteerien idea perustuu Akaiken vuonna 1979 esittämään malliin. Mallin parametrien hyvyyttä voidaan tutkia logaritmoidulla uskottavuusfunktiolla (Huuhtanen & Kallinen, 184). Kun esitellään keskiarvon logaritmoitu uskottavuusfunktio (mean expected log likelihood function), huomataan, että logaritmoitu uskottavuusfunktio on harhainen estimaattori keskiarvon suhteen logaritmoidulle uskottavuusfunktiolle. Tutkimalla harhan ja mallin vapaitten parametrien määrää, Akaike päätyi edellä esitettyyn informaatiokriteeriin. (Sakamoto, Ishiguro & Kitagawa, 56-57.). Akaiken informaatiokriteeri suunniteltiin alunperin yksiulotteiseen tapaukseen, mutta sitä voi soveltaa myös moniulotteisiin tapauksiin, kuten edellä. Tutkitaan viiveitä kahden vuoden ajalta, eli 24 viiveeseen saakka. Tulokset ovat taulukossa **15.1**.

Taulukko 15.1.

Viive	AIC	BIC
1	-608.3017	-626.0787
2	-580.4804	-610.0387
3	-568.8897	-610.1723
4	-570.3937	-623.3433
5	-572.0089	-636.5673
6	-573.6877	-649.7963
7	-576.4390	-664.0384
8	-578.2139	-677.2442
9	-580.3075	-690.7079
10	-583.7767	-705.4860
11	-549.7790	-682.7350
12	-522.5757	-666.7158
13	-504.7590	-660.0196
14	-397.3316	-563.6486
15	-396.2892	-573.5975
16	-399.4531	-587.6871
17	-402.3770	-601.4701
18	-401.1946	-611.0794
19	-401.9151	-622.5236
20	-406.0722	-637.3353
21	-404.1466	-645.9944
22	-402.1205	-654.4824
23	-404.7996	-667.6039
24	-406.7724	-679.9466

AIC pysyy suhteellisen samana viiveeseen 10 saakka, minkä jälkeen arvot nousevat nopeasti yli sadalla. BIC:n pienin arvo sattuu viiveen 10 kohdalle. Näiden kriteerien valossa viivepituutta ei kannata pistää yli vuodeksi.

Tutkitaan seuraavaksi viivepituutta LR –testillä. Valitaan ensin rajoittamattoman mallin viivepituudeksi 24, eli oletetaan, että kahden vuoden ajanjaksolla saadaan kuva systeemin dynamiikasta. Tällöin estimoitavia parametreja olisi $n+pn^2=2+24*2^2=98$ kappaletta. Tutkitaan vaihtoehtoa, että pudotettaisiin puoli vuotta viiveistä pois, eli saataisiin estimoitavien parametrien määräksi $2+18*2^2=78$. LR –testisuureen arvoksi saadaan 3.907652, joka $X^2(72)$ – jakaumassa saa arvon 0.99999934, eli H_0 jää voimaan, jolloin uudeksi rajoittamattoman mallin viivepituudeksi saadaan 18. Kun jatkettiin samaan malliin viivepituuksien pudottamista, saatiin taulukossa **15.2** olevat arvot hypoteeseille.

Taulukko 15.2

Viiveet	Testisuure	P –arvo
24 vs 18	28.029282	0.99999934
18 vs 12	20.846120	0.99978158
12 vs 6	41.634381	0.01417549
12 vs 8	34.138586	0.36520005
8 vs 6	6.022807	0.99992606

Eli kävi, kuten mainittiin kappaleessa 14.7. 12 viivettä olisi parempi kuin kuusi, mutta kahdeksan parempi kuin 12 ja kuusi parempi kuin kahdeksan. Nyt siis valitaan viivepituudeksi 12, eli yksi vuosi pitää sisällään tarvittavan systeemin dynamiikan.

Tutkitaan vielä vakiota mallissa, eli voidaanko malli kirjoittaa muodossa

$$\mathbf{x}_t = \mathbf{A}_1 \mathbf{x}_{(t-1)} + \mathbf{A}_2 \mathbf{x}_{(t-2)} + \dots + \mathbf{A}_{12} \mathbf{x}_{(t-12)} + \mathbf{e}_t$$

Muodostetaan moniulotteiset AIC ja BIC edellä mainitulla tavalla. Vakiotermississä mallissa arvot ovat -522.57569 ja -666.71578, kun taas mallissa, jossa vakio termiä ei ole, ovat arvot -530.27788 ja -674.41798. Näin ollen siis valitaan vakiotermitön malli. Taulukoissa **15.3a** ja **15.3b** ovat kaikki

PNS:llä lasketun mallin kertoimet keskivirheineen, t -arvoineen ja merkitsevyystasoinen.

Taulukko 15.3 (trs) selitettävänä)

Parametri	Estimaatti	Keskivirhe	T -arvo	P -arvo
1. TKTRS{1}	0.045717517	0.077622893	0.58897	0.55711163
2. TKTRS{2}	0.093177650	0.069126636	1.34793	0.18050244
3. TKTRS{3}	0.127066413	0.075009506	1.69400	0.09314660
4. TKTRS{4}	0.210071387	0.081400032	2.58073	0.01120086
5. TKTRS{5}	0.114874647	0.080649629	1.42437	0.15722344
6. TKTRS{6}	0.054234703	0.081694852	0.66387	0.50818873
7. TKTRS{7}	-0.123182648	0.083638119	-1.47281	0.14371202
8. TKTRS{8}	0.004803988	0.084090497	0.05713	0.95454816
9. TKTRS{9}	-0.040948048	0.083304566	-0.49155	0.62403766
10. TKTRS{10}	0.157607728	0.082939494	1.90027	0.06006377
11. TKTRS{11}	0.194790396	0.083816715	2.32400	0.02199675
12. TKTRS{12}	0.148595694	0.078780641	1.88620	0.06195483
13. SO21{1}	-0.108083325	0.041709282	-2.59135	0.01088093
14. SO21{2}	-0.000346708	0.046027377	-0.00753	0.99400377
15. SO21{3}	-0.045683539	0.045926750	-0.99470	0.32210349
16. SO21{4}	0.032556086	0.046437157	0.70108	0.48476193
17. SO21{5}	0.044976469	0.045973859	0.97831	0.33010970
18. SO21{6}	0.023422436	0.045256035	0.51755	0.60582806
19. SO21{7}	0.021445872	0.044470206	0.48225	0.63060246
20. SO21{8}	-0.113752641	0.044348400	-2.56498	0.01169093
21. SO21{9}	-0.056332715	0.044272782	-1.27240	0.20596346
22. SO21{10}	-0.002479068	0.039849142	-0.06221	0.95050953
23. SO21{11}	0.133237692	0.037390659	3.56340	0.00054642
24. SO21{12}	-0.039420348	0.028322521	-1.39184	0.16683142

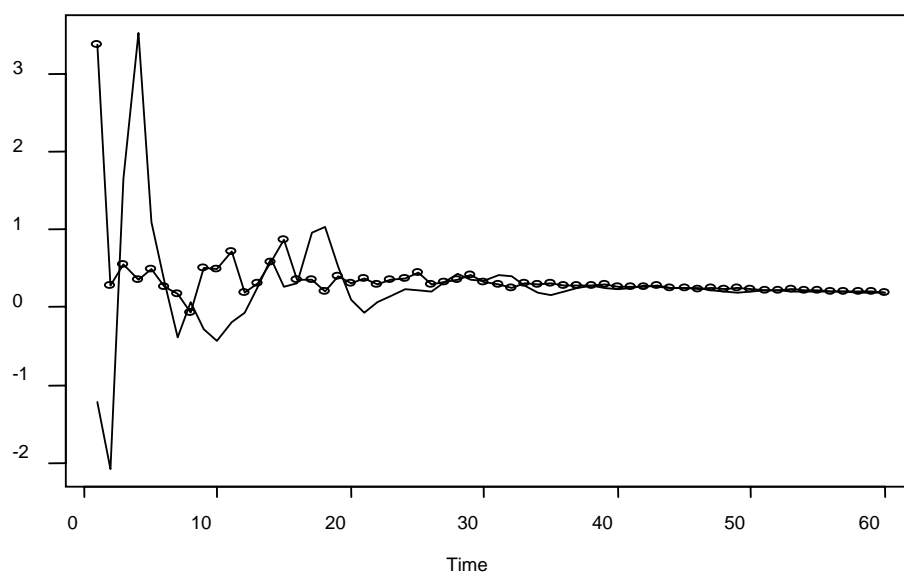
Taulukko 15.3b ({so21} selitettävänä)

Parametri	Estimaatti	Keskivirhe	T -arvo	P -arvo
1. TKTRS{1}	-0.418012478	0.178678943	-2.33946	0.02115040
2. TKTRS{2}	0.848477029	0.159121538	5.33226	0.00000054
3. TKTRS{3}	0.819768019	0.172663226	4.74779	0.00000635
4. TKTRS{4}	-0.298940916	0.187373479	-1.59543	0.11353875
5. TKTRS{5}	-0.332459021	0.185646139	-1.79082	0.07612338
6. TKTRS{6}	-0.427138451	0.188052122	-2.27138	0.02510605
7. TKTRS{7}	0.119153985	0.192525298	0.61890	0.53728523
8. TKTRS{8}	-0.102761085	0.193566619	-0.53088	0.59659041
9. TKTRS{9}	0.028448416	0.191757497	0.14836	0.88233830
10. TKTRS{10}	0.055080903	0.190917144	0.28851	0.77351178
11. TKTRS{11}	-0.092181883	0.192936404	-0.47778	0.63376972
12. TKTRS{12}	0.067959509	0.181343943	0.37475	0.70857772
13. SO21{1}	0.546445851	0.096009954	5.69155	0.00000011
14. SO21{2}	-0.027959896	0.105949710	-0.26390	0.79236156
15. SO21{3}	0.176037007	0.105718078	1.66516	0.09878020
16. SO21{4}	-0.035031425	0.106892975	-0.32772	0.74375458
17. SO21{5}	-0.094681398	0.105826518	-0.89468	0.37294509
18. SO21{6}	0.155595798	0.104174170	1.49361	0.13819289
19. SO21{7}	-0.125822889	0.102365281	-1.22916	0.22168468
20. SO21{8}	0.079103641	0.102084899	0.77488	0.44010327
21. SO21{9}	-0.090836940	0.101910836	-0.89134	0.37472968
22. SO21{10}	0.001018045	0.091728126	0.01110	0.99116533
23. SO21{11}	0.112395135	0.086068983	1.30587	0.19437055
24. SO21{12}	0.009581753	0.065195176	0.14697	0.88342949

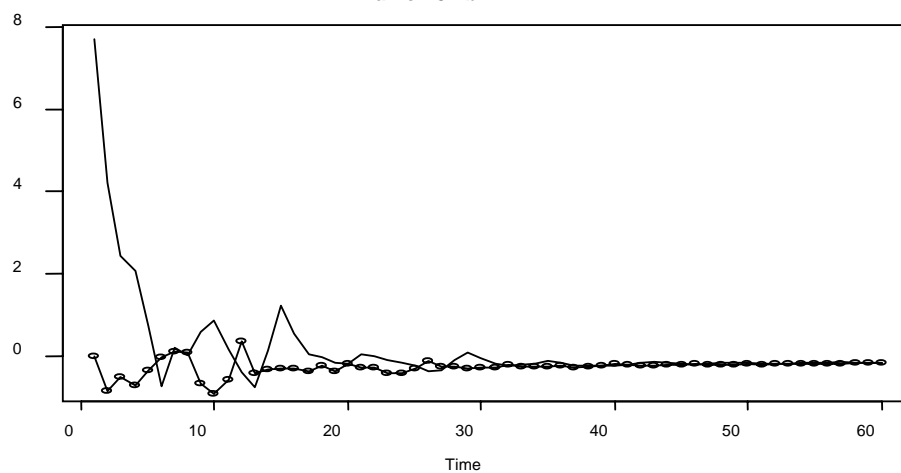
15.1.2. Innovaatioiden laskenta

Kuvioissa **15.2a** ja **15.2b** ovat graafiset esitykset impulssivastefunktioista edellä lasketulle vektori-autoregressiiviselle mallille siten että kuviossa **15.2a** on sokkien vaikutus $\{trs\}$ -sarjaan ja kuviossa **15.2b** taas sokkien vaikutus $\{so2I\}$ -sarjaan. Kuvioissa $\{so2I\}$ -sokkien vaikutukset on merkitty yhtenäisellä viivalla, kun $\{trs\}$ -sokkien huippuja on merkitty ympyröillä.

Kuvio 15.2a



Kuvio 15.2b



Kuvioista näkyy, kuinka molempien sarjojen sokit vaikuttavat aluksi hyvin epäsäännöllisesti itseensä ja toisiinsa, mutta asettuvat samoilte aikaurille vasta noin kahden vuoden jälkeen. Kuvioista näkyy myös Choleskin hajotelman vaikutus, eli välittömät sokit $\{trs\}$ –sarjassa eivät vaikuta $\{so2l\}$ –sarjaan. Molempiin sarjoihin vaikuttaa luonnollisesti oma sokki eniten lyhyellä aikavälillä, muuten vaihtelut näyttäisivät olevan aika satunnaisia.

Taulukoissa **15.4a** ja **15.4b** ovat varianssin dekomponoinnit 24 kuukauden ajalta siten että taulukossa **15.4a** ovat varianssin dekomponoinnit $\{trs\}$ –sarjalle ja taulukossa **15.4b** $\{so2l\}$ –sarjalle. Arvot konvergoivat, kun $t > 24$, joten taulukossa ei ole pitempää aikaväliä (arvoja oli laskettu 60 kuukauden ajalta).

Taulukko 15.4a

Askel	Keskivirhe	$\{trs\}$	$\{so2l\}$
1	3.384308387	100.000	0.000
2	3.496732479	94.343	5.657
3	3.574744267	92.669	7.331
4	3.663432785	89.202	10.798
5	3.712424129	88.595	11.405
6	3.721413499	88.644	11.356
7	3.726874648	88.594	11.406
8	3.728477360	88.546	11.454
9	3.818612284	86.174	13.826
10	3.957232200	81.806	18.194
11	4.060357838	80.800	19.200
12	4.081465507	80.170	19.830
13	4.114337527	79.462	20.538
14	4.168445994	79.399	20.601
15	4.268520460	79.903	20.097
16	4.293901100	79.644	20.356

17	4.324283331	79.185	20.815
18	4.334651103	79.014	20.986
19	4.367272389	78.649	21.351
20	4.382428289	78.604	21.396
21	4.406502120	78.480	21.520
22	4.424372726	78.305	21.695
23	4.456340958	77.809	22.191
24	4.489922884	77.302	22.698

Taulukko 15.4b

Askel	Keskivirhe	{trs}	{so21}
1	7.79028734	2.439	97.561
2	9.09355520	7.019	92.981
3	9.55612006	9.336	90.664
4	10.39508669	19.429	80.571
5	10.47874793	20.226	79.774
6	10.50890765	20.207	79.793
7	10.51811940	20.305	79.695
8	10.51832821	20.308	79.692
9	10.53946097	20.299	79.701
10	10.58402854	20.300	79.700
11	10.58694430	20.319	79.681
12	10.59450281	20.295	79.705
13	10.62430490	20.246	79.754
14	10.64405627	20.523	79.477
15	10.71708066	20.304	79.696
16	10.73521264	20.317	79.683
17	10.77755061	20.940	79.060
18	10.82703095	21.661	78.339

19	10.84105164	21.842	78.158
20	10.84303493	21.841	78.159
21	10.84333272	21.843	78.157
22	10.84358539	21.847	78.153
23	10.84481541	21.859	78.141
24	10.84847668	21.891	78.109

Varianssin dekomponointi tuottaa melko samanlaiset tulokset molemmille saastuttajille. Yhteenvetona voitaisiin sanoa, että kukin saastuttaja selittää omasta ennustevarianssistaan noin neljä viidesosaa, kun yksi viidesosa selittyy toisella.

15.1.3. Granger –kausaalisuus

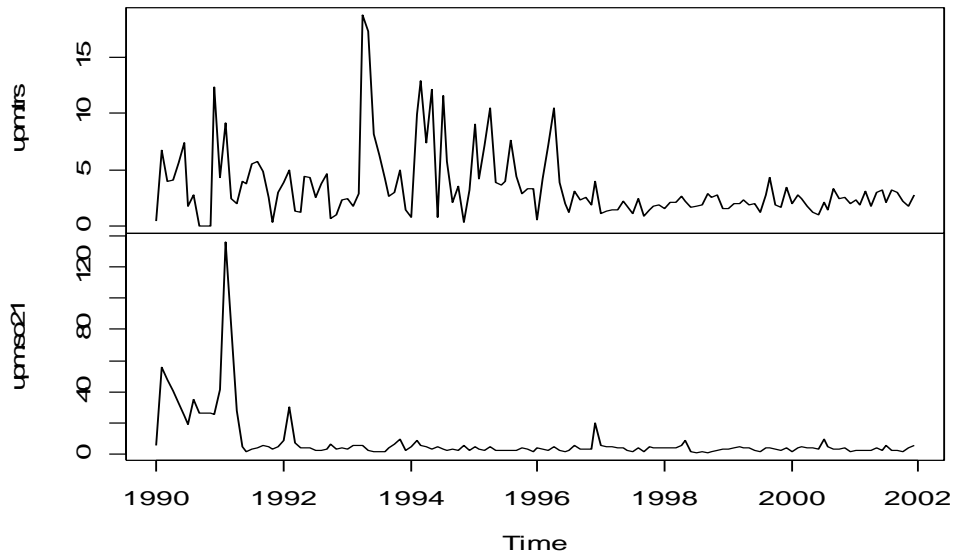
Granger –kausaalisuuden testaaminen suoritetaan Hamiltonin suosittamalla menetelmällä, joka on kuvailtu kappaleessa 14.6. Tarkastellaan ensin, onko $\{so2I\}$ Granger –kausaalinen $\{trs\}$:n suhteen. Rakennetaan lineaariset mallit, joissa toisessa $\{trs\}$:ää selitetään sekä sen omilla että $\{so2I\}$:n viiveillä, ja toisessa pelkästään omilla viiveillä. Näin ollen jäännöseliösummiksi tulee $SSR_1=1348.5128280$ ja $SSR_0=1835.3769781$. Kun p on 12 ja T on 144, saadaan S_1 :n arvoksi 3.219254, joka $F(12,144-2*12-1)$ –jakaumasta saa arvon 0.9994311. Nollahypoteesi voidaan siis hylätä, joten voidaan sanoa, että ainakin $\{so2I\}$:n arvot auttavat ennustamaan $\{trs\}$:n arvoja.

Tehdään seuraavaksi sama toisinpäin, eli tutkitaan, onko $\{trs\}$ Granger –kausaalinen $\{so2I\}$:een nähden. Jäännöseliösummiksi saadaan $SSR_1=7936.1493277$ ja $SSR_0=15538.720717$. Näin siis $S_1=8.541875$, joka samoilla vapausasteilla kuin edellä ~ 1 , eli myöskin $\{trs\}$:n voidaan sanoa auttavan ennustamaan $\{so2I\}$:tä.

15.2. UPM:n suunnalta tulevat päästöt

Tuulisektori UPM:n suunnalta tuleville päästöille on edelleen 225-251. Kaikki analyysit on suoritettu samalla tavoin kuin edellä Säterin suunnalta tulleita päästöjä tutkittaessa. Kuviossa **15.3** on kuvaajat UPM:n puolelta tulleista päästöistä.

Kuvio 15.3



15.2.1. Viivepituuden määrittäminen

Määritetään ensin viivepituus. Tarkastellaan aluksi mallille

$$\mathbf{x}_t = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{x}_{(t-1)} + \mathbf{A}_2 \mathbf{x}_{(t-2)} + \dots + \mathbf{A}_p \mathbf{x}_{(t-p)} + \mathbf{e}_t$$

, missä $\mathbf{x}_t = \{tktrs, so21\}$

sopivaa p :n arvoa informaatiokriteerien avulla samoin kuin edellisessä kappaleessa. Taulukossa 15.5 ovat AIC:n ja BIC:n arvot aina viiveeseen 24 saakka.

Taulukko 15.5

Viiveet	AIC	BIC
1	-498.6183	-516.3954
2	-488.4737	-518.0319
3	-485.2463	-526.5290
4	-490.4761	-543.4256
5	-495.0204	-559.5788

6	-498.6468	-574.7554
7	-502.5023	-590.1017
8	-503.3270	-602.3573
9	-507.1614	-617.5618
10	-501.2908	-623.0000
11	-502.1108	-635.0669
12	-499.4649	-643.6050
13	-485.0319	-640.2926
14	-335.4121	-501.7291
15	-322.1140	-499.4224
16	-323.9555	-512.1895
17	-328.2664	-527.3595
18	-334.4357	-544.3205
19	-339.6333	-560.2417
20	-341.3771	-572.6402
21	-346.2092	-588.0571
22	-350.9106	-603.2725
23	-356.3636	-619.1680
24	-359.6926	-632.8667

Luvut muistuttavat suuresti Säterin suunnalta saatuja arvoja. AIC pysyy samassa noin vuoden, minkä jälkeen tapahtuu suuri harppaus ylöspäin. BIC:n arvot pysyttelevät jälleen samoilla linjoilla. Näin ollen UPM:n tuulisektorilta voidaan tehdä Säterin sektorin kaltaisia johtopäätöksiä viivepituudesta.

Seuraavaksi suoritetaan LR –testit edellisen kappaleen tavoin. Taulukossa **15.6** ovat kaikki eri hypoteesien testisuureet p –arvoineen.

Taulukko 15.6

Viive	Testisuure	P –arvo
24 vs 18	19.276248	1.00000000
18 vs 12	28.302781	0.98949878

12 vs 6	36.707576	0.04678795
12 vs 8	27.285971	0.70417345
8 vs 6	8.913387	0.99777470

Taulukon **15.6** tulokset ovat hyvin samansuuntaisia Säteriltä tulleiden päästöjen kanssa. Valitaan samoin perustein sama viivepituus eli yksi vuosi.

Tarkastellaan myös tässä tapauksessa vakiotermin mukanaoloa informaatiokriteerien avulla. Vakiotermissä mallissa AIC ja BIC ovat -499.46487 ja -643.60497 sekä vakiotermitömmässä mallissa -500.23821 ja -644.37831. Näin ollen vakiotermitön malli antaisi hieman paremman tuloksen UPM:n puolelta. Taulukossa **15.7a** ja **15.7b** ovat kaikki PNS:llä lasketun mallin kertoimet keskihajontoineen sekä t- ja p -arvoineen siten että taulukossa **15.7a** {trs} toimii selitettävänä ja taulukossa **15.7b** {so21} on selitettävänä.

Taulukko 15.7a ({trs} selitettävänä)

Parametri	Estimaatti	Keskivirhe	T -arvo	P -arvo
1. TKTRS{1}	0.346518828	0.092112472	3.76191	0.00027437
2. TKTRS{2}	0.079586847	0.096682934	0.82317	0.41222320
3. TKTRS{3}	-0.002316578	0.098564093	-0.02350	0.98129220
4. TKTRS{4}	0.009587313	0.101174542	0.09476	0.92468095
5. TKTRS{5}	0.043709511	0.100076835	0.43676	0.66315773
6. TKTRS{6}	-0.038632711	0.100719306	-0.38357	0.70205263
7. TKTRS{7}	-0.040199401	0.099952770	-0.40218	0.68834350
8. TKTRS{8}	0.165079857	0.100035684	1.65021	0.10180544
9. TKTRS{9}	-0.049425346	0.101795392	-0.48554	0.62827960
10. TKTRS{10}	0.209647071	0.101946239	2.05645	0.04214979
11. TKTRS{11}	0.193184541	0.104419419	1.85008	0.06703531
12. TKTRS{12}	0.035840202	0.095917647	0.37366	0.70939277
13. SO21{1}	-0.019448615	0.026857822	-0.72413	0.47055025
14. SO21{2}	-0.023973327	0.034294524	-0.69904	0.48602825

15. SO21{3}	0.023717295	0.035687838	0.66458	0.50773798
16. SO21{4}	0.004062379	0.035886297	0.11320	0.91008104
17. SO21{5}	0.010926388	0.035822267	0.30502	0.76094050
18. SO21{6}	0.018958063	0.035497938	0.53406	0.59439695
19. SO21{7}	0.015059644	0.035147452	0.42847	0.66916200
20. SO21{8}	-0.034205087	0.034896870	-0.98018	0.32918968
21. SO21{9}	-0.018540788	0.034980657	-0.53003	0.59717935
22. SO21{10}	0.011348789	0.033984370	0.33394	0.73907131
23. SO21{11}	-0.001022907	0.030775285	-0.03324	0.97354620
24. SO21{12}	0.008058602	0.023402530	0.34435	0.73125459

Taulukko 15.7b ({so21} selitettävänä)

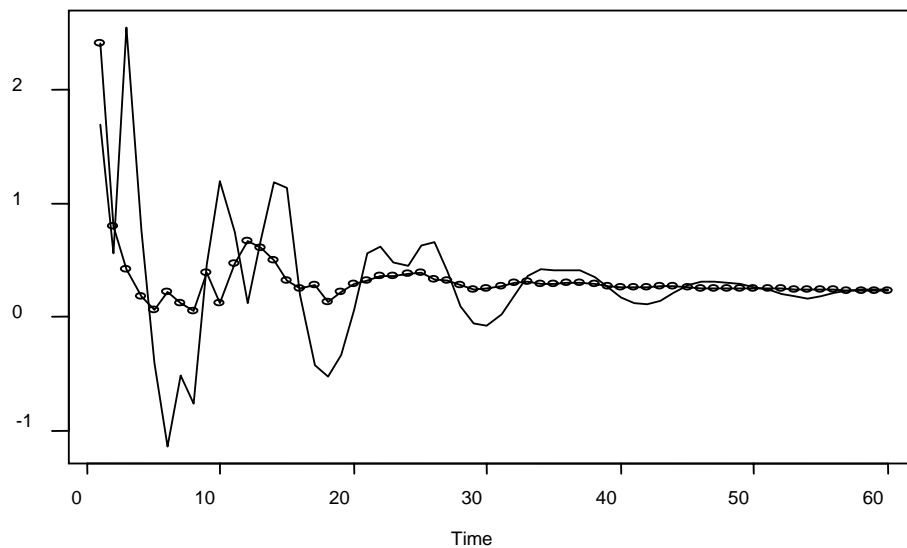
Parametri	Estimaatti	Keskivirhe	T -arvo	P -arvo
1. TKTRS{1}	-0.333869942	0.336044434	-0.99353	0.32267294
2. TKTRS{2}	1.264094654	0.352718376	3.58386	0.00050954
3. TKTRS{3}	-0.947929437	0.359581214	-2.63620	0.00961871
4. TKTRS{4}	0.099760423	0.369104646	0.27028	0.78746312
5. TKTRS{5}	-0.385068175	0.365099996	-1.05469	0.29391981
6. TKTRS{6}	0.228782297	0.367443855	0.62263	0.53483885
7. TKTRS{7}	-0.347921740	0.364647384	-0.95413	0.34214753
8. TKTRS{8}	0.537402912	0.364949868	1.47254	0.14378370
9. TKTRS{9}	0.281636559	0.371369632	0.75837	0.44987973
10. TKTRS{10}	-0.306336069	0.371919948	-0.82366	0.41194721
11. TKTRS{11}	0.036774022	0.380942598	0.09653	0.92327515
12. TKTRS{12}	0.032967928	0.349926460	0.09421	0.92511383
13. SO21{1}	0.802392196	0.097982621	8.18913	0.00000000
14. SO21{2}	-0.398209589	0.125113176	-3.18279	0.00190545
15. SO21{3}	0.202667799	0.130196259	1.55663	0.12248327

16. SO21{4}	-0.038413843	0.130920275	-0.29341	0.76976879
17. SO21{5}	0.018436266	0.130686683	0.14107	0.88807565
18. SO21{6}	0.031040463	0.129503467	0.23969	0.81102624
19. SO21{7}	-0.081466489	0.128224824	-0.63534	0.52654990
20. SO21{8}	0.032669669	0.127310652	0.25661	0.79796513
21. SO21{9}	-0.032858339	0.127616325	-0.25748	0.79730008
22. SO21{10}	0.079582370	0.123981673	0.64189	0.52230588
23. SO21{11}	0.134649157	0.112274298	1.19929	0.23304077
24. SO21{12}	0.039559128	0.085377038	0.46335	0.64404862

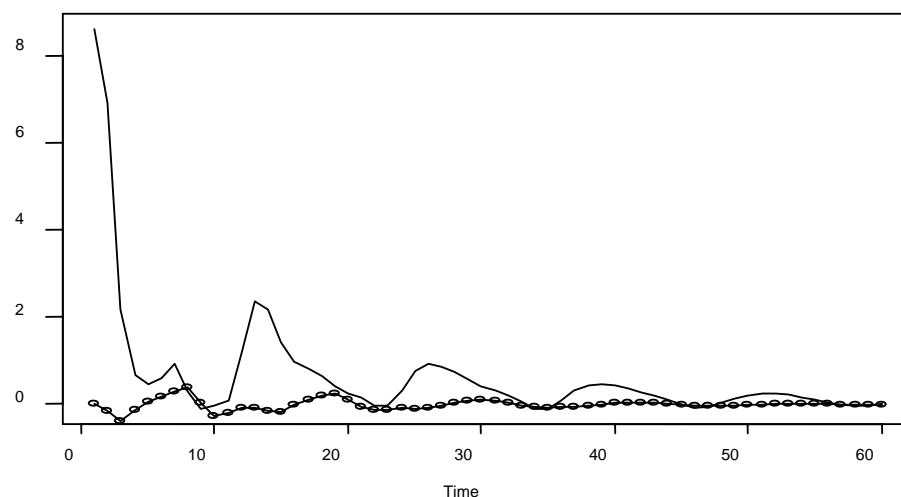
15.2.2. Innovaatioiden laskenta

Kuvioissa **15.4a** ja **15.4b** ovat impulssivastefunktiot edellä kuvatulla tavalla siten että kuviossa **15.4a** tarkastellaan sokkien vaikutusta $\{trs\}$ –sarjaan, kun taas kuviossa **15.4b** tarkastellaan $\{so2I\}$ –sarjaa.

Kuvio 15.4a



Kuvio 15.4b



UPM:n suunnalta tulleiden päästöjen impulssivastefunktiot käyttäytyvät huomattavasti rauhallisemmin ja säännönmukaisemmin kuin Säterin suunnalta tulleille päästöille. Molemmissa kuvioissa on havaittavissa, kuinka $\{trs\}$ –sarja tasaantuu nopeammin, kuin $\{so2I\}$. Ylemmästä kuvioista nähdään, kuinka $\{so2I\}$:n sokit vaikuttavat syklisesti $\{trs\}$ –sarjaan vielä pitkään sen jälkeen kun $\{trs\}$ –sokkien vaikutus omaan sarjaan on tasaantunut. Alemmassa kuviossa kannattaa kiinnittää huomiota siihen, kuinka $\{trs\}$ –sokit vaikuttavat $\{so2I\}$ –sarjaan aina, kun $\{so2I\}$ –sokit vaikuttavat omaan sarjaansa, mutta pienellä viiveellä.

Varianssin dekomponointi on esitetty taulukoissa **15.8a** ja **15.8b** siten että ensin tarkastellaan $\{trs\}$:n varianssia, minkä jälkeen otetaan $\{so2I\}$ tarkasteluun

Taulukko 15.8a

Askel	Keskivirhe	$\{trs\}$	$\{so2I\}$
1	2.407085448	100.000	0.000
2	2.542438925	99.566	0.434
3	2.607262327	97.244	2.756
4	2.618096263	96.915	3.085
5	2.619269293	96.883	3.117

6	2.632837457	96.575	3.425
7	2.651465713	95.444	4.556
8	2.676967674	93.666	6.334
9	2.705080060	93.795	6.205
10	2.724023831	92.698	7.302
11	2.771913245	92.360	7.640
12	2.853485676	92.674	7.326
13	2.918893657	92.905	7.095
14	2.967073298	92.762	7.238
15	2.989499193	92.495	7.505
16	3.000392507	92.539	7.461
17	3.015135292	92.501	7.499
18	3.024236016	92.146	7.854
19	3.040426738	91.694	8.306
20	3.055987292	91.683	8.317
21	3.073913773	91.719	8.281
22	3.099308727	91.601	8.399
23	3.122802295	91.527	8.473
24	3.147197841	91.538	8.462

Taulukko 15.8b

Askel	Keskivirhe	{trs}	{so21}
1	8.78152159	3.721	96.279
2	11.19042740	2.538	97.462
3	11.68031226	7.082	92.918
4	11.72327148	7.448	92.552
5	11.73867277	7.550	92.450
6	11.80784623	8.395	91.605
7	11.85425327	8.519	91.481
8	11.88209869	8.895	91.105

9	11.89160809	9.030	90.970
10	11.95200291	9.945	90.055
11	11.97541668	10.294	89.706
12	12.03432223	10.203	89.797
13	12.28194708	10.103	89.897
14	12.52857842	10.603	89.397
15	12.65808503	11.188	88.812
16	12.69589907	11.144	88.856
17	12.72837296	11.199	88.801
18	12.75443243	11.321	88.679
19	12.76470553	11.372	88.628
20	12.76699620	11.371	88.629
21	12.78017261	11.542	88.458
22	12.79541608	11.751	88.249
23	12.80451055	11.874	88.126
24	12.81610588	11.976	88.024

Suhteelliset varianssit asettuvat suhteellisen nopeasti uomiinsa myös UPM:n puolella. Ilmiö on sama kuin Säterin puolella, mutta nyt kumpikin sarja selittää noin 90% omasta varianssistaan, kun Säterin puolella vastaava luku jäi noin 80%:iin

15.2.3. Granger –kausaalisuus

Granger –kausaalisuuden testaus suoritetaan samalla tavoin kuten Säterin päästöjä tutkittaessa. Ensin tarkastellaan, onko $\{so2I\}$ Granger –kausaalinen $\{trs\}$:ään nähden. Näistä saadaan $SSR_1= 756.03989802$ ja $SSR_0= 820.82090506$, joista puolestaan $S_1= 0.7640214$, joka puolestaan antaa p –arvoksi 0.3142844. Nollahypoteesi siis säilyy, eli $\{so2I\}$:n arvot eivät auttaisi $\{trs\}$:n ennustamisessa. Koska arvot on mitattu samanaikaisesti Säterin suunnalta tulevien päästöjen kanssa, selitykseksi eri tuloksille jää tuulisektorin vaihto.

Testataan, onko puolestaan $\{trs\}$:n arvoista apua $\{so2I\}$:n käyttäytymisen ennustamisessa. Nyt $SSR_1=10167.695304$ ja $SSR_0=12140.377552$,

mistä saadaan $S_1=1.729964$, joka on vielä merkitsevä 10%:n riskitasolla. Granger –kausaalisuutta käsittelevien testien perusteella voidaan sanoa, että UPM:n suunnalta tulevat päästöt ovat Säteriä vaikeammin ennustettavia sarjojen keskinäisiä relaatioita tarkastelemalla.

16. Yhteisintegroituvuus

VAR –malleja käsittelevässä kappaleessa mainittiin, kuinka aikasarjojen oletettiin olevan stationaarisia, eikä yksiulotteisessa tapauksessa tutuksi tullutta differentia stationaarisuuden aikaansaamiseksi tehty. Näin ei kuitenkaan voida menetellä moniulotteisessa tapauksessa. On nimittäin mahdollista, että vaikka kaksi(tai useampi) muuttujaa olisikin epästationaarisia, löytyy niiden lineaarikombinaatio, joka puolestaan on stationaarinen. Tällaisia muuttujia sanotaan yhteisintegroituneiksi muuttujiksi.

16.1. Yhteisintegroituvuuden määritelmä

Engle ja Granger ovat osoittaneet yhteisintegraation periaatteen seuraavan tasapainorelaation avulla.

$$\beta_1 y_{1t} + \beta_2 y_{2t} + \dots + \beta_n y_{nt} = 0$$

, missä y_{it} , $i=1\dots n$, ovat aikasarjoja ja β_i , $i=1\dots n$, ovat vakioita.

Kun merkitään vektorit $\beta=(\beta_1,\dots,\beta_n)$ ja $y_t=(y_{1t},\dots,y_{nt})'$, niin systeemin sanotaan olevan pitkän aikavälin tasapainossa, kun $\beta y_t = 0$. Poikkeamaa pitkän aikavälin tasapainosta merkitään seuraavasti: $e_t = \beta y_t$. Nyt aikasarjojen y_{1t},\dots,y_{nt} sanotaan olevan yhteisintegroituneita astein b ja d , merk. $CI(b,d)$, jos 1) kaikki vektorin y_t komponentit(aikasarjat siis) ovat integroituneita astetta d , ja 2) on olemassa sellainen vektori β s.e. βy_t on integroitunut astetta $(d-b)$, $b>0$. Tällöin β :aa kutsutaan yhteisintegroituvuusvektoriksi. Useimmiten yhteisintegroituimisella

tarkoitetaan tapausta $C(1,1)$, jolloin yhteisintegroituvuus merkitään $C(1)$. (Banerjee, Dolado, Galbraith & Hendry, 2005, 255.)

Jos vektorissa y_t on n elementtiä, voi muuttujien välillä olla korkeintaan $n-1$ lineaarisesti riippumatonta yhteisintegraatiovektoria. Tässä tutkimuksessa vektorin y_t elementtien määrä on kuitenkin vain kaksi, joten muuttujien välillä voi olla korkeintaan yksi yhteisintegroituvuusvektori.

16.2. Virheenkorjaus

Yhteisintegroituvuus antaa kätevän työkalun kuvattaessa systeemin pitkän aikavälin dynamiikkaa; virheenkorjausmallin. Oletetaan, että meillä on kaksi aikasarjaa y_{1t} ja y_{2t} , joilla on pitkän aikavälin tasapainorelaatio, kuten edellä esitettiin. Nyt jos muuttujien välillä on suuri ero toisiinsa nähden, pakottaa tasapainorelaatio tuota eroa pienemmäksi. Havainnollistetaan asiaa seuraavan esimerkin avulla. Oletetaan, että aikasarjojen ensimmäisillä differensseillä (siis kahden peräkkäisen havainnon erotuksilla) on seuraavanlainen relaatio (mukailtu Enders, 366):

$$\begin{aligned}\nabla y_{1t} &= \alpha_1 (y_{2(t-1)} - \beta y_{1(t-1)}) + u_{1t} \\ \nabla y_{2t} &= -\alpha_2 (y_{2(t-1)} - \beta y_{1(t-1)}) + u_{2t}\end{aligned}$$

, missä u_{1t} ja u_{2t} ovat valkoista kohinaa, jotka voivat olla korreloituneita ja α_1 , α_2 ja β ovat positiivisia parametreja. Nyt $y_{2(t-1)} - \beta y_{1(t-1)}$ kuvastaa muuttujien poikkeamaa pitkän aikavälin tasapainosta. Joten jos tämä poikkeama sattuisi olemaan positiivinen, niin y_{1t} kasvaisi, kun taas y_{2t} pienenesi, jolloin mentäisiin kohti tasapainoa, joka saavutettaisiin, kun $y_{1t} = \beta y_{2t}$.

Yhteisintegroituvuuden ja virheenkorjauksen välinen yhteys tulee esiin siinä, että koska jos ∇y_{1t} :n oletetaan olevan stationaarinen $I(0)$ (aikasarjat olivat siis alun perin integroituneet astetta 1), niin myös ylemmän yhtälö oikean puolen on oltava stationaarinen. Koska u_{1t} määriteltiin stationaariseksi, on lineaarikombinaation $y_{2(t-1)} - \beta y_{1(t-1)}$ oltava myös stationaarinen, jolloin

Määritelmä 16.1:n mukaan muuttujat ovat yhteisintegroituneita yhteisintegroituuusvektorilla $(1, -\beta)$.

Samaa menetelmää voidaan soveltaa useamman muuttujan ja viiveen tapauksessa yleisemmin seuraavasti:

Määritelmä 16.2

Vektorilla $\mathbf{y}_t = (y_{1t}, \dots, y_{nt})'$ on virheenkorjausmuoto, jos se voidaan esittää muodossa

$$\nabla \mathbf{y}_t = \boldsymbol{\pi}_0 + \boldsymbol{\pi} \mathbf{y}_{t-1} + \boldsymbol{\pi}_1 \nabla \mathbf{y}_{t-1} + \dots + \boldsymbol{\pi}_p \nabla \mathbf{y}_{t-p} + \mathbf{u}_t$$

, missä $\boldsymbol{\pi}_0 = (n \times 1)$ –vakiovektori

$\boldsymbol{\pi}_i = (n \times n)$ –parametrimatriisi

$\boldsymbol{\pi}$ = matriisi, jonka elementit π_{jk} ovat s.e.

ainakin yksi $\pi_{jk} \neq 0$

$\mathbf{u}_t = (n \times 1)$ –vektori, jonka elementit ovat u_{it}

Nyt edellisen mukaan, jos kaikki \mathbf{y}_t :n muuttujat ovat $C(1)$, virheenkorjausmuodon olemassaolon ehtona on, että näiden muuttujien joukosta löytyy lineaarikombinaatio, joka on stationaarinen. Kun ratkaistaan Määritelmän 16.2 yhtälö $\boldsymbol{\pi} \mathbf{y}_{(t-1)}$:n suhteen, saadaan seuraavaa:

$$\boldsymbol{\pi} \mathbf{y}_{(t-1)} = \nabla \mathbf{y}_t - \boldsymbol{\pi}_0 - \sum_{i=1}^p \boldsymbol{\pi}_i \nabla \mathbf{y}_{(t-i)} - \mathbf{u}_t$$

Koska muuttujien oletettiin olevan $C(1)$, on $\nabla \mathbf{y}_t$ $C(0)$, joten yhtälön koko oikea puoli on $C(0)$, mistä seuraa, että myös $\boldsymbol{\pi} \mathbf{y}_{(t-1)}$:n on oltava myös stationaarinen.

Koska π koostuu pelkästään vakioista, on sen jokainen rivi yhteisintegroituuusvektori y_t :lle.

Nyt määritelmässä 16.2 pitää siis kiinnittää huomio matriisi π :hin. Jos sen kaikki elementit ovat nollia, palaa yhtälö tavalliseen ensimmäisten differenssien VAR –malliin. Näin ollen muuttujilla ei ole virheenkorjausesitystä. Jos taas useampi kuin yksi π :n elementeistä on erisuuri kuin nolla, ∇y_t reagoi edellisen periodin poikkeamaan tasapainotilasta. Näin ollen VAR –mallin käyttäminen olisi virhe, sillä nimenomaan termi $\pi y_{(t-1)}$ sisältää tarvittavan virheenkorjauksen.(Enders, 367.)

16.3. Yhteisintegroituisuuden testaus

Seuraavassa käsitellään muuttujien yhteisintegroituisuuden testaamista ns. Johansenin menetelmällä. Menetelmää on helpoin kuvailla lähtemällä liikkeelle n –muuttujaisesta ensimmäisen asteen VAR –mallista:

$$x_t = A_1 x_{(t-1)} + u_t \quad (16.1)$$

Kun käytetään differenssioperaattoria ∇ yhtälön (16.1) molemmille puolille, saadaan

$$\nabla x_t = -(I - A_1) x_{(t-1)} + u_t \quad (16.2)$$

Nyt matriisia $-(I-A_1)$ voidaan merkitä edellisestä tutulla merkinnällä π . Nyt tarkastellaan matriisin π astetta. Jos $\text{rank}(\pi) = 0$, on $A_1 = I$, jolloin $\nabla x_t = u_t$, eli $x_t = x_{(t-1)} + u_t$. Näin ollen kaikki muuttujat ovat yksikköjuuriprosesseja, eikä ole olemassa sellaista lineaarikombinaatiota, että ne olisivat stationaarisia. Toisessa ääripäässä taas lineaarisesti riippumattomien sarakkeiden lukumäärä on n , eli π on täysiasteinen. Tällöin taas kaikki muuttujat ovat stationaarisia(Enders s. 386). Jos kuitenkin $0 < \text{rank}(\pi) = r < n$, niin r on yhteisintegroituuusvektorien määrä systeemissä.(Enders, 372)

Edellä esitetty malli voidaan yleistää korkeamman asteen autoregressiivisiin prosesseihin seuraavasti: Olkoon n –muuttujainen p :n asteen vektoriautoregressiivinen malli

$$\mathbf{x}_t = \mathbf{A}_1 \mathbf{x}_{(t-1)} + \mathbf{A}_2 \mathbf{x}_{(t-2)} + \dots + \mathbf{A}_p \mathbf{x}_{(t-p)} + \mathbf{u}_t \quad (16.3)$$

Nyt, kun (16.3):sta vähennetään edellisen mukaan $\mathbf{x}_{(t-1)}$, jonka jälkeen lisätään ja vähennetään

$$(\mathbf{A}_1 - \mathbf{I})\mathbf{x}_{(t-1)}, (\mathbf{A}_1 + \mathbf{A}_2 - \mathbf{I})\mathbf{x}_{(t-2)}, \dots, (\mathbf{A}_p + \dots + \mathbf{A}_1 - \mathbf{I})\mathbf{x}_{(t-p)}$$

, saadaan

$$\nabla \mathbf{y}_t = \sum_{i=1}^{p-1} \boldsymbol{\pi}_i \nabla \mathbf{x}_{(t-i)} + \boldsymbol{\pi} \mathbf{x}_{(t-p)} + \mathbf{u}_t \quad (16.4)$$

$$\text{, missä } \boldsymbol{\pi} = -\left(\mathbf{I} - \sum_{i=1}^p \mathbf{A}_i \right)$$

$$\text{ja } \boldsymbol{\pi}_i = -\left(\mathbf{I} - \sum_{j=1}^i \mathbf{A}_j \right)$$

Kuten edellä, on matriisin $\boldsymbol{\pi}$ aste myös yhteisintegroituusvektorien määrä systeemissä (Enders s.389-390.). (16.4):n voi helposti laajentaa ottamaan mahdollisen vakiotermin malliin:

$$\nabla \mathbf{x}_t = \mathbf{A}_0 + \sum_{i=1}^{p-1} \boldsymbol{\pi}_i \nabla \mathbf{x}_{(t-i)} + \boldsymbol{\pi} \mathbf{x}_{(t-p)} + \mathbf{u}_t$$

Seuraavaksi tarkastellaan ns. Johansenin menetelmän mukaan matriisin $\boldsymbol{\pi}$ astetta kahden testin avulla. Koska matriisin aste on sama, kuin sen nolasta eroavien karakterististen juurien lukumäärä, voidaan tätä tietoa käyttää hyväksi. Järjestetään kaikki matriisin $\boldsymbol{\pi}$ karakteristiset juuret suuruusjärjestykseen

$\lambda_1 > \lambda_2 > \dots > \lambda_n$. Jos muuttujien välillä ei löydy yhteisintegraatiovektoreita, on edellä esitetyn mukaan π :n aste siis nolla, samoin kuin kaikki karakteristiset juuret. Tämä ehto voidaan kirjoittaa muodossa $\ln(1 - \lambda_i) = 0$ kaikilla $i = 1 \dots n$.

Käytännössä saatavilla on vain piin estimaatti, joten nolasta eroavien karakterististen juurien lukumäärää testataan seuraavilla testisuureilla:

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i)$$

$$\lambda_{max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1})$$

, missä $\hat{\lambda}_i$ on estimoitu karakteristisen juuren arvo

ja T on käytössä olevien havaintojen lukumäärä

λ_{trace} -testi testaa nollahypoteesia, jonka mukaan nolasta eroavien karakterististen juurien ja yhteisintegroituusvektoreitten määrä on $\leq r$ vastaan H_1 , jonka mukaan vektorien määrä $> r$. λ_{max} -testi puolestaan testaa nollahypoteesia, jonka mukaan nolasta eroavien karakterististen juurien ja yhteisintegroituusvektoreitten määrä on r vastaan H_1 , jonka mukaan vektorien määrä $= r+1$. Testeille löytyvät omat taulukkonsa sen mukaan, onko mallissa vakiotermejä tai vakiota yhteisintegroituusvektorissa. (Enders, 390-391.)

17. TRS:n ja SO21:n yhteisintegroituvuus

Tässä kappaleessa tutkitaan edellä esitetuin keinoin, onko $\{trs\}$:llä ja $\{so21\}$:llä mahdollisesti pitkän aikavälin tasapainorelaatiota, ja jos näin on, määritellään virheenkorjausmalli täydentämään kappaleessa 15 saatuja VAR -malleja. Edelliseen tapaan ensin käsitellään Säterin tuulisektori, minkä jälkeen käydään läpi UPM:n tuulisektori.

17.1. Säterin tuulisektori

Säterin tuulisektori on myös tässä yhteisintegraatiota tutkivassa kappaleessa sama kuin aikaisemmin. Ensimmäiseksi on määritelmän mukaan syytä tutkia sarjojen $\{trs\}$ ja $\{so2I\}$ integroituneisuuden astetta, sillä jos sarjat ovat integroituneet eri astetta, ei muuttujilla voi olla virheenkorjausmalliakaan.

Taulukoissa **17.1a** ja **17.1b** ovat $\{trs\}$:n ja $\{so2I\}$:n p –arvot kpss-, adf- ja pp –testeille

Taulukko 17.1a

Kpss	>0.1
Adf	<0.01
Pp	0.01

Taulukko 17.1b

Kpss	<0.01
Adf	0.01270
Pp	0.01

Taulukot **17.1a** ja **17.1b** antavat tuloksia, jotka viittaisivat siihen, että sarja $\{trs\}$ olisi stationaarinen $I(0)$. Kpss –testi $\{so2I\}$ –sarjalla antaa viitteitä, että sarja olisi stationaarinen, mutta muut testit eivät. Tarkastellaan asiaa vielä λ_{trace} - ja λ_{max} – testeillä. Edellä on jo määritelty sopiva viivepituus systeemille, ja päätelty, että malli on ilman vakiota. Näin ollen saadaan malli

$$\nabla x_t = \sum_{i=1}^{11} \pi_i \nabla x_{(t-1)} + \pi x_{(t-12)} + u_t$$

Nyt suoritetaan testaus RATS –ohjelman Johansen –proseduurilla. Taulukoissa **17.2a** ja **17.2b** ovat saadut tulostukset λ_{max} – ja λ_{trace} -testeille. Tässä tapauksessa n on siis kaksi.

Taulukko 17.2a (λ_{\max})

h0	h1	$\lambda_{\max}(0,1)$	95% arvo	
r=0	r=0+1=1	35.04355	14.595	h0 hylätään
h0	h1	$\lambda_{\max}(1,2)$	95% arvo	
r=1	r=1+1=2	9.07261	8.083	h0 hylätään

Taulukko 17.2b (λ_{trace})

h0	h1	$\lambda_{\text{trace}}(0)$	95% arvo	
r<=0	r>0	44.11616	17.844	h0 hylätään
h0	h1	$\lambda_{\text{trace}}(1)$	95% arvo	
r<=1	r>1	9.07261	8.083	h0 hylätään

Taulukot **17.2a** ja **17.2b** antavat lisää todisteita hypoteesille, että molemmat Säterin suunnalta tulevat päästösarjat ovat stationaarisia, sillä kun kaikki hypoteesit hylätään, jäävät jäljelle vaihtoehdot hypoteesit, joista toisen mukaan $r=2$ ja toisen mukaan $r>1$. Näin ollen edellisessä kappaleessa esitetty VAR -malli on pätevä.

17.2. UPM:n tuulisektori

Myös UPM:n tuulisektori on määritelty kuten edellä. Tehdään samat yksikköjuuritarkastelut kuin edellä taulukoihin **17.3a** ja **17.3b** ($\{trs\}$ ensimmäiseen ja $\{so2I\}$ jälkimmäiseen).

Taulukko 17.3a

Kpss	<0.01
Adf	<0.01
Pp	0.01

Taulukko 17.3b

Kpss	<0.01
Adf	0.0641
Pp	0.01

Nyt $\{trs\}$ näyttää olevan kpss –testin mukaan epästationaarinen. Vaikka muut testit eivät puollakaan yksikköjuuren olemassaoloa, tutkitaan kuitenkin $\{trs\}$:n integroituneisuuden astetta. Kun kerran differoituun sarjaan sovelletaan jälleen kpss –testiä, saadaan p –arvoksi >0.1 , joten tämän perusteella $\{trs\}$ –sarja olisi $C(1)$.

$\{so2I\}$:n kohdalla taas kpss –testi hylkää nollahypoteesin(sarja stationaarinen) 99% merkitsevyystasolla. Adf –testissä H_0 (yksikköjuuri) jää vielä voimaan 95% merkitsevyystasolla, joten tästä voidaan päätellä, että $\{so2I\}$ voisi hyvinkin olla integroitunut. Sovelletaan em. testejä vielä differoituun sarjaan, jotta saataisiin käsitys integroitumisen asteesta. Kpss –testi antoi p –arvoksi >0.1 ja adf –testi taas <0.01 , joten sarjan voidaan päätellä olevan integroitunut astetta 1.

$\{trs\}$ –sarjassa yksi kolmesta yksikköjuuritestistä antoi tuloksen $C(1)$, kun taas $\{so2I\}$:n tapauksessa suhde oli kaksi kolmesta. Seuraavassa tarkastellaan yhteisintegroituvuutta λ_{\max} - ja λ_{trace} –testeillä. Edellä on jo määritelty normaali VAR –malli jolloin määriteltiin sopiva viivepituus $p=12$. Nyt edellisen kappaleen mukaan saadaan malliksi

$$\nabla \mathbf{x}_t = \sum_{i=1}^{11} \boldsymbol{\pi}_i \nabla \mathbf{x}_{(t-1)} + \boldsymbol{\pi} \mathbf{x}_{(t-12)} + \mathbf{u}_t$$

Nyt suoritetaan λ_{\max} - ja λ_{trace} –testit RATSin Johansen –proseduurilla. Tulokset ovat taulukoissa **17.4a** ja **17.4b**.

Taulukko 17.4a (λ_{\max})

h0	h1	$\lambda_{\max}(0,1)$	95% arvo	
r=0	r=0+1	6.67824	14.595	h0 hyväksytään
h0	h1	$\lambda_{\max}(1,2)$	95% arvo	
r=1	r=1+1=2	2.05713	8.083	h0 hyväksytään

Taulukko 17.4b (λ_{trace})

h0	h1	$\lambda_{\text{trace}}(0)$	95% arvo	
r<=0	r>0	8.73536	17.844	h0 hyväksytään
h0	h1	$\lambda_{\text{trace}}(1)$	95% arvo	
r<=1	r>1	2.05713	8.083	h0 hyväksytään

λ_{trace} –testien mukaan piin aste olisi nolla, samoin kuin $\lambda_{\max}(0,1)$ –testin perusteella. Tällöin olisi siten, että molemmat sarjat olisivat yksikköjuuriprosesseja, joilla ei ole virheenkorjausmallia. Vaikka $\lambda_{\max}(1,2)$ –testi hyväksytäänkin, pitäydytään $\lambda_{\text{trace}}(0)$ –testin antamassa tuloksessa, jonka mukaan muuttujien välillä ei ole yhteisintegroituvuusvektoria. Näin ollen myös UPM:n puoleisille päästöille kappaleessa 15.2.1 esitetty VAR –malli on pätevä.

18. Loppusanat

Teollistumisen kasvaessa kaupunki-ilman hajuhaitat lisääntyvät samaa tahtia. Valkeakosken tapauksessa haitat eivät ole vielä saavuttaneet vakavia mittasuhteita (tämä osittain tietysti riippuu, keneltä kysytään). Kuitenkin luvun 13 tulokset osoittavat, kuinka saastuttavat tahot voivat omalta osaltaan ratkaisevasti parantaa ilman, ja samalla asukkaiden elämän, laatua.

Vuorokaudenaikojen mukaan vaihtelevat päästöt osoittavat kuormituksen jakoa. Hajurikkipäästöt ovat selvästi korkeimmillaan yöaikaan, ihmisten nukkuessa. Edes päivän liikenne ei ole saanut päästöjä yön lukemiin. Tämä on tietysti oikea suunta saastuttajien puolelta.

Vaikka tutkimuksissa ei havaittukaan hajurikkipitoisuuksien riippuvuutta lämpötilasta tai yhteisintegroitumista rikkidioksidin kanssa, saatiin kuitenkin vektori-autoregressiivisten mallien kautta suhde hajurikkipitoisuuksien ja rikkidioksidipäästöjen välille. Testit Granger –kausaalisuudelle osoittivat, että varsinkin Säterin puoleisilla päästöillä molemmat sarjat auttoivat toistensa ennustamisessa. Koska molemmat päästöt ovat pääosin tehdasperäisiä, voidaan tuloksia käyttää hyväksi suunniteltaessa tulevia toimenpiteitä saastuttavilla tehtailla. Tässä kuitenkin piilee eräs tutkimuksen heikkouksista. Tutkittavana on havaintoja, joiden takana piilevät tarkkaan harkitut prosessit. Toisin kuin esim. pörssikurssit, jotka pitkälti elävät ”omaa elämäänsä”, saastepitoisuuksia voidaan säädellä tehtaiden johdon toimesta moninaisilla tavoilla. Niinpä voidakseen sanoa päästöjen relaatioista tarkemmin, täytyisi tietää havaittujen arvojen lisäksi, miten tehtaissa tapahtuva toiminta vaikuttaa niihin.

Siltikin tutkimus on osoittanut, kuinka saastepitoisuudet riippuvat toisistaan jo havainnoitavalla tasolla (loppujen lopuksi tämä taso on se, minkä ”loppukäyttäjät” pelkästään havaitsevat), ja että pitoisuuksia on ylipäätään mahdollista vähentää. Nyt investointien jälkeen myös tiedetään, millä hinnalla niitä vähennettiin. Näitä tietoja hyväksikäyttäen päättävässä asemassa olevat ihmiset voivat tehdä tarkempia ja puhtaampia päätöksiä tulevaisuuden ilmanlaadun puolesta Valkeakoskella.

19. Lähdeluettelo

- An Introduction to R. R Development Core Team. 2004. Saatavilla [www-muodossa: URL: http://www.r-project.org/](http://www.r-project.org/)
- Andersen, T. & Bollerslev, T.:* Answering the Critics: Yes, Arch Models do Provide Good Volatility Forecasts. Working Paper Series. National Bureau of Economic Research, Inc. 1997
- Banerjee, A., Dolado, J., Galbraith, J. & Hendry, D.:* Co-Integration, Error-Correction, and the Econometric Analysis of non-Stationary Data. Oxford University Press 2003
- Bowerman, B. & O'Connell R.:* Time Series Forecasting. Unified Concepts and Computer Implementation. 2. painos. Duxbury 1987
- Box, G. & Jenkins, G.:* Time Series Analysis Forecasting and Control. 2. painos. Holden-Day 1976
- Box, G.:* Use and Abuse of Regression. 1966. The Collected Works of George E.P. Box. Volume I. Tiao(toim.) Wadsworth 1985
- Brockwell, P. & Davis, R.:* Introduction to Time Series and Forecasting. Springer 1996
- Doan, T.:* Rats Version 5 Users' Guide. Estima. 2000
- Enders, W.:* Applied Econometric Time Series. John Wiley & Sons, Inc. 1995
- Freund, R. & Minton, P.:* Regression Methods: A Tool for Data Analysis. Marcel Dekker, Inc. 1979
- Hamilton, J.:* Time Series Analysis. Princeton University Press. 1994
- Hansen, M., Hurwitz, W. & Madow, W.:* Sample Survey Methods and Theory, 3. painos. John Wiley & Sons, Inc. 1960
- Harvey, A.:* Time Series Models. Philip Allan. 1981
- Huhtanen, P. & Kallinen, A.:* Matemaattinen tilastotiede. 2. painos. Tampereen Yliopisto 1998
- Iversen, G. & Gergen, M.:* Statistics: the Conceptual Approach. Springer 1997
- Kahra, H. & Kanto, A.:* Aikasarja-analyysi II. Kurssimateriaali. Kevät 2004.
- Kendall, M. & Ord, J.:* Time Series. 3. painos. Edward Arnold 1990

- Kwiatkowski, D., Phillips, P., Schmidt, P. & Shin, Y.:* Testing the Null Hypothesis of Stationary Against the Alternative of a Unit Root. 1991. Journal of Econometrics 54 (1-3). 159-178. 1992
- McLeod, G.:* Box Jenkins in Practice. Gwilym Jenkins & Partners Ltd. 1983
- Opsis AR 500 –Series Opto-Analyzer. User’s Hardware Manual. Opsis AB. 1992
- Päällysaho, A.:* Hajukaasujen käsittelystä. Yksityinen sähköpostiviesti. Lähetetty 6.7.2004. atph.paallysaho@pp.inet.fi. Luettu 22.10.2004
- Pandit, S. & Wu, S.:* Time Series and System Analysis with Applications. John Wiley & Sons, Inc. 1983
- Pedhazur, E.:* Multiple Regression in Behavioral Research. Explanation and Prediction 2. painos. Holt, Rinehart and Winston 1982
- Puntanen, S.:* Regressioanalyysi I. Tampereen Yliopistopaino Oy. 1999
- Sakamoto, Y., Ishiguro, M. & Kitagawa G.:* Akaike Information Criterion Statistics. KTK Scientific Publishers. 1986
- Wei, W.:* Time Series Analysis. Univariate and Multivariate Methods. Addison-Wesley 1994
- Wetherill, B.:* Regression Analysis with Applications. Chapman and Hall. 1986
- World Health Organization:* WHO Air Quality Guidelines. 2000. Saatavilla [www-muodossa](http://www.muodossa): URL: http://www.euro.who.int/air/Activities/20020620_1

Liite 1. Tietokonekoodit

LUKU 3

KUVIO 3.1A

```
x<- rnorm(200, mean=1, sd=0.2)
x <- ts(x)
plot(x, ylim=c(0,2), main=c("Kuvio 3.1a"), ylab=c(""), xlab=c(""))
```

KUVIO 3.1B

```
y2 <- 0
a <- 0
for (i in 1:75){
  y2[i] <- rnorm(1, mean=1, sd=0.2)
}
for (i in 76:200){
  y2[i] <- a+rnorm(1, mean=1, sd=0.6)
  a <- a+0.01
}
y2 <- ts(y2)
plot(y2, main=c("Kuvio 3.1b"), ylab=c(""), xlab=c(""))
```

LUKU 4

KUVIO 4.1

```
trs.yö <- scan()
trs.aamu <- scan()
trs.päivä <- scan()
trs.ilta <- scan()

trs.yö <- ts(trs.yö, start=1990, frequency=12)
trs.aamu <- ts(trs.aamu, start=1990, frequency=12)
trs.päivä <- ts(trs.päivä, start=1990, frequency=12)
trs.ilta <- ts(trs.ilta, start=1990, frequency=12)

plot(cbind(trs.yö, trs.aamu, trs.päivä, trs.ilta), main=c("Kuvio 4.1"))
```

TAULUKKO 4.2

```
(SAS -koodi):
PROC GLM DATA = TKTRS;
MODEL TKTRS = DY DA DP DI/CLI;
ESTIMATE 'P1' DA-1 DP 1;
ESTIMATE 'P2' DA-1 DI 1;
ESTIMATE 'P3' DP-1 DI 1;
```

KUVIO 4.2

```
boxplot(trs.yö, trs.aamu, trs.päivä, trs.ilta, names=c("Yö","Aamu","Päivä","Ilta"), main=c("Kuvio 4.2"))
```

KUVIO 4.3

```
a <- c(rep(3.732, 100), rep(3.12, 100), rep(2.62, 100), rep(3.29, 100))
a <- ts(a)
plot(a, ylim = c(2,5), ylab = c("trs -pitoisuus"), xlab = c("Yö=1-100 Aamu=101-200 Päivä=201-300 Ilta=301-400"), main = c("Kuvio 4.3"))
```

LUKU 10

KUVIO 10.1(Heteroskedastinen aikasarja R:n tseries -paketista.)

```
data(EuStockMarkets)
data <-diff(EuStockMarkets[, "DAX"])
plot(data, main=c("Kuvio 10.1"))
```

LUKU 12

AIKASARJOJEN MUODOSTUS

```
trs <- ts(trs, start=1, frequency=24)
lämpö <- ts(lämpö, start=1, frequency=24)
```

KUVIO 12.1

```
plot(cbind(trs, lämpö), main=c("Kuvio 12.1"))
```

KUVIO 12.2

```
ccf(trs, lämpö, main=c("Kuvio 12.2"))
```

SARJAN DIFFEROINTI VIIVEELLÄ 24

```
d24.lämpö <- diff(lämpö, lag=24)
```

KUVIO 12.3

```
plot(d24.lämpö, main=c("Kuvio 12.3"))
```

SARJAN DIFFEROINTI

```
d.d24.lämpö <- diff(d24.lämpö)
```

KUVIO 12.4

```
plot(d.d24.lämpö, main=c("Kuvio 12.4"))
```

LAAJENNETTU DICKEY-FULLER -, PHILLIPS-PERRON - JA KIATKOWSKI-PHILLIPS-SCHMIDT-SHIN -TESTAUS

```
adf.test(d.d24.lämpö)
PP.test(d.d24.lämpö)
kpss.test(d.d24.lämpö)
```

KÄÄNNEPISTETESTI -FUNKTIO

```
turning.point.test<-function(x)
{
  DNAME<-deparse(substitute(x))
  n<-length(x)
  METHOD<-"Turning point test"
  X<-embed(x,3)
  STATISTIC<-sum((X[,2] > X[,1] & X[,2] > X[,3]) | (X[,2] < X[,1] & X[,2] < X[,3]))
  mu<-2*(n-2)/3
  sigma2<-(16*n-29)/90
  PVAL<-2*(1-pnorm(abs(STATISTIC-mu)/sqrt(sigma2)))
  PARAMETER<-c(mu,sigma2)
  names(STATISTIC)<-"normal"
  names(PARAMETER)<-c("mu","sigma2")

  structure(list(statistic=STATISTIC,parameter=PARAMETER,p.value=PVAL,method=METHOD,
  data.name=DNAME),class="htest")
}
```

BOX-LJUNG -, BOX-PIERCE -, JARQUE-BERA - JA KÄÄNNEPISTETESTAUS

```
Box.test(d.d24.lämpö, type=c("Ljung"))
Box.test(d.d24.lämpö)
jarque.bera.test(d.d24.lämpö)
turning.point.test(d.d24.lämpö)
```

KUVIOT 12.5A JA 12.5B

```
acf(d.d24.lämpö, lag=100, main=c("Kuvio 12.5a"))
pacf(d.d24.lämpö, lag=100, main=c("Kuvio 12.5b"))
```

SARIMA(1,1,0)(0,1,1)*12 -MALLI

```
m110.011.24 <- arima(lämpö, order=c(1,1,0), seasonal=list(order=c(0,1,1), period=24))
```

SARIMA(1,1,0)(0,1,1)*12 -MALLIN JÄÄNNÖSSARJAN RIIPPUMATTOMUUSTESTIT

```
Box.test(m110.011.24$residuals, type=c("Ljung"))
Box.test(m110.011.24$residuals)
jarque.bera.test(m110.011.24$residuals)
turning.point.test(m110.011.24$residuals)
```

PARAMETRIEN MERKITSEVYYDEN TESTAUS

```
1-pt(0.3523/0.0352, length(lämpö)-2)
1-pt(-0.8921/0.0236, length(lämpö)-2, lower.tail=F)
```

TRS -SARJAN DIFFEROINTI

```
d.trs <- diff(trs)
```

KUVIO 12.7

```
plot(d.trs, main=c("Kuvio 12.7"))
```

KUVIOT 12.8A JA 12.8B

```
acf(d.trs, main=c("Kuvio 12.8a"))  
pacf(d.trs, main=c("Kuvio 12.8b"))
```

MA(2) -MALLI

```
m02 <- arima(trs, order=c(0,1,2))
```

NELIÖITYJEN RESIDUAALIEN MUODOSTAMINEN

```
m02.res <- m02$residuals  
m02.res.sq <- m02.res^2  
m02.res.sq <- ts(m02.res.sq)
```

NELIÖITYJEN RESIDUAALIEN RIIPPUMATTOMUUDEN TESTAUS

```
Box.test(m02.res.sq)
```

GARCH(0,2) -JA (0,1) -MALLIT

```
g02 <- garch(d.trs, order=c(0,2))  
g01 <- garch(d.trs, order=c(0,1))
```

EDELLÄ SAATUJEN MALLIEN RESIDUAALITARKASTELUJA

```
summary(g01)  
summary(g02)  
Box.test(g02$residuals)  
Box.test(g01$residuals)
```

RESIDUAALISARJOJEN SAATTAMINEN SAMANMITTAISIKSI

```
resid.lämpö <- m110.011.24$residuals  
resid.trs <- g01$residuals
```

```
resid.trs <- resid.trs[-c(1,2)]  
resid.trs <- ts(resid.trs)
```

```
resid.lämpö <- resid.lämpö[-c(1:3)]  
resid.lämpö <- ts(resid.lämpö)
```

KUVIO 12.9

```
ccf(resid.lämpö, resid.trs, lag=100, main=c("Kuvio12.9"))
```

(Samat koodit toimivat kaikille tutkituille sarjoille)

LUKU 13

SÄTERIN PUOLEISEN AIKASARJAN TAAJUUDEN MÄÄRITTELY

```
säteri.tris <- ts(säteri.tris, start=1990, frequency=12)
```

KUVIO 13.1

```
plot(säteri.tris, main=c("Kuvio 13.1"))
```

AIKASARJAN JAKAMINEN KAHTEN OSAAN INTERVENTIOPISTEEN KOHDALTA

```
säteri.tris.alku <- säteri.tris[c(1:108)]
```

BOX-PIERCE -, JARQUE-BERA -SEKÄ STATIONAARISUUSTESTAUKSIA SARJAN ALKUPÄÄLLE

```
Box.test(säteri.tris.alku)  
jarque.bera.test(säteri.tris.alku)  
adf.test(säteri.tris.alku)  
PP.test(säteri.tris.alku)  
kpss.test(säteri.tris.alku)
```

KOLME ERI TESTATTAVAA MALLIA SARJAN ALKUPUOLISKOLLE

```
m100 <- arima(säteri.tris.alku, order=c(1,0,0))  
m001 <- arima(säteri.tris.alku, order=c(0,0,1))  
m101 <- arima(säteri.tris.alku, order=c(1,0,1))
```

AR(1) -MALLIN MUOTOINEN INTERVENTIOMALLI

```
z <- c(rep(0,108), rep(1,36))  
säteri.tris.l <- säteri.tris[-1]-diff(säteri.tris)  
malli <- lm(säteri.tris[-1]~säteri.tris.l + z[-1])
```

LINEAARINEN INTERVENTIOMALLI

```
z <- c(rep(0,108), rep(1,36))  
säteri.tris.l <- säteri.tris[-1]-diff(säteri.tris)  
l.malli <- lm(säteri.tris ~ int)
```

KUVIO 13.3

```
plot(säteri.tris, main=c("Kuvio 13.3"))  
lines(ts(predict(l.malli), start=1990, frequency=12))
```

UPM:N PUOLEISEN AIKASARJAN TAAJUUDEN MÄÄRITTELY

```
upm.tris <- ts(upm.tris, start=1990, frequency=12)
```

KUVIO 13.4

```
plot(upm.tris, main=c("Kuvio 13.4"))
```

AIKASARJAN JAKAMINEN KAHTEN OSAAN INTERVENTIOPISTEEN KOHDALTA

```
upm.tr.s.loppu <- upm.tr.s[-c(1:76)]
```

KUVIO 13.5

```
acf(upm.tr.s.loppu, main=c("Kuvio 13.5"))
```

BOX-PIERCE -, JARQUE-BERA -SEKÄ STATIONAARISUUSTESTAUKSIA SARJAN LOPPUPÄÄLLE

```
Box.test(upm.tr.s.loppu)
jarque.bera.test(upm.tr.s.loppu)
adf.test(upm.tr.s.loppu)
PP.test(upm.tr.s.loppu)
kpss.test(upm.tr.s.loppu)
```

LINEAARINEN INTERVENTIOMALLI, JOSSA YKSI INTERVENTIOPISTE

```
z <- c(rep(0,76), rep(1,68))
l.malli <- lm(upm.tr.s~z)
```

LINEAARINEN INTERVENTIOMALLI, JOSSA KAKSI INTERVENTIOPISTETTÄ

```
z1 <- c(rep(0,38), rep(1,106))
z2 <- c(rep(0,76), rep(1,68))
l2.malli <- lm(upm.tr.s ~ z1 + z2)
```

KUVIO 13.9

```
plot(upm.tr.s, main=c("Kuvio 13.9"))
lines(ts(predict(l2.malli), start=1990, frequency=12))
```

LUKU 15(RATS –koodi)

AINEISTON SISÄÄNLUKU

```
calendar 1990 1 12
allocate 2001:12
open data X:\Gradusälää\sasrtluvut1.xls
data(format=xls, org=obs) / TKTRS SO21
```

KUVIO 15.1

```
(R –koodi)
säteri.tr.s <- ts(säteri.tr.s, start=1990, frequency=12)
säteri.so21 <- ts(säteri.so21, start=1990, frequency=12)
plot(cbind(säteri.tr.s, säteri.so21), main=c("Kuvio 15.1"))
```

TAULUKKO 15.1

```
display
do lags=1,24
system 1 to 2
variables TKTRS SO21
lags 1 to lags
deterministic constant
```



```

end(system)

estimate(noprint, outsigma=cov)

compute aic=-0.5*%nobs*%logdet-2*(2*lags+1)*2
compute bic=-0.5*%nobs*%logdet-2*(2*lags+1)*log(%nobs)

if lags==1
  display @4 'lags' @20 'aic' @35 'bic'
  display @5 ##### lags @20 #####.##### aic @35 bic
end do lags

```

LR -TESTI VIIVEILLE 24 JA 18(samalla koodilla voi testata muita viiveitä)

```

smpl 1992:1 *

system 1 to 2
variables TKTRS SO21
lags 1 to 24
deterministic constant
end(system)
estimate(noprint, resids=res24)
compute logdetu=%logdet

system 1 to 2
variables TKTRS SO21
lags 1 to 18
deterministic constant
end(system)
estimate(noprint, resids=res18)
compute logdetr=%logdet

ratio(degrees=(2*2*18), mcorr=(2*24)+1)
#res24
#res18

cdf chisqr (%nobs(2*24)+1)*(logdetr-logdetu) (2*2*18)

```

VAKIOTERMIN TESTAAMINEN INFORMAATIOKRITTEERIEN AVULLA

```

system 1 to 2
variables TKTRS SO21
lags 1 to 12
deterministic
end(system)

estimate(noprint, outsigma=cov)

compute aic=-0.5*%nobs*%logdet-2*(2*12+1)*2
compute bic=-0.5*%nobs*%logdet-2*(2*12+1)*log(%nobs)

display @4 aic bic

system 1 to 2
variables TKTRS SO21
lags 1 to 12
deterministic constant
end(system)

```

```

estimate(noprint, outsigma=cov)

compute aic=-0.5*%nobs*%logdet-2*(2*12+1)*2
compute bic=-0.5*%nobs*%logdet-2*(2*12+1)*log(%nobs)

display @4 aic bic

```

VAR –MALLI ILMAN VAKIOTERMIÄ VIIVEELLÄ 12

```

system 1 to 2
variables TKTRS SO21
lags 1 to 12
deterministic
end(system)
estimate(resids=res12)

```

INNOVAATIOIDEN LASKENTA

```

system 1 to 2
variables TKTRS SO21
lags 1 to 12
deterministic
end(system)

```

```
estimate(outsigma=cov)
```

```

errors(impulses) 2 5*12 cov
#1
#2

```

KUVIOT 15.2A JA 15.2B

```

(R –koodi):
i.trs <- scan()
i.so21 <- scan()
plot(i.so21, main="Kuvio 15.2a", ylab="")
lines(i.trs, type="o")

```

(Kuviossa 15.2b sama koodi, mutta eri aikasarjat)

GRANGER –KAUSAALISUUDEN TESTAUS SO21:LLE

```

linreg TKTRS 1991:1 *
# constant TKTRS{1} TKTRS {2} TKTRS{3} TKTRS{4} TKTRS{5} TKTRS{6} TKTRS{7}
TKTRS{8} TKTRS{9} TKTRS{10} TKTRS{11} TKTRS{12} SO21{1} SO21{2} SO21{3}
SO21{4} SO21{5} SO21{6} SO21{7} SO21{8} SO21{9} SO21{10} SO21{11} SO21{12}

```

```

linreg TKTRS 1991:1 *
# constant TKTRS{1} TKTRS {2} TKTRS{3} TKTRS{4} TKTRS{5} TKTRS{6} TKTRS{7}
TKTRS{8} TKTRS{9} TKTRS{10} TKTRS{11} TKTRS{12}

```

```

(R –koodi):
rss1 <- 1348.5128280
rss0 <- 1835.3769781
p <- 12
T <- (144-12)
S1 <- ((rss0-rss1)/p)/(rss1/(T-2*p-1))

```

pf(S1, p, T-(2*p)-1)

GRANGER –KAUSAALISUUDEN TESTAUS TRS:LLE

```
linreg SO21 1991:1 *
# constant SO21{1} SO21{2} SO21{3} SO21{4} SO21{5} SO21{6} SO21{7} SO21{8}
SO21{9} SO21{10} SO21{11} SO21{12} TKTRS{1} TKTRS {2} TKTRS{3} TKTRS{4}
TKTRS{5} TKTRS{6} TKTRS{7} TKTRS{8} TKTRS{9} TKTRS{10} TKTRS{11}
TKTRS{12}
```

```
linreg SO21 1991:1 *
# constant SO21{1} SO21{2} SO21{3} SO21{4} SO21{5} SO21{6} SO21{7} SO21{8}
SO21{9} SO21{10} SO21{11} SO21{12}
```

```
(R –koodi):
rss1 <- 7936.1493277
rss0 <- 15538.720717
S1 <- ((rss0-rss1)/p)/((rss1/(T-2*p-1))
S1
pf(S1, p, T-(2*p)-1)
```

(Samat koodit käyvät UPM:n suuntaisille päästöille.)

LUKU 17(RATS –koodi)

TAULUKKO 17.1A

```
(R –koodi):
kpss.test(trs)
Adf.test(trs)
PP.test(trs)
```

TAULUKKO 17.1B

```
(R –koodi):
kpss.test(so21)
Adf.test(so21)
PP.test(so21)
```

TAULUKKO 17.2A JA 17.2B –LUVUT(TESTAUS JOHANSENIN MENETELMÄLLÄ)

```
source(noecho) johansen.src
@johansen(LAGS=12)
#TKTRS SO21
```

(Samat koodit käyvät myös UPM:n suuntaisille päästöille.)