

Jenni Sammaljoki

**AIKASARJA-ANALYYSI
VEKTORIAUTOREGRESSIIVISTA MALLIA
HYÖDYNTÄEN**

Energian kulutus, bruttokansantuote ja
hiilidioksidikaasupäästöt Suomessa

Kandidaatintyö
Tekniikan ja luonnontieteiden tiedekunta
Tarkastaja: Juho Kanninen
Lokakuu 2021

TIIVISTELMÄ

Jenni Sammaljoki: Aikasarja-analyysi vektoriautoregressiivista mallia hyödyntäen
Kandidaatintyö
Tampereen yliopisto
Tekniikka ja Luonnontieteet, TkK
Lokakuu 2021

Tässä kandidaatintyössä tarkastellaan kolmen tutkittavan suureen: hiilidioksidikaasupäästöjen suuruuden, energian kulutuksen ja bruttokansantuotteen yhteyttä toisiinsa aikasarja-analyysin avulla. Aihe on tiiviisti esillä keskustelussa ilmastonmuutoksen hillitsemisen vuoksi, ja tästä syystä aiheesta on tehty useita erilaisia tutkimuksia. Ilmastonmuutoksen hillitsemiseksi on tärkeää tunnistaa päästöjen suuruuteen vaikuttavat tekijät sekä kasvavan kulutuksen vaikutus. Tässä työssä suureiden yhteyttä tarkastellaan Suomessa vuosina 1975-2016. Tarkasteluväli on yksi vuosi, ja tarkastelussa hyödynnetään myös Suomen väkilukua, jotta eri vuosien data olisi vertailukelpoista keskenään.

Työssä esitellään aikasarja-analyysin peruskäsitteitä, kuten stationaarisuus, yksikköjuuri, yhteisintegroituvuus sekä Grangerin kausaalisuus ja esitellään näihin liittyviä testausmenetelmiä. Stationaarisuuteen eli sarjan paikallaan oloon liittyy oleellisesti yksikköjuuri. Mikäli tarkasteltavalla sarjalla on yksikköjuuri, se ei ole stationaarinen. Yhteisintegroituvuus kertoo sarjojen riippuvuudesta pitkällä aikavälillä, ja Grangerin kausaalisuus kertoo muuttujien hyödyllisyydestä toistensa ennustamisessa. Esitelyä teoriaa sekä testejä hyödynnetään vektoriautoregressiivisen mallin luomiseen. Mallia varten datan tulee olla stationaarista, joten mallinnuksessa hyödynnetään differentoitua dataa. Tällä pystytään poistamaan datasta mahdollinen kausivaihtelu. Mallinnus suoritetaan Matlab-ohjelmistolla ja apuna käytetään Mathworksin Econometrics Toolbox -laajennusta.

Malli rakennetaan kolmelle tarkasteltavalle aikasarjalle, ja se ottaa huomioon kaksi edeltävää datapistettä, tässä tapauksessa kahden edellisen vuoden arvot. Tarkasteltavan datan stationaarisuus mallin luontia varten tarkastetaan kolmella erilaisella yksikköjuuritestillä. Käytetyt testit ovat laajennettu Dickey–Fuller-testi, Kwiatkowski–Phillips–Schmidt–Shin-testi sekä Phillips–Perron-testi. Sarjojen yhteisintegroituvuutta testataan kahdella eri testillä, jotka ovat Engle–Granger–testi ja Johansenin testi. Näiden lisäksi Grangerin kausaalisuutta testataan Grangerin kausaalisuustestillä.

Vektoriautoregressiivisen mallin tarkastelussa havaittiin, että sarjat ennustavat itseään melko hyvin, mutta ne eivät juuri vaikuta toistensa ennustamiseen. Yhteisintegroituvuustesteillä differentoidulle datalle havaittiin erityisesti energian kokonaiskulutuksen vaikuttavan bruttokansantuotteeseen sekä hiilidioksidipäästöjen suuruuteen. Alkuperäiselle datalle tehtyjen testien perusteella nämä eivät kuitenkaan juuri vaikuta toisiinsa. Voidaan siis ajatella peräkkäisten arvojen erotuksien olevan riippuvaisia toisistaan tässä tapauksessa. Grangerin kausaalisuustestin avulla ei havaita merkittävää vaikutusta sarjojen välillä toisiinsa. Saatujen tuloksien p-arvoista voidaan kuitenkin päätellä hiilidioksidikaasupäästöjen suuruuden vaikuttavan bruttokansantuotteeseen merkittävästi.

Avainsanat: aikasarja-analyysi, VAR, yhteisintegroituvuus, Grangerin kausaalisuus, stationaarisuus

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

SISÄLLYSLUETTELO

1	Johdanto	1
2	Aikasarja-analyysi	2
2.1	Stationaarisuus	2
2.2	Autoregressiivinen malli	3
2.3	Vektoriautoregressiivinen malli	3
2.4	Yksikköjuuritestit	4
2.4.1	Yksikköjuuri	4
2.4.2	Dickey–Fuller-testi	4
2.4.3	Laajennettu Dickey–Fuller-testi	5
2.4.4	Kwiatkowski–Phillips–Schmidt–Shin-testi	5
2.4.5	Phillips–Perron-testi	5
2.5	Integrointijärjestys	6
2.6	Yhteisintegroituvuus	6
2.6.1	Engle–Granger-testi	7
2.6.2	Johansenin testi	7
2.7	Grangerin kausaalisuus	7
3	Mallintaminen	9
3.1	Aikasarjojen stationaarisuus	10
3.2	Aikasarjojen yhteisintegroituvuus	11
3.3	Vektoriautoregressiivisen mallin rakentaminen	13
3.4	Grangerin kausaalisuus	16
4	Tulokset	17
5	Kirjallisuudessa	18
6	Yhteenveto	19
	Lähteet	20
	Liite A Alkuperäinen data	22

LYHENTEET JA MERKINNÄT

ADF	Laajennettu Dickey-Fuller (<i>augmented Dickey-Fuller</i>)
AR	Autoregressiivinen (<i>autoregressive</i>)
DF	Dickey-Fuller
EG	Engle-Granger
IID	Riippumaton ja identtisesti jakautunut (<i>independent and identically distributed</i>)
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
MLE	Suurimman todennäköisyyden estimointimenttely (<i>maximum likelihood estimation</i>)
OLS	Pienimmän neliösumman menetelmä (<i>ordinary least squares</i>)
PP	Phillips-Perron
VAR	Vektoriautoregressiivinen (<i>vector autoregressive</i>)

1 JOHDANTO

Kasvihuonekaasupäästöjen, erityisesti hiilidioksidipäästöjen tiedetään olevan yksi suurimmista ilmastonmuutokseen vaikuttavista tekijöistä. Lisääntynyt kulutus kasvattaa kasvihuonekaasujen määrää ilmakehässä ja vauhdittaa ilmastonlämpenemistä, joka on tällä hetkellä yhteiskunnan vakavin ongelma. Hidastaakseen ilmastonlämpenemistä useat valtiot ovat sitoutuneet pienentämään päästöjään. Yleisessä tiedossa on, että talouskasvu on globaalisti lisännyt ympäristöhaittoja merkittävästi. Keskimääräisesti kulutus kasvaa tulojen kasvaessa ja kulutuksen kasvaessa päästöt luonnollisesti kasvavat. Jotta päästöjä pystyttäisiin pienentämään, on tärkeää tunnistaa päästöjen suuruuteen vaikuttavat tekijät sekä sitoutua yksilöityihin tavoitteisiin.

Tässä työssä tarkastellaan hiilidioksidikaasupäästöjen, energian kulutuksen ja bruttokansantuotteen yhteyttä toisiinsa Suomen tasolla hyödyntäen vektoriautoregressiivista mallia sekä esitellään tiivistetysti aikasarja-analyysin perusteet yleisellä tasolla ja tutustutaan käytettyihin menetelmiin tarkemmin. Työn tarkoituksena on rakentaa tarkasteltavasta datasta VAR-malli sekä tarkastella tutkittavien suureiden vaikutuksia toisiinsa erinäisillä testeillä. Tarkasteltava data on kerätty vuosilta 1975-2016. Tarkastelussa hyödynnetään tietoa vuosittaisesta väkiluvusta, jotta sarjat olisivat vertailukelpoisia keskenään.

Luvussa 2 käsitellään aikasarja-analyysin perusteita, kerrotaan mitä stationaarisuus ja yksikköjuuri tarkoittavat sekä esitellään näiden selvittämiseen käytettyjä testejä. Yksikköjuuren olemassaolon selvitykseen käytetyt testit ovat Dickey–Fuller-testi, Kwiatkowski–Phillips–Schmidt–Shin-testi ja Phillips–Perron-testi. Näiden lisäksi luvussa 2 käsitellään yhteisintegroituvuus, siihen liittyviä testejä sekä Grangerin kausaalisuus. yhteisintegroituvuutta testataan Engle–Granger-testillä ja Johansenin testillä. Grangerin kausaalisuutta testataan Grangerin kausaalisuustestillä. Luku 3 sisältää alkuperäisen datan esittelyn, käytetyt testit sekä niiden tulokset ja VAR-mallin luomisen sekä esittelyn. VAR-mallin luomisen jälkeen saadut tulokset esitellään luvussa 4 ja työn tuloksia verrataan muuhun kirjallisuuteen luvussa 5. Työn lopussa on yhteenveto sekä alkuperäinen data liitteenä.

2 AIKASARJA-ANALYYSI

Aikasarjalla tarkoitetaan ajassa t mitattuja perättäisiä havaintoja. Tyypillisesti aikasarjaa merkitään $X = X_1, X_2, \dots$, jossa perättäiset X arvot ovat havaintoja [1]. Näiden analysoinnin tavoitteena on selittää mahdollisen vaihtelun ja relaation syitä sekä ennustaa tulevaa. Mallinnustapoja on olemassa monia ja niiden tyypillisiä käyttötarkoituksia ovat esimerkiksi markkinoihin, talouteen sekä energiaan liittyvät mallinnukset. Aikasarjat itsessään voivat olla jatkuvia tai diskreettejä. [2]

2.1 Stationaarisuus

Stationaarisuudella tarkoitetaan paikallaan olevaa prosessia. Tällöin prosessin todennäköisyysjakauma ei muutu ajan kuluessa, eli sen varianssi ja keski-arvo eivät muutu, eikä sillä ei ole niin kutsuttua kausivaihtelua [3].

Määritelmä 2.1. Olkoon X_t aikasarja, jossa t on tarkasteltava ajanjakso väliltä $[1, T]$ ja olkoon $X_{t+\Delta}$ aikasarja, jossa t on tarkasteltava ajanjakso väliltä $[1, T+\Delta]$. Tällaisia aikasarjoja, joiden todennäköisyysjakaumat ovat yhtä suuret eli

$$P_{X_t}(x) = P_{X_{t+\Delta}}(x), \quad (2.1)$$

kutsutaan stationaarisiksi prosesseiksi. Tällöin ehdollinen yhteisjakauma ei ole riippuvainen ajasta.

Tietyin havainnon riippuvuus aiemmin tehtyjen havaintojen arvoihin säilyy samana [2]. Stationaarisuus itsessään on tärkeä oletus tilastollisia päätelmiä ja ennusteita tehdessä. Stationaarisuutta voidaan testata erilaisilla testeillä, jotka perustuvat yksikköjuuren löytämiseen. Yksikköjuuresta ja siihen liittyvistä testeistä lisää luvussa 2.4.

Monet olemassaolevista aikasarjoista eivät ole heti stationaarisia, vaan ne saattavat sisältää esimerkiksi kausivaihtelua tai trendejä. On kuitenkin olemassa lukuisia menetelmiä, joiden avulla pystytään muuttamaan ei-stationaarinen data stationaarisiksi. Mahdollisia käsittelytapoja ovat differensointi, varianssin stabiilisuus lineaarisen regression avulla sekä logaritointi. [2] Differensoimalla eli laskemalla peräkkäisten havaintojen erotuksia aikasarjasta pystytään poistamaan kausivaihtelua, kun taas ottamalla logaritmi pystytään pienentämään varianssin muutosta sekä eksponentiaalista trendiä. [3] Näihin käsittelytapoihin liittyy luonnollisesti myös epäkohtia. Differensoinnin lopettaminen sekä varianssin suuruus saattavat muodostua ongelmaksi.

2.2 Autoregressiivinen malli

Autoregressiivinen (AR) malli on yleisesti käytetty malli, minkä ennustukset pohjautuvat aiempien havaintojen perusteella. Tällöin tarkasteltavien suureiden tulee vaikuttaa toisiinsa eli korreloida keskenään. Yksinkertaisuudessaan malli on lineaarinen regressio aiempien arvojen pohjalta. [2]

Määritelmä 2.2. Olkoon

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \epsilon_t, \quad (2.2)$$

missä c on vakio, $\phi_1 \dots \phi_p$ ovat mallin parametrit, y_{t-1}, \dots, y_{t-p} ovat viivefunktioita, p on kertaluku, t on tarkasteltava ajanjakso väliltä $[1, T]$ ja ϵ_t on virhetermi. [1]

Virhetermi ϵ_t kuvaa valkoista kohinaa, joka kuvaa satunnaisten arvojen riippumattomuutta toisistaan. [1] Käsite autoregressiivinen tarkoittaa jokaisen tarkasteltavan suureen olevan mallinnettu aiempien arvojen funktiona. Kertaluku kuvaa kuinka pitkälle ajassa taaksepäin malli pohjautuu. Ensimmäisen kertaluvun mallia merkitään notaatiolla AR(1). Tämä tarkoittaa ennusteiden pohjautuvan takautuvasti ainoastaan yhden ajanjakson t päähän. Vastaavasti toisen ja kolmannen kertaluvun malleilla AR(2) ja AR(3), ennusteet pohjautuvat kahden tai kolmen ajanjakson t päähän. Näin ollen kertaluvun kasvaessa ennustuksen tarkkuus perustuu pidemmälle ajanjaksolle ja sen voidaan ajatella tarkentuvan. AR-mallia käytetään vain yhden sarjan tapauksissa.

2.3 Vektoriautoregressiivinen malli

Vektoriautoregressiivistä (*vector autoregressive*) mallia eli VAR-mallia käytetään kuvaamaan kahden tai useamman tarkasteltavan suureen vaikutusta toisiinsa ajan muuttuessa. VAR-mallin voidaan ajatella olevan yleistys AR-mallista, sillä se tunnistaa useamman aikasarjan vaikutuksen toisiinsa [1]. Tällaista mallia kutsutaan kaksisuuntaiseksi malliksi. VAR-mallissa käytettyjen sarjojen tulee olla stationaarisia.

Määritelmä 2.3. Olkoon

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \epsilon_t, \quad (2.3)$$

jossa c on vakio, Φ_1, \dots, Φ_p ovat autoregressiiviset $n \times n$ kertoimet parametreille kertaluvulla p , y_{t-1}, \dots, y_{t-p} ovat viivefunktioita ja ϵ_t on $1 \times n$ -virhetermi. Tarkasteltava ajanjakso merkitään muuttujan t siten, että $t = 1, \dots, T$. Tällöin väli $[1, T]$ kattaa koko tarkasteltavan ajanjakson annetulla havaintojen välillä. Kaikki muuttujat kerätään vektoriin y_t , jonka pituus on tarkasteltavien muuttujien määrä k . [1]

VAR-malli on erityisesti hyödyllinen käsiteltävän datan riippuvuussuhteiden tarkasteluun. Sen avulla saadaan helposti varianssijohdanto sekä impulssivaste. Varianssijohdanto pitää sisällään

yksittäisen suureen muutoksista aiheutuvat erot muihin suureisiin, kun taas impulssivaste kertoo kuinka odottamaton muutos yhdessä tarkasteltavassa suureessa vaikuttaa muihin tarkasteltaviin suureisiin eri aikajaksoilla. [4]

2.4 Yksikköjuuritestit

Yksikköjuuritesteillä pystytään testaamaan aikasarjan paikkansapitävyyttä [5]. Yksikköjuuri (*unit root*) itse on aikasarjan ominaisuus, joka tekee siitä epästationaarisen. Testeissä käytetty nollahypoteesi määritellään yleisesti yksikköjuuren olemassaolona ja vaihtoehdoisen hypoteesin määritelmä riippuu käytetystä testistä. [6]

2.4.1 Yksikköjuuri

Yksikköjuuri on tilastollisessa analyysissä käytetty stokastisten prosessien ominaisuus [2]. Sitä kutsutaan myös yksikköjuuriprosessiksi sekä erojen stationaariseksi prosessiksi. Yksikköjuuresta voi seurata epäkohtia aikasarjojen mallintamiseen ja tämän vuoksi ennustaminen hankaloituu. Perustasolla aikasarja voidaan kirjoittaa monomien (yksiterminen polynomi) sarjaksi. Jokainen monomi vastaa juurta ja mikäli jonkin näistä itseisarvon suuruus on yksi, se on yksikköjuuri [6].

Määritelmä 2.4. Olkoon

$$y_t = \phi_t y_{t-1} + \epsilon_t \quad (2.4)$$

stokastinen prosessi, missä $|\phi_i| \leq 1$. Jos $|\phi_i| = 1$, kyseessä on yksikköjuuri. [6]

Stokastisuus tarkoittaa satunnaisotoksen kuvaavan hyvin prosessia [2]. Mikäli $\phi_i = 1$, tilannetta kutsutaan satunnaiskuluksi. Tilanteessa, jossa $|\phi_i|$ on täsmälleen yksi, aikasarja ei ole stationaarinen. Muulloin kyseessä on stationaarinen sarja.

2.4.2 Dickey–Fuller-testi

Nollahypoteesia H_0 , jonka mukaan yksikköjuuri sisältyy autoregressiiviseen malliin, testataan Dickey-Fuller – testillä. Vaihtoehtoinen hypoteesi H_1 voidaan valita tilanteen mukaan sopivaksi, mutta yleisimmin käytetään stationaarisuutta. Testi on nimetty tilastotieteilijöiden David Dickey'n ja Wayne Fullerin mukaan, jotka kehittivät testin vuonna 1979. [5]

Testi pohjautuu peräkkäisten havaintojen eroon. Regressiomalli voidaan kirjoittaa

$$\Delta y_t = (\phi - 1)y_{t-1} + \epsilon_t. \quad (2.5)$$

Käyttämällä notaatiota $\beta = \phi - 1$ mallista saadaan lineaarinen regressioyhtälö. Testi mittaa nyt ϕ :n sijasta β :n suuruutta. Kyseessä on yksisuuntainen testi hypoteeseilla

$$\begin{aligned} H_0 : \beta &= 0 \text{ (vastaa } \phi = 1) \\ H_1 : \beta &< 0 \text{ (vastaa } \phi < 1). \end{aligned}$$

Tarkastelussa ei voida käyttää tavallista t-testiä, sillä nollahypoteesin ollessa totta, t-kerroin ei noudata normaalijakaumaa. Tällöin tulee käyttää vertailussa hyödyksi τ -jakaumaa. [4, 7] Mikäli määritetty τ :n arvo on pienempi kuin taulukoitu kriittinen arvo, vaihtoehtoinen hypoteesi jää voimaan. Muussa tapauksessa nollahypoteesi hyväksytään ja yksikköjuuri on olemassa. [4] Tällöin aikasarja ei ole stationaarinen.

2.4.3 Laajennettu Dickey–Fuller-testi

Laajennettu Dickey-Fuller –testi (*augmented Dickey-Fuller test*) eli ADF-testi on nimensä mukaan yksinkertaista DF-testiä kattavampi versio. Perusteet testauksen taustalla ovat täysin samat kuin suppeammassa, vain AR(1) -mallille pätevässä testissä. [4] ADF-testi mahdollistaa yksikköjuuren olemassaolon selvityksen suuremmille ja monimutkaisemmille malleille.

2.4.4 Kwiatkowski–Phillips–Schmidt–Shin-testi

Kwiatkowski-Phillips-Schmidt-Shin –testi kehitettiin vuonna 1992 täydentämään DF-testiä, ja se nimettiin tekijöidensä mukaan. Merkittävin ero KPSS-testin ja useimpien yksikköjuuri testien välillä on nollahypoteesin asettaminen päinvastaiseksi [8, 9]. KPSS-testissä nollahypoteesiksi asetetaan stationaarisuus ja vaihtoehtoiseksi hypoteesiksi yksikköjuuren olemassaolo.

KPSS-testi perustuu lineaariseen regressioon ja pienimmän neliösumman menetelmään. PNS-menetelmässä regressiokertoimien estimaattorit määrätään minimoimalla jäännös- eli virhetermien ϵ_t neliösumma regressiokertoimien suhteen. Minimointi voidaan tehdä derivoimalla neliösumma regressiokertoimien suhteen ja merkitsemällä derivaatat nollassi. [10] Aikasarja ilmaistaan deterministisen trendin, satunnaiskulun ja virhetermin summana seuraavasti

$$y_t = \beta t + r_t + \epsilon_t, \quad (2.6)$$

missä βt on deterministinen trendi, $r_t = r_{t-1} + u_t$ satunnaiskulku, $u_t = \text{IID}(0, \sigma_u^2)$ ja ϵ_t stationaarinen virhetermi. [8] Merkintä IID tarkoittaa riippumatonta ja identtisesti jakautunutta tilastollista jakaumaa. IID on yksi regressioanalyysin perusoletuksista.

Tällöin hypoteeseiksi asetetaan

$$\begin{aligned} H_0 : \sigma_u^2 &= 0 \text{ (vastaa } \phi < 1) \\ H_1 : \sigma_u^2 &> 0 \text{ (vastaa } \phi = 1). \end{aligned} \quad [9]$$

2.4.5 Phillips–Perron-testi

Phillips-Perron –testi on myös yleistys DF-testistä. PP-testi arvioi yksikköjuuren olemassaoloa yksimuuttujaisessa aikasarjassa y . Testi pohjautuu malliin

$$y_t = c + \beta t + a y_{t-1} + \epsilon_t, \quad (2.7)$$

missä βt on deterministinen trendi, a on autoregressiivinen kerroin, y_{t-1} on ensimmäinen viive-funktio ja ϵ_t stationaarinen virhetermi.

PP-testissä nollahypoteesiksi H_0 asetetaan yksikköjuuren olemassaolo. [11] Tällöin yhtälön 2.7 autoregressiivinen kerroin $a = 1$. Vaihtoehtoinen hypoteesi H_1 kertoo sarjan stationaarisuudesta.

2.5 Integrointijärjestys

Integrointijärjestys (*order of integration*) $I(d)$ on yhteenvetotilasto, jota käytetään kuvaamaan aikasarjan yksikköjuuri-prosessia. Merkinnällä d kuvataan differensointien lukumäärää eli kertalukua. Integrointijärjestys kertoo vähimmäismäärän differensoinneille, joita tarvitaan stationaarisen sarjan saavuttamiseksi. [5]

Määritelmä 2.5. Olkoon y_t ei-stationaarinen prosessi. Prosessista saadaan stationaarinen prosessi x_t differensoimalla sitä kertaluvulla d , jolloin

$$x_t = \Delta^d y_t \quad (2.8)$$

on stationaarinen prosessi.

Yleisesti stationaarisuus saavutetaan käyttämällä integrointijärjestyksiä $I(0)$ tai $I(1)$. Tällöin sarjaa differensoidaan enintään kerran. [12] Korkeamman kertaluvun differensoinnit ovat harvinaisia, mutta kuitenkin mahdollisia.

2.6 Yhteisintegroituvuus

Yhteisintegroituvuudella (*cointegration*) tarkoitetaan kahden tai useamman sarjan riippuvuutta toisistaan pitkällä aikavälillä. Ollakseen yhteisintegroituja, sarjojen tulee itsessään olla ei-stationaarisia. Tällöin sarjojen lineaarikombinaatio on stationaarinen. [13, 14] Lineaarikombinaation integrointijärjestyksen tulee siis olla pienempi kuin alkuperäisten sarjojen integrointijärjestysten. Esimerkiksi kaksi integrointijärjestyksen $I(1)$ sarjaa ovat yhteisintegroituja, mikäli niiden lineaarikombinaatio on integroitu järjestyksellä $I(0)$ [12].

Määritelmä 2.6. Olkoon $y_{i,t}$ aikasarja, missä $i = 1, 2, \dots, n$ on kertaluvulla d integroituja muuttujia ja olkoon v_i vektori, missä $i = 1, 2, \dots, n$ on vakioita. Mikäli näiden välillä on olemassa lineaarikombinaatio

$$\eta_t = \sum_{i=1}^n v_i y_{i,t} \quad (2.9)$$

kertaluvulla $d - b$, kun $0 < b = d$, niin aikasarjan $y_{i,t}$ sanotaan olevan yhteisintegroitu kertaluvulla (b, d) . Kerrointa v_i kutsutaan yhteisintegrointivektoriksi ja summaa $\sum_{i=1}^n v_i y_{i,t}$ kutsutaan yhteisintegroituneeksi kombinaatioksi. [12]

Määritelmään 2.6 liittyy erityistapaus, jossa $d = b$. Tällöin lineaarikombinaatio on stationaarinen. Yhteisintegroituvuudella ja siihen liittyvillä testeillä pyritään selvittämään muuttujien herk-

kyyttä suhteessa keskiarvoon tietyllä ajanjaksolla. Riippuvuus useamman aikasarjan välillä on usein havaittavissa suoraan niiden kuvaajista, mutta tämän olemassaolo varmistetaan testaamalla.

2.6.1 Engle–Granger-testi

Aikasarjojen yhteisintegroituvuutta voidaan testata Engle-Granger –testillä. Käsiteltävien aikasarjojen tulee olla ei-stationaarisia. Tämä voidaan helposti testata luvussa 2.4 esitellyillä yksikköjuuritesteillä. Yksikköjuuren olemassaolon lisäksi sarjojen integrointijärjestyksen kertaluvun d tulee olla yksi. Tällöin sarjojen lineaarikombinaatio on stationaarinen. [1, 14]

Lause 2.7. Jos aikasarjat y_t ja x_t ovat ei-stationaarisia ja niiden integrointijärjestyksen kertaluku $d = 1$, niin lineaarikombinaation $y_t - \beta x_t = u_t$ täytyy olla stationaarinen, kun $\beta \in \mathbb{R}$.

EG-testin nollahypoteesiksi H_0 asetetaan ettei yhteisintegroituvuutta ole olemassa. Tällöin vaihtoehtoinen hypoteesi H_1 kertoo riippuvuuden olemassaolosta. Testillä voidaan testata vain kahden sarjan riippuvuutta toisistaan [14].

2.6.2 Johansenin testi

Johansenin testiä käytetään selvittämään yhteisintegroituvuutta kahden tai useamman sarjan välillä. Tämä on testin merkittävin ero EG-testiin, jonka avulla pystytään selvittämään vain kahden aikasarjan riippuvuutta toisistaan. Testi perustuu suurimman todennäköisyyden estimointimenettelyyn (*maximum likelihood estimation*). Tällä tarkoitetaan menetelmää, joka maksimoi todennäköisyydet sarjojen parametrien suhteen. [14] Testillä pystytään testaamaan sekä riippuvuuksien olemassaoloa että niiden suhteiden suuruutta.

Testi on mahdollista suorittaa kahdella tavalla, joko hyödyntämällä jälkeä (*trace*) tai hyödyntämällä ominaisarvojen maksimeja (*maximum eigen value*) [15]. Tässä työssä käytetään ominaisarvojen avulla tapahtuvaa testiä, joten jäljen hyödyntäminen jätetään tarkastelussa huomiotta. Molemmilla tavoilla nollahypoteeseiksi asetetaan ettei yhteisintegroituvuutta ole. Mikäli tarkastelussa on yhteensä n sarjaa, Johansenin testissä hypoteeseiksi asetetaan

$$\begin{aligned} H_0 : r' < n, \quad r' = 1, 2, \dots \\ H_1 : r' + 1, \quad r' = 1, 2, \dots, \end{aligned}$$

jossa r' tarkoittaa tarkasteltavista sarjoista muodostetun matriisin astetta (*rank*). Matriisin aste kertoo lineaarisesti riippumattomien rivien tai sarakkeiden eli yhteisintegrointivektoreiden v_i lukumäärän. Testin edetessä ensimmäinen nollahypoteesin hylkäämättä jättäminen otetaan v_i lukumäärän arviona [14, 15].

2.7 Grangerin kausaalisuus

Eräs VAR-mallin tärkeimpiä käyttötarkoituksia on ennustaa tulevaa. Grangerin kausaalisuus kertoo onko jostakin yksittäisestä tai useammasta muuttujasta hyötyä toisen muuttujan ennustamisessa.

Se tutkii näin kahden muuttujan välistä kausaliteettia hyödyntämällä empiiristä dataa korrelaatiomallien löytämiseen. [3, 12]

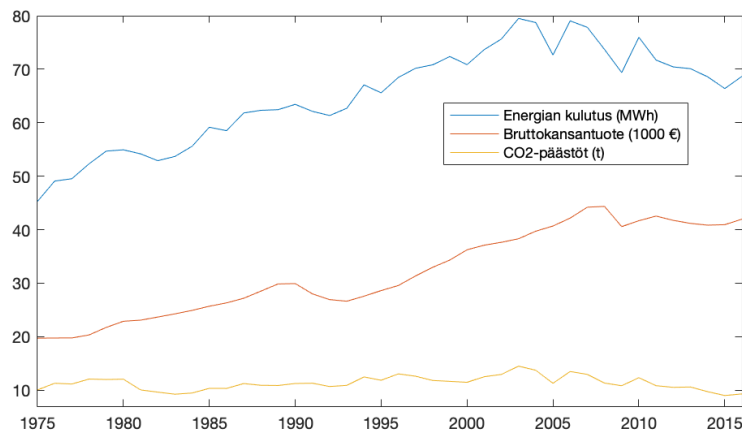
Määritelmä 2.8. Muuttujan X sanotaan olevan Granger-kausaalinen muuttujan Y kanssa, mikäli muuttujan Y ennustaminen tarkentuu hyödyntämällä molempien muuttujien X ja Y aikaisempia arvoja verrattuna ainoastaan muuttujan Y arvojen hyödyntämiseen. [16]

Grangerin kausaalisuutta tulee huomioida, ettei se tarkoita muuttujien välillä olevan todellista kausaalisuutta. Grangerin kausaalisuus kuvaa ainoastaan muuttujien välistä yhteyttä ennustuksessa [12]. Nollahypoteesi H_0 olettaa, ettei edeltävät arvot selitä syntyvää vaihtelua. Toisin sanoen nollahypoteesi tarkoittaa, ettei sarjojen välillä ole Grangerin kausaalisuutta. Vaihtoehtoinen hypoteesi H_1 olettaa tämän olemassaolon [17].

3 MALLINTAMINEN

Tässä työssä tarkastellaan diskreettejä aikasarjoja ja havaintojen välisenä aikana käytetään yhtä vuotta. Tarkasteluväli voisi olla lyhyempi, mutta tilastot julkaistaan vain vuositasolla. Tarkastelun kohteena ovat Suomen hiilidioksidikaasupäästöt [18], energian kulutus [19] sekä bruttokansantuote (BKT) [20] vuosina 1975-2016. Suureiden välisten yhteyksien tarkastelussa hyödynnetään myös väkilukua [21]. Alkuperäisen datan tunnusluvut ovat esillä yhteenvetona taulukossa 3.1.

Merkintä *ekk* tarkoittaa energian kokonaiskulutusta megawattitunneissa, *bkt* bruttokansantuotetta euroissa ja *co₂* hiilidioksidikaasupäästöjen kokonaismäärää tonneissa. Data esitellään graafisesti kuvassa 3.1. Laajemmin se on saatavilla liitteessä 1. Tarkastelu suoritetaan Matlab-ohjelmistolla ja apuna käytetään Mathworksin Econometrics Toolbox -laajennusta. Saadut tulokset esitellään luvussa 4.



Kuva 3.1. Energian kulutus, bruttokansantuote ja CO₂-päästöt asukasta kohti. [18, 19, 20]

Taulukko 3.1. Yhteenveto alkuperäisestä datasta asukasta kohti tunnusluvuin esitettynä.

	ekk	bkt	co ₂
keskiarvo	65.0802	32037	11.3262
varianssi	82.3120	64292000	1.6319
keskihajonta	9.0726	8018.2	1.2774
minimi	45.298	19720	9.0017
maksimi	79.494	44354	14.495
mediaani	66.73	29894	11.267

3.1 Aikasarjojen stationaarisuus

VAR-mallia varten aikasarjan tulee olla stationaarinen. Tämän ominaisuuden selvittämiseksi tulee tehdä yksikköjuuritestejä, joita on esitelty tarkemmin kappaleessa 2.4. Tuodaan käsittelyä varten data Matlabiin käyttämällä komentoa

```
values = xlsxread(filename),
```

jossa tiedostossa olevat arvot tallentuvat numeeriseen muuttujaan *values*. Eri tilastolliset suureet ovat eri sarakkeissa ja näin dataa on helppo käsitellä suoritetuissa testeissä.

Testataan ensimmäisenä yksikköjuuren olemassaoloa ADF-testillä. ADF-testi suoritetaan komennolla

```
[h,pValue] = adftest(Y),
```

missä *h* on testin arvo, *pValue* käytetty riskitaso ja *Y* aikasarja. Riskitaso kuvaa kuinka suuri riski ollaan valmiita ottamaan. Tietty riskitaso *p* tarkoittaa saadun lopputuloksen olevan *p* % todennäköisyydellä virheellinen.

KPSS-testi saadaan käyttämällä komentoa

```
[h,pValue] = kpsstest(Y),
```

missä sijoitettavat parametrit ovat täsmälleen samat kuin ADF-testissä.

Vastaavasti PP-testi suoritetaan komennolla

```
[h,pValue] = pptest(Y).
```

Yksikköjuuritestistä saadut arvot ja testien p-arvot ovat esillä taulukossa 3.2. Nämä arvot on laskettu käyttämällä alkuperäistä dataa asukasta kohti.

Taulukko 3.2. Yksikköjuuritestien tulokset ja p-arvot alkuperäiselle datalle.

		ADF	KSS	PP
ekk	h	0	1	0
	p-arvo	0.9134	0.01	0.9134
bkt	h	0	1	0
	p-arvo	0.9981	0.01	0.9981
co ₂	h	0	1	0
	p-arvo	0.5111	0.01	0.5111

ADF-testin tulos $h = 0$ tarkoittaa testin epäonnistuvan hylkäämään nollahypoteesi, eli nollahypoteesi H_0 jää voimaan [4]. Tällöin sarjalla on yksikköjuuri ja sarja ei ole stationaarinen. KPSS-testissä tulos $h = 0$ tarkoittaa nollahypoteesin hylkäämistä eli vaihtoehtoinen hypoteesi jää voimaan [9]. Tällöin sarjalla on yksikköjuuri. PP-testissä testisuureen h saadessa arvokseen nolla, testi epäonnistuu hylkäämään nollahypoteesin H_0 ja tällöin se jää voimaan [11]. Tällöin sarjalla on yksikköjuuri eikä se siten ole stationaarinen.

Taulukoitujen arvojen mukaan ADF-testi tai PP-testi eivät kumpikaan anna luotettavia tuloksia hiilidioksidikaasupäästöjen stationaarisuudesta. Tehtyjen testien mukaan kuitenkin energian kokonaiskulutus sekä bruttokansantuote vaikuttavat stationaarisilta. Tarkastellaan kuitenkin sarjojen stationaarisuutta logaritmoimalla mittausdata. Taulukossa 3.3 esitellään logaritmoinnin jälkeen saadut tulokset. Logaritointi kasvattaa luottamusta energiankulutuksen ja bruttokansantuotteen stationaarisuuteen.

Taulukko 3.3. Yksikköjuuritestien tulokset ja p-arvot logaritmoinnin jälkeen.

		ADF	KSS	PP
ekk	h	0	1	0
	p-arvo	0.9651	0.01	0.9651
bkt	h	0	1	0
	p-arvo	0.999	0.01	0.999
CO ₂	h	0	1	0
	p-arvo	0.5572	0.01	0.5572

Valitettavasti logaritmoinnin jälkeen hiilidioksidikaasupäästöt eivät vaikuta stationaariselta ADF- tai PP-testin mukaan. Tarkastellaan stationaarisuutta käyttämällä datan differensointia hyödyksi. Näin saadut tulokset esitellään taulukossa 3.4.

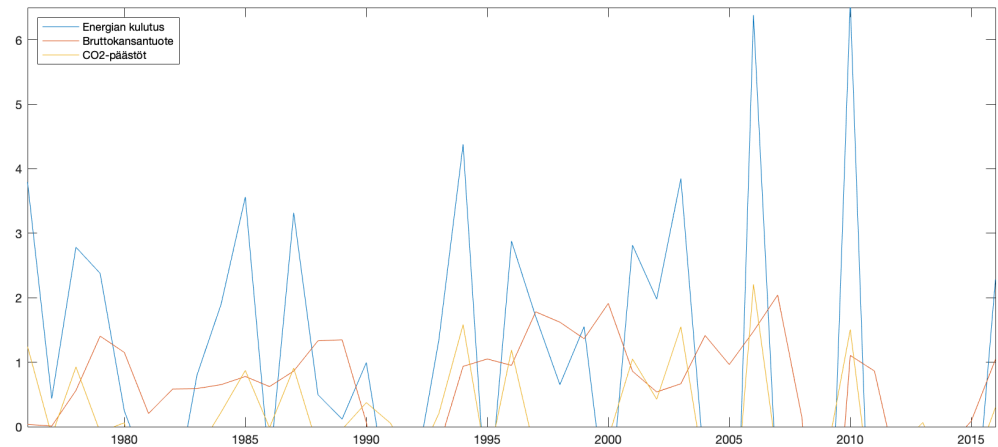
Taulukko 3.4. Yksikköjuuritestien tulokset ja p-arvot differensoinnin jälkeen.

		ADF	KSS	PP
ekk	h	1	1	1
	p-arvo	0.001	0.01	0.001
bkt	h	1	1	1
	p-arvo	0.001	0.0962	0.001
CO ₂	h	1	1	1
	p-arvo	0.001	0.01	0.001

Peräkkäisten havaintojen erotuksia hyödyntämällä sarjat saavat kaikilla käytetyillä yksikköjuuritestillä luotettavan tuloksen niiden stationaarisuudesta. Tämän vuoksi jatkossa käytetään sarjojen differensointeja hyödyksi sarjojen analysoinnissa.

3.2 Aikasarjojen yhteisintegroituvuus

Kuten kuvasta 3.1 voidaan havaita, aikasarjat näyttävät korreloivan keskenään. Erityisesti energian kulutus ja bruttokansantuote vaikuttavat kasvavan suhteessa. CO₂-päästöt pysyvät melko tasaisena ajan kuluessa. Tarkastellaan aikasarjojen riippuvuuksia toisistaan hyödyntämällä differensoitua dataa sekä graafisesti että testien avulla. Kuvassa 3.2 esitetään käsiteltävä data differensoinnin jälkeen. Siitä erottuvat sarjojen yhteiset piikit sekä laskut selkeämmin. Graafisesta esityksestä voidaan näin ollen arvella aikasarjojen riippuvan toisistaan.



Kuva 3.2. Energian kulutus, bruttokansantuote ja CO₂-päästöt asukasta kohti differentoituna.

Testataan sarjojen yhteisintegroituvuutta alkuperäisellä datalla ilman differensointia. Suoritetaan ensimmäiseksi EG-testi, jolla voidaan tarkastella ainoastaan kahden sarjan välistä riippuvuutta. [14] Tämä testi voidaan suorittaa komennolla

$$[h, pValue] = \text{egcitest}(Y).$$

Johansenin testin avulla voidaan tarkastella useamman sarjan riippuvuutta. Johansenin testi voidaan suorittaa vastaavasti komennolla

$$[h, pValue] = \text{jcitest}(Y).$$

Yhteisintegroituvuuden testien tulokset alkuperäiselle datalle on näkyvillä taulukossa 3.5. Taulukossa sarjojen riippuvuuksien välinen arvo $h = 1$ tarkoittaa yhteisintegroituvuuden olemassaoloa eli vaihtoehdoisen hypoteesin hyväksyntää, kun taas arvo $h = 0$ tarkoittaa ettei sitä ole olemassa ja tällöin nollihypoteesi jää voimaan.

Taulukko 3.5. Yhteisintegroituvuustestien tulokset alkuperäiselle datalle.

		rank	h	p-arvo
EG	ekk, bkt		0	0.4037
	ekk, CO ₂		0	0.8166
	bkt, CO ₂		0	0.9693
Johansen	ekk, bkt	0	1	0.0193
	ekk, bkt	1	1	0.0396
	ekk, CO ₂	0	0	0.1682
	ekk, CO ₂	1	1	0.0160
	bkt, CO ₂	0	0	0.1178
	bkt, CO ₂	1	0	0.1608
	ekk, bkt, CO ₂	0	0	0.1197
	ekk, bkt, CO ₂	1	0	0.1832
	ekk, bkt, CO ₂	2	1	0.0189

Suoritetaan vastaavat testit myös differensoidulle datalle, jota hyödynnetään stationaarisuutensa vuoksi VAR-mallin rakentamiseen. Näin saadut tulokset esitellään taulukossa 3.6.

Taulukko 3.6. Yhteisintegroituvuustestien tulokset differensoidulle datalle.

		rank	h	p-arvo
EG	ekk, bkt		1	0.001
	ekk, CO ₂		1	0.001
	bkt, CO ₂		1	0.0052
Johansen	ekk, bkt	0	1	0.001
	ekk, bkt	1	1	0.001
	ekk, CO ₂	0	1	0.001
	ekk, CO ₂	1	1	0.001
	bkt, CO ₂	0	1	0.001
	bkt, CO ₂	1	1	0.001
	ekk, bkt, CO ₂	0	1	0.001
	ekk, bkt, CO ₂	1	1	0.001
	ekk, bkt, CO ₂	2	1	0.001

Kuten edellä, arvo $h = 1$ tarkoittaa yhteisintegroituvuuden olemassaoloa, kun taas arvo $h = 0$ tarkoittaa ettei sitä ole olemassa ja tällöin nollahypoteesi jää voimaan.

3.3 Vektoriautoregressiivisen mallin rakentaminen

VAR-malli voidaan rakentaa stationaarisista sarjoista, kuten luvussa 2.3 kerrotaan. Matlabissa VAR-mallin rakentaminen aloitetaan käyttämällä komentoa

$$\text{Mdl} = \text{varm}(\text{numseries}, \text{numlags}),$$

jossa *numseries* sisältää tarkasteltavien muuttujien määrän k ja *numlags* vastaavasti viiveiden lukumäärän p . Komento luo annetuilla muuttujilla VAR-mallin, johon voidaan sijoittaa tarkasteltavat aikasarjat. Tässä tapauksessa $k = 3$ ja viiveiden lukumääräksi valitaan $p = 2$. Tällöin malli ottaa huomioon kahden edellisen kuukauden mittaukset. Koska tarkasteltavien sarjojen tulee olla stationaarisia, muodostetaan VAR-malli logaritmoidusta datasta. Tämä suoritetaan Matlabissa seuraavasti

$$\text{EstMdl} = \text{estimate}(\text{Mdl}, Y),$$

jossa matriisi Y sisältää tarkasteltavat aikasarjat.

Matlab antaa autoregressiivisiksi kerroinmatriiseiksi

$$\text{Est.Mdl}\{1, 1\} = \begin{bmatrix} 0.2487 & -4.5196 \cdot 10^4 & 8.7453 \cdot 10^4 \\ 6.2875 \cdot 10^{-7} & 1.195 & -0.0029 \\ -2.4079 \cdot 10^{-6} & -0.4155 & 1.1393 \end{bmatrix}$$

sekä

$$\text{Est.Mdl}\{1, 2\} = \begin{bmatrix} 0.7002 & 3.395 \cdot 10^4 & -7.9194 \cdot 10^4 \\ -2.7311 \cdot 10^{-7} & -0.4072 & -0.0087 \\ 3.5956 \cdot 10^{-6} & 0.3486 & -0.3978 \end{bmatrix},$$

missä ensimmäisellä rivillä olevat kertoimet kertovat energiankulutuksen riippuvuuden tarkasteltavista sarjoista, toisella rivillä olevat kertoimet bruttokansantuotteen riippuvuuden tarkasteltavista sarjoista ja viimeisellä rivillä olevat kertoimet vastaavasti hiilidioksidikaasupäästöjen riippuvuuden tarkasteltavista sarjoista. Vastaavaa merkintää käytetään jatkossa. Matriisissa $\text{Est.Mdl}\{1, 2\}$ on esitetty nykyisen arvon riippuvuus takautuvasti yhdestä kuukaudesta (edellinen kuukausi vaikuttaa uuteen), kun taas matriisissa $\text{Est.Mdl}\{1, 1\}$ on esitetty kahden edeltävän aikapisteen (kuukauden) vaikutus nykyiseen arvoon.

VAR-malliin kuuluviksi vakioiksi c sekä aikariippuvuuksiksi t saadaan seuraavat arvot

$$c = \begin{bmatrix} 0 \\ 1.933 \\ 5.025 \end{bmatrix} \text{ ja } t = \begin{bmatrix} 42.9739 \\ 0.0025 \\ -0.0042 \end{bmatrix}.$$

Kovarianssimatriisiksi olettaen keskiarvon olevan nolla saadaan

$$\Sigma = \begin{bmatrix} 1.5567 \cdot 10^8 & 188.3572 & 769.7493 \\ 188.3572 & 5.9794 \cdot 10^{-4} & 6.7272 \cdot 10^{-4} \\ 769.7493 & 6.7272 \cdot 10^{-4} & 0.0050 \end{bmatrix}.$$

Hyödyntämällä Matlabin komentoa

`summarize(EstMD1),`

saadaan selville malleihin liittyvät virheet sekä p-arvot. Nämä ovat esillä yhdessä autoregressiivisten kertoimien kanssa taulukossa 3.7.

Taulukko 3.7. Saadut autoregressiiviset kertoimet, virheet sekä p-arvot.

	kerroin	virhe	p-arvo
Constant (1)	0	$1.1531 \cdot 10^6$	1
Constant (2)	1.933	2.2599	0.39229
Constant (3)	5.0205	6.5531	0.4436
AR{1}(1,1)	0.24871	0.36521	0.49586
AR{1}(2,1)	$6.2875 \cdot 10^{-7}$	$7.1577 \cdot 10^{-7}$	0.37971
AR{1}(3,1)	$-2.4079 \cdot 10^{-6}$	$2.0756 \cdot 10^{-6}$	0.24599
AR{1}(1,2)	-45196	89614	0.61402
AR{1}(2,2)	1.195	0.17563	$1.0185 \cdot 10^{-11}$
AR{1}(3,2)	-0.1554	0.5093	0.41455
AR{1}(1,3)	87453	58810	0.137
AR{1}(2,3)	-0.0028937	0.11526	0.97997
AR{1}(3,3)	1.1393	0.33423	0.0065304
AR{2}(1,1)	0.70022	0.37731	0.063477
AR{2}(2,1)	$-2.7311 \cdot 10^{-7}$	$7.3948 \cdot 10^{-7}$	0.71189
AR{2}(3,1)	$3.5956 \cdot 10^{-6}$	$2.1443 \cdot 10^{-6}$	0.093585
AR{2}(1,2)	33158	80508	0.68044
AR{2}(2,2)	-0.40724	0.15779	0.0098522
AR{2}(3,2)	0.3486	0.45755	0.44613
AR{2}(1,3)	-79194	59266	0.18147
AR{2}(2,3)	0.086335	0.11616	0.94041
AR{2}(3,3)	-0.39783	0.33683	0.23756
Trend (1)	42.974	791.52	0.9567
Trend (2)	0.0024595	0.0015513	0.11286
Trend (3)	-0.004302	0.0044984	0.34702

Taulukossa 3.7. kolme ensimmäistä constant-riviä kertovat tulokset samat tulokset kuin edellä esitettyssä matriisissa c . Nämä vakiot kuuluvat siis VAR-malliin. Vakioiden jälkeen taulukossa

esitetään AR(1)-malleja eri sarjojen kesken ja vastaavasti näiden jälkeen AR(2)-malleja. Suluissa olevat merkinnät mallien jälkeen kertovat käsiteltävän sarjan sekä termin järjestyksen. Esimerkiksi merkintä $AR\{1\}(1,1)$ tarkoittaa yhden viiveen mallia, jossa tarkasteltava sarja on ensimmäinen (energian kokonaiskulutus) ja kertoimen järjestysluku on yksi.

3.4 Grangerin kausaalisuus

VAR-mallin tärkeimpiä käyttötarkoituksia on ennustaa tulevaa. Tarkastellaan seuraavaksi juuri luodun mallin ennustusmahdollisuuksia hyödyntämällä Grangerin kausaalisuustestiä. Testi suoritetaan Matlabissa käyttämällä komentoa

$$[h, Summary] = gctest(EstMdl),$$

jossa h antaa testin tuloksen ja $Summary$ antaa tulokseksi taulukon, joka sisältää myös p-arvot.

Luodulle mallille saadut tulokset esitellään taulukossa 3.8.

Taulukko 3.8. Grangerin kausaalisuus luodulle VAR-mallille.

H_0	h	p-arvo
bkt \rightarrow ekk	0	0.9032
co ₂ \rightarrow ekk	0	0.3835
ekk \rightarrow bkt	0	0.7061
co ₂ \rightarrow bkt	0	0.9972
ekk \rightarrow co ₂	0	0.3247
bkt \rightarrow co ₂	0	0.7559

Tulos $h = 1$ hylkää nollahypoteesin, sillä sarjojen välillä havaitaan Grangerin kausaalisuus. Vastaavasti tulos $h = 0$ tarkoittaa, ettei nollahypoteesia onnistuta hylkäämään, jolloin riippuvuutta ei ole. [17]

4 TULOKSET

Luvussa 3 tehdyistä yhteisintegroituvuustesteistä voidaan havaita riippuvuuksia aikasarjojen välillä. Jos tilastollisen merkitsevyyden rajan ajatellaan olevan $p < 0.05$, voidaan havaita alkuperäisen datan, esillä taulukossa 3.5, riippuvan toisistaan ainoastaan energian kokonaiskulutuksen ja bruttokansantuotteen sekä energian kokonaiskulutuksen ja hiilidioksidikaasupäästöjen suuruuden osalta. Alkuperäiselle datalle tehtyjen testien perusteella bruttokansantuote ja hiilidioksidipäästöt eivät vaikuta toisiinsa. Vastaavat testit suoritettiin myös differensoidulle datalle ja saadut tulokset ovat esillä taulukossa 3.6. Saadut tulokset eroavat merkittävästi alkuperäiselle datalle tehdyistä testeistä, sillä nyt kaikkien sarjojen välillä havaitaan riippuvuutta. Erityisesti Johansenin testi antaa merkittäviä tuloksia, sillä se kertoo kaikkien sarjojen välillä olevan riippuvuutta merkitsevyydellä $p = 0.001$, jota voidaan pitää tilastollisesti erittäin merkitseväenä. Erot tuloksissa selittyvät differensoinnin vaikutuksesta, sillä tarkastelu suoritetaan näin peräkkäisten havaintojen erotuksille.

Kun tarkastellaan luotua VAR-mallia ja taulukoituja p-arvoja, voidaan havaita, että eniten tulevien arvojen ennustamiseen vaikuttavat kunkin sarjan omat aikaisemmat arvot. Erityisesti p-arvot puoltavat tätä AR $\{1\}$ (2,2), AR $\{1\}$ (3,3), AR $\{2\}$ (1,1) ja AR $\{2\}$ (2,2) tapauksissa. Tällöin myös mahdollisen tehdyn virheen suuruus on pieni. Näistä suurinta riippuvuutta aikaisempiin arvoihinsa kokee sarja 2, joka on bruttokansantuote. Näin ollen sarjat ennustavat itseään melko hyvin, mutta eivät juuri vaikuta toistensa ennustamiseen.

Luodulle VAR-mallille tehdystä Grangerin kausaalisuustestistä ja sen tuloksista ei voida havaita merkittävää vaikutusta sarjojen välillä toisiinsa. Toisaalta, saadut merkittävyydet eivät ole luotettavia. Tuloksista voidaan ajatella hiilidioksidipäästöjen suuruuden vaikuttavan bruttokansantuoteseen, sillä tämän kausaalisuuden hylkäämiseen merkitsevyyden $p = 0.9972$, jolloin nollassa hypoteesi voidaan hylätä olettaen vaihtoehdoisen hypoteesin olevan merkittävämpi. Grangerin kausaalisuustestillä saadut tulokset eivät kuitenkaan ole tilastollisesti kovin merkittäviä, sillä p-arvot ovat erityisen suuria.

Tulosten luotettavuus kasvaisi käytettäessä enemmän mittauspisteitä. Nyt käytetty tarkasteluväli, yksi vuosi, on melko pitkä ajanjakso VAR-mallille sekä yhteisintegroituvuustesteille. Havaintojen vaikutus toisiinsa tulisi todennäköisesti paremmin esiin, mikäli tarkasteluvälinä käytettäisiin lyhyempää ajanjaksoa, kuten yhtä kuukautta. Työssä käytetyn tarkasteluvälin pituus johtui ainoastaan datan saatavuudesta.

5 KIRJALLISUUDESSA

Yhteiskunta ja sen rakenne vaikuttavat suurelta osin saatuihin tuloksiin. Energian kokonaiskulutuksen sekä hiilidioksidikaasupäästöjen yhteyteen vaikuttaa merkittävästi tarkasteltavan maan energiantuotannon rakenne. Tilastokeskuksen mukaan Suomessa uusiutuvien energialähteiden käyttö kasvoi vuonna 2019 samalla, kun energian kokonaiskulutus vähentyi. Eriteltynä hiilen käyttö väheni 20 % ja käytettyjen polttoaineiden hiilidioksidipäästöt vähenivät 7 %, kun uusiutuvien lähteiden käyttö kasvoi prosentoin. [22] Mikäli työssä tarkasteltua tilannetta verrataan Kiinan datasta tehtyyn tutkimukseen [23], jossa on hyödynnetty pääosin Grangerin kausaalisuutta sekä laajennettua VAR-mallia, voidaan nähdä tuloksissa yhtäläisyyksiä.

Tutkimuksen mukaan [23] Grangerin kausaalisuutta havaitaan erityisesti bruttokansantuotteen sekä energian kokonaiskulutuksen välillä. Tämän lisäksi energian kokonaiskulutus vaikuttaa pitkällä aikavälillä hiilidioksidipäästöjen suuruuteen, mutta korrelaatiota ei havaita toiseen suuntaan. Vertaillen tutkimuksen [23] tuloksia tässä työssä saatuihin tuloksiin tulee huomioida, että maat eivät ole energiantuotannon osalta vertailukelpoisia. Kiinassa on tuotettu lähes 70 % käytetystä energiasta hiilivoimalla, mikä eroaa Suomen tilanteesta merkittävästi. [23]

Erään tutkimuksen tulokset [24] indikoivat pitkän aikavälin vaikutusta energian kokonaiskulutuksen, hiilidioksidikaasupäästöjen sekä taloudellisen kasvun välillä. Tässä tutkimuksessa tarkasteltavana kohdemaana oli Saudi-Arabia ja tutkimus pohjautui yhteisintegroituvuuteen. Uusiutuvien energialähteiden käyttö vähentää fossiilisten polttoaineiden kulutusta ja näin hiilidioksidipäästöjä.

6 YHTEENVETO

Tässä kandidaatin työssä tarkasteltiin hiilidioksidikaasupäästöjen, energian kulutuksen ja bruttokansantuotteen yhteyttä toisiinsa hyödyntäen vektoriautoregressiivista mallia. Tarkastelussa käytettiin Suomen dataa vuosilta 1975-2016 ja tarkasteluvälinä oli yksi vuosi. Vuosittaiset datapisteet saatiin vertailukelpoisiksi hyödyntämällä tietoa vuosikohtaisesta väkiluvusta. Työssä hyödynnettiin tietoa sarjan stationaarisuudesta tehtyjen yksikköjuuritestien avulla, testattiin sarjojen yhteisintegroituvuutta sekä niiden välistä Grangerin kausaalisuutta ja luotiin tarkasteltavasta datasta vektoriautoregressiivinen malli.

Vektoriautoregressiivista mallia varten käsiteltävän datan tulee olla stationaarista. Tämä mahdollistettiin käyttämällä differentoitua dataa eli tarkastelussa hyödynnettiin peräkkäisten havaintojen erotusta. Käsitellyn datan stationaarisuus tarkistettiin erilaisilla yksikköjuuritesteillä, jotka ovat esitelty teoriaosuudessa. Käytetyt yksikköjuuritestit ovat laajennettu Dickey–Fuller-testi, Kwiatkowski–Phillips–Schmidt–Shin-testi sekä Phillips–Perron-testi. Sarjojen yhteisintegroituvuuden testaamiseen käytettiin Engle–Granger-testiä ja Johansenin testiä. Grangerin kausaalisuutta testattiin Grangerin kausaalisuustestillä. Kaikki suoritettut testit sekä vektoriautoregressiivinen malli luotiin Matlab-ohjelmistolla, ja tarkastelussa hyödynnettiin Econometrics Toolbox -laajennusta.

Vektoriautoregressiivisen mallin tarkastelussa havaittiin sarjojen ennustavan itseään melko hyvin, mutta niistä ei ole kuitenkaan juuri hyötyä toistensa ennustamiseen. Yhteisintegroituvuustesteillä differentoidulle datalle havaittiin erityisesti energian kokonaiskulutuksen vaikuttavan bruttokansantuotteeseen sekä hiilidioksidikaasupäästöjen suuruuteen. Alkuperäiselle datalle tehtyjen testien perusteella nämä eivät kuitenkaan juuri vaikuta toisiinsa. Grangerin kausaalisuustestin avulla tarkasteltavien sarjojen välillä ei havaita merkittävää vaikuttavuutta toisiinsa, mutta tulosten p-arvoista voidaan kuitenkin päätellä hiilidioksidikaasupäästöjen suuruuden vaikuttavan bruttokansantuotteeseen.

LÄHTEET

- [1] J. D. Hamilton. *Distribution of the estimators for autoregressive time series with a unit root*. Princeton University Press, 1994.
- [2] W. Palma. *Time Series Analysis*. Hoboken, New Jersey: Wiley, 2016.
- [3] J. L. Braams. *Modelling Non-Stationary Economic Time Series: a Multivariate Approach*. Basingstoke: Palgrave Macmillan, 2005.
- [4] D. D. A ja F. W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Society* 12.2 (1979), 427–431.
- [5] K. D. Patterson. *Unit Root Tests in Time Series Volume 1, Key Concepts and Problems*. New York: Palgrave Macmillan, 2011.
- [6] L. M, L. H ja S. P. Comparison of unit root tests for time series with level shifts. *Journal of Time Series Analysis* 23.6 (2002), 667–685.
- [7] W. A. Fuller. *Introduction to statistical time series*. Hoboken: John Wiley Sons, Incorporated. Print., 1995.
- [8] K. Denis, P. P. C.B, S. Peter ja S. Yongcheol. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54.1–3 (1992), 159–178.
- [9] S. Peter ja S. Yongcheol. The KPSS stationarity test as a unit root test. *Economincs Letters* 38.4 (1992), 387–392.
- [10] *Tilastolliset menetelmät: Lineaarinen regressioanalyysi*. Aalto yliopisto, 2006.
- [11] P. P. C.B ja S. Peter. Testing for a unit root in time series regression. *Biometrika* 75.2 (1988), 335–346.
- [12] S. P. Burke ja H. John. *Modelling non-stationary time series: A Multivariate Approach*. Basingstoke : Palgrave Macmillan, 2005.
- [13] S. Johansen. *Statistical Analysis of Cointegration Vectors* (1987).
- [14] D. J. J, G. Jesús ja M. Francesc. *Blackwell Companions to Contemporary Economics: A Companion to Theoretical Econometrics*. eng. 2001. ISBN: 063121254X.
- [15] MathWorks. *Johansen cointegration test*. Saantitapa: <https://se.mathworks.com/help/econ/jcitest.html>. Econometrics Toolbox™. 2021.
- [16] H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag Berlin Heidelberg, 2005.
- [17] MathWorks. *gcitesti*. Saantitapa: <https://se.mathworks.com/help/econ/varm.gctest.html>. Econometrics Toolbox™. 2021.
- [18] Worldometer. *Finland CO2 Emissions*. Saantitapa: <https://www.worldometers.info/co2-emissions/finland-co2-emissions/>. 2021.
- [19] *Suomen virallinen tilasto (SVT): Energian hankinta ja kulutus [verkkojulkaisu]*. ISSN=1799-795X. Helsinki: Tilastokeskus [viitattu: 3.7.2021]. Saantitapa: <http://www.stat.fi/til/ehk/tau.html>. 2021.

- [20] *Suomen virallinen tilasto (SVT): Kansantalouden tilinpito [verkkójulkaisu]*. ISSN=1795-8881. Helsinki: Tilastokeskus [viitattu: 3.7.2021]. Saantitapa: <http://www.stat.fi/til/vtp/index.html>. 2021.
- [21] *Suomen virallinen tilasto (SVT): Väestörakenne [verkkójulkaisu]*. ISSN=1797-5379. Helsinki: Tilastokeskus [viitattu: 3.7.2021]. Saantitapa: <http://www.stat.fi/til/vaerak/tau.html>. 2021.
- [22] *Suomen virallinen tilasto (SVT): Energian hankinta ja kulutus [verkkójulkaisu]*. ISSN=1799-795X. Helsinki: Tilastokeskus [viitattu: 16.10.2021]. Saantitapa: http://www.stat.fi/til/ehk/2019/ehk/2019_020-12-21_ie_01_fi.html. 2019.
- [23] X.-P. Zhang ja X.-M. Cheng. Energy consumption, carbon emissions, and economic growth in China. eng. *Ecological economics*. Ecological Economics 68.10 (2009), 2706–2712. ISSN: 0921-8009.
- [24] A. S. Alshehry ja M. Belloumi. Energy consumption, carbon dioxide emissions and economic growth: The case of Saudi Arabia. eng. *Renewable sustainable energy reviews* 41 (2015), 237–247. ISSN: 1364-0321.

A ALKUPERÄINEN DATA

Vuosi	Energian kulutus (GWh)	Väkiluku	BKT (€)	CO ₂ -päästöt (t)
1975	213829	4 720 492	19 720	47437360
1976	232245	730 836	19 756	53428142
1977	235115	4 746 967	19 765	52946366
1978	248909	4 758 088	20 327	57478433
1979	260947	4 771 292	21 732	57238645
1980	263018	4 787 778	22 883	57730716
1981	260718	4 812 150	23 088	48361818
1982	256128	4 841 715	23 672	46648483
1983	261500	4 869 858	24 264	45029320
1984	272083	4 893 748	24 916	46352799
1985	290511	4 910 664	25 695	50796763
1986	288169	4 925 644	26 317	50872302
1987	305305	4 938 602	27 181	55493982
1988	308742	4 954 359	28 512	54079191
1989	310575	4 974 383	29 860	54095870
1990	317044	4 998 478	29 928	56225383
1991	312421	5 029 002	28 009	56839415
1992	310096	5 054 982	26 936	54029130
1993	318353	5 077 912	26 630	55322190
1994	341986	5 098 754	27 566	63620252
1995	335586	5 116 826	28 616	60674530
1996	351368	5 132 320	29 567	66962787
1997	361186	5 147 349	31 348	64927425
1998	365413	5 159 646	32 969	60952782
1999	374260	5 171 302	34 333	60230860
2000	367094	5 181 115	36 245	59429611
2001	382692	5 194 901	37 105	65038995

Vuosi	Energian kulutus (GWh)	Väkiluku	BKT (€)	CO ₂ -päästöt (t)
2002	393839	5 206 295	37 644	67402922
2003	414938	5 219 732	38 310	75661758
2004	412195	5 236 611	39 725	71851086
2005	381845	5 255 580	40 689	59307059
2006	417087	5 276 955	42 167	71195992
2007	412530	5 300 484	44 209	68655492
2008	392548	5 326 314	44 354	60363132
2009	371225	5 351 427	40 574	57954217
2010	408356	5 375 276	41 679	66309965
2011	387336	5 401 267	42 543	58559105
2012	382205	5 426 674	41 747	57144268
2013	382155	5 451 270	41 180	57758459
2014	375275	5 471 753	40 850	53205640
2015	364293	5 487 308	40 937	49395297
2016	377919	5 503 297	41 981	51183960