

Harjoitustyöraportti

Lasse Rintakumpu
63555

Aikasarjat ja indeksit

Kevät 2014

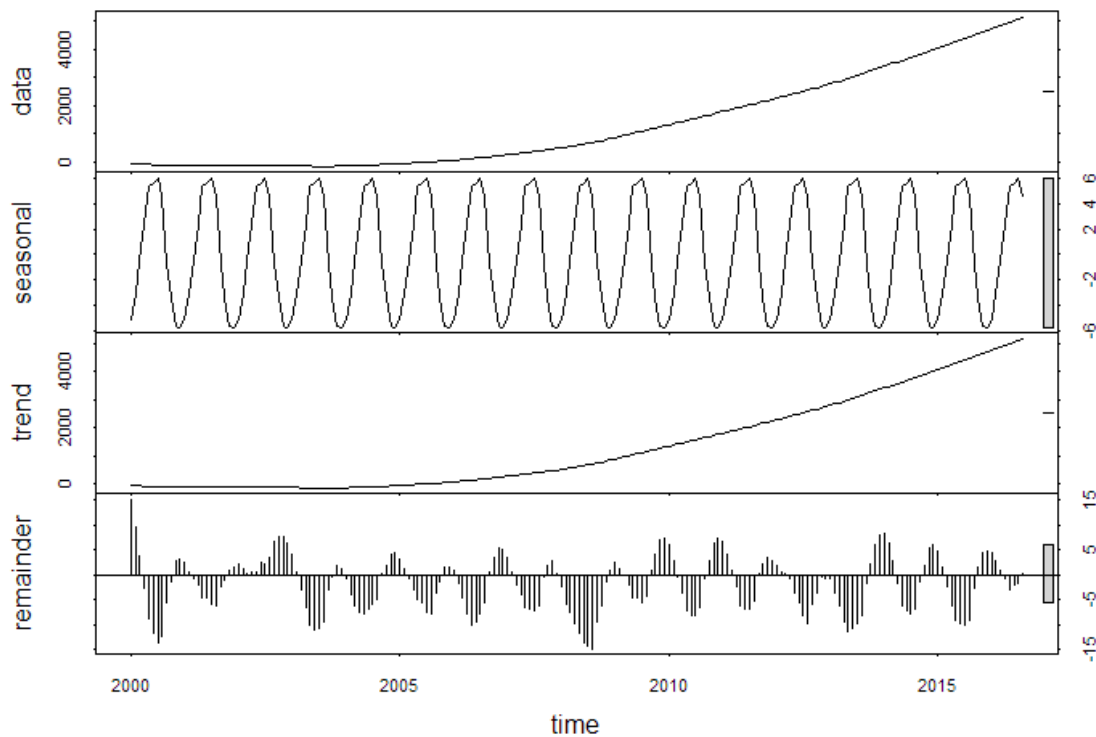
Turun yliopisto

Tilastotiede

Käsitellään aikasarjaa x_t , joka koostuu $T_{xt} = 200$:sta generoidusta havainnosta. Pyritään identifioimaan ja sovittamaan sarjalle ARIMA-malli. Kuvailaan aluksi sarjaa graafisesti, keskittyen sarjassa mahdollisesti havaittavaan trendiin, periodisuuteen ja heteroskedastisuuteen. Tämän jälkeen siirrytään sarjan stationaarisuuden tarkasteluun, mallin identifiointiin, parametrien estimointiin sekä lopuksi sovitetun mallin diagnostiseen tarkasteluun.

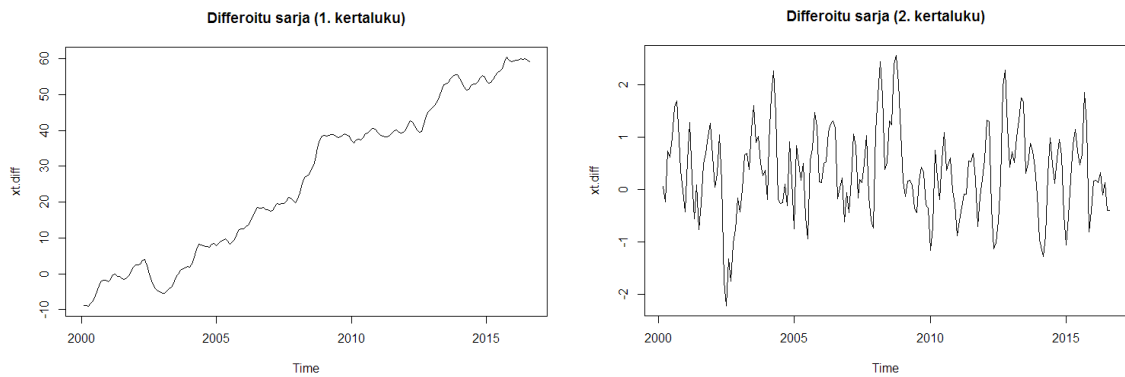
1. Sarjan kuvaileva tarkastelu

Tarkastellaan sarjan x_t kuvaajaa. Ollaan ensisijaisesti kiinnostuneita näkyykö kuvaajassa selkeää trendiä.



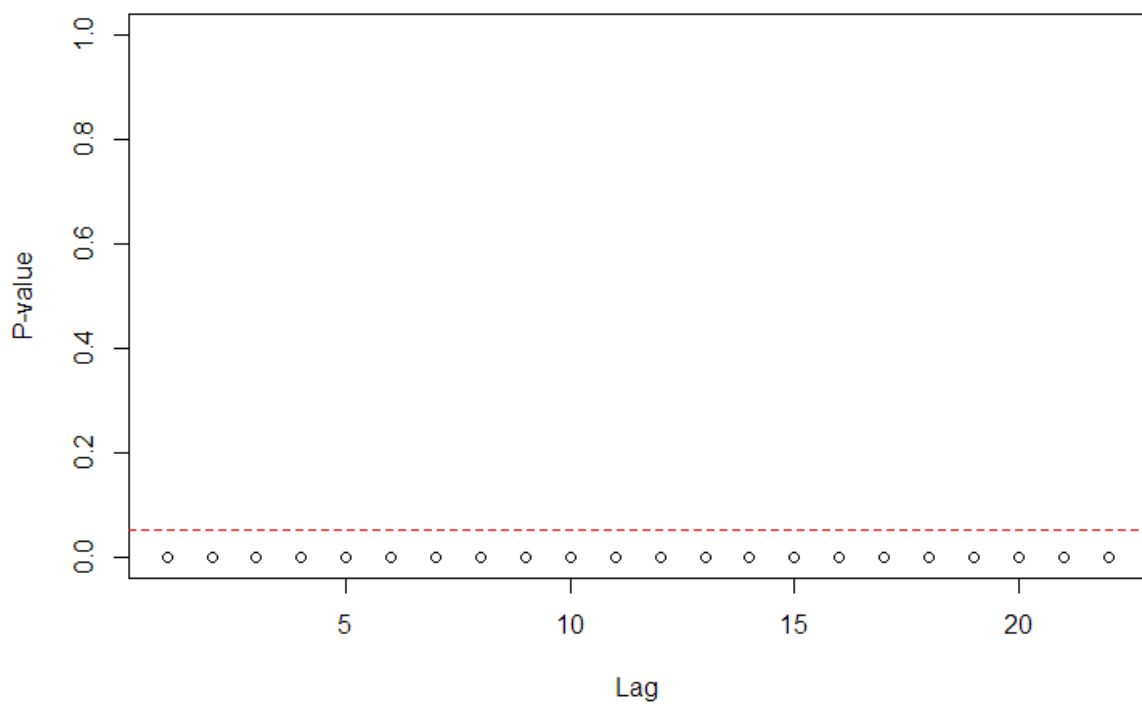
Kuva 1.1.

Sarjassa havaitaan selkeä nouseva trendi (kuva 1.1., paneelit 1 ja 3). Pyritään eliminoidaan havaittu trendi differoimalla. Ensimmäisen kertaluvun differenssi säilyttää nousevan trendin (kuva 1.2.), mutta toisen kertaluvun differenssillä trendi onnistutaan eliminoidaan (kuva 1.3.).



Kuvat 1.2. ja 1.3.

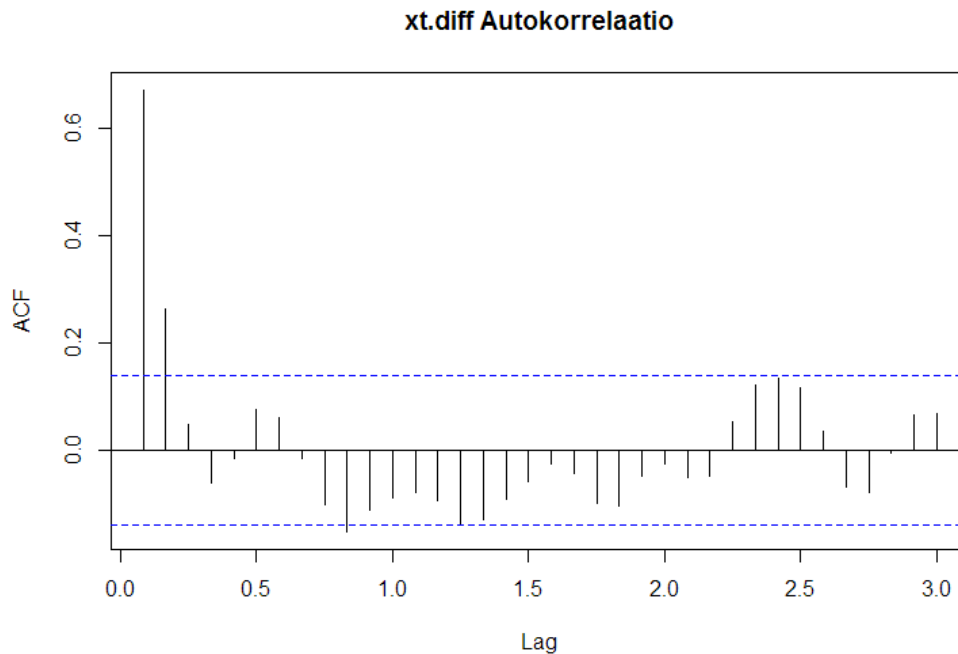
Silmämääräisesti tarkasteltuna kuvan 1.3. kuvaaja ei myöskään sisällä periodisuutta tai heteroskedastisuutta. Oletusta homoskedastisuudesta tukevat myös McLeod-Li-testin kuvaajan ilmaisemat merkitsevyydet (kuva 1.4.). Voimme jatkaa mallin sovitusta differoidulla aikasarjalla `xt.diff`.



Kuva 1.4.

2. Stationaarisuustarkastelu

Tarkastellaan sarjan stationaarisuutta autokorrelaatiofunktion kuvaajan (kuva 2.1.) perusteella. Kuvaaja kuolee viipeen kaksi jälkeen, mikä viittaa stationaariseen sarjaan.



Kuva 2.1.

Varmennetaan stationaarisuus laajennetulla Dickey-Fuller-testillä. Tehdään testi tasolla 0.05. Asetetaan hypoteesipariksi:

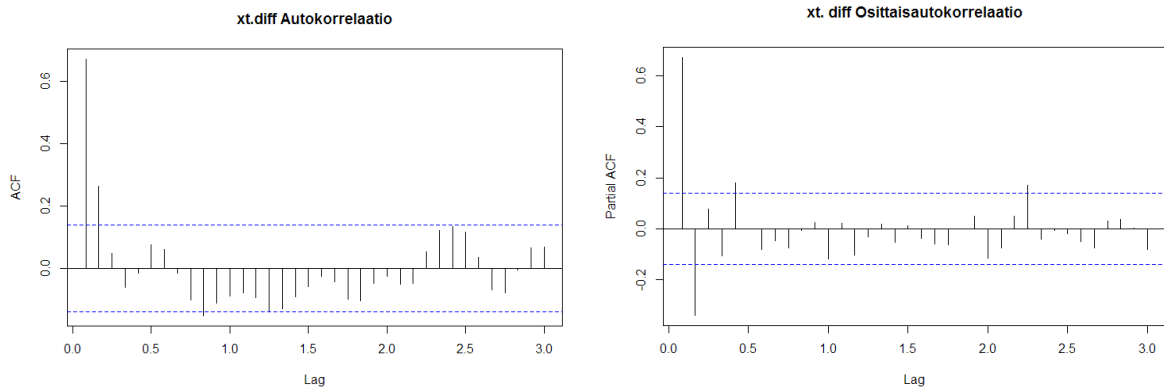
H_0 : Sarja ei ole stationaarinen

H_A : Sarja on stationaarinen.

Saadaan testisuureen arvoksi -4.5022 ja sitä vastaavaksi p-arvoksi < 0.01 . Hylätään nollahypoteesi tasolla 0.05. Myös testin perusteella sarja on stationaarinen.

3. Mallin identifiointi, sovitus ja diagnostinen tarkastelu

Pyritään sovittamaan sarjaan `xt.diff` ARIMA-mallia. Arima mallin sovittaminen on järkevää, sillä sarjan pituus $T_{xt.diff} = 198 > 50$. Käytetään mallin parametrien valinnassa apuna otoksesta laskettuja auto- ja osittaisautokorrelaatiofunktioita (kuvaajat kuvissa 3.1. ja 3.2.).



Kuvat 3.1. ja 3.2.

Kuvien 3.1. ja 3.2. kuvaajien perusteella sekä prosessin autokorrelaatio että osittaisautokorrelaatio kuolevat viipeen kaksi jälkeen. Tämä viittaa joko AR2 tai MA2 -prosessiin. Sovitetaan mallin viiveparametrit (p,q) sekä arvoilla (2,0) että (0,2). Vaihtoehtoista (0,2) antaa mallille lievästi paremman AIC-suureen $369.4 > 365.33$. Tarkastellaan siis jälkimmäistä mallia lähemmin.

Mallin parametrien tarkastelussa havaitaan, etteivät parametrit `sma1` ja `sma2` ole merkitseviä:

<code>ma1</code>	<code>ma2</code>	<code>sma1</code>	<code>sma2</code>
0.000000e+00	3.048916e-07	7.572915e-01	5.738827e-01.

Yksinkertaistetaan mallia poistamalla ensin parametri `sma2`, jolloin parametrin `sma1` p-arvoksi saadaan $8.478926e-01$, joka ei ole tilastollisesti merkitsevä. Poistetaan mallista myös parametri `sma1`, jolloin jäljelle jäävät tilastollisesti merkitsevät parametrit (AIC-suureen arvo nyt 365.75):

<code>ma1</code>	<code>ma2</code>
0.000000e+00	3.067038e-07.

Päädytään seuraavaan malliin (`malli1`, kuva 3.3.):

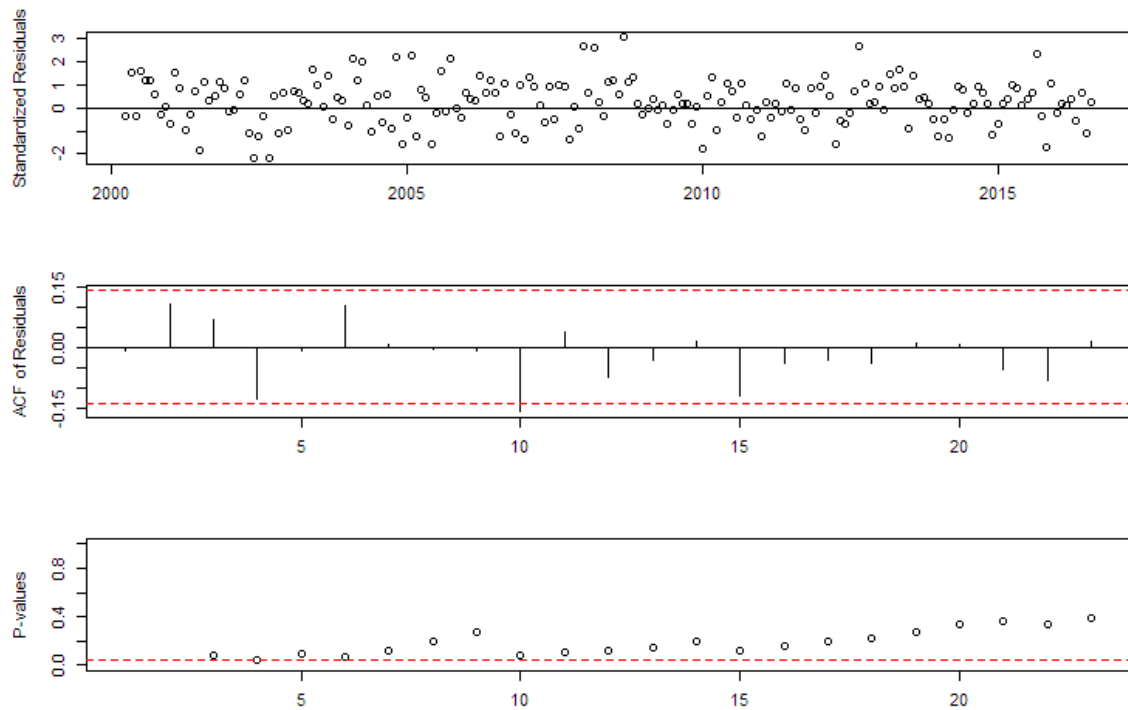
Call:

```
arima(x = xt, order = c(p, d, q), seasonal = list(order = c(p, 0, q - 2)),
method = "ML")
```

Coefficients:

	<code>ma1</code>	<code>ma2</code>
	0.9606	0.3034
s.e.	0.0650	0.0593

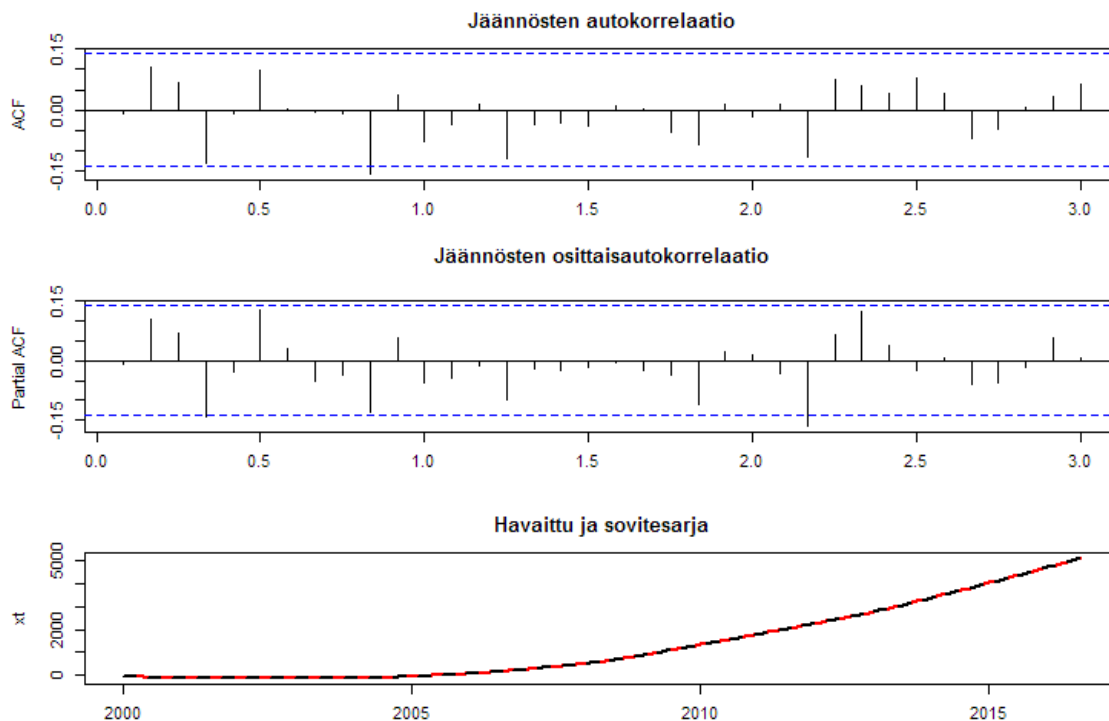
sigma^2 estimated as 0.3621: log likelihood = -180.88, aic = 365.75



Kuva 3.3.

Jos sovitettu malli on sopiva, tulee jäännössarjan (`malli1$residuals`) autokorrelaatio- ja osittaisautokorrelaatiofunktioiden käyttäytyä yhtenevällä tavalla ja olla sisältämättä tilastollisesti merkitseviä arvoja. Kuvasta 3.4. voidaan päätellä, että molemmat kriteerit täyttyvät. Lisäksi kuvasta havaitaan, että sovitesarja vastaa hyvin havaittua sarjaa. Sovitettu malli on siis yhteensopiva havaintosarjan kanssa.

Kuvailevan tarkastelun pohjalta tehtyä päätelmää mallin sopivuudesta tukee myös jäännössarjalle tehty Box-Ljung-testi, jonka testisuure antaa havaitun merkitsevyystason 0.9005.



Kuva 3.4.

Liitteet

Liitteenä harjoitustyössä käytetty R-koodi. Koodi on ladattavissa R-muodossa osoitteesta https://raw.githubusercontent.com/rintakumpu/tilm3508/master/tilm3508_harjoitustyö_rintakumpu.R.

```
#####
# TILM3508 -harjoitustyö, R-koodi #
# Lasse Rintakumpu, 63555          #
# 31.5.2014                        #
#####

# Asetetaan työhakemisto

wd <- ""
setwd(wd)

# Funktio kirjastojen asentamiselle / lataamiselle

lataa_kirjasto <- function(kirjasto) {
```

```

    if(kirjasto %in% rownames(installed.packages()) == FALSE)
    {install.packages(kirjasto)}
    library(kirjasto, character.only = TRUE)
}

# Ladataan/asennetaan käytetyt kirjastot

lapply(c("TSA", "timeSeries"), lataa_kirjasto)

# Ladataan havaintoaineisto

aikasarjat <-
read.table('https://raw.githubusercontent.com/rintakumpu/tilm3508/master/sa
rjat.txt', header = TRUE)

# Valitaan oma sarja (63555)
x <- aikasarjat[, 'X63555']

## Luodaan havaintoaineistosta hypoteettinen kuukausisarja (f=12)
xt <- ts(x, start = c(2000, 1), frequency = 12)
t <- length(xt) # Sarjan pituus == 200

# Edetään varsinaiseen harjoitustyöhön ladatun ja aikasarjaksi muunnetun
aineiston pohjalta =>

#####
# 1. Sarjan kuvaileva tarkastelu #
#####

plot(stl(xt, s.window = 'periodic'))

pdf('kuvaaja_dekomp.pdf')
dev.off()

# Kuvaajan perusteella sarja sisältää nousevan trendin
# => Yritetään eliminoida trendi 1. kertaluvun differenssillä

xt.diff <- diff(xt, differences = 1) # Ei kausidifferointia, so. lag = 1
ts.plot(xt.diff, main = 'Differoitu sarja (1. kertaluku)')

pdf('kuvaaja_sarja_diff1.pdf')
dev.off()

# Kuvaajan perusteella sarja sisältää edelleen nousevan trendin
# => Yritetään eliminoida trendi 2. kertaluvun differenssillä

xt.diff <- diff(xt, differences = 2)
ts.plot(xt.diff, main = 'Differoitu sarja (2. kertaluku)')

```



```

pdf('kuvaaja_sarja_diff2.pdf')
dev.off()

plot(stl(xt.diff, s.window = 'periodic')) # Huom kausivaihtelu!

# => Trendi häviää toisen kertaluvun differoinnilla
# Testataan differoidun sarjan heteroskedastisuutta McLeod.Li.testillä:

McLeod.Li.test(y = xt.diff)

#####
# 2. Stationaarisuustarkastelu #
#####

# Tarkstellaan stationaarisuutta

acf(xt.diff, lag = 36, main = 'xt.diff Autokorrelaatio')
pdf('kuvaaja_acf.pdf')
dev.off()

# => Otosautokorrelaatiofunktio vaimenee toisella viipeellä => Sarja
vaikuttaa stationaariselta.

# Testataan stationarisuutta vielä laajennetulla Dickey-Fuller -testillä

# Valitaan hypoteesipariksi:
# H0: Sarja ei ole stationaarinen
# Hv: Sarja on stationaarinen

# Tehdään testi tasolla 0.05.

adf.test(xt.diff, alternative = c("stationary"))

# Saadaan testisuureen arvoksi -4.5022, joka antaa p-arvon < 0.01.
# Hylätään H0 tasolla 0.05.

#####
# 3. Mallin identifiointi, sovitus ja diagnostinen tarkastelu #
#####

pacf(xt.diff, lag = 36, main = 'xt. diff Osittaisautokorrelaatio')
pdf('kuvaaja_pacf.pdf')
dev.off()

# acf Kuolee viipeen kaksi jälkeen
# pacf Kuolee viipeen kaks jälkee

```

```

# => Viittaa AR2 tai MA2 -prosessiin

# Sovitetaan sarjaan ARIMA-mallia. Sarjan pituus t > 50, joten mallin
sovitus on järkevää.

d <- 2 # 2. kertaluvun differenssi, aiemman päättelyn mukaan
p <- 0
q <- 2

malli1 <- arima(xt, order = c(p,d,q), seasonal = list(order = c(0,0,q-2)),
method = 'ML') # Parametrit smal ja sma2 poistettu
tsdiag(malli1)

# Mallin parametrien merkitsevyys
(1-pnorm(abs(malli1$coef)/sqrt(diag(malli1$var.coef))))*2

# Jäännösten (osittais)autokorrelaatiokuvaajat

acf(malli1$residual, lag = 36, main = 'Jäännösten autokorrelaatio')
pdf('kuvaaja_acf_residual.pdf')
dev.off()

pacf(malli1$residual, lag = 36, main = 'Jäännösten
osittaisautokorrelaatio')
pdf('kuvaaja_pacf_residual.pdf')
dev.off()

sovite <- xt - malli1$residuals # Luodaan sovite poistamalla sarjasta
sovitetun mallin jäännökset
ts.plot(xt, col = 'red', lwd = 2, main = 'Havaittu ja sovitesarja')
lines(sovite, col = 'black', lwd = 2, lty = 2)

Box.test(malli1$residuals, lag = 1, type = c("Ljung-Box"), fitdf = 0)

# Box-Ljung test
# data: malli1$residuals
# X-squared = 0.0156, df = 1, p-value = 0.9005

```