

TILASTOTIETEEN PRO GRADU -TUTKIELMA

Ari Perälä

**Eksponentiaalisen tasoituksen sovellus
alkoholikuolleisuuden ennustamiseksi**

JYVÄSKYLÄN YLIOPISTO
Matematiikan ja tilastotieteen laitos
Tilastotiede
22.04.2018

JYVÄSKYLÄN YLIOPISTO Matematiikan ja tilastotieteen laitos

Perälä, Ari: Eksponentiaalisen tasoituksen sovellus alkoholikuolleisuuden ennustamiseksi

Tilastotieteen pro gradu -tutkielma, 29 sivua, 13 liitesivua

Tilastotiede

Huhtikuu 2018

Tiivistelmä

Eksponentiaalisen tasoituksen menetelmillä voidaan ennustaa tulevia havain-toja. Yksinkertaisen eksponentiaalisen tasoituksen menetelmällä ennustetaan tulevien havaintojen tasoa, kun taas Holtin lokaalin lineaarisen trendin mene-telmä mahdollistaa myös aikasarjan trendin ennustamisen. Näiden menetelmien tarvitsemat tasoitusparametrit saatiin uskottavuusfunktioita optimoimalla.

Esiteltyjä menetelmiä sovellettiin vertailemalla suomalaisten miesten ja nais-ten alkoholikuolleisuutta eri 5-vuotislukissa vuosina 1969-2016. Havaintojen oletettiin tässä olevan Poisson-jakautuneita, joten tasoitusparametrit saatiin optimoimalla Poisson-jakauman uskottavuusfunktioita. Nämä optimoimalla saa-dut tasoitusparametrit olivat usein huomattavasti kirjallisuuden suosittelimia arvoja suurempia.

Kokonaisuutena suomalaisten alkoholikuolleisuus on laskussa sekä miehissä että naisissa. Tämä ei kuitenkaan päde jokaisessa sukupuolittain jaotellussa ikäluokassa. Alkoholikuolleisuus tasaantuu ja kääntyy jopa nousuun eläkeiän kynnyksellä molemmissa sukupuolissa.

Vuosien 2012-2016 alkoholikuolleisuudelle laskettiin ennusteet Holtin lokaa-lisen lineaarisen trendin menetelmällä. Kun näitä alkoholikuolleisuuden ennus-teita verrattiin kyseisinä vuosina toteutuneisiin arvoihin, niin ennusteiden ha-vaittiin vastaavan varsin hyvin toteutuneita arvoja.

Asiasanat: Alkoholikuolleisuus, eksponentiaalinen tasoitus, ennustaminen, Poisson-jakauma, suurimman uskottavuuden estimointi.

Sisältö

1	Johdanto	3
2	Yksinkertainen eksponentiaalinen tasoitus	5
3	Holtin lokaalin lineaarisen trendin menetelmä	8
4	Ehdollinen Poisson-jakauma	10
4.1	Yksinkertainen eksponentiaalinen tasoitus Poisson-jakautuneille havainnoille	10
4.2	Holtin lokaalin lineaarisen trendin menetelmä Poisson-jakautuneille havainnoille	10
4.3	Ehdollinen uskottavuusfunktio Poisson-jakaumalle	11
5	Mallin ja ennusteiden arviointi	12
5.1	Luottamusvälien laskeminen	12
5.2	Ennustevälien laskeminen	12
5.3	Ljung-Box menetelmä	12
6	Optimointi	14
6.1	Muunnokset	14
7	Aineisto	15
7.1	Alkoholikuolleisuusaineisto	15
7.2	Väestötilastoaineisto	15
7.3	Alkoholilainsäädännön historia Suomessa	16
7.4	Alkoholikuolleisuus Suomessa	16
8	Sovellus alkoholikuolleisuuteen Suomessa	18
8.1	Alkoholikuolleisuuden ennustaminen Holtin lokaalin lineaarisen trendin menetelmällä	18
8.2	Alkoholikuolleisuuden tason ennustaminen eksponentiaalisen tasoituksen menetelmällä	21
8.3	Jäännöstarkastelut	26
8.4	Muut ikäluokat	30
	Lähteet	32

A Liite A: R-koodit	33
B Liite B: Kuvat	36

1 Johdanto

Aikasarjojen ennustamisen perusidea on yksinkertaisimmillaan sovittaa aineistoon jokin ajan funktio ja ennustaa tämän avulla tulevia arvoja (Harvey 1989, s. 22). Eräs esimerkki aikasarjojen ennustusmenetelmistä ovat eksponentiaalisen tasoituksen menetelmät. Eksponentiaalisen tasoituksen menetelmät ovat joukko sellaisia ennustusmenetelmiä, jotka painottavat ennustusajankohdan lähellä olevia havaintoja kaukaisia havaintoja enemmän. Tässä kappaleessa eksponentiaalisen tasoituksen menetelmien historiaa käsittelevät kohdat perustuvat pääosin artikkeliin *Exponential smoothing: The state of the art Part II* (Gardner 2006).

Robert G. Brownin työ Yhdysvaltain laivastossa muodosti pohjan eksponentiaalisen tasoituksen menetelmille. Brown esitteli ideansa kirjassaan *Statistical Forecasting for Inventory Control* (1959) ja edelleen kehitti näitä ideoita kirjassa *Smoothing, Forecasting and Prediction of Discrete Time Series* (1963).

Samanaikaisesti, Brownista riippumatta, Charles Holt kehitti eksponentiaalisen tasoituksen menetelmiä. Nämä menetelmät dokumentoitiin Yhdysvaltain laivaston tutkimusviraston sisäisessä muistiossa (1957). Tämä muistio kuitenkin julkaistiin vasta 2004. Holtin menetelmät saavuttivat suosiota vuonna 1960, kun hänen oppilaansa Peter Winters kokeili Holtin esittelemiä menetelmiä empiirisellä aineistolla (1960). Holtin kehittämiä kausivaihtelua sisältävän aineiston ennustamiseen soveltuvia menetelmiä kutsutaankin usein Holt-Wintersin menetelmiksi (Hyndman 2008).

John Muth perusteli ensimmäisenä eksponentiaalisen tasoituksen hyvin määritellyllä tilastollisella mallilla ja tutki eksponentiaalisesti painotettujen ennusteiden optimaalisia ominaisuuksia (1960).

Tässä tutkielmassa keskitytään yksinkertaisen eksponentiaalisen tasoituksen menetelmään ja Holtin lokaalin lineaarisen trendin menetelmään. Yksinkertainen eksponentiaalinen tasoitus on eksponentiaalisen tasoituksen menetelmistä yksinkertaisin ja sen avulla on mahdollista estimoida havaintojen tasoa. Yksinkertainen eksponentiaalinen tasoitus sopii siis sellaisille aineistoille, jotka eivät sisällä trendiä tai kausivaihtelua. Kun tason lisäksi halutaan huomioida aineiston trendi, niin voidaan käyttää Holtin lokaalin lineaarisen trendin menetelmää.

Eksponentiaalisen tasoituksen menetelmät eivät kuitenkaan mitenkään huomioi aineiston jakaumaa. Tarkoituksena olisikin muodostaa R-sovellus, jonka avulla olisi mahdollista ottaa huomioon aineiston jakauma Poisson-jakauman tapauksessa.

Luvussa 2 esitellään tarkemmin käytettävistä menetelmistä ensimmäinen eli yksinkertainen eksponentiaalinen tasointus. Kun tason ennustaminen ei riitä, niin voidaan käyttää luvussa 3 esiteltävää Holtin lokaalin lineaarisen trendin menetelmää. Näiden menetelmien avulla päästään lukuun 4, jossa muutetaan aiemmissa luvuissa esiteltyjä menetelmiä Poisson-jakautuneen aineiston tapaukseen ja muodostetaan uskottavuusfunktio, jonka avulla voidaan estimoida tuntemattomia parametreja. Luku 5 keskittyy mallin ja ennusteiden arviointimenetelmiin ja luku 6 siihen, miten esiteltyä uskottavuusfunktiota tulee optimoida tuntemattomien parametrien suhteen. Kappale 7 esittelee alkoholikuolleisuus- ja väestötilastoaineistot sekä sen miten näitä käytetään tutkielmassa. Kappaleessa 8 käydään läpi soveltava esimerkki alkoholikuolleisuudesta Suomessa eri ikäluokissa käyttäen esiteltyjä menetelmiä.

2 Yksinkertainen eksponentiaalinen tasoitus

Aikasarjan y_1, \dots, y_T hetken T tason estimaatti voidaan määrittellä painotettuna summana aiemmista havainnoista (Harvey 1989, s. 25)

$$m_T = \sum_{j=0}^{T-1} w_j y_{T-j}, \quad (2.1)$$

missä kertoimet w_j ovat aiempien havaintojen painoja, joiden summan on oltava 1. Merkinnällä $\hat{y}_{T+h|T}$ tarkoitetaan havainnon y_{T+h} ennustetta ehdolla aiemmat havainnot y_1, \dots, y_T , missä h on ennustettava askel. Hetken T estimaatti määritellään nyt jokaisen uuden havainnon ennusteeksi

$$\hat{y}_{T+h|T} = m_T, \quad h = 1, 2, \dots, \quad (2.2)$$

eli ennuste $\hat{y}_{T+h|T}$ on sama jokaisella h :n arvolla ja nämä ennusteet muodostavat vakioisen tason.

Yksinkertaisen eksponentiaalisen tasoituksen menetelmässä ennustetta laskettaessa uusille havainnoille annetaan enemmän painoa kuin vanhoille. Antamalla nyt kaavan (2.1) painojen w_j pienentyä eksponentiaalisesti saadaan (Harvey 1989, s. 25-26)

$$\begin{aligned} m_T &= \phi y_T + \phi(1 - \phi)y_{T-1} + \phi(1 - \phi)^2 y_{T-2} + \dots + \phi(1 - \phi)^{T-1} y_1 \\ &= \phi \sum_{j=0}^{T-1} (1 - \phi)^j y_{T-j}, \quad 0 < \phi \leq 1, \end{aligned} \quad (2.3)$$

missä ϕ on tasoitusparametri. Mikäli ϕ saa arvon 1, niin kaava (2.3) supistuu muotoon

$$m_T = y_T, \quad (2.4)$$

jolloin nykyinen havainto sisältää kaiken seuraavan havainnon ennustamiseen tarvittavan tiedon, eikä aiempia havaintoja huomioida seuraavaa havaintoa ennustettaessa. Jos taas ϕ voisi saada arvon 0, niin

$$m_T = 0, \quad (2.5)$$

eli jokaisen havainnon ennuste olisi 0.

Kaavan (2.3) havaintojen painot eivät nyt kuitenkaan summaudu ykköseksi. Painot kuitenkin muodostavat geometrisen sarjan ja suurilla T :n arvoilla

$$\lim_{T \rightarrow \infty} \phi \sum_{j=0}^{T-1} (1-\phi)^j = \lim_{T \rightarrow \infty} \frac{\phi - \phi(1-\phi)^T}{1 - (1-\phi)} = \lim_{T \rightarrow \infty} 1 - (1-\phi)^T = 1, \quad (2.6)$$

joten vaatimus painojen summautumisesta ykköseksi voidaan likimain katsoa täytetyksi.

Kirjoittamalla kaava (2.3) auki komponenteittain

$$\begin{aligned} m_1 &= \phi y_1, \\ m_2 &= \phi y_2 + \phi(1-\phi)y_1 = \phi y_2 + (1-\phi)m_1, \\ m_3 &= \phi y_3 + \phi(1-\phi)y_2 + \phi(1-\phi)^2 y_1 = \phi y_3 + (1-\phi)m_2, \\ &\vdots \\ m_T &= \phi y_T + \phi(1-\phi)y_{T-1} + \phi(1-\phi)^2 y_{T-2} + \dots + \phi(1-\phi)^{T-1} y_1 \\ &= \phi y_T + (1-\phi)m_{T-1}, \end{aligned} \quad (2.7)$$

havaitaan, että mikäli tämä kaava on määritelty kaikilla $t = 1, \dots, T$, niin se voidaan esittää rekursiokaavana

$$m_t = (1-\phi)m_{t-1} + \phi y_t, \quad t = 2, \dots, T. \quad (2.8)$$

Tällöin kuitenkin ongelmaksi muodostuu alkuarvon asettaminen, sillä ennustetta m_0 ei ole määritelty eikä ennustetta m_1 voida täten laskea. Tämän johtaa siihen, ettei kaavan (2.8) rekursiota pystytä aloittamaan. Alkuarvon valitsemiseksi on kuitenkin useita keinoja. Asetetaan nyt tasoksi $m_1 = y_1$ ja aloitetaan rekursio kohdasta $t = 2$, joka johtaa siihen, että kaavassa (2.3) havainto y_1 saa painon $(1-\phi)^{T-1}$, eikä $\phi(1-\phi)^{T-1}$ kuten aiemmin ja siten kaava (2.1) summautuu ykköseksi myös pienillä otoskoilla (Harvey 1989, s. 26). Alkuarvo voidaan myös asettaa esimerkiksi laskemalla keskiarvo muutamasta ensimmäisestä havainnosta tai optimoimalla se havainnoista.

Havainnon ennuste on painotettu keskiarvo edellisestä havainnosta ja sen ennusteesta. On myös mahdollista esittää kaava (2.3) käyttäen edellisen havainnon ennustusvirhettä. Tällöin

$$\begin{aligned} m_t &= m_{t-1} + \phi(y_t - m_{t-1}) \\ &= m_{t-1} + \phi e_t, \quad 0 < \phi \leq 1. \end{aligned} \quad (2.9)$$

missä

$$e_t = y_t - m_{t-1} = y_t - \hat{y}_{t|t-1}, \quad t = 1, \dots, T. \quad (2.10)$$

e_t on siis yhden askeleen ennusteen ennustevirhe ajanhetkellä t . Parametri ϕ määrittelee, missä määrin aiemmat havainnot vaikuttavat ennusteeseen. Suurilla ϕ :n arvoilla ennustettavaa ajankohtaa lähellä olevat havainnot saavat isomman painoarvon kuin pienillä ϕ :n arvoilla, joilla painotukset jakaantuvat tasaisemmin aiempien havaintojen kesken.

Parametrin ϕ arvon valitsemiseksi on useita keinoja. Yksi vaihtoehto on antaa ϕ :lle prioriarvo, yleensä 0.05:n ja 0.30:n väliltä. Mikäli havainnot on tarpeeksi, niin voidaan ϕ :n arvo voidaan myös estimoida minimoimalla jännösneliösumma (Hyndman ja Athanasopoulos 2013)

$$SSE = \sum_{t=1}^T (y_t - \hat{y}_{t|t-1})^2 = \sum_{t=1}^T (e_t)^2. \quad (2.11)$$

Tässä tutkielmassa ϕ :n arvo tullaan saamaan optimoimalla uskottavuusfunktiota R:n optim-paketin avulla. Tämä esitellään myöhemmässä kappaleessa.

3 Holtin lokaalin lineaarisen trendin menetelmä

Holt (1957) ja Winters (1960) yleistivät yksinkertaisen eksponentiaalisen tasoituksen koskemaan trendin sisältäviä aikasarjoja. Tämä menetelmä sisältää kaksi tuntematonta parametria ϕ_0 ja ϕ_1 , joiden avulla voidaan määrittellä sarjan taso a_t ja kulmakerroin b_t . Tällöin h :n askeleen ennuste voidaan laskea kaavalla (Harvey 1989, s. 27)

$$\hat{y}_{T+h|T} = a_t + hb_t, \quad h = 1, 2, \dots, \quad (3.1)$$

jolloin yhden askeleen ennuste on

$$\hat{y}_{T|T-1} = a_{t-1} + b_{t-1}. \quad (3.2)$$

Havaintojen y_t ja yhden askeleen ennusteen $\hat{y}_{T|T-1}$ avulla voidaan nyt muodostaa rekursiokaavat tason a_t ja kulmakertoimen b_t päivittämiseksi

$$\begin{aligned} a_t &= \phi_0 y_t + (1 - \phi_0)(a_{t-1} + b_{t-1}), & 0 < \phi_0 < 1, t = 3, \dots, T, \\ b_t &= \phi_1(a_t - a_{t-1}) + (1 - \phi_1)b_{t-1}, & 0 < \phi_1 < 1, t = 3, \dots, T. \end{aligned} \quad (3.3)$$

Ongelmaksi muodostuu jälleen tämän rekursiokaavan aloittaminen. Valitaan tässä alkuarvoksi tasolle $a_2 = y_2$, kulmakertoimelle $b_2 = y_2 - y_1$ ja ajetaan rekursiot väliltä $t = 3, \dots, T$.

Pienellä ϕ_0 :n arvolla edellisellä havainnolla on vain vähän merkitystä tason nykyisen ennusteen laskemisessa. Tällöin tason ja kulmakertoimen edellistä ennustetta painotetaan enemmän. Vastaavasti pienellä ϕ_1 :n arvolla kulmakertoimen laskettaessa painotetaan enemmän edellisen kulmakertoimen ennustetta. Suurella ϕ_0 edellisen havainnon merkitys tason ennusteessa on suuri, ja vastaavasti suurella ϕ_1 :n arvolla kulmakertoimeksi muodostuu kahden edellisen tason ennusteiden erotus. (Harvey 1989, s.27).

Kuten yksinkertaisen eksponentiaalisen tasoituksen yhteydessä, niin tässäkin parametrien ϕ_0 ja ϕ_1 arvot saadaan esimerkiksi harkinnan mukaan tai minimoimalla jäännösneliösumma. Tässä tutkielmassa ϕ_0 :n ja ϕ_1 :n arvot saadaan optimoimalla uskottavuusfunktioita R:n optim-funktion avulla.

Vaihtoehtoisesti rekursiokaava (3.3) voidaan jälleen esittää myös ennustevirheiden avulla

$$\begin{aligned} e_t &= y_t - (a_t + b_t), \\ a_t &= a_{t-1} + b_{t-1} + \phi_0 e_t, \\ b_t &= b_{t-1} + \phi_0 \phi_1 e_t, \end{aligned} \quad (3.4)$$

missä jälleen e_t :n avulla voidaan minimoida jäännösneliösumma SSE ja estimoida ϕ_0 :n ja ϕ_1 :n arvot kaavan (2.11) avulla (Hyndman ja Athanasopoulos 2013).

4 Ehdollinen Poisson-jakauma

4.1 Yksinkertainen eksponentiaalinen tasoitus Poisson-jakautuneille havainnoille

Havainto y_1 on kiinnitetty vakio. Oletetaan, että

$$y_t|y_1, \dots, y_{t-1} \sim \text{Po}(u_t m_{t-1}), \quad t = 2, \dots, T. \quad (4.1)$$

missä u_t on tunnettu verrannollisuuskerroin, esimerkiksi altistuneiden määrä, ja m_{t-1} on intensiteetti. Yksinkertaisen eksponentiaalisen tasoituksen rekursiokaava (2.8) voidaan kirjoittaa nyt logaritmfunktion avulla Poisson-jakautuneille havainnoille muotoon

$$\log(m_{t+1}) = \log(m_t) + \phi(\log(y_t/u_t) - \log(m_t)), \quad 0 < \phi < 1, t = 1, \dots, T. \quad (4.2)$$

Mikäli havainnon y_t arvo on 0, niin rekursiokaavassa (4.2) asetetaan y_t :n arvoksi esimerkiksi 0.5 tai jokin muu sopiva luku. Poisson-jakaumalle havainnon odotusarvo ja varianssi ovat samat, eli

$$E(y_t|y_1, \dots, y_{t-1}) = \text{Var}(y_t|y_1, \dots, y_{t-1}) = u_t m_{t-1}. \quad (4.3)$$

Parametrin ϕ alkuarvo saadaan maksimoimalla Poisson-jakauman log-uskottavuusfunktio (4.7).

4.2 Holtin lokaalin lineaarisen trendin menetelmä Poisson-jakautuneille havainnoille

Holtin menetelmän rekursiokaavat (3.3) voidaan kirjoittaa logaritmfunktion avulla Poisson-jakautuneille havainnoille muotoon

$$\begin{aligned} \log(m_t) &= a_t + b_t, & t &= 3, \dots, T, \\ a_{t+1} &= \phi_0 \log(y_t/u_t) + (1 - \phi_0)(a_t + b_t), & 0 < \phi_0 < 1, \\ b_{t+1} &= \phi_1(a_{t+1} - a_t) + (1 - \phi_1)b_t, & 0 < \phi_1 < 1. \end{aligned} \quad (4.4)$$

Rekursiokaavat voidaan alustaa asettamalla $a_2 = \log(y_2/u_2)$ ja $b_2 = \log(y_2/u_2) - \log(y_1/u_1)$. Jos havainnon y_t arvo on 0, niin myös tässä korvataan tämä arvolla 0.5. Ennusteet tuleville havainnoille m_{T+h} viiveellä h saadaan kaavalla

$$m_{T+h} = a_T + hb_T, \quad h = 1, 2, \dots. \quad (4.5)$$

Parametrien ϕ_0 ja ϕ_1 alkuarvot estimoidaan jälleen maksimoimalla jakauksen log-uskottavuusfunktio (4.7).

4.3 Ehdollinen uskottavuusfunktio Poisson-jakaumalle

Koska havainnot riippuvat toisistaan, niin uskottavuusfunktiota laskettaessa on huomioitava aiemmat havainnot. Yksinkertaisen eksponentiaalisen tasoituksen tapauksessa $f(y_1)$ on kiinnitetty ja uskottavuusfunktioksi saadaan

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{t=2}^T f(y_t | y_{t-1}, \dots, y_1; \boldsymbol{\theta}) \\ &= \prod_{t=2}^T \frac{(u_t m_{t-1}(\boldsymbol{\theta}))^{y_t} e^{-u_t m_{t-1}(\boldsymbol{\theta})}}{y_t!}, \end{aligned} \quad (4.6)$$

jolloin logaritminen uskottavuusfunktio on

$$\log L = \sum_{t=2}^n y_t \log u_t + y_t \log m_{t-1}(\boldsymbol{\theta}) - \log y_t! - u_t m_{t-1}(\boldsymbol{\theta}), \quad (4.7)$$

missä $m_{t-1}(\boldsymbol{\theta})$ saadaan yksinkertaisen eksponentiaalisen tasoituksen tapauksessa kaavasta (4.2), kun $\boldsymbol{\theta} = \phi$.

Holtin menetelmän tapauksessa kiinnitetään $f(y_1)$ ja $f(y_2)$, jolloin uskottavuusfunktio (4.6) aloitetaan kohdasta $t = 3$. Nyt $m_{t-1}(\boldsymbol{\theta})$ saadaan kaavalla (4.4), kun $\boldsymbol{\theta} = (\phi_0, \phi_1)$.

5 Mallin ja ennusteiden arviointi

Ennustettaessa aikasarjoja jäännökset ovat havainnon ja sen yhden askeleen ennusteen välinen erotus, eli

$$e_t = y_t - \hat{y}_{t|t-1}, \quad t = 1, \dots, T. \quad (5.1)$$

Hyvän ennustemenetelmän jäännösten tulisi olla korreloimattomia ja niiden keskiarvon tulisi olla lähellä nollaa (Hyndman ja Athanasopoulos 2013). Tässä jäännösten autokorrelaatiota ja siten mallin hyvyttä pyritään graafisten tarkastelujen lisäksi arvioimaan Ljung-Box menetelmän avulla. Ennusteita puolestaan arvioidaan luottamus- ja ennustevälien avulla.

5.1 Luottamusvälien laskeminen

Parametreille ϕ , ϕ_0 ja ϕ_1 saadaan 95%:n luottamusvälit kaavalla

$$\hat{\theta} \pm 1.96 \times se(\hat{\theta}) \quad (5.2)$$

missä $\hat{\theta}$ on se parametrin ϕ , ϕ_0 tai ϕ_1 estimaatti, jolle ollaan laskemassa 95%:n luottamusväliä. Kahden parametrin tapauksessa keskivirhe voidaan laskea Fisherin informaatiomatriisin avulla.

5.2 Ennustevälien laskeminen

Ennustetuille arvoille m_t lasketaan ennustusväli seuraavalla algoritmilla.

1. Otetaan kokoa N oleva otos Poisson-jakaumasta odotusarvolla m_t .
2. Järjestetään otos, ja otetaan talteen arvo 5%:n alarajan sekä 95%:n ylärajan kohdalta.
3. Toistetaan jokaiselle ennustetulle arvolle.

5.3 Ljung-Box menetelmä

Ljung-Box menetelmän avulla voidaan tutkia eroavatko aikasarjan autokorrelaatiot tilastollisesti merkitsevästi nolasta. Menetelmä perustuu suureeseen

$$Q^* = T(T+2) \sum_1^h (T-k)^{-1} r_k^2, \quad (5.3)$$

missä T on havaintojen määrä, h suurin tutkittava viive ja r_k on autokorrelaatio viiveellä k . Suuret Q^* :n arvot viittaavat autokorrelaatioon jäännösten välillä. Näitä arvoja voidaan testata χ^2 -testisuuretta vastaan vapausastein $h-K$, missä K on mallin parametrien lukumäärä (Hyndman ja Athanasopoulos 2013).

6 Optimointi

Uskottavuusfunktion 4.7 maksimointi tapahtuu \mathbb{R} :ssä optim-funktion avulla. Tämän funktion avulla on mahdollista optimoida funktioita useilla eri menetelmillä. Tämän tutkielman esimerkissä yksinkertaisen eksponentiaalisen tasoituksen ϕ ja Holtin menetelmän arvot ϕ_0 ja ϕ_1 optimoidaan SANN-metodilla, jonka avulla voidaan saada toimivia arvoja vaikka optimoitavassa funktiossa olisi suurtakin vaihtelua.

6.1 Muunnokset

Koska optim-funktio estimoii parametreja välillä $(-\infty, \infty)$, ja koska parametrit ϕ , ϕ_0 ja ϕ_1 puolestaan voivat saada arvoja väliltä $(0, 1)$, niin kukin näistä parametreista joudutaan muuntamaan optimoinnin jälkeen halutulle välille monotonisella muunnoksella

$$\theta = \frac{1}{1 + \exp(-\delta)}, \quad (6.1)$$

missä θ on vuorotellen ϕ , ϕ_0 tai ϕ_1 ja δ vastaava arvo, jonka suhteen optimointi toteutetaan. Koska

$$\lim_{\delta \rightarrow \infty} \frac{1}{1 + \exp(-\delta)} = 1, \quad (6.2)$$

ja

$$\lim_{\delta \rightarrow -\infty} \frac{1}{1 + \exp(-\delta)} = 0, \quad (6.3)$$

niin kaavan 6.1 avulla voidaan suorittaa haluttu muunnos. Lisäksi optim-funktiolle annettavat δ :n alkuarvot tulee muuntaa oikeaan skaalaan kaavalla

$$\delta = -\log\left(\frac{1}{\theta} - 1\right). \quad (6.4)$$

7 Aineisto

Esiteltyjen menetelmien soveltamiseksi käytetään esimerkissä kahta aineistoa, jotka molemmat ovat Tilastokeskuksen keräämiä ja ne on poimittu Tilastokeskuksen kotisivuilta. Väestötilastoaineisto sisältää Suomessa virallisesti asuvien miesten ja naisten lukumäärät. Alkoholikuolleisuusaineistossa sen sijaan on Suomessa alkoholiperäisiin tauteihin ja tapaturmaisiin alkoholimyrkytyksiin kuolleiden ihmisten lukumäärät. Molemmat aineistot ovat jaoteltu sukupuolen ja ikäluokan mukaan. Ikäluokat ovat jaoteltu viiden ikävuoden välein siten, että nuorimmassa ikäluokassa ovat 0-4 vuotiaat ja vanhimmassa, eli yli 85-vuotiaiden luokassa ovat yhdistettyinä kaikki 85-vuotiaat ja tätä vanhemmat ihmiset. Havaintojen vähyden takia ei kuitenkaan tarkastella alle 15-vuotiaiden ikäluokkia.

7.1 Alkoholikuolleisuusaineisto

Alkoholikuolleisuusaineisto sisältää alkoholiperäisistä taudeista ja tapaturmaisista alkoholimyrkytyksistä johtuvat Suomessa tapahtuneet kuolemat vuosina 1969-2016. Tämä aineisto on osa laajempaa kansallista kuolemansyytilastoa, joka perustuu kuolintodistusten tietoihin ja jota on täydennetty väestörekisterikeskuksen väestötietojärjestelmän kuolleiden tiedoilla. Tässä tilastossa ovat mukana sellaiset Suomessa tai ulkomailla kalenterivuoden aikana kuolleet ihmiset, joiden kotipaikka kuolinhetkellä oli Suomi (Tilastokeskus 2017a).

7.2 Väestötilastoaineisto

Lisäksi käytetään vuosien 1968-2016 väestötilastoaineistoa, joka sisältää tiedot vuoden vaihteessa Suomessa vakinaisesti asuvista henkilöistä iän- ja sukupuolen mukaan. Suomessa vakituisesti asuviksi lasketaan sellaiset Suomen ja ulkomaiden kansalaiset, joiden vakituinen asuinpaikka on Suomi, vaikka he asuisivatkin väliaikaisesti ulkomailla. Ulkomaiden kansalaisen katsotaan asuvan Suomessa vakituisesti, jos hän on asunut tai aikoo asua Suomessa vähintään vuoden (Tilastokeskus 2017b).

Väestöluvuista on muodostettu elinvuosia, joiden avulla pyritään arvioimaan Suomessa kunkin vuoden aikana eläneitä henkilöitä ja samalla huomioida se, että osa ihmisistä kuolee vuoden kuluessa. Elinvuosien approksimaation laskemisessa ei kuitenkaan ole huomioitu vuoden aikana tapahtuvaa maahan-

muuttoa tai maastamuuttoa.

Elinvuosia on nyt approksimoitu kaavalla

$$K(t) \approx \frac{V(t) + V(t-1)}{2}, \quad t = 1969, \dots, 2016, \quad (7.1)$$

missä $K(t)$ on elinvuosien määrä ajanhetkellä t ja $V(t)$ on kiinnostavan populaation koko ajanhetkellä t . Elinvuosien laskentavasta johtuen vuoden 1968 tuloksia ei ole tämän aineiston perusteella mahdollista laskea ja koska alkoholikuolleisuusainestokin on saatavilla vuodesta 1969 alkaen, niin tarkastelu rajoittuu vuosiin 1969-2016. Elinvuosien avulla alkoholikuolemat standardisoidaan kohti sataatuhatta vuoden aikana elettyä elinvuotta kaavalla

$$\text{alkoholikuolleisuus} = \frac{\text{alkoholikuolemat}}{\text{elinvuodet}} \times 100000 \quad (7.2)$$

7.3 Alkoholilainsäädännön historia Suomessa

Eräs alkoholin saatavuuteen ja siten mahdollisesti alkoholikuolleisuuteen vaikuttavista tekijöistä on lainsäädäntö. Tarkastelujakson alussa, 1.1.1969 voimaan astunut alkoholilaki (1968/459) vähensi alkoholin säätelyä erityisesti keskioluen osalta. Tämä laki salli muun muassa keskioluen myynnin elintarvikeliikkeissä ja mahdollisti Alkon myymälöiden perustamisen kaupunkien lisäksi myös maaseudulle. Lisäksi mietojen alkoholijuomien ikärajaksi tuli 18 vuotta ja kaikille alkoholijuomille 20 vuotta.

Tammikuussa 1995 voimaan tullut alkoholilaki (1994/1143) purki Alko Oy:n täysmonopolin alkoholijuomien valmistuksen, tuonnin, viennin, tukkumyynnin ja anniskelun osalta. Alko Oy kuitenkin säilytti vähittäismyyntimonopolin yli 4.7 tilavuusprosenttisten juomien osalta. Tämä lakimuutos lisäsi erityisesti alkoholin saatavuutta, sillä alkoholijuomien sallittuja myyntipaikkoja lisättiin ja myyntiaikoja pidennettiin. Lisäksi alle 22 tilavuusprosenttisten alkoholijuomien mainonta sallittiin (Karlsson ym. 2013, s.48-50).

Maaliskuussa 2004 alkoholijuomien veroja alennettiin väkevien alkoholijuomien osalta 44%, oluen osalta 32% ja viinien osalta 10%. Tällä veronalennuksella pyrittiin reagoimaan 1.1.2004 EU:n sisäisten alkoholijuomien rajoitusten poistumisesta ja ennakoimaan Viron liittymistä EU:hun 1.5.2004. Vuoteen 2012 mennessä alkoholiverotuksen kiristykset ovat nostaneet oluiden ja viinien osalta verotuksen vuoden 2004 tasoa korkeammalle (Karlsson ym. 2013, s. 51).

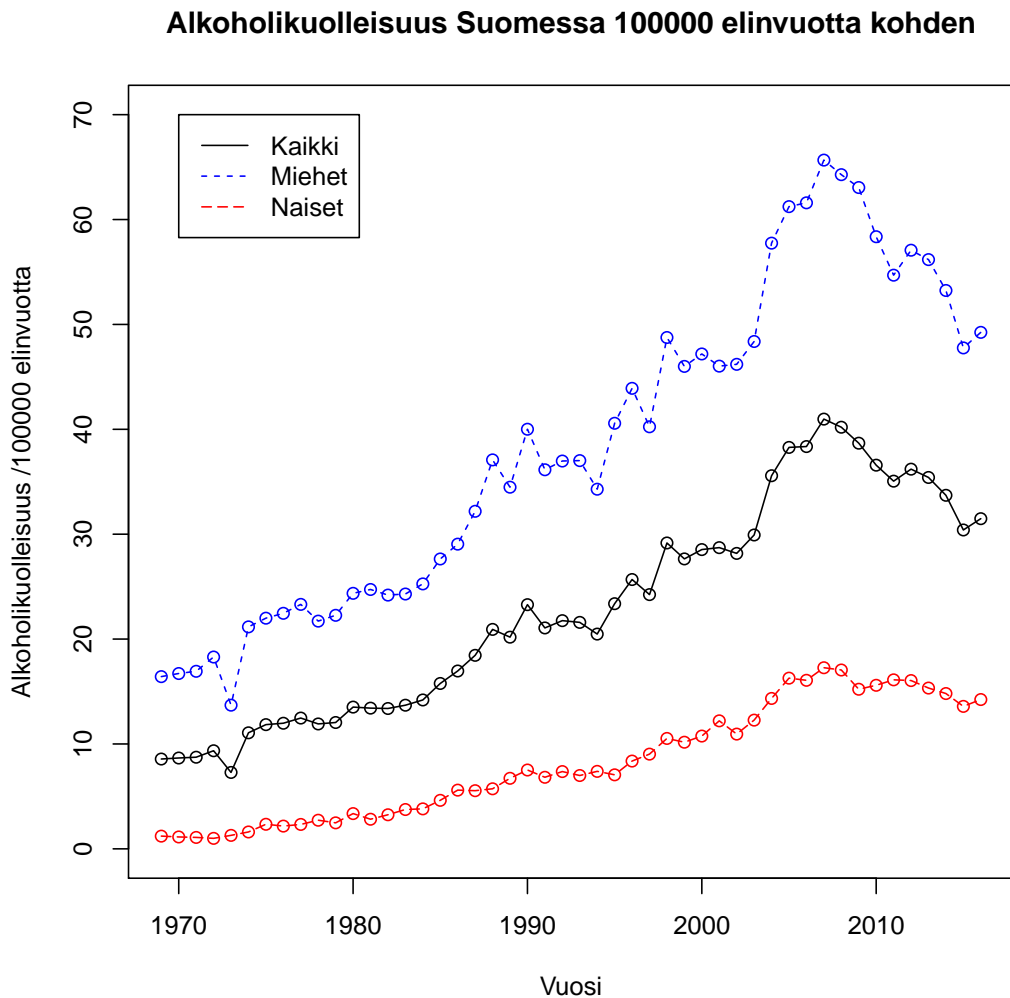
Maaliskuussa 2018 voimaan tullut alkoholilaki (2017/1102) kumosi aieman alkoholilain mutta tämän uuden lain muutokset eivät ehdi vaikuttaa tähän tutkielmaan.

7.4 Alkoholikuolleisuus Suomessa

Tarkasteltaessa kuvassa 7.1 näkyviä vuosien 1969-2016 alkoholikuolleisuuslukuja kokonaisuutena, voidaan havaitaan että molemmassa sukupuolissa alko-

holikuolleisuus sataatuhatta elinvuotta kohden on noussut merkittävästi. Tarkasteluvälin viimeisen vuoden, eli vuoden 2016 alkoholikuolleisuusluvut ovat moninkertaisia vuoden 1969 lukuihin verrattuna.

Alkoholikuolleisuuden kehitys on ollut samansuuntaista molemmilla sukupuolilla. Miehillä kuolleisuusluvut ovat koko tarkasteluvälillä selvästi naisten lukuja suurempia. Alkoholikuolleisuuden huippu on saavutettu molemmissa sukupuoliissa 2000-luvulla ja tämän jälkeen lasku on ollut voimakasta erityisesti miesten osalta. Naisillakin alkoholikuolleisuus on kääntynyt laskuun mutta lasku on ollut selvästi alhaisempaa.



Kuva 7.1. Alkoholikuolleisuus 100000 elinvuotta kohden 1969-2016.

Vaikka alkoholikuolleisuuden kokonaiskehitys on samansuuntaista molemmilla sukupuolilla, niin tämä kehitys ei kuitenkaan päde kaikissa ikäluokissa. Seuraavassa, esiteltyjä menetelmiä soveltavassa kappaleessa tutkimme alkoholikuolleisuuden ikäluokittaista kehitystä.

8 Sovellus alkoholikuolleisuuteen Suomessa

Tässä kappaleessa sovelletaan jo esiteltyjä menetelmiä alkoholikuolleisuusaineistoon. Tämän esimerkin pääpaino on Holtin menetelmää ja esiteltyä Poisson-jakauman uskottavuusfunktioita käyttävässä mallissa, jonka avulla optimoidaan parametrien ϕ_0 ja ϕ_1 arvot. Vuosien 1969-2011 alkoholikuolleisuuslukujen avulla pyritään ennustamaan vuosien 2012-2016 alkoholikuolleisuutta. Saatuja tuloksia tullaan vertaamaan sekä perinteisemmän neliösumman minimointiin perustuvan mallin tuloksiin että vuosina 2012-2016 oikeasti tapahtuneisiin alkoholikuolleisuuslukuihin. Aineiston tulkinta tapahtuu alkoholikuolemina sataatuhatta elinvuotta kohden. Lisäksi sovelletaan yksinkertaisen eksponentiaalisen tasoituksen menetelmää ja ennustetaan sen avulla sarjojen tasoa, vaikkakaan kappaleen pääpaino ei tule olemaan näiden tulosten esittelyssä.

Tarkastelu rajoittuu 35-39 ja 65-69 vuotiaiden miesten ja naisten ryhmiin. Tarkasteltavat ikäluokat on pyritty valitsemaan siten, että niistä olisi nähtävillä erot alkoholikuolleisuuden kehityksessä eri ikä- ja sukupuoliryhmissä. Holtin menetelmän osalta alkoholikuolleisuudet ja niiden ennusteet molemmille sukupuolille 15-19 vuotiaista, aina yli 85-vuotiaisiin ovat nähtävillä liitteessä B. Tätä nuorempia ikäluokkia ei esitellä, sillä onnekkaisesti näissä luokissa alkoholikuolleisuus oli hyvin vähäistä. Yksinkertaisen eksponentiaalisen tasoituksen mallin tasoa ei ole näille muille ikä- ja sukupuoliluokille ennustettu.

8.1 Alkoholikuolleisuuden ennustaminen Holtin lokaalin lineaarisen trendin menetelmällä

Taulukossa 8.1 olevien ϕ_0 :n ja ϕ_1 :n arvoista havaitaan että ennustaminen tuotti kulmakertoimelle ϕ_1 hyvin pienen arvon 65-69 vuotiaille naisille. Tämä tarkoittaa että kulmakertoimen ennuste saa suuren painoarvon edellisen havainnon kulmakertoimen ennusteelta, eikä havaintojen tasojen ennusteiden erolle juuri anneta painoarvoa. Huomion arvoista lienee myös se, että kaikki ϕ_0 :n arvot ovat selvästi suositeltua 0.30:n ylärajaa suurempia. Kuvassa 8.1 näkyvät vuosien 1969-2016 havaittujen alkoholikuolleisuuslukujen lisäksi punaisella katkoviivalla 35-39 vuotiaiden miesten ja naisten vuosien 1969-2011 alkoholikuolleisuuden yhden askeleen ennusteet, sekä vuosien 2012-2016 useamman askeleen ennusteet. Ennustetut alkoholikuolleisuusluvut näyttäisivät 35-39 vuotiaiden ikäluokassa vastaavan varsin hyvin toteutuneita arvoja. Sekä miehissä

että naisissa kyseisessä ikäluokassa alkoholikuolleisuus on kääntynyt laskuun, ja tämä menetelmä kykeni ennustamaan kehityksen suunnan. Kahtena viime vuotena tosin on 35-39 vuotiaiden naisten osalta tapahtunut alkoholikuolleisuudessa selvää nousua, jota malli ei ole kyennyt aiempien havaintojen perusteella ennustamaan. Näin lyhyestä ja pienestä nousujaksosta kuitenkin ei pystytä tekemään johtopäätöksiä mihinkään suuntaan.

Alkoholikuolleisuus 35-39 vuotiaissa miehissä näyttäisi saavuttaneen huippunsa 1990-luvulla ja laskeneen hieman sen jälkeen, kunnes vuoden 2008 paikkeilla alkoholikuolleisuus näyttäisi romahtaneen. Myös 35-39 vuotiailla naisilla on havaittavissa hieman samaa muotoa. Naisilla alkoholikuolleisuus näyttäisi saavuttaneen huippunsa hieman myöhemmin, eikä alkoholikuolleisuuden lasku ole yhtä huomattavaa. On kuitenkin syytä huomata, että 35-39 vuotiaiden ikäluokassa miesten alkoholikuolleisuus on moninkertainen naisiin verrattuna. Verrattaessa suurimpia kuolleisuusvuosia miesten alkoholikuolleisuus on yli nelinkertainen naisiin verrattuna.

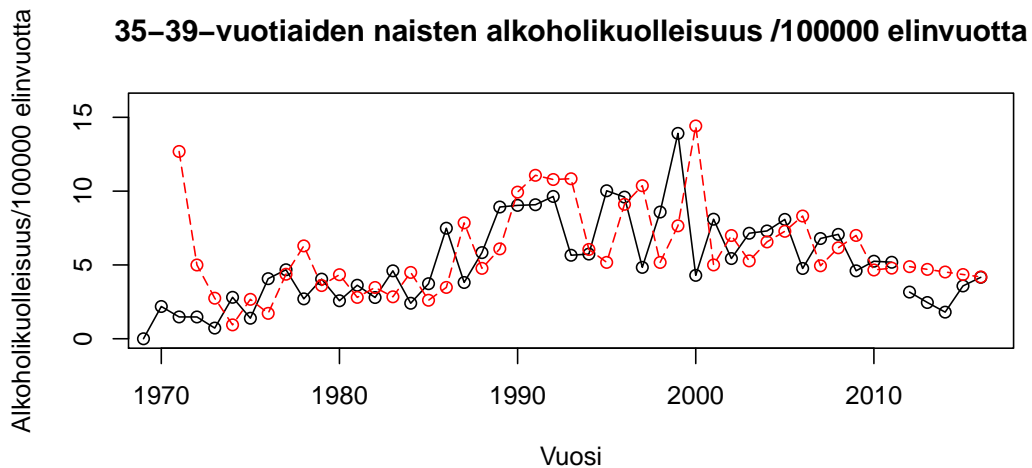
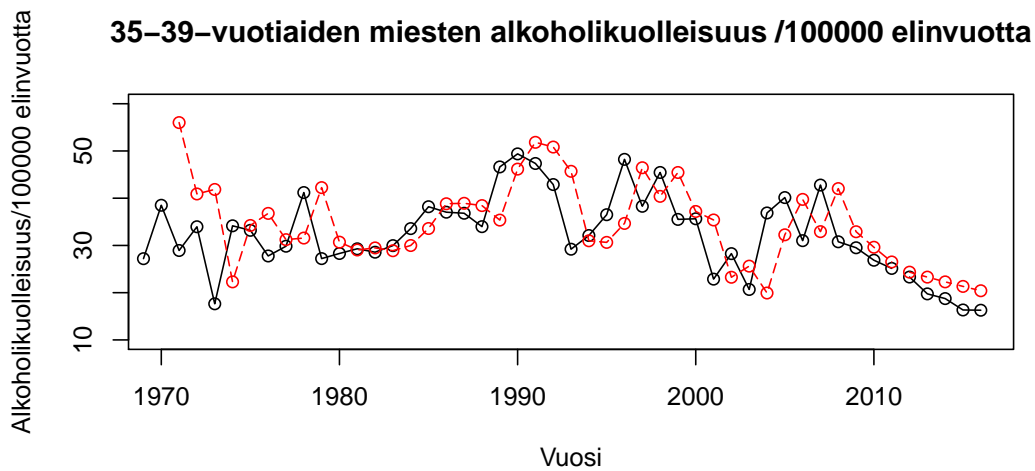
65-69 vuotiaille miehille ja naisille lasketut ennusteet ovat nähtävillä kuvassa 8.2. Verrattaessa näiden ikäluokkien alkoholikuolleisuuksia 35-39 vuotiaisiin, havaitaan että aikasarjat ovat selvästi eri muotoisia. Siinä missä 35-39 vuotiailla miehillä ja naisilla alkoholikuolleisuus oli tasaantunut ja jopa laskenut, niin 65-69 vuotiailla tällaista muutosta ei ole juurikaan nähtävissä.

Kasvu alkoholikuolleisuudessa on 65-69 vuotiaiden ikäluokassa ollut jatkuvaa käytännössä koko tarkasteluvälin ajan niin miehillä kuin naisilla. Molemmilla sukupuolilla on kuitenkin nähtävissä merkkejä alkoholikuolleisuuden tasaantumisessa. Miehillä alkoholikuolleisuuden ennusteet vuosille 2012-2016 ovat selvästi nousussa, mutta toteutuneet kuolemat vaikuttaisivat tasaantuneen. Alkoholikuolleisuuden ennusteet 65-69 naisten ryhmässä ovat kääntyneet lievään laskuun. Estimoitu, varsin korkea ϕ_0 :n ja erittäin alhainen ϕ_1 :n arvo kyseisessä ikä- ja sukupuoliryhmässä yhdessä aiheuttavat sen, että vuosien 1969-2011 ennusteet seuraavat voimakkaasti toteutuneita alkoholikuolleisuuden arvoja yhden vuoden viiveellä. Alkoholikuolleisuus 65-69 vuotiaiden miesten ja naisten ryhmissä on moninkertainen verrattaessa sukupuolittain 35-39 vuotiaisiin.

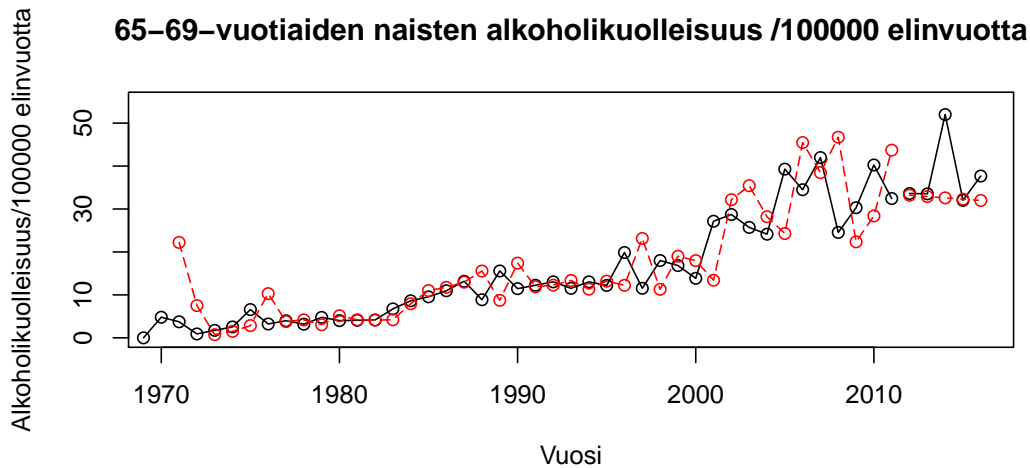
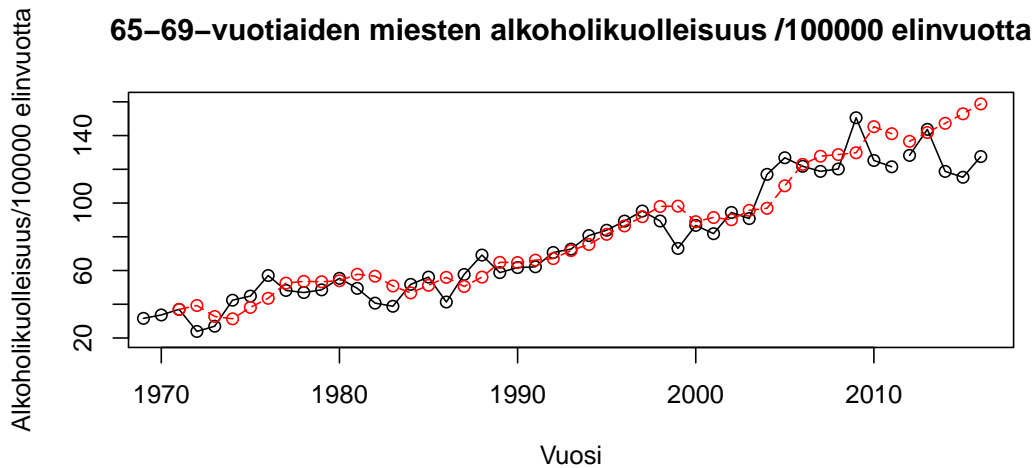
Ikäluokan 35-39 miesten ja naisten osalta sekä ennusteiden, että toteutuneiden alkoholikuolleisuuksien tarkat arvot ovat nähtävillä taulukossa 8.2. Tämä taulukko sisältää myös 90%:n ennustusvälin vuosien 2012-2016 ennusteille. Vastaavasti taulukossa 8.3 on nähtävillä ikäluokan 65-69 miesten ja naisten ennusteet, ennustusvälit ja toteutuneet alkoholikuolleisuusluvut.

Taulukko 8.1. Holtin menetelmällä estimoidut ϕ_0 :n ja ϕ_1 :n arvot sekä 95%:n luottamusvälit.

	ϕ_0	95%:n luottamusväli	ϕ_1	95%:n luottamusväli
Miehet 35-39	0.770	(0.751,0.788)	0.197	(0.184,0.211)
Naiset 35-39	0.740	(0.716,0.763)	0.394	(0.362,0.427)
Miehet 65-69	0.469	(0.387,0.552)	0.039	(0.032,0.049)
Naiset 65-69	0.889	(0.867,0.908)	0.492	(0.460,0.524)



Kuva 8.1. Holtin menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden 35-39 vuotiaille. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.



Kuva 8.2. Holtin menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden 65-69 vuotiaille. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.

8.2 Alkoholikuolleisuuden tason ennustaminen eksponentiaalisen tasoituksen menetelmällä

Alkoholikuolleisuuden tasoa ennustetaan yksinkertaisen eksponentiaalisen tasoituksen menetelmällä. Parametri ϕ on saatu uskottavuusfunktiota optimoimalla. Taulukossa 8.5 on nähtävillä eri ryhmille estimoidut ϕ :n arvot ja näiden 95%:n luottamusvälit. Kuten aiemmin on mainittu, niin pienillä ϕ :n arvoilla ajassa kaukaisemmat havainnot saavat enemmän painoarvoa. Vastaavasti isoilla ϕ :n arvoilla ennustettavaa ajankohtaa lähellä olevat arvot saavat isomman painoarvon. Kaikkien ikä- ja sukupuoliryhmien osalta ϕ :n arvot ovat suositeltua prioriarvojen ylärajaa 0.30 korkeampia. Erityisesti 65-69 vuotiaiden miesten

Taulukko 8.2. Holtin menetelmän ennusteet vuosille 2012-2016, näiden 90%:n ennustusväli 35-39 vuotiaiden miesten ja naisten alkoholikuolleisuuden trendille 100000 elinvuotta kohden.

	Vuosi	Toteutunut	35-39 vuotta	
			Ennuste	90%:n ennustusväli
Miehet	2012	23.314	24.350	(17,33)
	2013	19.752	23.300	(16,31)
	2014	18.756	22.294	(15,30)
	2015	16.342	21.332	(14,29)
	2016	16.288	20.412	(13,28)
Naiset	2012	3.163	4.890	(2,9)
	2013	2.456	4.701	(1,9)
	2014	1.804	4.523	(1,8)
	2015	3.579	4.352	(1,8)
	2016	4.163	4.187	(1,8)

Taulukko 8.3. Holtin menetelmän ennusteet vuosille 2012-2016, näiden 90%:n ennustusväli 65-69 vuotiaiden miesten ja naisten alkoholikuolleisuuden trendille 100000 elinvuotta kohden.

	Vuosi	Toteutunut	65-69 vuotta	
			Ennuste	90%:n ennustusväli
Miehet	2012	128.247	138.588	(118,156)
	2013	143.679	141.824	(123,162)
	2014	118.743	147.260	(128,168)
	2015	115.277	152.906	(133,174)
	2016	127.569	158.767	(138,180)
Naiset	2012	37.682	33.214	(27,43)
	2013	32.017	32.902	(27,43)
	2014	51.992	32.593	(27,42)
	2015	33.532	32.288	(23,42)
	2016	33.618	31.985	(32,42)

Taulukko 8.4. Holtin menetelmän ennusteiden vertailua toteutuneisiin arvoihin 2012-2016. Tasoitusparametrit ϕ_0 ja ϕ_1 optimoitu maksimimomilla suurin uskottavuus(SU), tai minimimomilla jäännösneliösumma(SSE)

	Holtin menetelmän ennuste vuodelle				
	2012	2013	2014	2015	2016
Miehet 35-39 SU	24.350	23.299	22.294	21.332	20.412
Miehet 35-39 SSE	23.858	22.330	20.800	19.271	17.742
Miehet 35-39 Toteutunut	23.314	19.752	18.756	16.342	16.288
Naiset 35-39 SU	4.885	4.701	4.523	4.352	4.187
Naiset 35-39 SSE	4.808	4.510	4.211	3.912	3.614
Naiset 35-39 Toteutunut	3.163	2.456	1.804	3.579	4.163
Miehet 65-69 SU	136.588	141.824	147.261	152.906	158.767
Miehet 65-69 SSE	129.857	131.872	133.888	135.903	137.919
Miehet 65-69 Toteutunut	128.247	143.679	118.743	115.277	127.569
Naiset 65-69 SU	33.214	32.902	32.593	32.288	31.985
Naiset 65-69 SSE	32.858	31.899	30.939	29.979	29.019
Naiset 65-69 Toteutunut	37.682	32.017	51.992	33.532	33.618

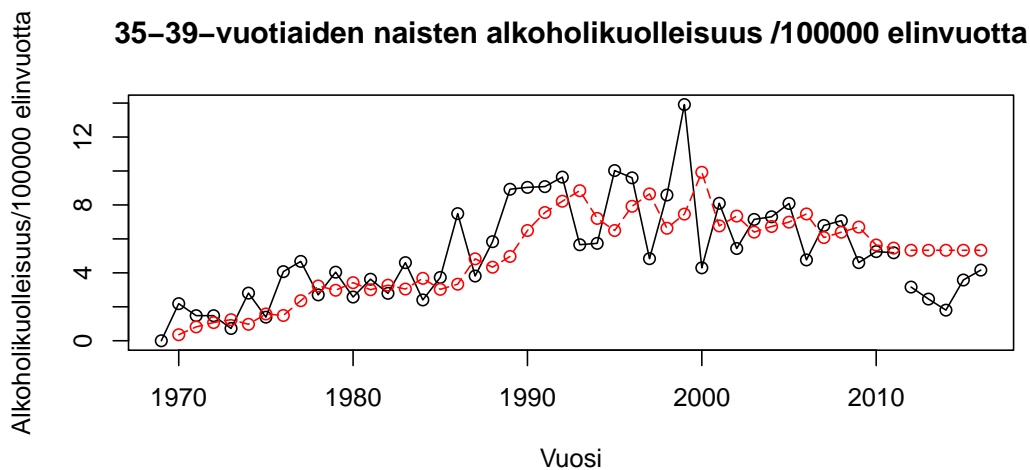
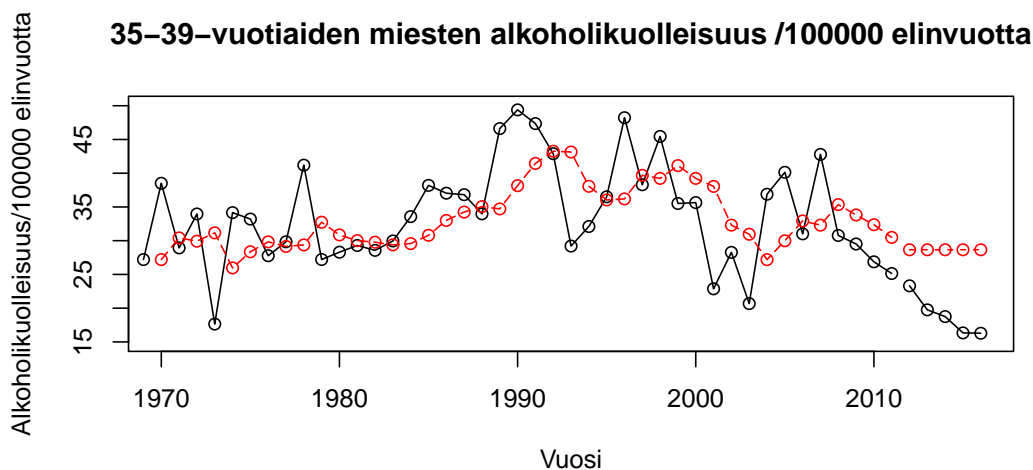
ϕ :n arvo 0.785 on korkea.

Kuvassa 8.3 on nähtävillä yksinkertaisella eksponentiaalisella menetelmällä lasketut tason ennusteet 35-39 vuotiaiden miesten ja naisten alkoholikuolleisuudelle. Kuvassa punaisella katkoviivalla merkityt ennusteet ovat vuosina 1969-2011 yhden askeleen ennusteita tason arvolle. Vuosien 2012-2016 ennusteet sen sijaan ovat kaavan (2.2) mukaisia useamman askeleen ennusteita, jotka muodostavat tason. Ennustettu taso näyttäisi sopivan varsin hyvin erityisesti naisten osalta. Miehistä tason ennuste ei ole kyennyt ennakoimaan alkoholikuolleisuuden jyrkkää pudotusta. Toteutuneet alkoholikuolleisuudet ovat molemmissa sukupuolissa 35-39 vuotiaiden osalta ennustettua tasoa alhaisempia, jonka voidaan kansanterveyden kannalta ajatella hyvänä asiana. Tässä on kuitenkin syytä muistaa, että yksinkertaisen eksponentiaalisen tasoituksen menetelmä ei sovellu trendiä sisältävän aineiston analyysiin, eli esimerkiksi tässä kappaleessa käytettävään alkoholikuolleisuutta käsittelevään aineistoon.

Vastaavasti kuvassa 8.4 on nähtävillä tason ennuste ikäluokan 65-69 miehille ja naisille. Tason ennusteet 65-69 vuotiaille näyttäisivät järkeviltä, tosin taaskaan ei tasoa ennustava menetelmä pysty ennustamaan aineiston trendiä. Tarkasteltujen sukupuolien ja ikäluokkien tason ennusteet ovat nähtävillä taulukossa 8.6.

Taulukko 8.5. Estimoitu ϕ :n arvo eksponentiaalisien tasoituksen menetelmällä ja 95%:n luottamusväli.

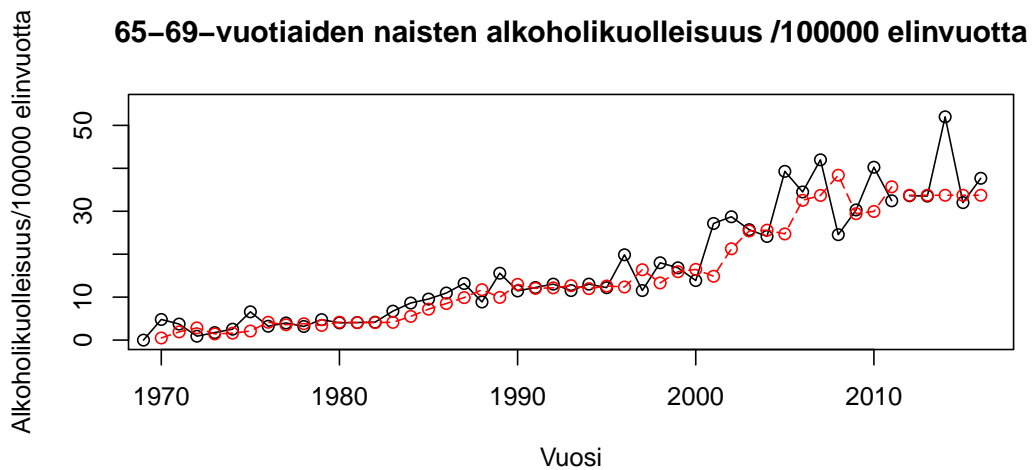
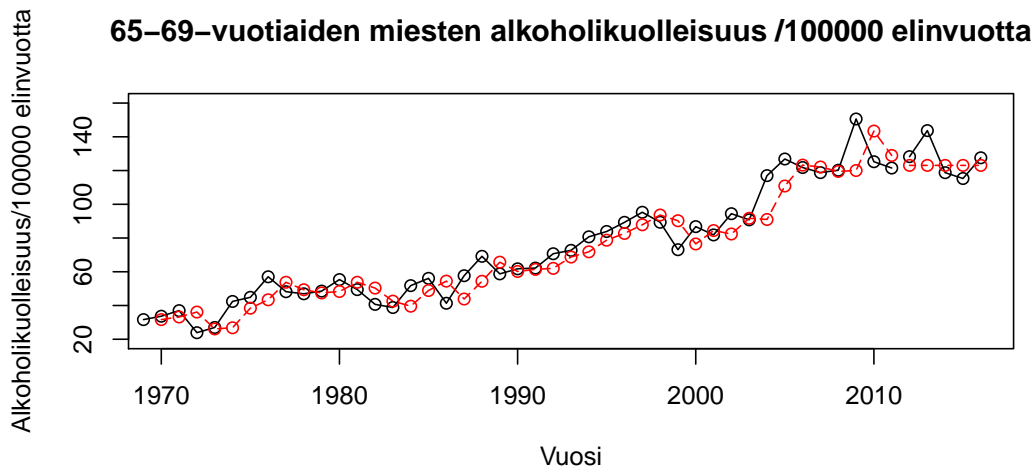
	ϕ	95%:n luottamusväli
Miehet 35-39	0.321	(0.293,0.350)
Naiset 35-39	0.458	(0.434,0.481)
Miehet 65-69	0.785	(0.742,0.822)
Naiset 65-69	0.594	(0.569,0.619)



Kuva 8.3. Eksponentiaalisien tasoituksen menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden 35-39 vuotiaille. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.

Taulukko 8.6. Eksponentiaalisen tasoituksen ennusteet vuosille 2012-2016, näiden 90%:n ennustusväli 35-39 ja 65-69 vuotiaiden miesten ja naisten alkoholikuolleisuuden tasolle 100000 elinvuotta kohden.

	35-39 vuotta		65-69 vuotta	
	ennuste	90%:n ennustusväli	ennuste	90%:n ennustusväli
Miehet	28.660	(20,38)	123.057	(105,142)
Naiset	5.331	(2,9)	33.720	(24,44)



Kuva 8.4. Eksponentiaalisen tasoituksen menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden 65-69 vuotiaille. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.

Taulukko 8.7. Eksponentiaalisen menetelmän tason ennusteiden vertailu 2012-2016. Tasoitusparametri ϕ optimoitu maksimoimalla suurin uskottavuus(SU), tai minimoimalla jäännösneliösumma(SSE)

	SU-maksimointi	SSE-minimointi
Miehet 35-39	28.660	29.264
Naiset 35-39	5.331	5.579
Miehet 65-69	123.057	123.715
Naiset 65-69	33.727	33.953

8.3 Jäännöstarkastelut

Jäännöksiä, eli toteutuneiden arvojen ja niiden yhden askeleen ennusteiden välistä erotusta tarkastelemalla voidaan tutkia mallin hyvyttä. Tässä esimerkiksi jäännökset on Holtin menetelmälle laskettu vuosien 1970-2011 havaintojen ja ennusteiden avulla. Holtin menetelmän alkuarvojen asettamisesta johtuen vuoden 1969 jäännöksiä ei ole mahdollista laskea. Yksinkertaiselle eksponentiaaliselle tasoitukselle sen sijaan jäännökset on ollut mahdollista laskea vuosille 1969-2011.

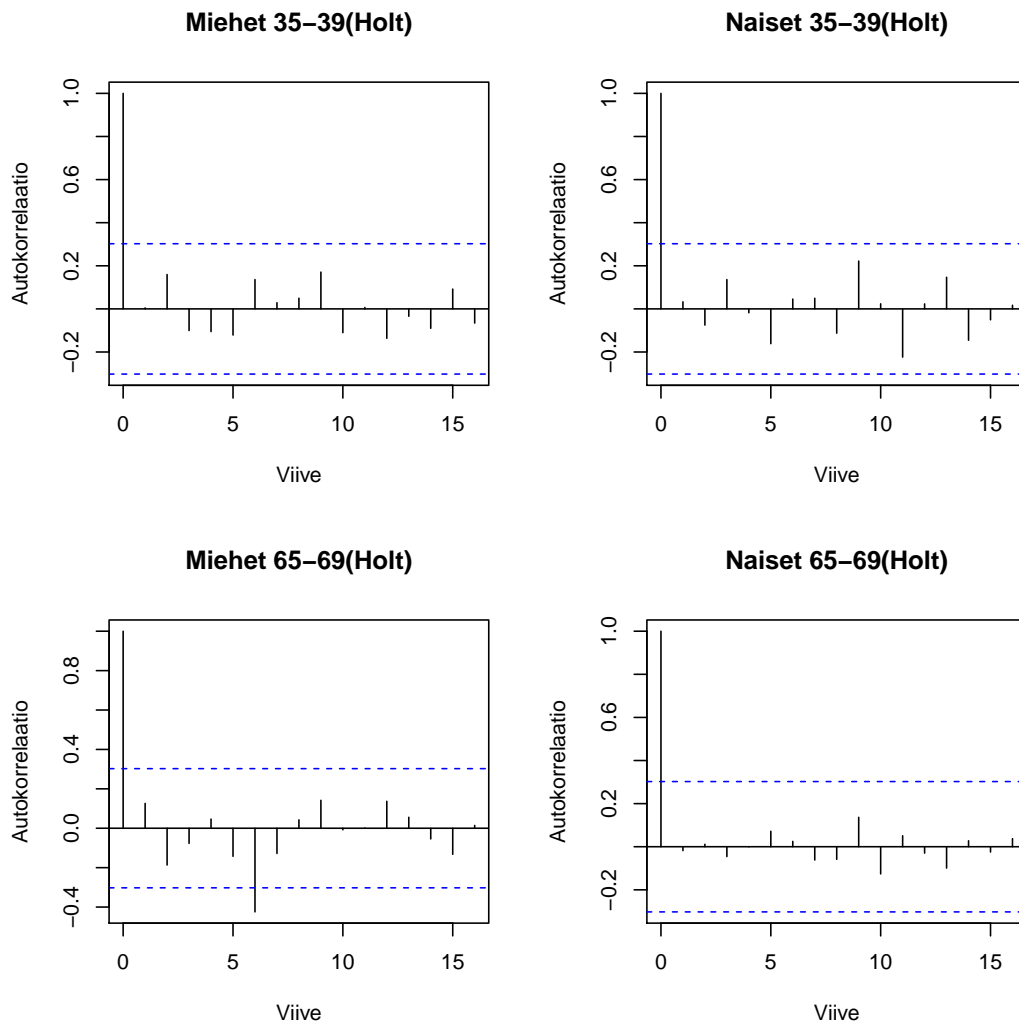
Holtin menetelmän autokorrelaatiot kuvassa 8.5 pysyvät molempien sukupuolten osalta 35-39 vuotiaiden ikäluokassa merkitsevyysrajojen sisäpuolella, mutta 65-69 vuotiaiden miesten ryhmässä merkitsevyysraja ylittyy viiveen ollessa 6. Myöskään 65-69 vuotiailla naisilla autokorrelaatiot eivät ylitä merkitsevyysrajoja. Kuvan 8.6 Holtin menetelmän jäännökset näyttäisivät jotakuinkin keskittyneen nollan ympärille, joskin vaihtelu on erityisesti 65-69 vuotiaiden miesten ryhmässä suurta, ja vaihtelu vaikuttaisi kasvavan ajan myötä.

Yksinkertaisen eksponentiaalisen tasoituksen autokorrelaatioissa kuvassa 8.7 merkitsevyysraja ylittyy 35-39 naisten ja 65-69 vuotiaiden miesten ryhmissä. Kuvan 8.8 yksinkertaisen eksponentiaalisen tasoituksen jäännöksiä tarkastelemalla havaitaan jälleen mahdollista vaihtelun kasvamista ajan myötä 65-69 vuotiaiden ryhmissä. Lisäksi 35-39 vuotiaiden naisten osalta jäännösten vaihtelu kasvaa kuvaajan keskiosassa suureksi.

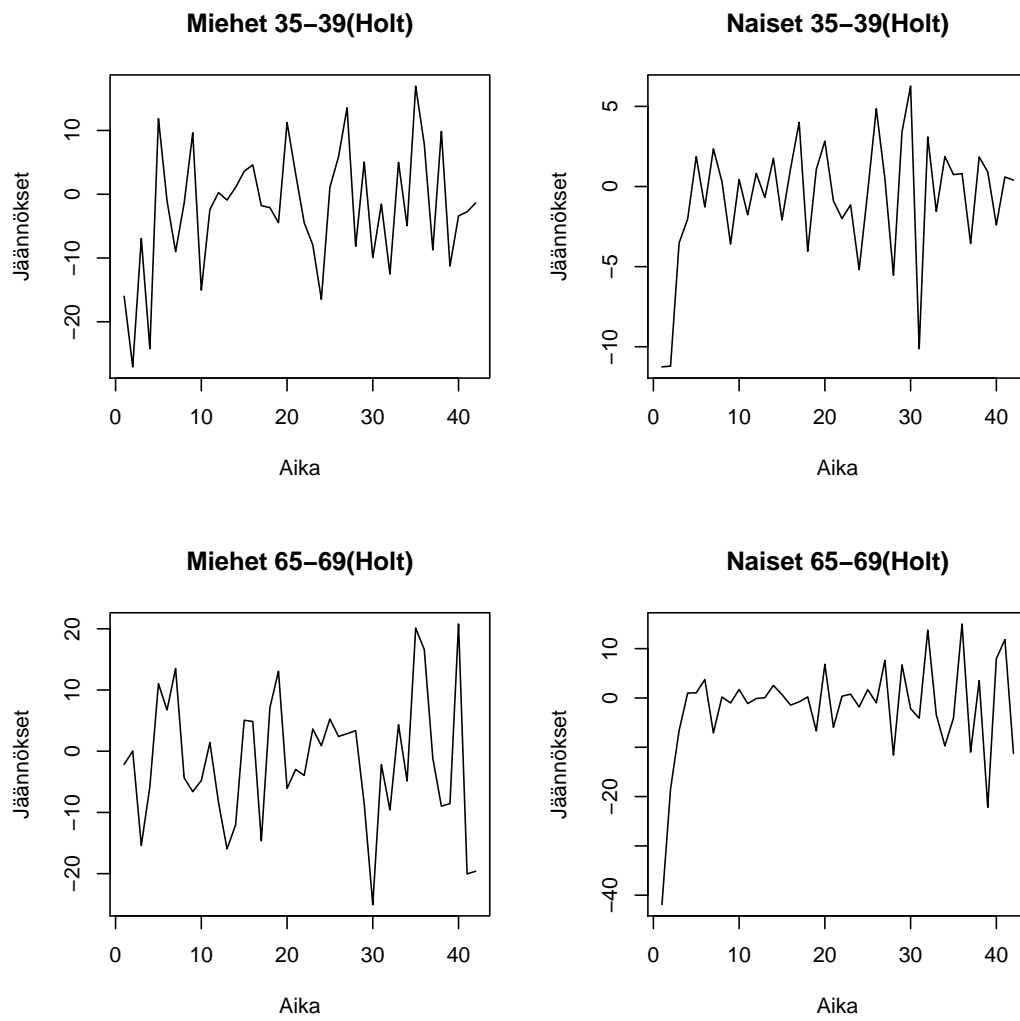
Taulukossa 8.8 olevien Holtin menetelmän jäännösten Ljung-Box testin tulosten perusteella mikään näistä ryhmistä ei eroaisi satunnaiskohinasta, eikä niissä siten olisi autokorrelaatiota ensimmäisen 10 viiveen aikana. Sen sijaan taulukossa 8.9 olevien yksinkertaisen eksponentiaalisen tasoituksen Ljung-Box testin tulosten perusteella 35-39 vuotiaiden naisten ryhmässä olisi autokorrelaatiota, sillä sen $p < 0.05$.

Taulukko 8.8. Ljung-Boxin testi Holtin menetelmälle.

	χ^2	vapausasteet	p-arvo
Miehet 35-39	6.391	10	0.781
Naiset 35-39	6.219	10	0.797
Miehet 65-69	15.075	10	0.130
Naiset 65-69	2.757	10	0.987



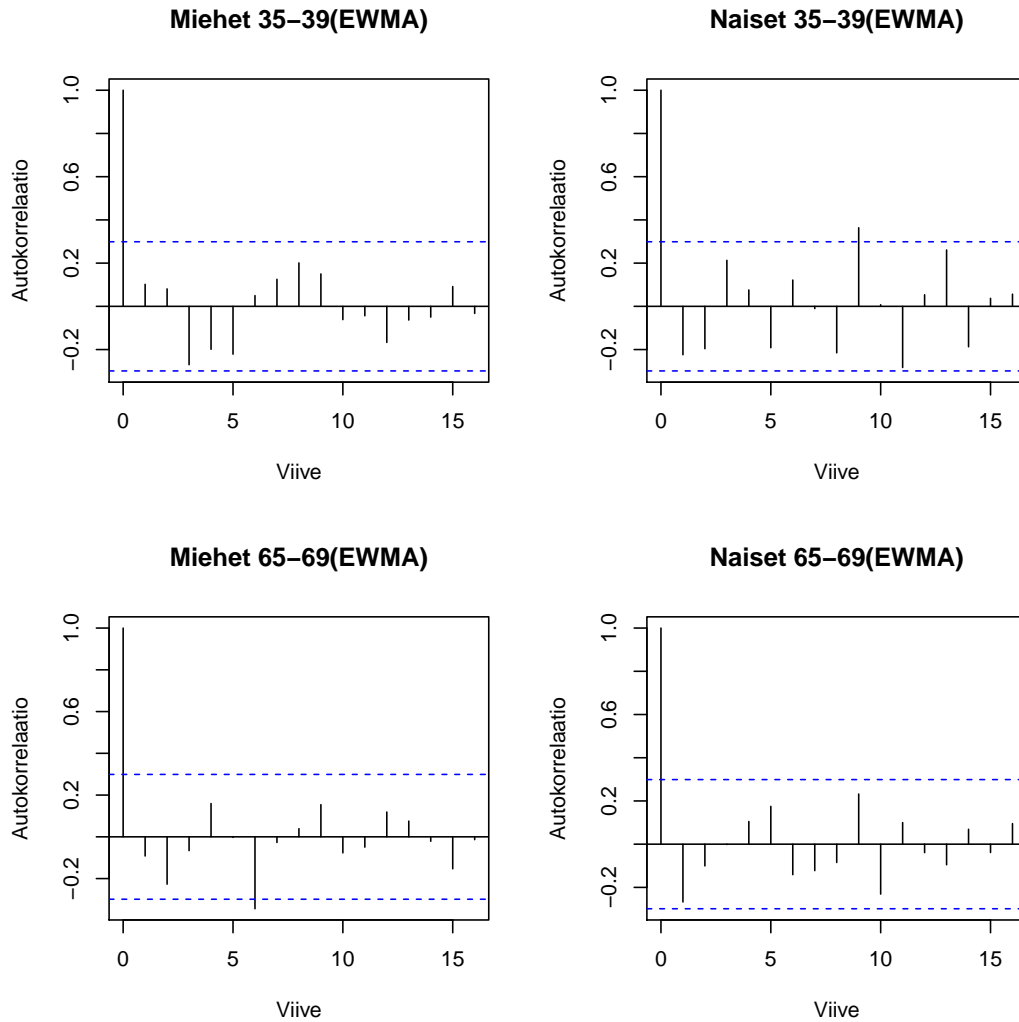
Kuva 8.5. Autokorrelaatiogrammi Holtin menetelmälle.



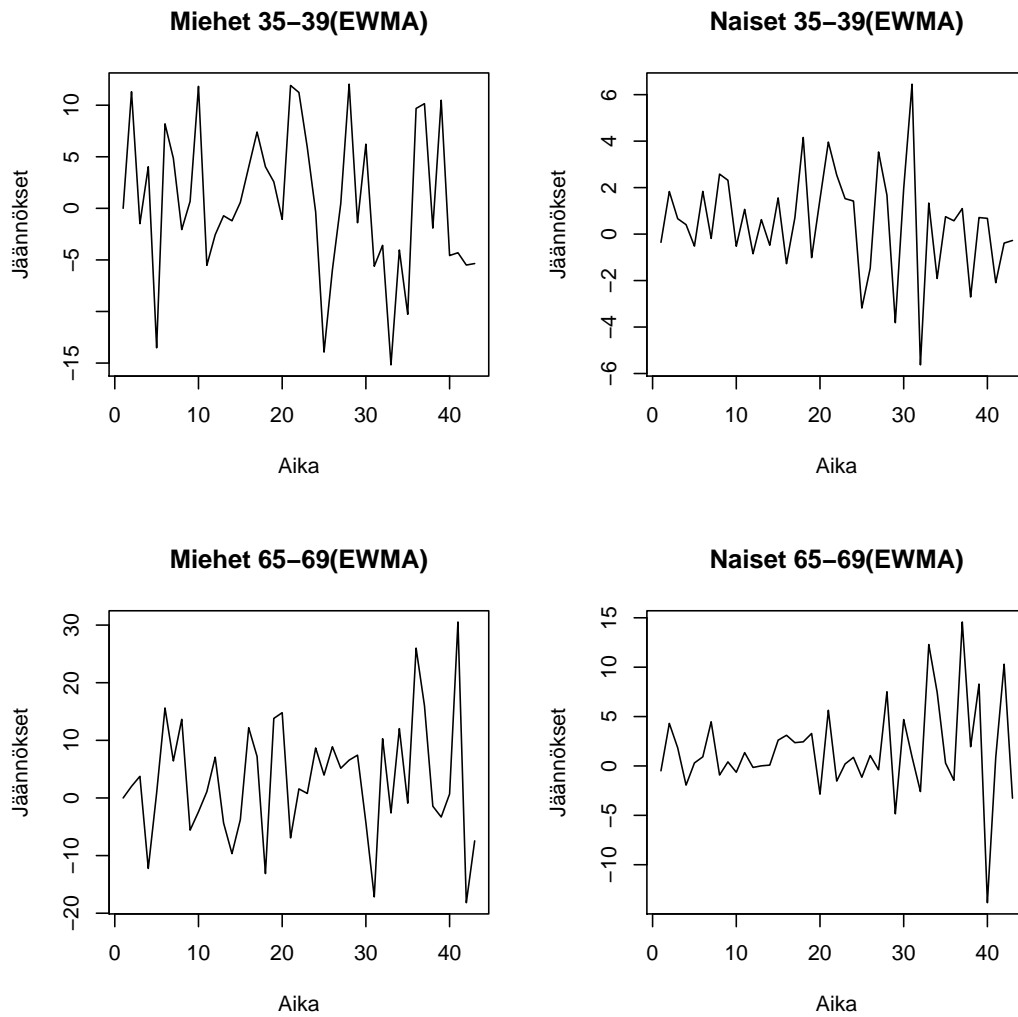
Kuva 8.6. Jäännökset Holtin menetelmälle miehet ja naiset 35-39 sekä 65-69 vuotta.

Taulukko 8.9. Ljung-Boxin testi yksinkertaiselle eksponentiaaliselle ta-
soitukselle.

	χ^2	vapausasteet	p-arvo
Miehet 35-39	13.476	10	0.198
Naiset 35-39	19.341	10	0.036
Miehet 65-69	12.322	10	0.264
Naiset 65-69	14.319	10	0.159



Kuva 8.7. Autokorrelaatiogrammi yksinkertaiselle eksponentiaaliselle ta-
soitukselle.



Kuva 8.8. Jäännökset yksinkertaiselle eksponentiaaliselle tasoitukselle miehet ja naiset 35-39 sekä 65-69 vuotta.

8.4 Muut ikäluokat

Liitteenä olevissa kuvissa B.1, B.2, B.3, B.4 ja B.5 ovat nähtävillä miesten alkoholikuolleisuudet 5-vuotislukot vuosille 1969-2016 ja näille Holtin menetelmällä lasketut ennusteet. Vertailujen helpottamiseksi kuvat sisältävät myös jo esiteltyt ikä- ja sukupuoliluokat. Osassa ikäluokista raju vaihtelu aikasarjan alussa aiheuttaa rekursiokaavoille asetettujen alkuarvojen suuren poikkeaman havaituista alkoholikuolleisuuksista. Poikkeavillakin alkuarvoilla alustetut rekursiokaavat lähestyvät nopeasti toteutuneita alkoholikuolleisuuslukuja. Ikäluokassa 15-19 havaintoja alkoholikuolemista ei ole joka vuodelta, mutta myös tässä ikäluokassa on havaittavissa alkoholikuolleisuuden tasaantumista tai jopa laskua. Vastaava trendi on nähtävissä lähes jokaisessa miesten ikäluokassa, ai-

na kuvassa B.4 oleviin yli 60-vuotiaiden ikäluokkiin asti. Näyttäisikin siltä, että vaikka kuvan 7.1 perusteella miesten alkoholikuolleisuus on selvässä laskussa, niin tämä ei ulotu jokaiseen ikäluokkaan.

Harvassa yli 60-vuotiaiden miesten ryhmässä on alkoholikuolleisuudessa nähtävissä laskua tai edes tasaantumista. Eläkeiän lähellä olevien ja sen jo ylittäneissä miesikäluokissa alkoholikuolleisuus näyttäisikin usein olevan huolestuttavan jyrkässä kasvussa.

Naisten alkoholikuolleisuus 5-vuotislukissa ja Holtin menetelmällä vuosille 2011-2016 lasketut ennusteet ovat nähtävissä kuvissa B.6, B.7, B.8, B.9 ja B.10. Kuvista voidaan nähdä, että naisilla alkoholikuolleisuus on nousussa myös useassa eläkeikää nuoremmissa ikäluokassa. Eläkeiän lähestyminen ei naisissa jaa alkoholikuolleisuuden trendin kehittymistä niin selvästi kuin miehissä. Kokonaisuutena naisten alkoholikuolleisuus on kuitenkin jokaisessa ikäluokassa selvästi miesten vastaavia ikäluokkia alhaisempaa.

Lähteet

- Alkoholilaki (1968/459). Annettu Helsingissä 26.7.1968. Viitattu 22.4.2018. Saatavissa: <https://www.finlex.fi/fi/laki/ajantasa/kumotut/1968/19680459>
- (1994/1143). Annettu Helsingissä 8.12.1994. Viitattu 22.4.2018. Saatavissa: <https://www.finlex.fi/fi/laki/ajantasa/1994/19941143>
- (2017/1102). Annettu Helsingissä 28.12.2017. Viitattu 22.4.2018. Saatavissa: <https://www.finlex.fi/fi/laki/alkup/2017/20171102>
- Brown, R.G. (1959), "Statistical Forecasting for Inventory Control". New York: McGraw-Hill.
- (1963), "Smoothing, Forecasting and Prediction of Discrete Time Series". Englewood Cliffs: Prentice-Hall.
- Gardner, E.S.Jr. (2006), "Exponential smoothing: The state of the art - Part II". International Journal of Forecasting, 22 (2006): 637-666. Amsterdam : Elsevier.
- Harvey, A.C. (1989), "Forecasting, structural time series models and the Kalman filter". Cambridge; New York: : Cambridge University Press.
- Holt, C.C. (1957), "Forecasting seasonals and trends by exponentially weighted moving averages". ONR Research Memorandum 52, Carnegie Institute of Technology: Pittsburgh, Pennsylvania.
- Hyndman, R. J. (2008), "Forecasting with exponential smoothing: The state space approach". Berlin; London: Springer.
- Hyndman, R.J. & Athanasopoulos, G. (2013), "Forecasting: principles and practice". Saatavissa: <https://www.otexts.org/fpp>. Viitattu 28.3.2018.
- Karlsson, T., Kotovirta, T., Tigerstedt, C. & Warpenius, K. (toim.) (2013), "Alkoholi Suomessa: Kulutus, haitat ja politiikkatoimet". Helsinki: Terveysten ja hyvinvoinnin laitos.
- Muth, J.F. (1960), "Optimal properties of exponentially weighted forecasts". Journal of the American Statistical Association, 55: 299-306.
- R Development Core Team (2012), "R: A language and environment for statistical computing". R Foundation for Statistical Computing, Vienna, Austria. Saatavissa: <http://www.R-project.org/>.
- Suomen virallinen tilasto (2017a), "Kuolemansyyt". Helsinki: Tilastokeskus. Saantitapa: <http://www.stat.fi/til/ksyyt/>. Viitattu: 29.3.2018.
- (2017b), "Väestörakenne". Helsinki: Tilastokeskus. Saantitapa: <http://www.stat.fi/til/vaerak/>. Viitattu: 29.3.2018.
- Winters, P.R. (1960), "Forecasting sales by exponentially weighted moving averages". Management Science 6: 324-42.

Liite A

Liite A: R-koodit

```
### holt-wintersin menetelmä log-uskottavuuden optimointi
pois.logL.hw <- function(par,y,ut) {
  y[y==0] <- 0.5
  phi0 <- 1/(1+exp(-par[1]))
  phi1 <- 1/(1+exp(-par[2]))
  at <- log(y[2]/ut[2])
  bt <- log(y[2]/ut[2]) - log(y[1]/ut[1])
  log_mt <- NULL

  for(i in 2:length(y)) {
    at[i] <- phi0*log(y[i]/ut[i])
      +(1-phi0)*(at[i-1]+bt[i-1])
    bt[i] <- phi1*(at[i]-at[i-1])
      +(1-phi1)*bt[i-1]
    log_mt[i-1] <- at[i]+bt[i]
  }
  y <- y[-(1:2)]
  ut <- ut[-(1:2)]
  log_mt <- log_mt[-length(log_mt)]
  -sum(y*log(ut)+y*log_mt-sum(log(y))-ut*exp(log_mt))
}
```

```
##### Tason optimointi, yksinkertainen
##### eksponentiaalisen tasoitus
```

```
pois.logL <- function(par,y,ut) {
  y[y==0] <- 0.5
  phi <- 1/(1+exp(-par[1]))
  log_mt <- log(y[1]/ut[1])

  for(i in 2:length(y)) {
    log_mt[i] <- log_mt[i-1]+phi*(log(y[i-1]/ut[i-1])
      -log_mt[i-1])
  }
}
```

```

    }
    -sum(y*log(ut)+y*log_mt-sum(log(y))-ut*exp(log_mt))
}

```

```

### Ennusteiden laskeminen optimoiduilla arvoilla
# holtin menetelmä

```

```

pois.pred.HW <- function(par, n.pred=5, y, ut) {
  y[y==0] <- 0.5
  phi0 <- par[1]
  phi1 <- par[2]
  at <- log(y[2]/ut[2])
  bt <- log(y[2]/ut[2]) - log(y[1]/ut[1])

  h <- 2:n.pred
  log_mt <- at+bt

  for(i in 2:length(y)){
    at[i] <- phi0*log(y[i]/ut[i])+(1-phi0)*(at[i-1]
      +bt[i-1])
    bt[i] <- phi1*(at[i]-at[i-1])+(1-phi1)*bt[i-1]
    log_mt[i] <- at[i]+bt[i]
  }
  temp <- exp(at[length(at)]+h*bt[length(bt)])
  enn <- exp(log_mt)
  enn <- c(enn, temp)
  return(enn)
}

```

```

# ewma

```

```

pois.pred.ewma <- function(phi, n.pred, y, ut) {
  y[y==0] <- 0.5
  log_mt <- log(y[1]/ut[1])

  for(i in 2:(length(y)+1)) {
    log_mt[i] <- log_mt[i-1]+phi*(log(y[i-1]/ut[i-1])
      -log_mt[i-1])
  }
  log_mt <- c(log_mt, rep(log_mt[length(log_mt)], 4))
  enn <- exp(log_mt)
  return(enn)
}

```

```

### 90% ennusteväli Poisson-jakaumalle, toimii sekä
### ewma että h-w.

```

```

rajat.p <- function(sim, ennusteet){
  temp <- NULL
  yla <- NULL
  ala <- NULL
  enn <- ennusteet
  for(i in 1:length(enn)){
    temp <- rpois(n=sim, lambda=enn[i])
    temp <- sort(temp)
    ala[i] <- temp[.050*length(temp)]
    yla[i] <- temp[.950*length(temp)]
  }
  list(enn=ennusteet, ala=ala, yla=yla)
}

### phi:n 95%:n luottamusväli

philv.e <- function(phi, hesse, n) {
  fisherI <- solve(hesse)
  s <- sqrt(diag(fisherI))
  yla <- phi+1.96*s[1]/sqrt(n)
  ala <- phi-1.96*s[1]/sqrt(n)
  phi <- 1/(1+exp(-phi))
  yla <- 1/(1+exp(-yla))
  ala <- 1/(1+exp(-ala))

  return(data.frame(phi=phi, ala=ala, yla=yla))
}

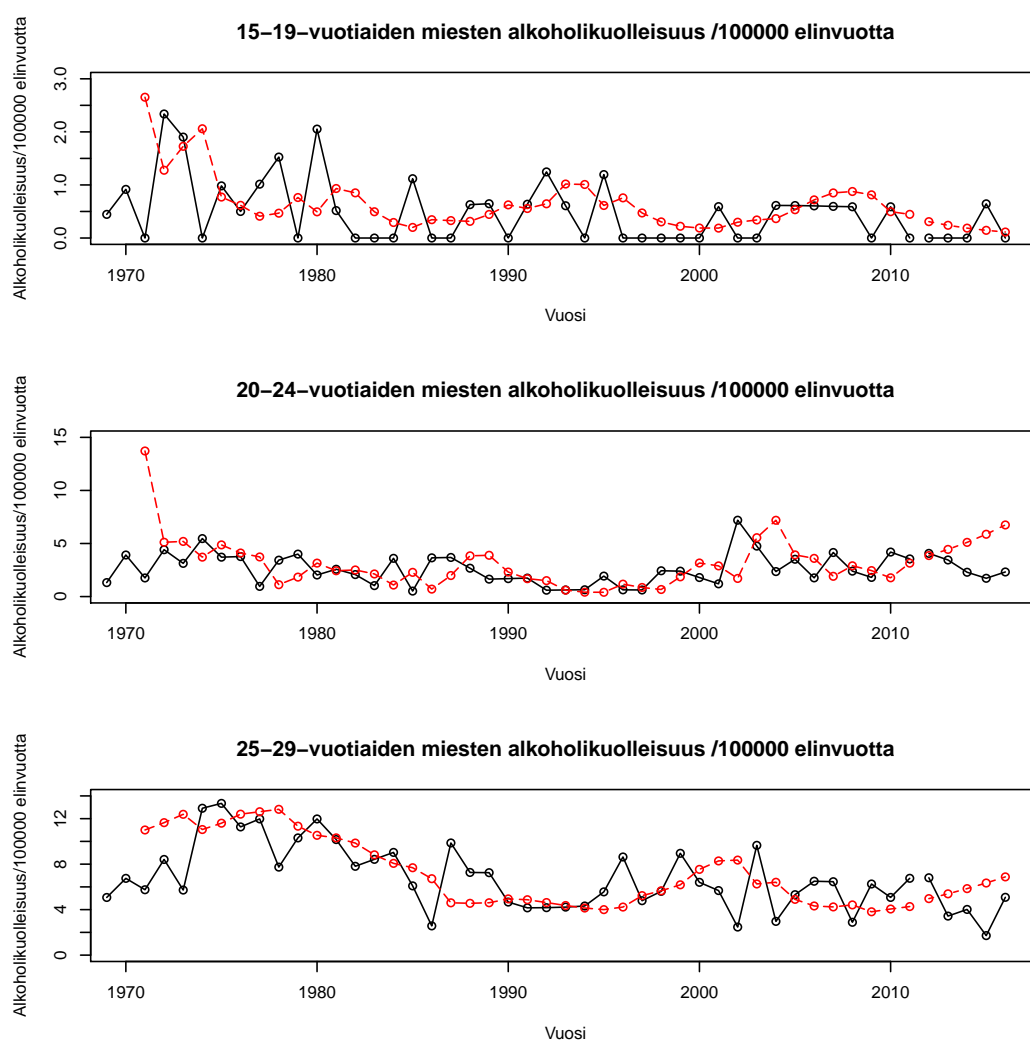
philv.hw <- function(phi, hesse, n) {
  fisherI <- solve(hesse)
  s <- sqrt(diag(fisherI))
  yla <- phi+1.96*s/sqrt(n)
  ala <- phi-1.96*s/sqrt(n)
  phi <- 1/(1+exp(-phi))
  yla <- 1/(1+exp(-yla))
  ala <- 1/(1+exp(-ala))

  return(data.frame(phi=phi, ala=ala, yla=yla))
}

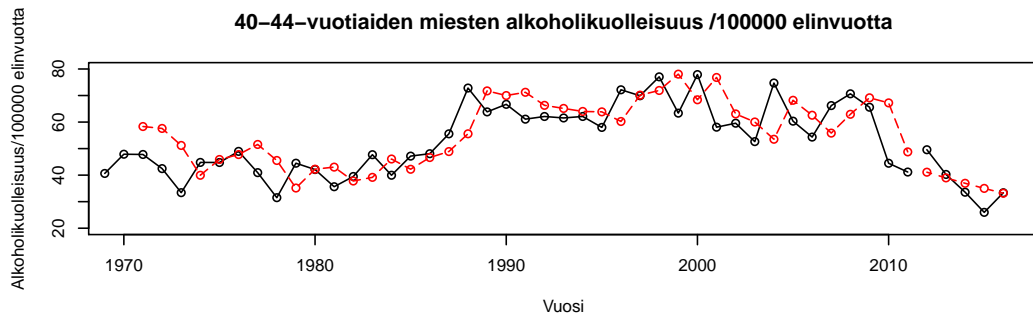
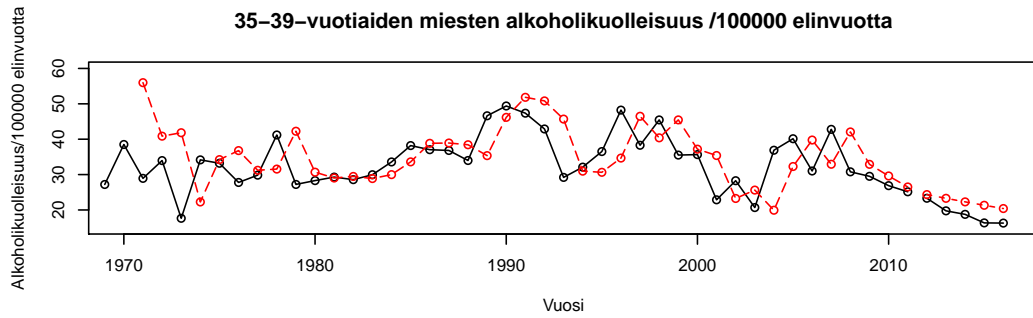
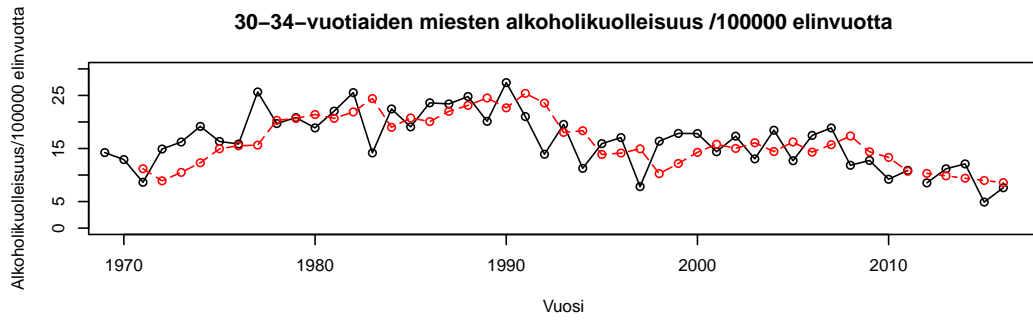
```

Liite B

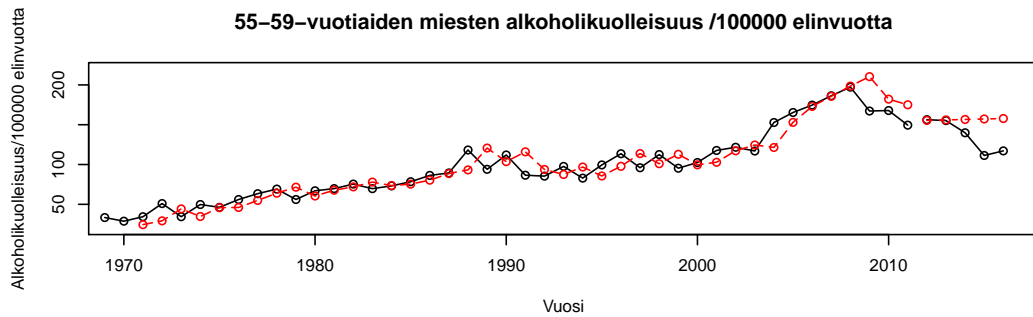
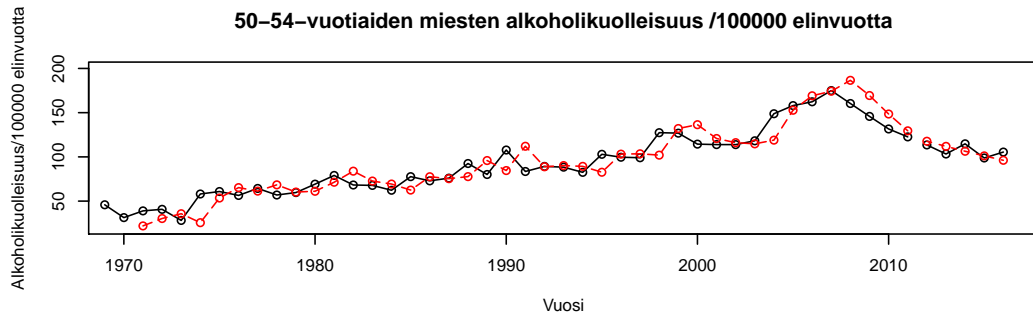
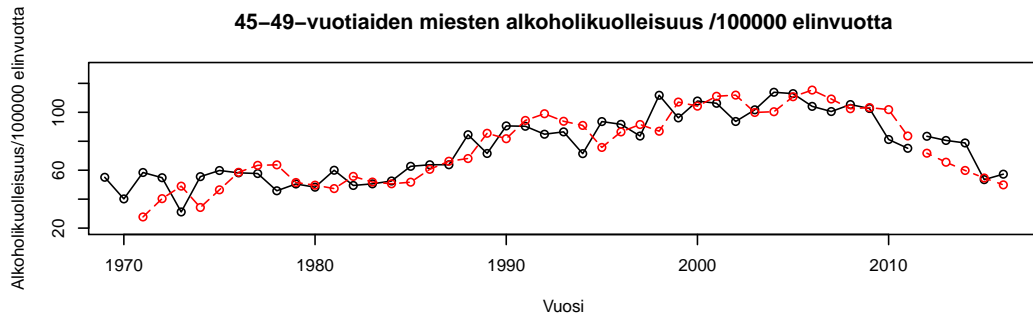
Liite B: Kuvat



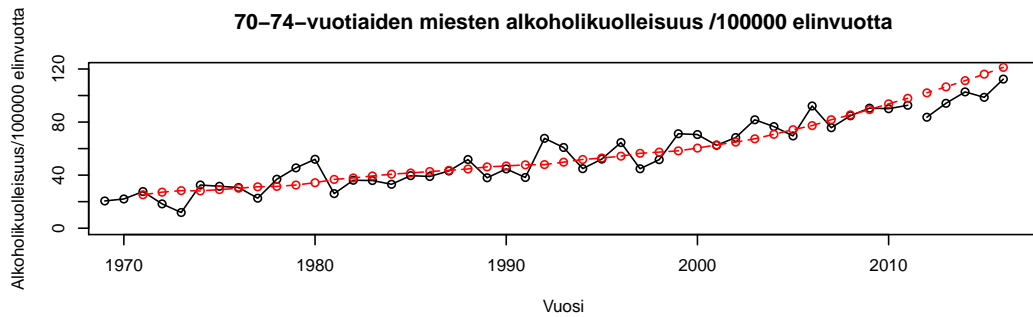
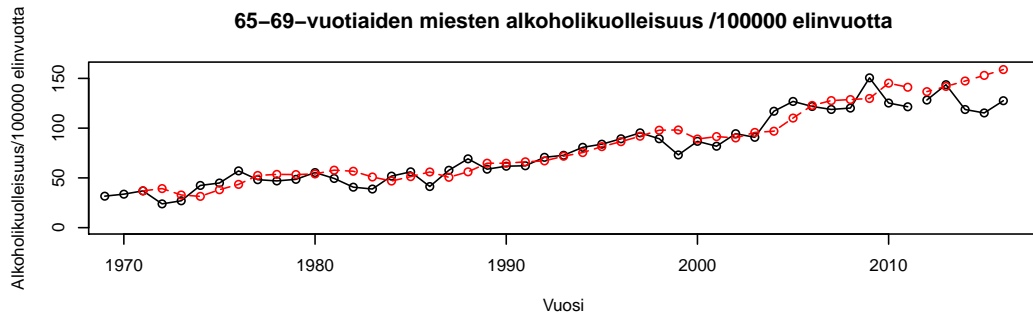
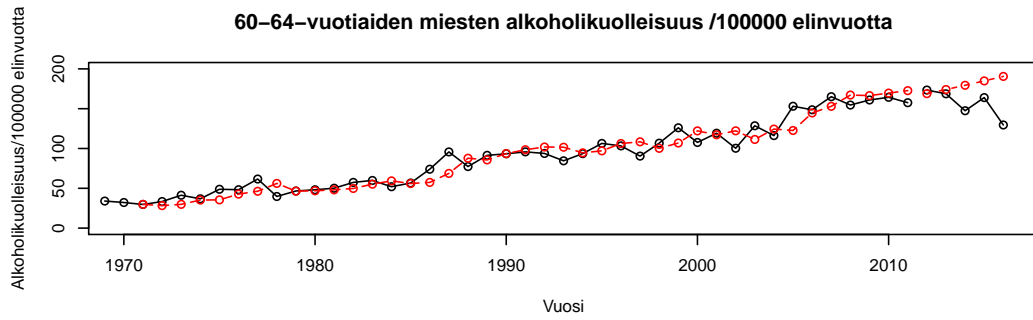
Kuva B.1. Holtin menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden miesten ikäluokille 15-19, 20-24, 25-29 vuotta. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.



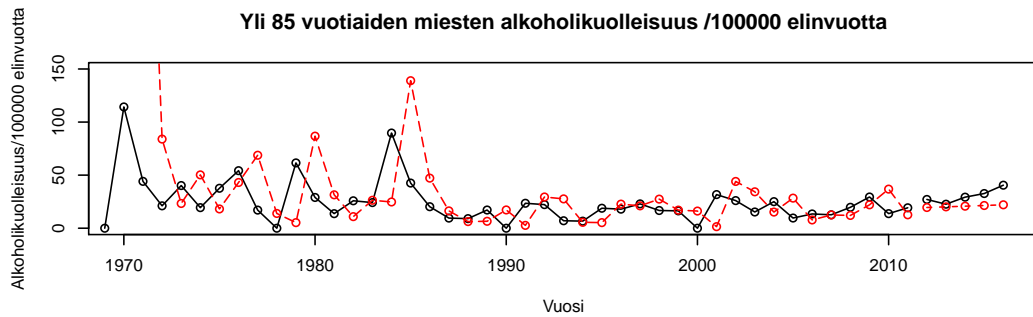
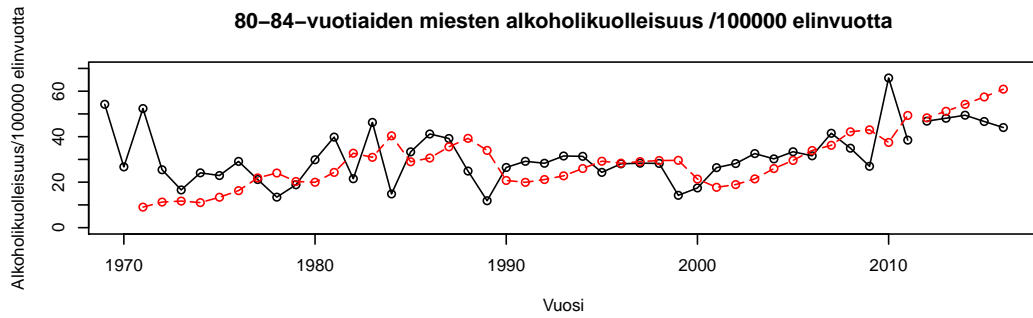
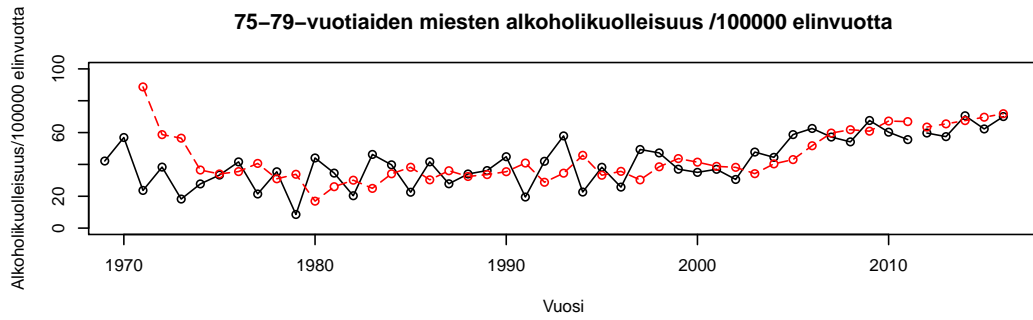
Kuva B.2. Holtin menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden miesten ikäluokille 30-34, 35-39, 40-44 vuotta. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.



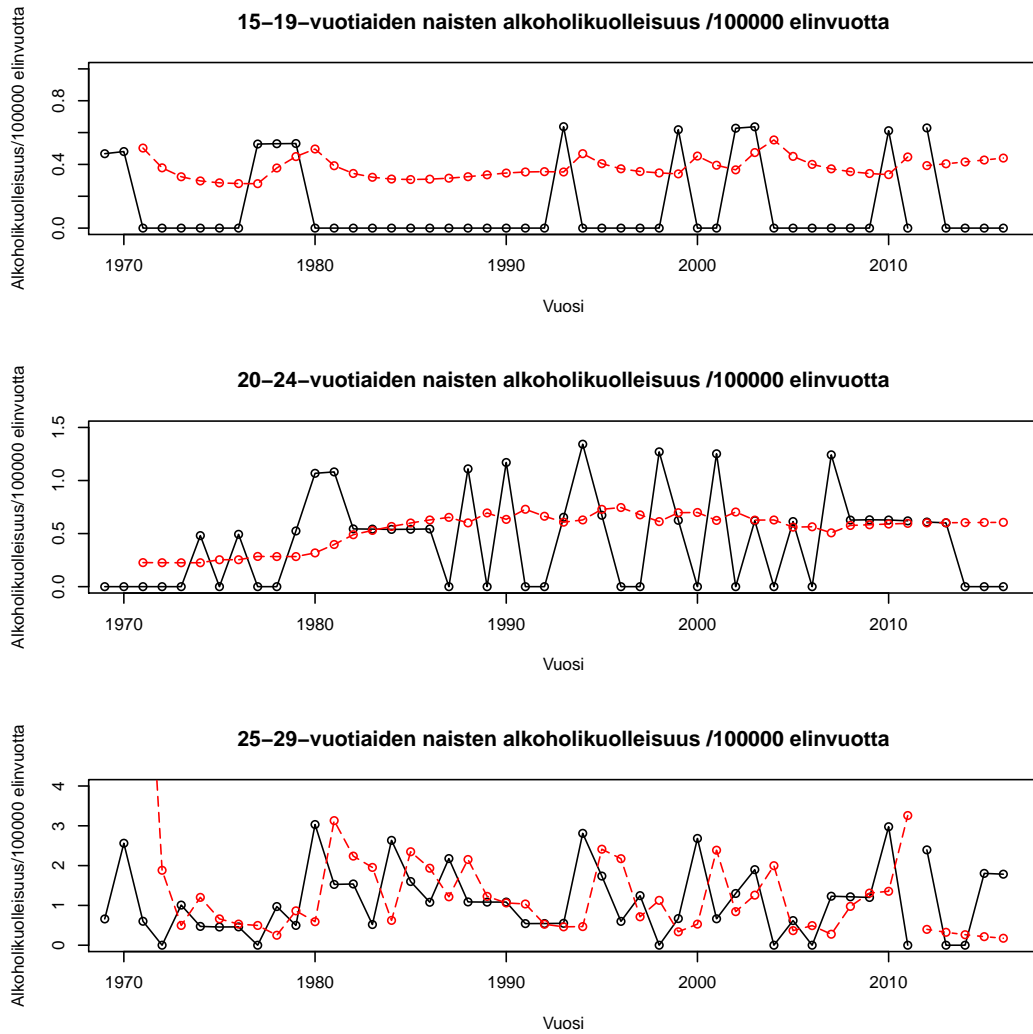
Kuva B.3. Holtin menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden miesten ikäluokille 45-49, 50-54, 55-59 vuotta. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.



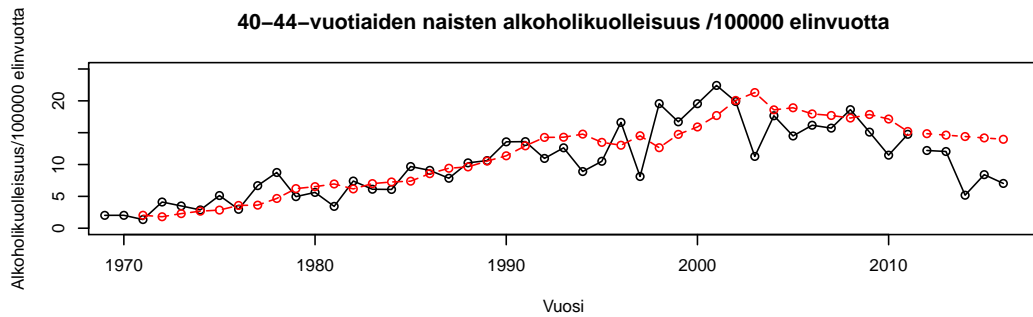
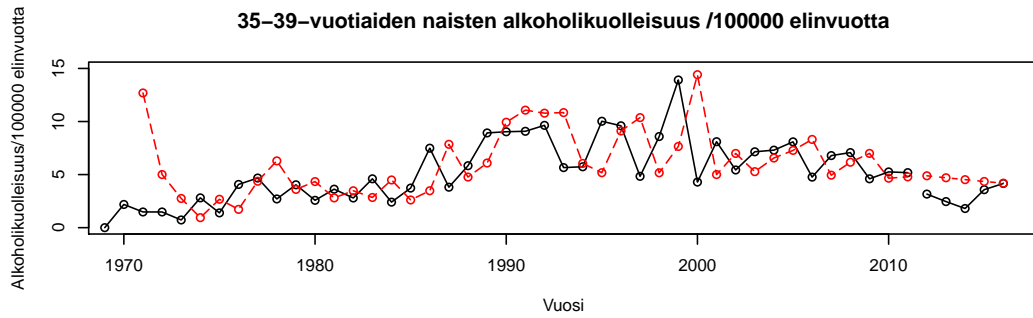
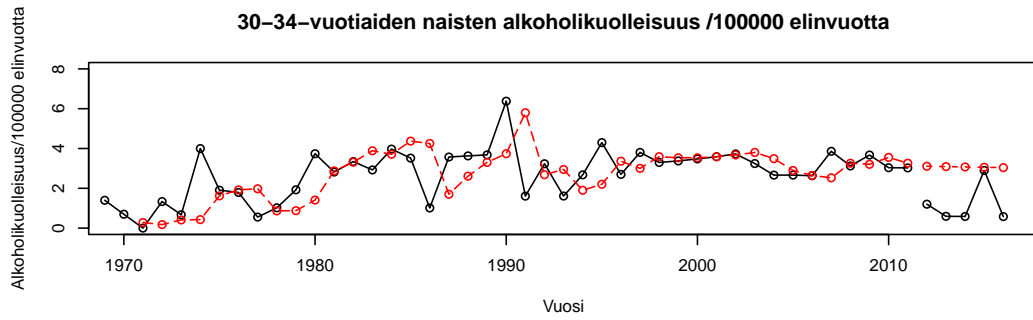
Kuva B.4. Holtin menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden miesten ikäluokille 60-64, 65-69, 70-74 vuotta. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.



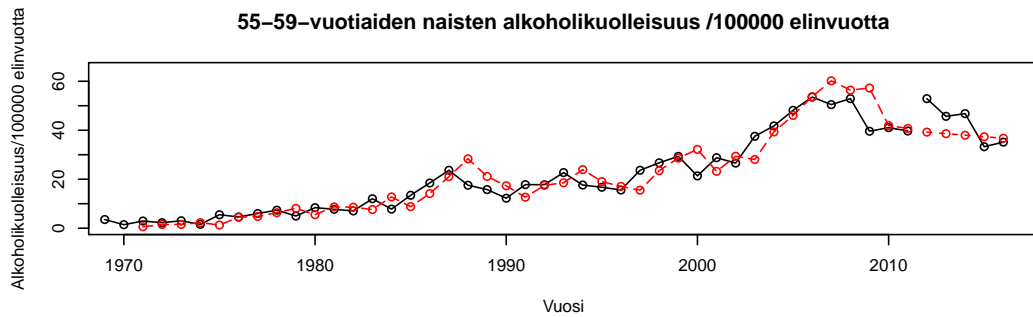
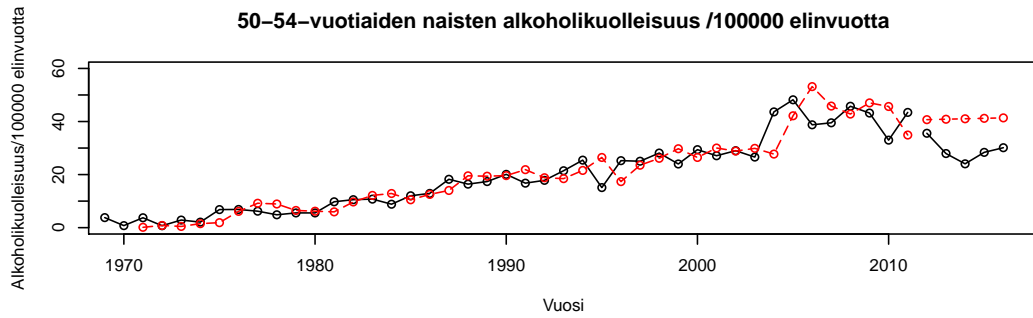
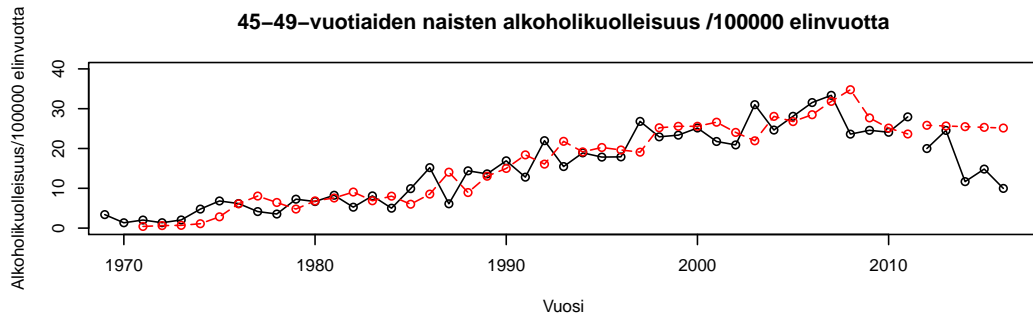
Kuva B.5. Holtin menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden miesten ikäluokille 75-79, 80-84 ja yli 85-vuotiaat. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.



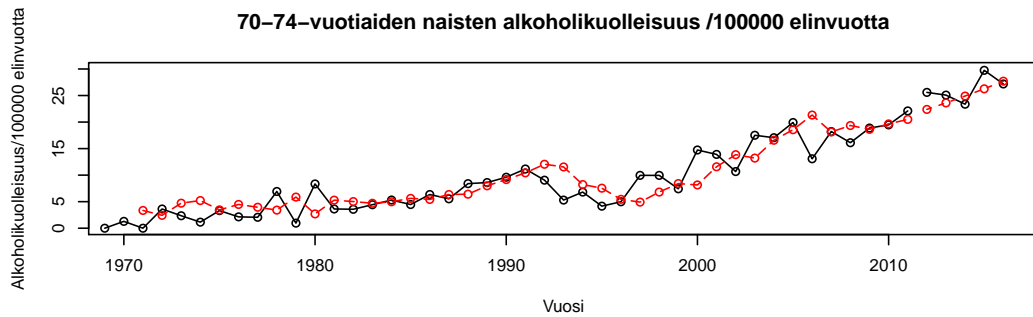
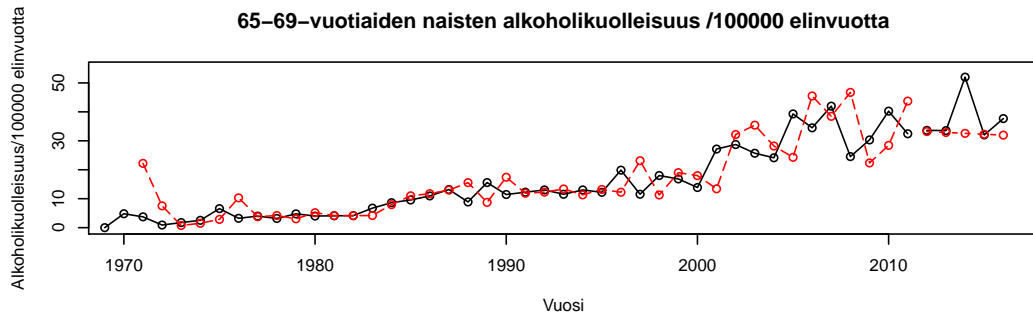
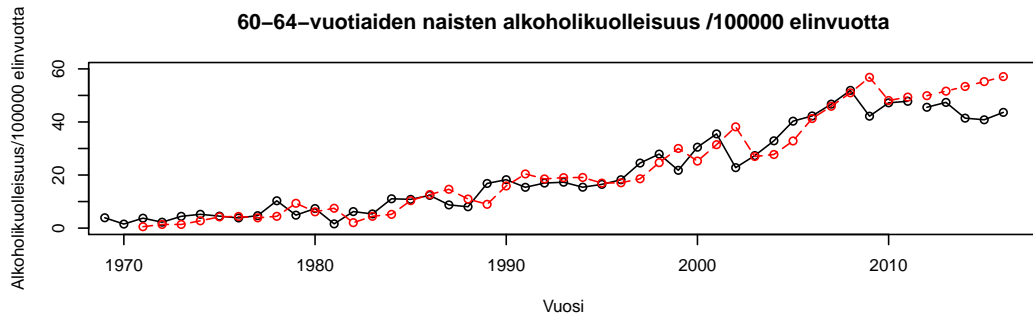
Kuva B.6. Holtin menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden naisten ikäluokille 15-19, 20-24, 25-29 vuotta. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.



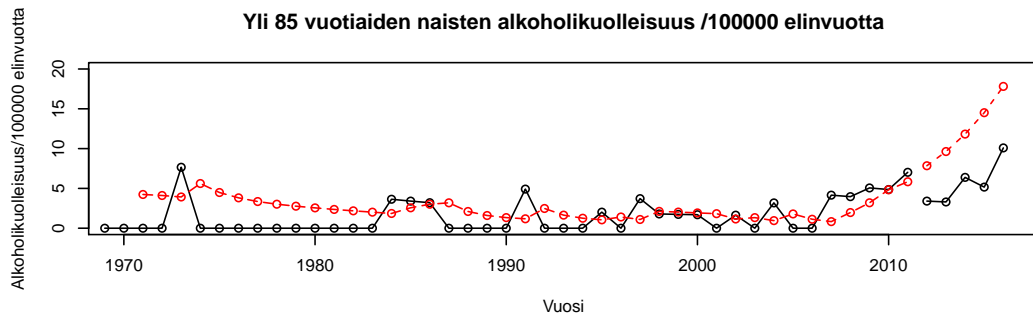
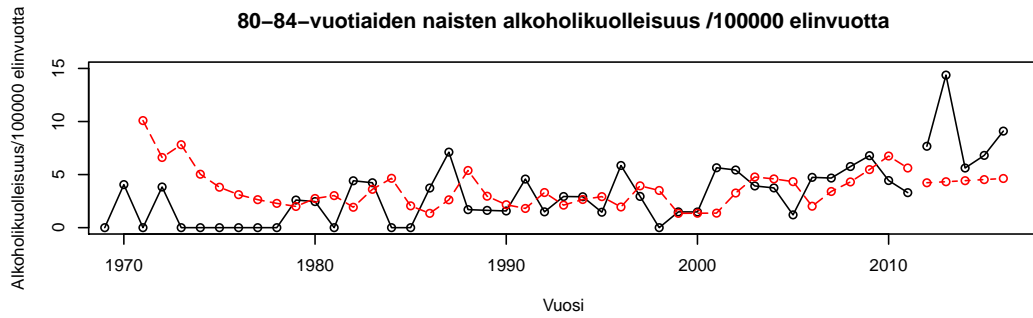
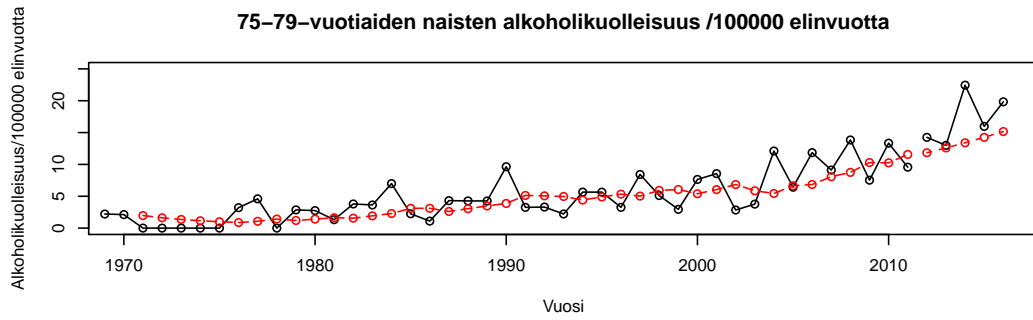
Kuva B.7. Holtin menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden naisten ikäluokille 30-34, 35-39, 40-44 vuotta. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.



Kuva B.8. Holtin menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden naisten ikäluokille 45-49, 50-54, 55-59 vuotta. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.



Kuva B.9. Holtin menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden naisten ikäluokille 60-64, 65-69, 70-74 vuotta. Ennustettut arvot punaisella katkoviivalla ja havaitut arvot mustalla.



Kuva B.10. Holtin menetelmällä estimoitu alkoholikuolleisuus 100000 elinvuotta kohden naisten ikäluokille 75-79, 80-84 ja yli 85-vuotiaat. Ennustetut arvot punaisella katkoviivalla ja havaitut arvot mustalla.