

Hirvipopulaation koon muutoksen ennustaminen saalisrakenneindeksien avulla

Tilastotieteen pro gradu -tutkielma

27.6.2014

Viivi Nuojua

Matematiikan ja tilastotieteen laitos

Jyväskylän yliopisto

Tiivistelmä

Suomen merkittävimmän riistaeläimen, hirven, populaation kokoa kontrolloidaan tarkasti luvanvaraisella metsästyksellä. On jo pitkään arveltu, että metsästyspaine heijastuisi saalisrakenneindekseihin, erityisesti saaliin vasanaarassuhteeseen ja aikuissaaliin urososuuteen, ja että näillä indekseillä olisi yhteys populaation koon muutokseen vuodesta toiseen. Tässä työssä osoitamme, että populaation koon muutoksen ja saalisrakenneindeksien välillä on tilastollinen yhteys ja että indeksejä voi käyttää populaation koon muutoksen ennustamisessa. Tarkastelemme myös populaation koon muutoksen ja indeksien välisen riippuvuuden alueellista vaihtelua Suomessa.

Populaation koon muutoksen ja saalisrakenneindeksien välisen yhteyden mallinnuksessa käytämme kiinteiden vaikutusten lineaarisia regressiomalleja ja lineaarisia sekamalleja, joissa populaation koon muutoksen selittäjinä toimivat joko vasanaarassuhde, urososuus tai molemmat. Aloitamme populaation koon muutoksen ja saalisrakenneindeksien välisen yhteyden tutkimisen valtakunnallisten regressiomallien sovittamisella. Näistä jatkamme riistanhoitopiirikohtaisiin regressiomalleihin, joista näemme, kuinka suuria eroja piirien välillä on. Otamme alueellisen vaihtelun huomioon laajentamalla regressiomallin sekamalliksi lisäämällä siihen piirien välistä vaihtelua kuvaavan satunnaisvaikutuksen.

Valtakunnallisesti ilman piirierottelua toimivat hyvin varsinkin regressiomallit, joissa populaation koon muutosta selittävät vasanaarassuhde tai vasanaarassuhde ja urososuus. Piiritasolla vasanaarassuhde yksinään toimii populaation koon muutoksen selittäjänä suuremmalle määrälle piirejä kuin urososuus tai molemmat yhdessä. Piirien välillä näyttää olevan enemmän eroa urososuuden kuin vasanaarassuhteen selittäessä populaation koon muutosta. Sekamalleihin siirryttäessä huomaamme yllättäen, että urososuuden tapauksessa sekamalli osoittautuu tarpeettomaksi, ja regressiomalli on riittävä. Neljän parhaaksi arvioidun mallin antamat piirikohtaiset kannanmuutosennusteet vuosille 2001–2012 ovat yhtä mallia lukuun ottamatta hyvin samanlaiset. Diagnostiset tarkastelut osoittavat, että paras sekamallimme ja erityisesti paras regressiomallimme ovat onnistuneita.

Avainsanat: Hirvi, populaation koon muutos, kannanmuutoksen ennustaminen, saalisrakenneindeksi, metsästyspaine, urososuus, vasanaarassuhde, Riistaweb, riistanhoitopiiri, kiinteiden vaikutusten lineaarinen regressiomalli, lineaarinen sekamalli.

SISÄLLYS

1	JOHDANTO.....	1
2	HIRVISTÄ JA NIIDEN METSÄSTYKSESTÄ.....	3
3	SAALISRAKENNEINDEKSIT JA KANNANMUUTOS.....	4
4	AINEISTO.....	5
5	LINEAARISTEN MALLIEN TEORIAA.....	7
5.1	Kiinteiden vaikutusten lineaarinen regressiomalli.....	7
5.2	Lineaarinen sekamalli.....	10
5.2.1	Kiinteiden vaikutusten estimointi.....	10
5.2.2	Satunnaisvaikutusten ennustaminen.....	11
5.2.3	Kovarianssimatriisin estimointi.....	12
5.2.4	Sisäkorrelaatio.....	13
5.3	Mallien vertailu.....	14
5.4	Käytetystä ohjelmistosta.....	14
6	KIIENTEIDEN VAIKUTUSTEN LINEAARISTEN REGRESSIOMALLIEN SOVITTAMINEN.....	15
6.1	Valtakunnalliset regressiomallit ilman piirierottelua.....	15
6.2	Regressiomallit piireittäin.....	17
6.3	Yhdysvaikutusten testaus.....	21
7	LINEAARISTEN SEKAMALLIEN SOVITTAMINEN.....	25
7.1	Paras vasanaarassuhteen sekamalli.....	25
7.2	Paras urososuuden sekamalli.....	26
7.3	Paras kahden selittäjän sekamalli.....	28
7.4	Paras sekamalli.....	30
8	YHTEENVETO TULOKSISTA.....	31
8.1	Parhaiksi arvioidut mallit.....	31
8.2	Kannanmuutoksen ennustaminen.....	32
8.3	Tasa-arvokäyrät.....	36
8.4	Mallien diagnostiikka.....	40
9	POHDINTAA.....	47
	LÄHTEET.....	50

1 JOHDANTO

Suomen hirvipopulaatiota pyritään kontrolloimaan jokasyksyisellä luvanvaraisella metsästyksellä. Suuret vuosittaiset saalismäärät tekevätkin hirvestä maamme merkittävimmän riistaeläimen. Tarkka populaation koon arviointi on hankalaa, mutta tässä auttavat vähintään 2000-luvun alusta asti tarkasti saatavilla olleet vuosittaiset saalismäärät ja jäävän kannan (metsästyskauden jälkeen eloonjääneet hirvet) arviot. Tässä tutkimuksessa käytämme aineistoa, joka on peräisin Suomen riistakeskuksen ja riistanhoitoyhdistysten riistatiedoista koostuvasta valtakunnallisesta tietokannasta, Riistawebistä (<http://www.riistaweb.riista.fi>, viitattu 2.6.2014). Tarvitsemamme saalistiedot ja jäävän kannan arviot ovat siellä täydelliset 13 riistanhoitopiirin osalta, ja tarkasteltava aikaväli rajoittuu vuosiin 2001–2012.

Riistantutkijoiden aiempien julkaisemattomien tutkimusten perusteella on arveltu, että hirvipopulaation koon muutoksen (tästä eteenpäin lyhyesti hirvikannan muutos tai pelkästään kannanmuutos) ja niin sanottujen saalisrakenneindeksien välillä saattaa olla yhteys. Voidaan nimittäin olettaa, että metsästyspaine heijastuu suoraan saalisrakenneindekseihin, vasanaarassuhteeseen ja urososuuteen (ks. luku 3). Tämän seurauksena saalisrakenneindeksit ja kannanmuutos edellisvuodesta näyttävät korreloivan positiivisesti. Tarkoituksenamme onkin selvittää, voidaanko saalisrakenneindeksejä käyttää kannanmuutoksen ennustamiseen. Lisäksi tutkimme kannanmuutoksen ja saalisrakenneindeksien välisen yhteyden alueellista vaihtelua Suomessa.

Käytämme tutkimusongelman ratkaisemiseen kiinteiden vaikutusten lineaarisia regressiomalleja ja lineaarisia sekamalleja, joissa selitämme kannanmuutosta joko yhdellä selittäjällä (vasanaarassuhde tai urososuus) tai kahdella selittäjällä (sekä vasanaarassuhde että urososuus). Aloitamme kannanmuutoksen ja saalisrakenneindeksien välisen yhteyden tutkimisen valtakunnallisista regressiomalleista ilman piirierottelua. Näin saamme viitettä siitä, kuinka vasanaarassuhde ja urososuus tulevat käyttäytymään kannanmuutoksen selittäjinä. Sen jälkeen sovitamme piirikohtaiset regressiomallit, jotta näemme, onko piirien välillä eroja. Alueellisen vaihtelun voi ottaa huomioon laajentamalla regressiomallin sekamalliksi. Täten lisäämme regressiomalleihin piirien välistä vaihtelua kuvaavan satunnaisvaikutuksen. Lisäksi tarkoituksenamme on tarkastella parhaita malleja hieman tarkemmin.

Seuraavaksi kerromme yleisesti hirvistä ja niiden metsästyksestä. Sen jälkeen luvussa 3 esittelemme käytettävät saalirakenneindeksit ja sen, millaisella suhdeluvulla kuvaamme hirvikannan muutosta. Luvussa 4 esittelemme aineiston: mistä se on peräisin, ja mitä tietoja siitä on otettu tutkimukseen. Tämän jälkeen käymme läpi kiinteiden vaikutusten lineaaristen regressiomallien ja lineaaristen sekamallien teoriaa sekä yleistä asiaa mallien vertailusta ja käytetystä ohjelmistosta. Lopuksi esitämme tutkimuksemme tulokset ja pohdintaa sekä menetelmistä että tuloksista.

2 HIRVISTÄ JA NIIDEN METSÄSTYKSESTÄ

Hirvi on Suomen merkittävin riistaeläin. Vuosina 2001–2011 saalismäärä on vaihdellut välillä 57097–84525 yksilöä/vuosi (<http://www.metsastajaliitto.fi>, viitattu 2.6.2014). Suomessa hirvikanta kontrolloidaan hyvin tarkoin lähinnä hirvien aiheuttamien liikennevahinkojen ja metsätuhojen vuoksi. Metsästys alkaa suurimmassa osassa maata syyskuun viimeisenä lauantaina ja kestää tauotta vuoden loppuun saakka. Viime vuosina pohjoisimmassa Lapissa ja käsivarressa metsästys on aloitettu kokeilumielessä jo syyskuun ensimmäisenä päivänä, syys-lokakuun vaihteessa parhaana kiima-aikana pidetään kolmen viikon mittainen tauko, ja metsästys lopetetaan jo marraskuun viimeisenä päivänä (<http://www.riista.fi>, viitattu 2.6.2014).

Hirvien vasominen ajoittuu loppukevääseen tai alkukesään, yleensä toukokuulle. Jos naaras on tullut syksyllä tiineeksi, se synnyttää kahdeksan kuukauden kantoajan jälkeen yleensä yhden tai kaksi vasaa. Vasat kulkevat emänsä mukana vajaan vuoden, jonka jälkeen emä häätää ne pois pari viikkoa ennen seuraavaa synnytystä. Aikuisista hirvistä voidaan käyttää myös nimityksiä sonni ja lehmä erottamaan vasat aikuisista. Tässä tutkielmassa käytämme kuitenkin urosta ja naarasta.

Hirviä saa metsästä vain Suomen riistakeskuksen myöntämällä asianmukaisella luvalla. Luvat jaetaan sen hetkisen hirvikannan suuruuden ja metsästysseurueelle kuuluvien metsästysalueiden yhteiskoon perusteella. Yksi lupa oikeuttaa joko yhteen aikuiseen hirveen tai kahteen vasaan. Metsästysseurueet saavat itse päättää, minkä verran ne ampuvat aikuisia ja minkä verran vasoja, mutta lupaan on usein merkitty aikuisten enimmäismäärä. Metsästäjien on otettava huomioon myös niin sanottu vasasuoja. Koska vasat seuraavat emäänsä vajaan vuoden synnytyksen jälkeen, ne ovat mukana myös metsästyskauden aikana. Vasasuoja tarkoittaa sitä, että ei saa ampua naarashirveä, jolla on vasa tai vasoja. Näin ollen erityisesti metsästyskauden alussa naaraita saatetaan ampua vähemmän kuin uroksia ja vasoja. Lisäksi metsästäjät arvostavat uroksia enemmän kuin naaraita niiden komeiden sarvien sekä usein myös suuremman koon ja paremman lihan vuoksi. Vasasuoja ja urosten suosiminen heijastuvat suoraan seuraavaksi esiteltäviin saalisrakenneindekseihin.

3 SAALISRAKENNEINDEKSIT JA KANNANMUUTOS

Pyrimme selittämään hirvikannan muutosta vuodesta toiseen seuraavanlaisilla saalisrakennelindekseillä:

1. aikuisten urosten osuus koko aikuissaaliista (aikuiset urokset/aikuissaalis)
2. vasoja suhteessa aikuisiin naaraisiin saaliissa (vasoja/aikuinen naaras)

Käytämme ensimmäisestä indeksistä lyhyesti nimitystä urososuus ja toisesta vasanaarassuhde. Urososuus on aina välillä 0–1, ja hieman vaikeammin hahmotettava vasanaarassuhde vaihtelee käyttämässämme aineistossa suunnilleen välillä 1.2–3.2.

Hirvikannan muutoksen ja saalisrakennelindeksien välisestä mahdollisesta yhteydestä on saatu viitteitä suomalaisten riistantutkijoiden julkaisemattomissa tutkimuksissa. Heidän mukaansa voidaan olettaa, että metsästyspaine heijastuu saalisrakennelindekseihin. Metsästyspaine on matala, kun lupia on vähän suhteessa kannan kokoon. Tällöin ammutaan vasasuojan vuoksi yleensä enemmän uroksia ja vasoja kuin naaraita. Näin ollen hirvikanta jää usein edellisvuotta suuremmaksi, ja myös saalisrakennelindeksit ovat suuria. Saalisrakennelindeksit ja kannanmuutos edellisvuodesta näyttävät siis korreloivan positiivisesti. Täten saalisrakennelindeksejä voidaan käyttää ennustamaan kannanmuutosta.

Hirvikannan muutosta kuvattaessa käytetään termiä jäävä kanta tai lisääntyvä kanta, jolla tarkoitetaan metsästyskauden jälkeen eloonjääneitä hirviä. Kuvaamme kannanmuutosta suhdeluvulla

$$\frac{(\text{jäävä kanta vuonna } t) - (\text{jäävä kanta vuonna } t - 1)}{\text{jäävä kanta vuonna } t - 1}$$

Tällöin otetaan huomioon väliin sijoittuva metsästys, joka aiheuttaakin lähestulkoon kaiken hirvien kuolleisuuden. Riista- ja kalatalouden tutkimuslaitoksen tutkija Tuire Nygrénin (2010) mukaan arviolta peräti 97–98 % hirvien kuolemista aiheutuu metsästyksestä. Muu kuolleisuus on niin pientä ja vaikeasti selvitettävissä, ettei sillä ole kovin suurta merkitystä.

4 AINEISTO

Käyttämämme aineisto on peräisin Suomen riistakeskuksen ja riistanhoitoyhdistysten riistatiedoista koostuvasta valtakunnallisesta tietokannasta, Riistawebistä. Se on kaikkien saatavilla Internet-sivustolta <http://www.riistaweb.riista.fi> (viitattu 2.6.2014). Tähän tutkimukseen olemme valinneet tarkasteltavaksi kaikki muut alueet paitsi Lapin ja Kainuun eli yhteensä 13 riistanhoitopiiriä. Riistawebistä puuttuvat Kainuun jäävän kannan tiedot kokonaan, ja Lapin jäävän kannan tiedot ovat täydelliset vain parilta viime vuodelta. Näin ollen nämä kaksi riistanhoitopiiriä jäävät automaattisesti pois tutkimuksesta, koska kannanmuutoslukuja ei pystytä laskemaan. Tutkimukseen mukaan otetut 13 riistanhoitopiiriä ovat Etelä-Häme, Etelä-Savo, Kaakkois-Suomi, Keski-Suomi, Oulu, Pohjanmaa, Pohjois-Häme, Pohjois-Karjala, Pohjois-Savo, Rannikko-Pohjanmaa, Satakunta, Uusimaa ja Varsinais-Suomi (ks. Kuva 1).

Tarkasteltavat 13 riistanhoitopiiriä koostuvat yhteensä 270 riistanhoitoyhdistyksestä. Käytännössä riistanhoitopiirit vastaavat maakuntia ja riistanhoitoyhdistykset niiden sisällä kuntia. Riistanhoitoyhdistykset edelleen koostuvat metsästysseurueista, joten kyseessä on kolmitasoinen hierarkia. Olemme ottaneet Riistawebistä tähän tutkimukseen tiedot saaliista ja jäävästä kannasta vuosilta 2000–2012 sekä valtakunnallisesti ilman piirierottelua että piireittäin. Koska tarvitsemme kannanmuutoksen laskemiseksi edellisen vuoden jäävän kannan tiedot, tarkasteltava aikaväli rajoittuu todellisuudessa vuosiin 2001–2012. Ensimmäistä vuotta käytämme vain kannanmuutoksen laskemiseen. Riistawebin jäävän kannan tiedot ovat peräisin metsästysseurueiden täyttämistä hirvihavaintokorteista, joten niihin liittyy hieman epävarmuutta.



Kuva 1 Suomi jaettuna riistanhoitopiireihin (<http://www.riistaweb.riista.fi>, viitattu 2.6.2014).

5 LINEAARISTEN MALLIEN TEORIAA

Selittäjämuuttujien vaikutusta selitettävään muuttujaan eli vasteeseen tutkitaan regressiomallien avulla. Kun vastetta selittävät vain kiinteät vaikutukset, kyseessä on kiinteiden vaikutusten regressiomalli. Kun selittäjiksi lisätään kiinteiden vaikutusten lisäksi satunnaisvaikutuksia, regressiomalli laajenee sekamalliksi. Lineaarisilla malleilla tarkoitetaan malleja, joissa vastemuuttuja voidaan kirjoittaa selittäjämuuttujien lineaarisena lausekkeena. Seuraavien lukujen 5.1 ja 5.2 teoria perustuu teoksiin Brown & Prescott (1999) sekä McCulloch & Searle (2001).

5.1 Kiinteiden vaikutusten lineaarinen regressiomalli

Yleinen kiinteiden vaikutusten lineaarinen regressiomalli on matriisimerkinnöin muotoa

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

missä \mathbf{y} on n -havaintovektori (vaste), \mathbf{X} on selittäjämuuttujien tunnettu $n \times p$ -design-matriisi (määrittelee jokaisen havainnon jokaista parametria vastaavat kiinteiden vaikutusten arvot), $\boldsymbol{\beta}$ on estimoitavien kiinteiden vaikutusten p -parametrivektori ja \mathbf{e} on n -jäännösvektori, jolle oletetaan $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, missä \mathbf{I} on $n \times n$ -identiteettimatriisi.

Merkitään vastemuuttujan \mathbf{y} odotusarvovektoria ja kovarianssimatriisiä

$$E(\mathbf{y}) = \boldsymbol{\mu} \quad \text{ja} \quad \text{cov}(\mathbf{y}) = \mathbf{V}$$

ja jakaumaa

$$\mathbf{y} \sim (\boldsymbol{\mu}, \mathbf{V}).$$

Tässä hyvin yleisessä vastemuuttujan \mathbf{y} jakaumassa odotusarvovektorissa $\boldsymbol{\mu}$ on n alkia ja kovarianssimatriisissa $\mathbf{V} = \mathbf{V}^T$ taas $n(n + 1)/2$ eri alkia.

Oletusten perusteella yleiselle kiinteiden vaikutusten lineaariselle regressiomallille pätee

$$E(\mathbf{y}) = \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \quad \text{ja} \quad \text{cov}(\mathbf{y}) = \mathbf{V} = \sigma^2 \mathbf{I}.$$

Näin ollen oletuksesta $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ seuraa, että

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}).$$

Havainnot \mathbf{y} ovat siis toisistaan riippumattomia (niiden väliset kovarianssit ovat nollija) ja niiden varianssit ovat samat, σ^2 .

Parametrien estimointi

Suurimman uskottavuuden menetelmällä saamme estimaattorit samalla kertaa molemmille parametreille $\boldsymbol{\beta}$ ja σ^2 . Kun oletamme, että vastemuuttuja \mathbf{y} noudattaa moniulotteista normaalijakaumaa, uskottavuusfunktio on muotoa

$$L(\boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right],$$

ja siten logaritminen uskottavuusfunktio

$$l(\boldsymbol{\beta}, \sigma^2) = \log L(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Kun derivoimme logaritmisen uskottavuusfunktion estimoitavien parametrien $\boldsymbol{\beta}$ ja σ^2 suhteen, saamme

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} = \frac{\mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}}{\sigma^2}$$

ja

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^4}.$$

Kun asetamme yllä olevat derivaatat nolliksi, saamme ratkaistua yhtälöistä suurimman uskottavuuden estimaattorit

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y},$$

jos käänteismatriisi $(\mathbf{X}^T \mathbf{X})^{-1}$ on olemassa, ja

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n}.$$

Käänteismatriisin olemassaolo edellyttää, että selittäjämuuttujien $n \times p$ -design-matriisi \mathbf{X} on astetta p eli täysiasteinen. Näin ei kuitenkaan usein ole. Silloin käytämme käänteismatriisin $(\mathbf{X}^T \mathbf{X})^{-1}$ tilalla yleistettyä käänteismatriisia $(\mathbf{X}^T \mathbf{X})^{-}$, jolle pätee

$$(\mathbf{X}^T \mathbf{X})(\mathbf{X}^T \mathbf{X})^{-}(\mathbf{X}^T \mathbf{X}) = \mathbf{X}^T \mathbf{X}.$$

Yleistetty käänteismatriisi ei ole yksikäsitteinen, minkä seurauksena ei myöskään ole sitä käyttämällä saatu estimaattori

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-} \mathbf{X}^T \mathbf{y}.$$

Sen sijaan $\mathbf{X}\hat{\boldsymbol{\beta}}$ on yksikäsitteinen ja lisäksi vastemuuttujan \mathbf{y} odotusarvon harhaton estimaattori, sillä

$$E(\hat{\boldsymbol{\mu}}) = E[\widehat{E(\mathbf{y})}] = E(\mathbf{X}\hat{\boldsymbol{\beta}}) = E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}.$$

Jos suurimman uskottavuuden menetelmän vaatimaa normaalisuusoletusta ei voida tehdä, parametrien $\boldsymbol{\beta}$ estimaattoreiden johtaminen onnistuu pienimmän neliösumman menetelmällä. Siinä tarkoituksena on minimoida aiemmin uskottavuusfunktiossakin esiintynyt virheneliösumma $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ parametrin $\boldsymbol{\beta}$ suhteen. Tuloksena saamme parametrille $\boldsymbol{\beta}$ saman estimaattorin kuin suurimman uskottavuuden menetelmälläkin.

Suurimman uskottavuuden menetelmällä parametrille σ^2 saatu estimaattori on alaspäin harhainen. Kun vaihdamme sen nimittäjäksi $n - p$, missä $p = \text{rank}(\mathbf{X})$, saamme harhattoman estimaattorin

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - p},$$

sillä on tiedossa, että

$$E[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] = (n - p)\sigma^2,$$

jolloin

$$E(\hat{\sigma}^2) = \frac{E[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]}{n - p} = \frac{(n - p)\sigma^2}{n - p} = \sigma^2.$$

5.2 Lineaarinen sekamalli

Lisäämällä luvussa 5.1 esiteltyyn kiinteiden vaikutusten lineaariseen regressiomalliin satunnaisvaikutuksia malli laajenee lineaariseksi sekamalliksi. Sekä kiinteistä vaikutuksista että satunnaisvaikutuksista koostuva lineaarinen sekamalli näyttää matriisimuotoon kirjoitettuna seuraavalta:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

missä \mathbf{y} , \mathbf{X} , $\boldsymbol{\beta}$ ja \mathbf{e} ovat kuten regressiomallissa, \mathbf{Z} on satunnaisvaikutusten tunnettu $n \times q$ -designmatriisi (määrittelee jokaista havaintoa vastaavat satunnaisvaikutukset) ja \mathbf{u} on satunnaisvaikutusten q -parametrivektori. Oletetaan lisäksi $\mathbf{u} \sim N(\mathbf{0}, \mathbf{D})$ ja $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$. Näin ollen vastemuuttujan \mathbf{y} jakauma on sekamallien tapauksessa

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}),$$

missä $\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R}$. Kiinteitä vaikutuksia käytetään siis odotusarvon mallintamiseen, kun taas satunnaisvaikutukset määräävät kovarianssirakenteen. Lisäksi kiinteiden vaikutusten lineaarisesta regressiomallista poiketen havainnot \mathbf{y} eivät ole nyt toisistaan riippumattomia. Lineaaristen sekamallien tapauksessa eri luokkiin (esimerkiksi alueisiin) kuuluvien havaintojen ajatellaan olevan toisistaan riippumattomia, mutta saman luokan sisällä ne riippuvat toisistaan.

5.2.1 Kiinteiden vaikutusten estimointi

Kun oletamme, että vastemuuttuja \mathbf{y} on normaalijakautunut ja sen kovarianssimatriisi \mathbf{V} tunnettu, kiinteiden parametrien $\boldsymbol{\beta}$ estimointi onnistuu jälleen suurimman uskottavuuden menetelmällä. Tällöin uskottavuusfunktio on muotoa

$$L(\boldsymbol{\beta}) = (2\pi)^{-\frac{n}{2}} |\mathbf{V}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right],$$

ja siten logaritminen uskottavuusfunktio

$$l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Derivoimalla logaritminen uskottavuusfunktio estimoitavien parametrien $\boldsymbol{\beta}$ suhteen, asettamalla derivaatta nolaksi ja ratkaisemalla yhtälöstä $\boldsymbol{\beta}$, saamme suurimman uskottavuuden estimaattoriksi

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}.$$

Edellytyksenä on kuitenkin käänteismatriisin $(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$ olemassaolo. Jos sitä ei ole olemassa, voimme käyttää yleistettyä käänteismatriisia $(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-}$ kuten aiemmin kiinteiden vaikutusten regressiomallin tapauksessa. Yleistetty käänteismatriisi ja sitä käyttämällä saatu estimaattori

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$$

eivät ole yksikäsitteisiä, mutta kiinteiden vaikutusten regressiomallin tapaan $\mathbf{X}\hat{\boldsymbol{\beta}}$ sitä vastoin on.

5.2.2 Satunnaisvaikutusten ennustaminen

Satunnaisvaikutuksia käsitellään eri tavalla kuin kiinteitä vaikutuksia. Kun kiinteät vaikutukset ajatellaan estimoitavina vakioina, satunnaisvaikutukset taas nähdään realisaatioina jostain jakaumasta, yleensä normaalijakaumasta. Estimoinnin sijaan käytetäänkin termiä ennustaminen. Oletamme edelleen vasteen \mathbf{y} kovarianssimatriisin \mathbf{V} tunnetuksi. Paras ennuste $BP(\mathbf{u})$ (*best predictor*) satunnaisvaikutuksille \mathbf{u} on ehdollinen odotusarvo

$$\tilde{\mathbf{u}} = BP(\mathbf{u}) = E(\mathbf{u}|\mathbf{y}),$$

missä ”parhaasta” puhuttaessa tarkoitetaan pienintä keskineliövirhettä. Oletamme edelleen, että $\mathbf{u} \sim N(\mathbf{0}, \mathbf{D})$ ja $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$ ja lisäksi, että \mathbf{u} ja \mathbf{y} noudattavat yhteistä moniulotteista normaalijakaumaa, jolloin

$$E(\mathbf{u}|\mathbf{y}) = \mathbf{D}\mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

(ks. Morrison 1990, 92). Kun korvaamme estimoitavat parametrit $\boldsymbol{\beta}$ suurimman uskottavuuden estimaattorilla $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$, saamme satunnaisvaikutusten \mathbf{u} parhaan lineaarisen harhattoman ennusteen $BLUP(\mathbf{u})$ (*best linear unbiased predictor*)

$$\tilde{\mathbf{u}}^0 = BLUP(\mathbf{u}) = E(\mathbf{u}|\mathbf{y}) = \mathbf{DZ}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

Tällöin kovarianssimatriisi saa muodon

$$\text{cov}(\tilde{\mathbf{u}}^0) = \mathbf{DZ}^T \mathbf{PZD}$$

(ks. McCulloch & Searle 2001, 170), missä $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X}(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}$. Lisäksi satunnaisvaikutusten ennusteiden $\tilde{\mathbf{u}}$ ja varsinaisten satunnaisvaikutusten \mathbf{u} variansseille pätee

$$\text{var}(\tilde{\mathbf{u}}) \leq \text{var}(\mathbf{u}),$$

ts. satunnaisvaikutusten ennusteiden $\tilde{\mathbf{u}}$ vaihtelu on pienempää tai yhtä suurta kuin varsinaisten satunnaisvaikutusten \mathbf{u} vaihtelu. Tätä ilmiötä kutsutaan kutistumiseksi (*shrinkage*).

5.2.3 Kovarianssimatriisin estimointi

Jos vastemuuttujan \mathbf{y} kovarianssimatriisi \mathbf{V} on tuntematon, se täytyy estimoida. Estimointi voidaan toteuttaa joko suurimman uskottavuuden menetelmällä tai rajoitetulla suurimman uskottavuuden menetelmällä eli lyhyesti REML-menetelmällä (*restricted/residual maximum likelihood*). Perinteisessä suurimman uskottavuuden menetelmässä maksimoimme logaritmisen uskottavuusfunktion

$$l(\mathbf{V}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

parametrien $\boldsymbol{\beta}$ suhteen, jolloin tuloksena on estimaattori

$$\hat{\boldsymbol{\beta}}_{ML} = (\mathbf{X}^T \hat{\mathbf{V}}_{ML}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}_{ML}^{-1} \mathbf{y},$$

missä $\hat{\mathbf{V}}$ on saatu maksimoimalla sama uskottavuus kovarianssimatriisin \mathbf{V} parametrien suhteen. Maksimointi voidaan tehdä numeerisesti esimerkiksi Newton-Raphsonin algoritmilla.

REML-menetelmän ideana on soveltaa suurimman uskottavuuden menetelmää vastemuuttujan \mathbf{y} lineaarisiin funktioihin $\mathbf{K}^T \mathbf{y}$, joissa ei esiinny lainkaan parametria $\boldsymbol{\beta}$. Matriisi \mathbf{K}^T valitaan siten, että

sillä on mahdollisimman monta lineaarisesti riippumatonta riviä ja sille pätee $\mathbf{K}^T \mathbf{X} = \mathbf{0}$. Näin ollen matriisin \mathbf{K}^T rivit ovat täyttä astetta $n - p$. Kun oletamme edelleen, että vastemuuttuja \mathbf{y} noudattaa moniulotteista normaalijakaumaa, saamme

$$\mathbf{K}^T \mathbf{y} \sim \mathbf{N}(\mathbf{0}, \mathbf{K}^T \mathbf{V} \mathbf{K}).$$

Logaritminen uskottavuusfunktio on nyt muotoa

$$l(\mathbf{V}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{K}^T \mathbf{V} \mathbf{K}| - \frac{1}{2} \mathbf{y}^T \mathbf{K} (\mathbf{K}^T \mathbf{V} \mathbf{K})^{-1} \mathbf{K}^T \mathbf{y}.$$

Funktion maksimointi tuottaa estimaattorin vastemuuttujan \mathbf{y} kovarianssimatriisille \mathbf{V} . Tätä estimaattoria $\hat{\mathbf{V}}$ käyttämällä saamme myös parametrille $\boldsymbol{\beta}$ estimaattorin

$$\hat{\boldsymbol{\beta}}_{REML} = (\mathbf{X}^T \hat{\mathbf{V}}_{REML}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}_{REML}^{-1} \mathbf{y}.$$

5.2.4 Sisäkorrelaatio

Kiinteiden vaikutusten lineaarisessa regressiomallissa vasteen \mathbf{y} sisältämät havainnot y_{ij} ovat toisistaan riippumattomia. Lineaaristen sekamallien tapauksessa havainnot on jaettu esimerkiksi alueisiin, joiden ajatellaan olevan toisistaan riippumattomia, mutta saman alueen eri havainnot riippuvat toisistaan. Saman alueen eri havaintojen välisen korrelaation kertoo sisäkorrelaatio

$$\rho = \text{cor}(y_{ij}, y_{il}) = \frac{\text{cov}(y_{ij}, y_{il})}{\sqrt{\text{var}(y_{ij})\text{var}(y_{il})}}, \quad j \neq l,$$

joka on aina välillä 0–1. Tämä kaava on tarkoitettu sisäkorrelaation laskemiseen sekamallille, jossa on kiinteiden vaikutusten lisäksi satunnaisena vain vakio. Kun sisäkorrelaatio on lähellä nollaa, havainnot saman alueen sisällä eivät riipu toisistaan, jolloin tilanne ei ole lineaarisille sekamalleille tyypillinen. Tällaiselle aineistolle kannattaa kokeilla myös vaihtoehtoisia malleja.

5.3 Mallien vertailu

Olemme pitäneet mallien vertailussa periaatteena sitä, että sisäkkäisiä malleja vertaillaan ANOVA-lalla (*Analysis of variance*) ja ei-sisäkkäisiä malleja AIC-luvuilla (*Akaike's information criterion*). AIC lasketaan seuraavasti:

$$AIC = -2 \log L + 2n_{par},$$

missä n_{par} on mallin parametrien lukumäärä (Pinheiro & Bates 2000, 10). Sekamallien tapauksessa se on kiinteiden vaikutusten β , satunnaisvaikutusten kovarianssimatriisin D ja jäännösten kovarianssimatriisin R kaikkien parametrien yhteislukumäärä. Sääntönä on, että malleista parempi on se, jolla on pienempi AIC-luku.

Malleja vertailtaessa käytämme termejä täysi ja laajempi malli. Täysi malli esiintyy erityisesti sekamallien tapauksessa, ja sillä tarkoitetaan mallia, joka sisältää kaikki mahdolliset termit, joita kyseiseen malliin voi kuulua. Esimerkiksi sekamalli, jossa kannanmuutosta selittävät kiinteä urosuus sekä satunnaiset vakio ja urosuus, on täysi malli, koska siinä on molemmat mahdolliset satunnaisvaikutukset mukana. Laajemmalla mallilla tarkoitetaan mallia, joka sisältää enemmän termejä kuin verrattava malli. Sitä käytetään erityisesti ANOVA-vertailujen yhteydessä, koska tällöin on kyseessä sisäkkäiset mallit, joista toinen on aina laajempi.

5.4 Käytetystä ohjelmistosta

Tutkielman tekninen puoli on toteutettu R-ohjelmistolla (R Core Team 2013). Olemme tehneet osan sekamallivertailuista perinteisen *lme*-funktion lisäksi myös *lmer*-funktiolla, jonka käyttämiseksi R-ohjelmistoon täytyy ladata *lme4*-paketti. Tasa-arvokäyrien piirtämisessä olemme käyttäneet R-ohjelmiston *akima*-paketin *interp*-funktiota.

6 KIINTEIDEN VAIKUTUSTEN LINEAARISTEN REGRESSIOMALLIEN SOVITTAMINEN

6.1 Valtakunnalliset regressiomallit ilman piirierottelua

Päätimme lähestyä tutkimuskysymystä valtakunnallisesti ilman piirierottelua saadaksemme jonkinlaisen käsityksen siitä, kuinka vasanaarassuhde ja urososuus käyttäytyvät kannanmuutoksen selittäjinä. Sovitimme regressiomallit, joissa vasteena on kannanmuutos ja selittäjinä vasanaarassuhde, urososuus tai molemmat:

$$muutos_j = \beta_0 + \beta_1 vasanaarassuhde_j + e_j, \quad (1)$$

$$muutos_j = \beta_0 + \beta_2 urososuus_j + e_j, \quad (2)$$

$$muutos_j = \beta_0 + \beta_1 vasanaarassuhde_j + \beta_2 urososuus_j + e_j, \quad (3)$$

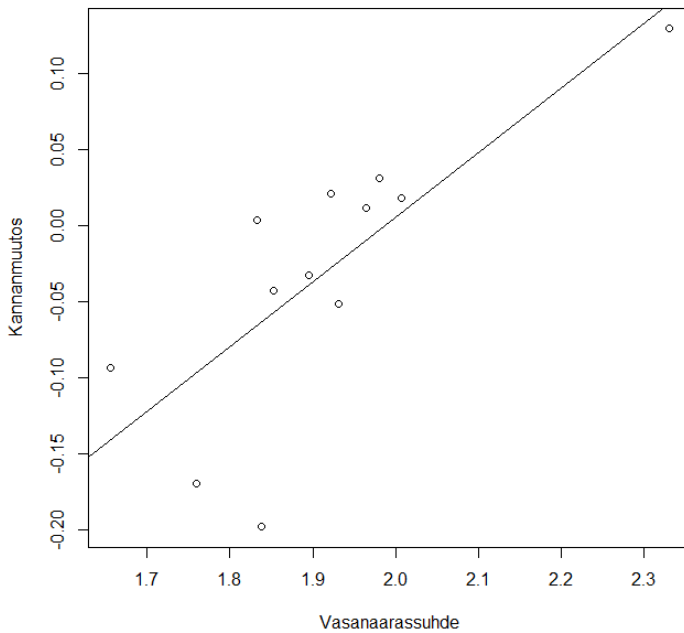
joissa j = vuosi. Taulukossa 1 on näiden mallien regressioparametrien estimaatit keskivirheineen.

Taulukko 1 Mallien (1), (2) ja (3) regressioparametrien estimaatit keskivirheineen.

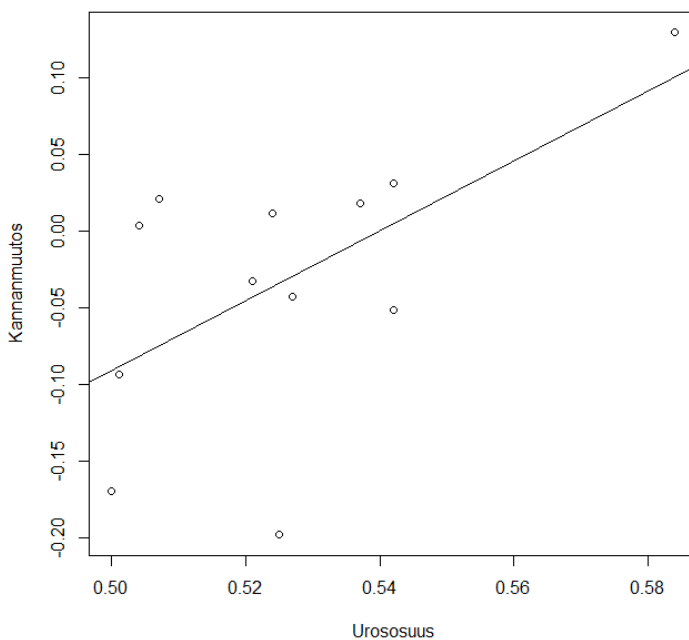
Malli	$\hat{\beta}_0$ (s.e.)	$\hat{\beta}_1$ (s.e.)	$\hat{\beta}_2$ (s.e.)
(1)	-0.85 (0.21)	0.43 (0.11)	-
(2)	-1.23 (0.51)	-	2.27 (0.97)
(3)	-0.27 (0.52)	0.70 (0.25)	-2.10 (1.74)

Kuvista 2 ja 3 huomaamme, että sekä vasanaarassuhde että urososuus korreloivat kannanmuutoksen kanssa oletetusti eli positiivisesti. Tosin molemmissa kuvissa on hieman muista poikkeavia vuosihavaintoja, jotka määräävät pitkälti regressiosuoran jyrkkyyden. Vasanaarassuhde ($R^2 = 0.60$, $F = 14.85$, $df1 = 1$, $df2 = 10$, $p = 0.0032$) toimii paremmin kannanmuutoksen selittäjänä kuin urososuus ($R^2 = 0.36$, $F = 5.53$, $df1 = 1$, $df2 = 10$, $p = 0.041$), sillä urososuuden tapauksessa selitysaste jää melko pieneksi. Vasanaarassuhteen regressiomalliakin parempi selitysaste on kahden selittäjän regressiomallilla ($R^2 = 0.65$, $F = 8.49$, $df1 = 2$, $df2 = 9$, $p = 0.0085$), jossa selittäjinä ovat siis sekä vasanaarassuhde että urososuus. Kun vertaamme urososuuden regressiomallia kahden selittäjän regressiomalliin, huomaamme, että kahden selittäjän regressiomalli on merkitsevästi pa-

rempi kuin urososuuden regressiomalli (ANOVA, $p = 0.021$). Kahden selittäjän regressiomalli ei kuitenkaan ole merkitsevästi vasanaarassuhteen regressiomallia parempi (ANOVA, $p = 0.26$), joten valtakunnallisesti kannanmuutosta selittää parhaiten vasanaarassuhde yksinään.



Kuva 2 Vasanaarassuhde kannanmuutoksen selittäjänä vuosina 2001-2012 valtakunnallisella aineistolla.

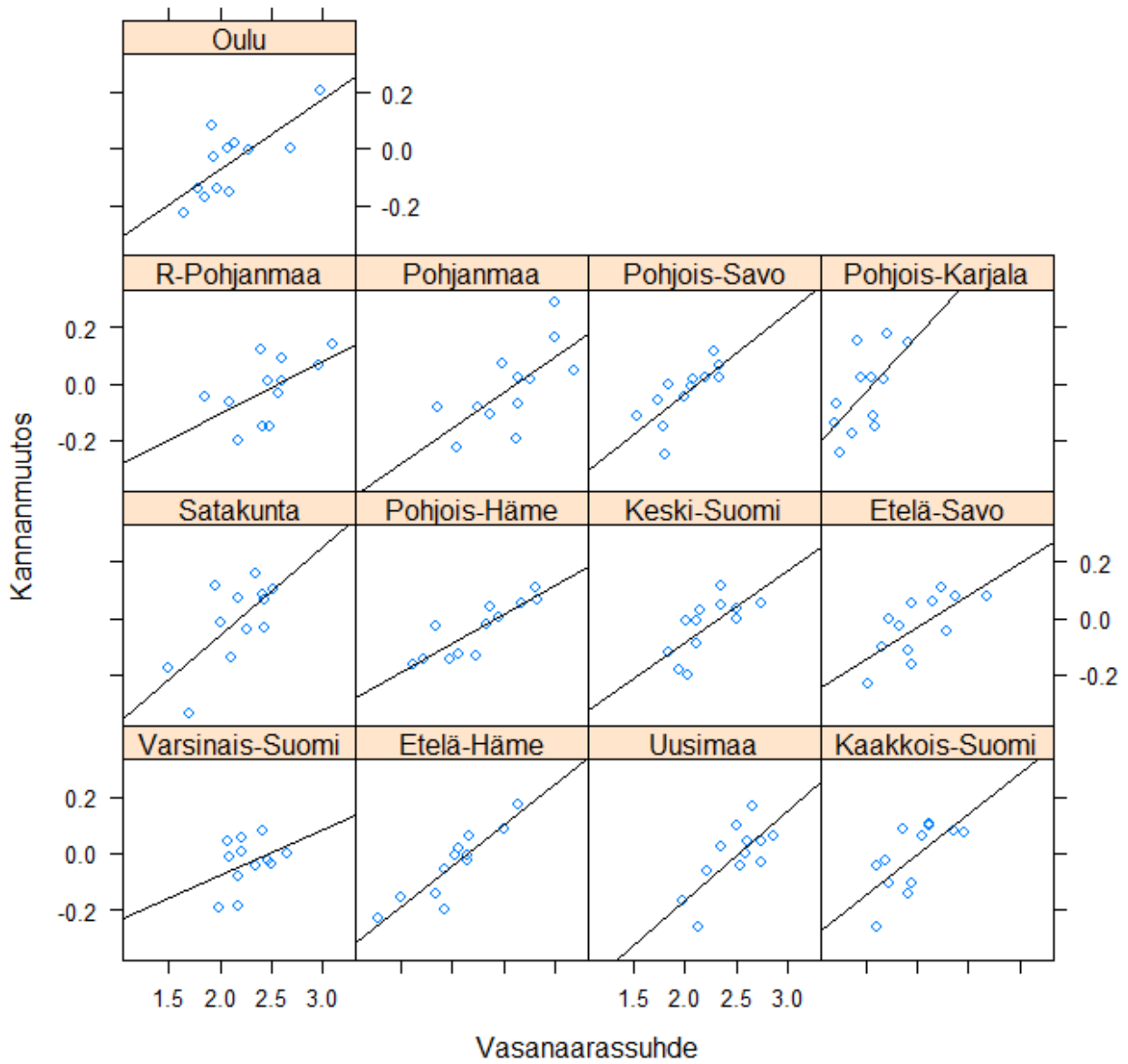


Kuva 3 Urososuus kannanmuutoksen selittäjänä vuosina 2001-2012 valtakunnallisella aineistolla.

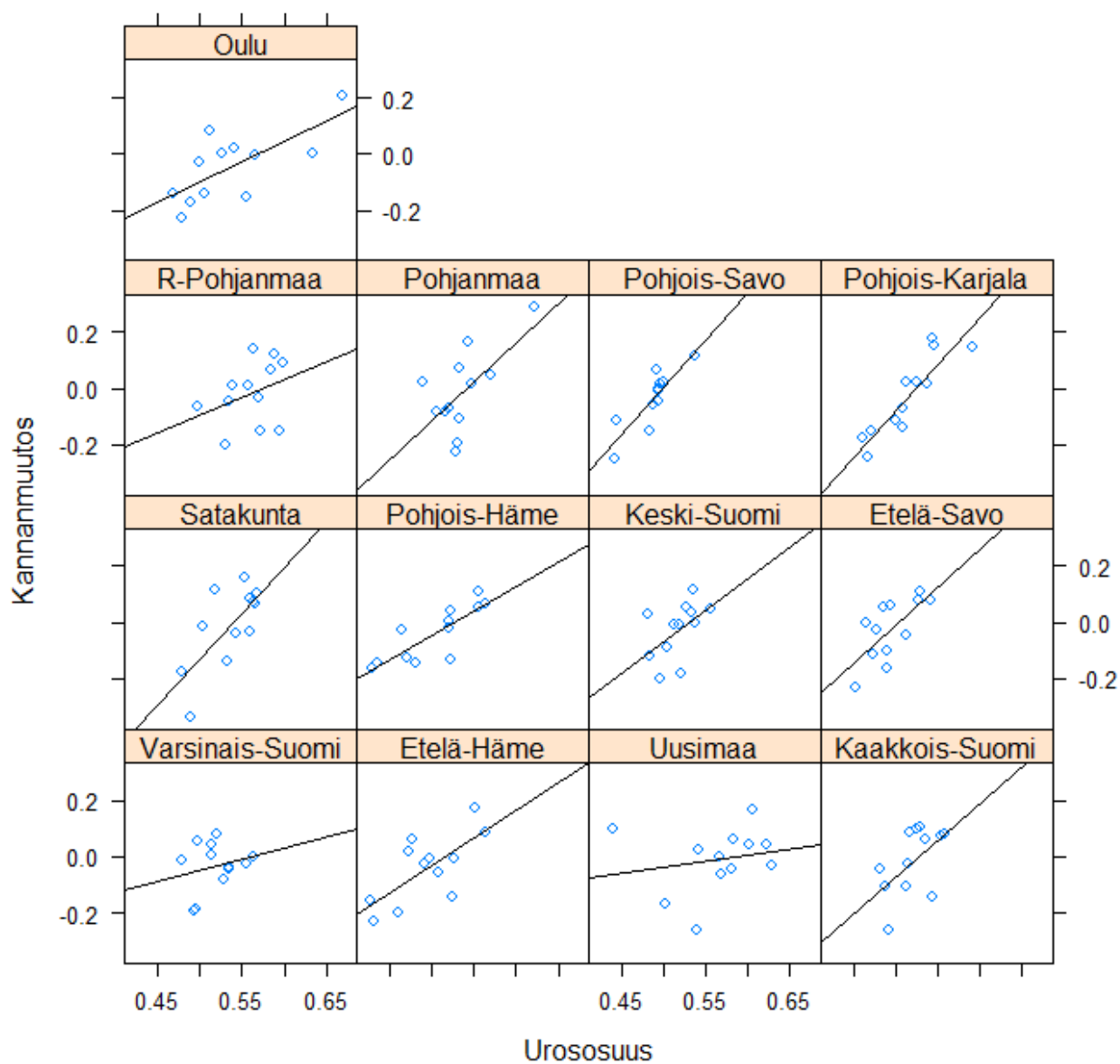
6.2 Regressiomallit piireittäin

Valtakunnallisesti ilman piirierottelua vasanaarassuhde ja urososuus näyttävät siis korreloivan oletusti (positiivisesti) kannanmuutoksen kanssa. Lisäksi ainakin vasanaarassuhteen ja sekä vasanaarassuhteen että urososuuden selittäessä kannanmuutosta mallien selitysasteet ovat hyviä. Seuraavaksi etenemme piiritasolle. Sovitamme kolme regressiomallia (kannanmuutoksen selittäjinä vasanaarassuhde, urososuus ja molemmat) jokaiselle 13 piirille erikseen. Kuvista 4 ja 5 näemme, että urososuuden ollessa selittäjänä eri piirien regressiosuorat poikkeavat enemmän toisistaan kuin vasanaarassuhteen tapauksessa. Näin ollen urososuuden vaikutus kannanmuutokseen näyttää vaihtelevan piireittäin. Vasanaarassuhteen tapauksessa piirien väliset erot eivät ole niin suuria. Lisäksi maantieteellisessä mielessä erityisesti urososuuden toimiessa selittäjänä (Kuva 5) regressiosuorien jyrkkyydessä esiintyy jonkinlaista trendiä. Pääsääntöisesti loivimmat regressiosuorat ovat piirikartan länsi-, lounais- ja eteläosissa ja ne jyrkkenevät itään ja koilliseen mentäessä.

Taulukoista 2–4 voimme päätellä, kuinka monella piirillä vasanaarassuhde, urososuus tai molemmat toimii hyvin ($p < 0.05$) kannanmuutoksen selittäjänä, kuinka monella huonosti ($p > 0.10$), ja moniko piiri jää niin sanotuksi rajatapaukseksi ($0.05 < p < 0.10$). Taulukosta 2 näemme, että vasanaarassuhde näyttää toimivan kannanmuutoksen selittäjänä hyvin 11 piirillä, huonosti Varsinais-Suomella ja että Rannikko-Pohjanmaa on rajatapaus. Urososuus vastaavasti toimii kannanmuutoksen selittäjänä hyvin kahdeksalla piirillä, huonosti kolmella, ja rajatapauksia on kaksi (Taulukko 3). Taulukon 4 perusteella vasanaarassuhde ja urososuus yhdessä toimivat kannanmuutoksen selittäjinä hyvin kymmenellä piirillä, huonosti kahdella ja Kaakkois-Suomi on rajalla. Näin ollen vasanaarassuhde yksinään toimii kannanmuutoksen selittäjänä suurimmalle määrälle piirejä (11).



Kuva 4 Vasanaarassuhde kannanmuutoksen selittäjänä vuosina 2001–2012 eri piireissä maantieteellisesti Suomen piirikarttaa mukailevassa järjestyksessä (R-Pohjanmaa on lyhenne Rannikko-Pohjanmaasta).



Kuva 5 Urososuus kannanmuutoksen selittäjänä vuosina 2001–2012 eri piireissä maantieteellisesti Suomen piirikarttaa mukailevassa järjestyksessä (R-Pohjanmaa on lyhenne Rannikko-Pohjanmaasta).

Taulukko 2 Vasanaarassuhde kannanmuutoksen selittäjänä: regressioparametrien estimaatit keskivirheineen (β_1 = vasanaarassuhteen parametri), selityasteet ja merkitsevyydet.

	$\hat{\beta}_0$ (s.e.)	$\hat{\beta}_1$ (s.e.)	R^2	p
Etelä-Häme	-0.62 (0.09)	0.29 (0.04)	0.81	< 0.001
Etelä-Savo	-0.48 (0.15)	0.23 (0.07)	0.49	0.011
Kaakkois-Suomi	-0.57 (0.20)	0.28 (0.10)	0.45	0.017
Keski-Suomi	-0.59 (0.19)	0.25 (0.08)	0.48	0.013
Oulu	-0.57 (0.15)	0.25 (0.07)	0.57	0.004
Pohjanmaa	-0.65 (0.22)	0.25 (0.09)	0.45	0.016
Pohjois-Häme	-0.50 (0.08)	0.20 (0.04)	0.75	< 0.001
Pohjois-Karjala	-0.61 (0.24)	0.39 (0.16)	0.38	0.034
Pohjois-Savo	-0.61 (0.16)	0.29 (0.08)	0.58	0.004
Rannikko-Pohjanmaa	-0.47 (0.21)	0.18 (0.09)	0.32	0.057
Satakunta	-0.69 (0.23)	0.31 (0.10)	0.48	0.013
Uusimaa	-0.79 (0.23)	0.31 (0.09)	0.54	0.007
Varsinais-Suomi	-0.40 (0.29)	0.16 (0.13)	0.14	0.227

Taulukko 3 Urosuus kannanmuutoksen selittäjänä: regressioparametrien estimaatit keskivirheineen (β_2 = urosuuden parametri), selityasteet ja merkitsevyydet.

	$\hat{\beta}_0$ (s.e.)	$\hat{\beta}_2$ (s.e.)	R^2	p
Etelä-Häme	-1.01 (0.31)	1.96 (0.62)	0.50	0.011
Etelä-Savo	-1.36 (0.44)	2.71 (0.90)	0.48	0.013
Kaakkois-Suomi	-1.37 (0.65)	2.59 (1.25)	0.30	0.065
Keski-Suomi	-1.19 (0.60)	2.26 (1.16)	0.27	0.081
Oulu	-0.81 (0.24)	1.43 (0.45)	0.50	0.010
Pohjanmaa	-1.50 (0.57)	2.78 (1.05)	0.41	0.025
Pohjois-Häme	-0.91 (0.19)	1.73 (0.37)	0.68	< 0.001
Pohjois-Karjala	-1.73 (0.28)	3.31 (0.54)	0.79	< 0.001
Pohjois-Savo	-1.68 (0.32)	3.38 (0.65)	0.73	< 0.001
Rannikko-Pohjanmaa	-0.73 (0.61)	1.28 (1.08)	0.12	0.265
Satakunta	-1.75 (0.55)	3.24 (1.02)	0.50	0.010
Uusimaa	-0.25 (0.38)	0.43 (0.67)	0.04	0.537
Varsinais-Suomi	-0.45 (0.54)	0.81 (1.04)	0.06	0.456

Taulukko 4 Vasanaarassuhde ja urososuus kannanmuutoksen selittäjinä: regressioparametrien estimaatit keskivirheineen ($\beta_1 =$ vasanaarassuhteen parametri, $\beta_2 =$ urososuuden parametri), selityssasteet ja merkitsevyydet.

	$\hat{\beta}_0$ (s.e.)	$\hat{\beta}_1$ (s.e.)	$\hat{\beta}_2$ (s.e.)	R^2	p
Etelä-Häme	-0.52 (0.23)	0.32 (0.08)	-0.36 (0.71)	0.82	< 0.001
Etelä-Savo	-0.87 (0.80)	0.14 (0.19)	1.14 (2.29)	0.51	0.041
Kaakkois-Suomi	-0.25 (0.92)	0.35 (0.22)	-0.87 (2.45)	0.46	0.064
Keski-Suomi	-0.85 (0.56)	0.22 (0.11)	0.66 (1.30)	0.49	0.047
Oulu	-0.12 (0.49)	0.61 (0.38)	-2.25 (2.34)	0.61	0.014
Pohjanmaa	-1.32 (0.53)	0.17 (0.10)	1.63 (1.20)	0.55	0.028
Pohjois-Häme	-0.23 (0.43)	0.32 (0.19)	-1.06 (1.65)	0.76	0.002
Pohjois-Karjala	-1.71 (0.29)	0.06 (0.12)	3.11 (0.73)	0.79	< 0.001
Pohjois-Savo	-1.49 (0.34)	0.11 (0.09)	2.52 (0.89)	0.78	0.001
Rannikko-Pohjanmaa	-0.55 (0.58)	0.17 (0.11)	0.18 (1.22)	0.32	0.178
Satakunta	-1.50 (1.15)	0.08 (0.34)	2.44 (3.39)	0.51	0.042
Uusimaa	-0.61 (0.28)	0.38 (0.11)	-0.62 (0.55)	0.59	0.018
Varsinais-Suomi	-0.15 (0.61)	0.26 (0.24)	-0.90 (1.90)	0.16	0.448

6.3 Yhdysvaikutusten testaus

Kun lisäämme regressiomalliin selittäjäksi myös piirifaktorin, voimme tutkia, vaihtelee ko vasanaarassuhteen ja urososuuden vaikutus kannanmuutokseen piireittäin. Toisin sanoen haluamme tietää, onko saalisrakenneindeksien ja piirifaktorin välillä yhdysvaikutusta. Kun regressiomallissa on kannanmuutoksen selittäjinä vasanaarassuhde ja piirifaktori, ja tutkimme, tarvitseeko malliin myös niiden välisen yhdysvaikutuksen, huomaamme, että se ei ole merkitsevä (ANOVA, $p = 0.96$). Tämä viittaa siihen, että vasanaarassuhteen vaikutus kannanmuutokseen ei vaihtele piireittäin. Kun varmistamme vielä päävaikutusten merkitsevyyden, saamme tulokseksi, että vasanaarassuhde (ANOVA, $p < 0.001$) ja piirifaktori (ANOVA, $p < 0.001$) ovat merkitseviä ja ne jäävät kumpikin malliin selittämään kannanmuutosta. Jatkossa nimitämme vasanaarassuhteen päävaikutusmalliksi tätä vasanaarassuhteen ja piirifaktorin muodostamaa regressiomallia

$$muutos_{ij} = \beta_0 + \beta_{0i} + \beta_1 vasanaarassuhde_{ij} + e_{ij}, \quad (4)$$

missä i = piiri ja j = vuosi. Tälle mallille vakiotermin β_0 estimaatti on -0.55 ja keskivirhe 0.05 sekä vasanaarassuhteen regressiokertoimen β_1 estimaatti 0.25 ja keskivirhe 0.02. Taulukossa 5 on mallin piirikohtaiset regressioparametrien estimaatit keskivirheineen.

Taulukko 5 Mallin (4) piirikohtaiset regressioparametrien estimaatit keskivirheineen. Etelä-Häme on vertailupiiri, jolle $\beta_{0i} = 0$ ja jonka parametrien arvoon muiden piirien parametrien arvot vertautuvat.

	$\hat{\beta}_{0i}$ (s.e.)
Etelä-Savo	0.01 (0.03)
Kaakkois-Suomi	0.04 (0.03)
Keski-Suomi	-0.04 (0.04)
Oulu	-0.03 (0.03)
Pohjanmaa	-0.11 (0.04)
Pohjois-Häme	-0.06 (0.04)
Pohjois-Karjala	0.15 (0.04)
Pohjois-Savo	0.01 (0.03)
Rannikko-Pohjanmaa	-0.10 (0.04)
Satakunta	-0.01 (0.04)
Uusimaa	-0.09 (0.04)
Varsinais-Suomi	-0.06 (0.04)

Urososuuden ja piirifaktorin välinen yhdysvaikutus taas on melko merkitsevä (ANOVA, $p = 0.074$), kun regressiomallissa on kannanmuutoksen selittäjinä urososuus ja piirifaktori. Näin ollen urososuuden vaikutus kannanmuutokseen saattaisi vaihdella piireittäin. Tämän varmistamiseksi selvitämme ensin, mihin päävaikutusmalliin päädyimme, jos pudotamme yhdysvaikutustermiä mallista pois. Sen jälkeen vertaamme päävaikutusmallia yhdysvaikutuksen sisältävään malliin ja saamme näin selville yhdysvaikutuksen tarpeellisuuden. Pelkkien päävaikutusten regressiomallissa urososuus on merkitsevä (ANOVA, $p < 0.001$), mutta piirifaktori ei (ANOVA, $p = 0.15$). Täten paras päävaikutusmalli koostuu pelkästä urososuudesta kannanmuutoksen selittäjänä. Kun vertaamme tätä mallia yhdysvaikutukset (ja siten myös päävaikutukset) sisältävään malliin, tuloksena on, että laajempi malli on merkitsevä (ANOVA, $p = 0.047$). Koska urososuuden ja piirifaktorin välillä on yhdysvaikutus, urososuuden vaikutus kannanmuutokseen vaihtelee piireittäin. Malli on muotoa

$$muutos_{ij} = \beta_0 + \beta_{0i} + (\beta_2 + \beta_{2i})urososuus_{ij} + e_{ij}, \quad (5)$$

missä i = piiri ja j = vuosi, ja nimitämme sitä jatkossa urososuuden yhdysvaikutusmalliksi. Tälle mallille vakiotermin β_0 estimaatti on -1.01 ja keskivirhe 0.32 sekä urososuuden regressiokertoimen β_2 estimaatti 1.96 ja keskivirhe 0.64. Taulukossa 6 on mallin piirikohtaiset regressioparametrien estimaatit keskivirheineen.

Taulukko 6 Mallin (5) piirikohtaiset regressioparametrien estimaatit keskivirheineen. Etelä-Häme on vertailupiiri, jolle $\beta_{0i} = 0$ ja $\beta_{1i} = 0$ ja jonka parametrien arvoihin muiden piirien parametrien arvot vertautuvat.

	$\hat{\beta}_{0i}$ (s.e.)	$\hat{\beta}_{2i}$ (s.e.)
Etelä-Savo	-0.35 (0.59)	0.75 (1.20)
Kaakkois-Suomi	-0.36 (0.66)	0.63 (1.29)
Keski-Suomi	-0.19 (0.71)	0.30 (1.38)
Oulu	0.20 (0.40)	-0.53 (0.79)
Pohjanmaa	-0.50 (0.55)	0.82 (1.05)
Pohjois-Häme	0.10 (0.44)	-0.23 (0.88)
Pohjois-Karjala	-0.72 (0.50)	1.35 (0.98)
Pohjois-Savo	-0.68 (0.63)	1.42 (1.28)
Rannikko-Pohjanmaa	0.27 (0.60)	-0.68 (1.12)
Satakunta	-0.74 (0.57)	1.28 (1.09)
Uusimaa	0.76 (0.43)	-1.53 (0.83)
Varsinais-Suomi	0.55 (0.65)	-1.15 (1.27)

Tutkimme myös kolmen selittäjämuuttujan (vasanaarassuhde, urososuus ja piirifaktori) välisiä yhdysvaikutuksia laittamalla kaikki samaan malliin kannanmuutosta selittämään. Laajimmassa mallissamme selittäjinä ovat vasanaarassuhde, urososuus, piirifaktori, vasanaarassuhteen ja piirifaktorin välinen yhdysvaikutus, urososuuden ja piirifaktorin välinen yhdysvaikutus sekä vasanaarassuhteen, urososuuden ja piirifaktorin välinen yhdysvaikutus. Ensimmäisessä selittäjän välisen yhdysvaikutuksen merkitsevyyden. Se ei selvästikään ole merkitsevä (ANOVA, $p = 0.79$), joten vasanaarassuhteen, urososuuden ja piirifaktorin muodostama yhdysvaikutustermi putoaa mallista pois. Seuraavaksi tutkimme, tarvitaanko mallissa vasanaarassuhteen ja piirifaktorin välistä yhdysvaikutusta ja toisaalta urososuuden ja piirifaktorin välistä yhdysvaikutusta. Vasanaarassuhteen ja piirifaktorin välinen yhdysvaikutus ei ole merkitsevä (ANOVA, $p = 0.82$), joten se putoaa mallista pois. Myöskään urososuuden ja piirifaktorin välinen yhdysvaikutus ei ole merkitsevä (ANOVA, $p = 0.11$), joten malliin jäävät pelkät päävaikutukset. Lopuksi selvitämme, mitkä päävaikutuksista (vasanaarassuhde, urososuus ja piirifaktori) ovat merkitseviä. Sekä vasanaarassuhde että piirifaktori ovat merkitseviä (ANOVA, $p < 0.001$ ja $p < 0.001$), mutta urososuus ei ole (ANOVA, $p = 0.12$). Päädyimme siis samaan malliin (4), eli malli sisältää selittäjinä vain vasanaarassuhteen ja piirifaktorin.

Yhteenvetona voimme todeta, että urososuuden ja piirifaktorin välillä on yhdysvaikutus silloin, kun vasanaarassuhde ei ole mukana mallissa. Mitään yhdysvaikutuksia ei tule malliin, jos myös vasanaarassuhde on mallissa. Tulokset tukevat kuvista 4 ja 5 aiemmin tehtyjä pohdintoja, joiden mu-

kaan urososuuden ollessa kannanmuutoksen selittäjänä piirien välillä on eroja ja vasanaarassuhteen tapauksessa ei juurikaan.

Jatkoa varten selvitämme vielä, kumpi vasanaarassuhteen päävaikutusmallista ja urososuuden yhdysvaikutusmallista on parempi. Koska mallit eivät ole sisäkkäisiä, selvitämme niiden välisen paremmuusjärjestyksen vertailemalla mallien AIC-lukuja. Vasanaarassuhteen päävaikutusmallilla on pienempi AIC-luku ($AIC_{\text{main}} = -309.61$, $AIC_{\text{inter}} = -273.15$), joten se on parempi malli.

7 LINEAARISTEN SEKAMALLIEN SOVITTAMINEN

Aiemmin luvussa 6.2 päättelimme kuvista 4 ja 5, että kannanmuutoksen ja saalisrakenneindeksien välinen riippuvuus näyttäisi vaihtelevan piireittäin, erityisesti urososuuden ollessa selittäjänä. Lisäksi luvussa 6.3 tulimme siihen tulokseen, että urososuuden ja piirifaktorin välillä on yhdysvaikutus. Tämän vuoksi kokeilemme laajentaa kiinteiden vaikutusten regressiomallit sekamalleiksi lisäämällä niihin piirien välistä vaihtelua kuvaavan satunnaisvaikutuksen.

Käytämme nimitystä vasanaarassuhteen sekamalli, kun mallissa kannanmuutoksen selittäjänä on vasanaarassuhde. Vastaavasti urososuuden ollessa kannanmuutoksen selittäjänä käytämme termiä urososuuden sekamalli. Kun kannanmuutoksen selittäjinä ovat sekä vasanaarassuhde että urososuus, kutsumme mallia kahden selittäjän sekamalliksi. Lisäksi käytämme yksinkertaisuuden vuoksi urososuuden satunnaisesta kulmakertoimesta nimitystä satunnainen urososuus ja vasanaarassuhteen satunnaisesta kulmakertoimesta nimitystä satunnainen vasanaarassuhde.

7.1 Paras vasanaarassuhteen sekamalli

Aloitamme mallivertailut täydestä mallista, jossa kannanmuutosta selittävät kiinteä vasanaarassuhde sekä satunnainen vakio ja satunnainen vasanaarassuhde

$$\text{muutos}_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})\text{vasanaarassuhde}_{ij} + e_{ij},$$

missä i = piiri ja j = vuosi. Kun selvitämme ensin satunnaisen vasanaarassuhteen ja sitten satunnaisen vakion tarpeellisuuden, huomaamme, että laajempi malli ei ole kummassakaan tapauksessa merkitsevä (ANOVA, $p = 1$ ja $p = 0.38$). Jotta saamme selville, onko parempi, että mallissa on kiinteiden vaikutusten lisäksi satunnainen vakio vai satunnainen vasanaarassuhde, vertailemme mallien AIC-lukuja. Pienempi AIC-luku on mallilla, jossa satunnaisena on vain vakio ($AIC_{\text{vakio}} = -281.52$, $AIC_{\text{vasa}} = -279.58$), joten se on parempi malli. Näin ollen mallimme koostuu nyt kiinteästä vasanaarassuhteesta ja satunnaisesta vakiosta kannanmuutoksen ollessa vasteena.

Tutkitaan vielä, tarvitaanko satunnaisvaikutuksia (ja siten sekamallia) ensinkään. Vertaamme sekamallia, jossa kannanmuutosta selittämässä ovat kiinteä vasanaarassuhde ja satunnainen vakio

kiinteiden vaikutusten regressiomalliin, jossa on vain kiinteä vasanaarassuhde eikä lainkaan satunnaistermejä. Laajempi malli osoittautuu merkitsevästi paremmaksi (ANOVA, $p < 0.001$). Näin ollen vasanaarassuhteen ollessa kannanmuutoksen selittäjänä on aiheellista käyttää sekamallia, ja parhaiten toimii malli, jossa on kiinteä vasanaarassuhde ja satunnainen vakio:

$$muutos_{ij} = \beta_0 + u_{0i} + \beta_1 vasanaarassuhde_{ij} + e_{ij}, \quad (6)$$

missä i = piiri ja j = vuosi. Tälle mallille vakiotermin β_0 estimaatti on -0.52 ja keskivirhe 0.05 sekä vasanaarassuhteen regressiokertoimen β_1 estimaatti 0.23 ja keskivirhe 0.02. Taulukossa 7 on mallin piirikohtaiset satunnaisen vakion ennusteet. Koska vasanaarassuhteen sekamallissa on kiinteiden vaikutusten lisäksi satunnaisena vain vakio, voimme laskea sille sisäkorrelaation. Saamme sisäkorrelaatioksi 0.30, mikä tarkoittaa, että saman piirin eri havainnot korreloivat keskenään. Tämäkin tukee sekamallien käyttöä.

Taulukko 7 Mallin (6) piirikohtaiset satunnaisen vakion ennusteet.

	\hat{u}_{0i}
Etelä-Häme	0.015
Etelä-Savo	0.027
Kaakkois-Suomi	0.048
Keski-Suomi	-0.013
Oulu	-0.009
Pohjanmaa	-0.068
Pohjois-Häme	-0.028
Pohjois-Karjala	0.125
Pohjois-Savo	0.024
Rannikko-Pohjanmaa	-0.054
Satakunta	0.013
Uusimaa	-0.051
Varsinais-Suomi	-0.030

7.2 Paras urososuuden sekamalli

Urososuuden tapauksessa parhaan sekamallin selvittäminen osoittautuu vasanaarassuhdetta hankalammaksi sekä R-ohjelmiston koodin että mallivertailujen osalta. Aloitamme vertailun täydestä mallista, jossa kannanmuutosta selittävät kiinteä urososuus sekä satunnaiset vakio ja urososuus

$$muutos_{ij} = \beta_0 + u_{0i} + (\beta_2 + u_{2i})urososuus_{ij} + e_{ij},$$

missä i = piiri ja j = vuosi (konvergenssiongelmien välttämiseksi R-ohjelmistossa *lme*-funktion sisään on lisättävä *control = lmeControl(opt = "optim", msMaxIter=500)* aina, kun sekamallissa on satunnaispuolella yhtä aikaa sekä vakio että urososuus). Kun testaamme ensin satunnaisen urososuuden ja sitten satunnaisen vakion tarpeellisuutta, huomaamme, että kummassakaan tapauksessa laajempi (täysi) malli ei ole merkitsevä (ANOVA, $p = 0.11$ ja $p = 0.13$). Kumpikaan merkitsevyyksistä ei ole yhtä selkeä kuin vasanaarassuhteen mallivertailuissa saadut, mutta molemmat viittaavat siihen suuntaan, että täysi malli ei ole merkitsevä. Näin ollen seuraavaksi on ratkaistava, onko sekamallissa parempi olla satunnaisena vakio vai urososuus kiinteän urososuuden lisäksi. AIC-lukujen ($AIC_{vakio} = -269.02$, $AIC_{uros} = -269.36$) perusteella hieman parempi on malli, jossa kannanmuutoksen selittäjänä on kiinteiden vaikutusten lisäksi satunnaisena vain urososuus. Tätä mallivalintaa tukee myös se, että urososuuden tapauksessa eri piirien regressiosuorien kulmakerrointen välillä näytti olevan eroa (ks. Kuva 5).

Seuraavaksi selvitämme, tarvitaanko kannanmuutosta selittämään lainkaan satunnaista urososuutta kiinteän urososuuden lisäksi. Kun vertaamme satunnaisen urososuuden sisältävää sekamallia regressiomalliin, huomaamme, ettei laajempi malli ole merkitsevästi parempi kuin pelkät kiinteät vaikutukset sisältävä malli (ANOVA, $p = 0.54$). Tämä tarkoittaa, että urososuuden ollessa kannanmuutoksen selittäjänä ei tarvitse mennä sekamalleihin asti. Regressiomalli, jossa selittäjänä on vain kiinteä urososuus, osoittautuu riittäväksi. Aiemmin luvussa 6.3 vertasimme tätä päävaikutusmallia urososuuden ja piirifaktorin muodostamaan yhdysvaikutusmalliin ja totesimme yhdysvaikutusmallin olevan parempi malli. Näin ollen emme ota tätä mallia mukaan jatkotarkasteluihin.

Urososuuden tapauksessa sekamallin tarpeettomuus tuli hieman yllätyksenä, sillä piirien välillä näytti olevan eroja (luku 6.2), ja yhdysvaikutustarkasteluissa (luku 6.3) päädyimme siihen tulokseen, että urososuuden ja piirifaktorin välillä on yhdysvaikutus. On kuitenkin muistettava, että sekamalleihin siirryttäessä riippuvuusoletukset muuttuvat. Kiinteiden vaikutusten lineaarisissa regressiomalleissa oletetaan, että havainnot ovat toisistaan riippumattomia. Lineaaristen sekamallien tapauksessa piirien ajatellaan olevan toisistaan riippumattomia, mutta saman piirin eri havainnot riippuvat toisistaan.

7.3 Paras kahden selittäjän sekamalli

Yhden selittäjän sekamallien tutkimisesta jatkamme kahden selittäjän sekamalliin, jossa kannanmuutoksen selittäjinä ovat sekä vasanaarassuhde että urososuus. Täysi sekamalli on muotoa: kiinteät vasanaarassuhde ja urososuus sekä satunnaiset vakio, vasanaarassuhde ja urososuus

$$muutos_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i})vasanaarassuhde_{ij} + (\beta_2 + u_{2i})urososuus_{ij} + e_{ij},$$

missä i = piiri ja j = vuosi (konvergenssiongelmiin välttämiseksi R-ohjelmistossa *lme*-funktion sisään on lisättävä `control = lmeControl(opt = "optim", msMaxIter=500)` aina, kun sekamallissa on satunnaispuolella sekä vakio että urososuus). Aloitamme selvittämällä, tarvitaanko mallissa satunnaista vasanaarassuhdetta. Laajempi malli ei ole merkitsevä (ANOVA, $p = 0.40$), joten satunnainen vasanaarassuhde putoaa mallista pois. Seuraavaksi tutkimme, tarvitaanko satunnaista urososuutta, kun kiinteät vaikutukset ovat ennallaan, ja malli sisältää satunnaisen vakion. Huomaamme, että laajempi malli on merkitsevä (ANOVA, $p = 0.0092$), joten satunnainen urososuus on hyvä olla mallissa mukana. Myös satunnaisen vakion tarpeellisuus on tutkittava, kun kannanmuutosta selittämässä ovat lisäksi kiinteät vasanaarassuhde ja urososuus sekä satunnaisena urososuus. Laajempi malli on merkitsevä (ANOVA, $p = 0.034$), joten satunnainen vakio pysyy mallissa.

Seuraavaksi tutkimme, tarvitaanko kahden selittäjän sekamallissa lainkaan satunnaisvaikutuksia (vakiota ja urososuutta) kiinteiden vasanaarassuhteen ja urososuuden lisäksi. Laajempi malli on selvästi merkitsevä (ANOVA, $p < 0.001$). Täten sekamalli on tarpeellinen, ja päädyimme kahden selittäjän sekamalliin, jossa kannanmuutoksella on seuraavat selittäjät: kiinteät vasanaarassuhde ja urososuus sekä satunnaiset vakio ja urososuus.

Koska edellä olevissa mallivertailuissa tuli vastaan virheilmoituksia, joiden pois saamiseksi jouduimme käyttämään R-ohjelmistossa lisäkoodia, päätimme tehdä kahden selittäjän sekamallivertailut *lme*-funktion lisäksi myös *lmer*-funktioilla. R-ohjelmiston *lmer*-funktio on siinä mielessä parempi kuin perinteinen *lme*-funktio, että koodiin ei tarvitse lisätä mitään ylimääräistä välttääkseen virheilmoitukset, sillä sekamalleja vertailtaessa virheilmoituksia ei tule vastaan. Etenemme mallien vertailussa samalla tavalla kuin aiemminkin eli aloitamme täydestä mallista, jossa kannanmuutosta selittävät kiinteät vasanaarassuhde ja urososuus sekä satunnaiset vakio, vasanaarassuhde ja urososuus. Sekamallien vertailu osoittautuu yllättäen hankalaksi, sillä mikään satunnaistermi ei näytä olevan merkitsevä. Näin ollen parhaaksi sekamalliksi on valittava jokin yhden satunnaisvai-

kutuksen malleista. Vertailemme keskenään sekamalleja, joissa kiinteiden vasanaarassuhteen ja urososuuden lisäksi selittäjänä on vain joko satunnainen vakio, satunnainen vasanaarassuhde tai satunnainen urososuus. Pienin AIC-luku on sekamallilla, jossa satunnaisena on vain urososuus ($AIC_{\text{vakio}} = -283.62$, $AIC_{\text{vasa}} = -282.34$, $AIC_{\text{uros}} = -286.21$), joten kahden selittäjän sekamalli muodostuu nyt kiinteistä vasanaarassuhteesta ja urososuudesta sekä satunnaisesta urososuudesta kannanmuutoksen selittäjinä.

R-ohjelmiston *lme*- ja *lmer*-funktioilla saadut sekamallit eroavat toisistaan siten, että satunnaispuoli koostuu ensimmäisessä sekä vakiosta että urososuudesta ja toisessa vain urososuudesta. Koska selvisimme *lmer*-funktioilla tehdyissä mallivertailuissa ilman virheilmoituksia ja R-lisäkoodia, on perustellumpaa pitää sen tuottamaa sekamallia parempana. Valitettavasti *lmer*-sekamallin ja regressiomallin vertaaminen ei onnistu (syy, miksi emme tehneet myös urososuudelle *lmer*-mallivertailuja luvussa 7.2), joten emme saa vietyä testausta aivan loppuun asti. Funktiota *lmer* käyttämällä saatu kahden selittäjän sekamalli, jossa kannanmuutosta selittävät kiinteät vasanaarassuhde ja urososuus sekä satunnaisena vain urososuus, ei kuitenkaan vaadi *lme*-funktiossakaan lisäkoodia (sillä siinä ei ole satunnaisvaikutuksena sekä vakio että urososuus), joten voimme toteuttaa regressiomalliin vertaamisen alkuperäisellä *lme*-funktioilla. Sekamalli on selvästi tarpeellinen (ANOVA, $p < 0.001$), joten lopullisessa kahden selittäjän sekamallissa kannanmuutoksen selittäjinä ovat kiinteät vasanaarassuhde ja urososuus sekä satunnainen urososuus

$$\text{muutos}_{ij} = \beta_0 + \beta_1 \text{vasanaarassuhde}_{ij} + (\beta_2 + u_{2i}) \text{urososuus}_{ij} + e_{ij}, \quad (7)$$

missä i = piiri ja j = vuosi. Tälle mallille vakiotermin β_0 estimaatti on -0.72 ja keskivirhe 0.10, vasanaarassuhteen regressiokertoimen β_1 estimaatti 0.18 ja keskivirhe 0.03 sekä urososuuden regressiokertoimen β_2 estimaatti 0.61 ja keskivirhe 0.27. Taulukossa 8 on mallin piirikohtaiset urososuuden satunnaiskertoimen ennusteet.

Taulukko 8 Mallin (7) piirikohtaiset urososuuden satunnaiskertoimen ennusteet.

	\hat{u}_{2i}
Etelä-Häme	0.043
Etelä-Savo	0.066
Kaakkois-Suomi	0.073
Keski-Suomi	-0.016
Oulu	-0.038
Pohjanmaa	-0.104
Pohjois-Häme	-0.032
Pohjois-Karjala	0.195
Pohjois-Savo	0.068
Rannikko-Pohjanmaa	-0.110
Satakunta	0.011
Uusimaa	-0.109
Varsinais-Suomi	-0.047

7.4 Paras sekamalli

Jatkoa varten täytyy vielä selvittää, kumpi parhaaksi arvioiduista vasanaarassuhteen ja kahden selittäjän sekamalleista on parempi. Koska mallit eivät ole sisäkkäisiä, vertaamme niiden AIC-lukuja. Sekamalleista paras on kahden selittäjän sekamalli ($AIC_{\text{vasa}} = -293.76$, $AIC_{\text{kaksi}} = -299.43$), jossa kannanmuutosta selittävät kiinteät vasanaarassuhde ja urososuus sekä satunnainen urososuus.

8 YHTEENVETO TULOKSISTA

8.1 Parhaiksi arvioidut mallit

Lukujen 6 ja 7 perusteella voimme koota yhteen piiritason parhaiksi arvioidut mallit, joita on neljä:

1. vasanaarassuhteen päävaikutusmalli (4), jossa kannanmuutosta piirissä i selittää vasanaarassuhde

$$muutos_{ij} = \beta_0 + \beta_{0i} + \beta_1 vasanaarassuhde_{ij} + e_{ij},$$

missä i = piiri ja j = vuosi

2. urososuuden yhdysvaikutusmalli (5), jossa kannanmuutosta piirissä i selittävät urososuus sekä piirin ja urososuuden välinen yhdysvaikutus

$$muutos_{ij} = \beta_0 + \beta_{0i} + (\beta_2 + \beta_{2i}) urososuus_{ij} + e_{ij},$$

missä i = piiri ja j = vuosi

3. vasanaarassuhteen sekamalli (6), jossa kannanmuutoksen selittäjinä ovat kiinteä vasanaarassuhde sekä satunnainen vakiotermi

$$muutos_{ij} = \beta_0 + u_{0i} + \beta_1 vasanaarassuhde_{ij} + e_{ij},$$

missä i = piiri ja j = vuosi

4. kahden selittäjän sekamalli (7), jossa kannanmuutoksen selittäjinä ovat kiinteät urososuus ja vasanaarassuhde sekä satunnainen urososuus

$$muutos_{ij} = \beta_0 + \beta_1 vasanaarassuhde_{ij} + (\beta_2 + u_{2i}) urososuus_{ij} + e_{ij},$$

missä i = piiri ja j = vuosi

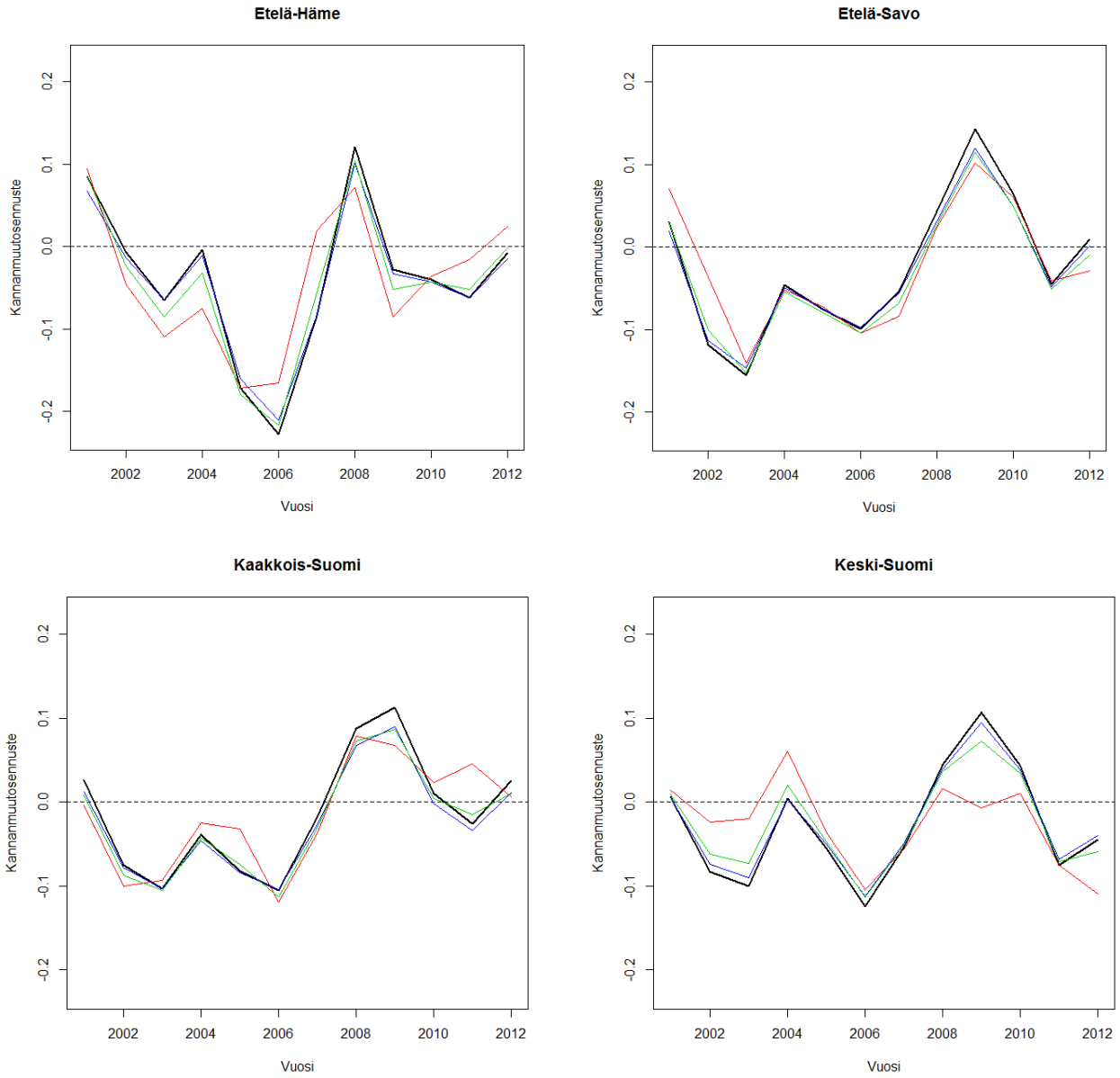
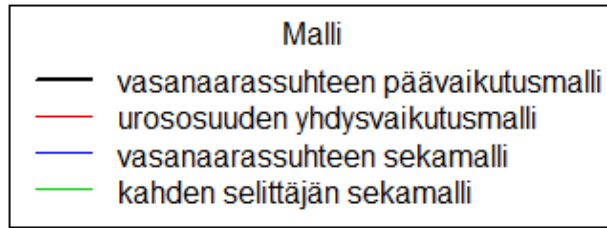
Listassa on kaksi vaihtoehtoista mallia vasanaarassuhteelle riippuen siitä, haluaako käyttää regressiomallia vai sekamallia. Urososuudelle on vain yksi yhdysvaikutusmalli, sillä sille sekamalli todettiin tarpeettomaksi. Kahdelle selittäjälle on vain sekamalli, sillä luvun 6.3 kahden selittäjän yhdysvaikutustarkasteluissa päädyimme listassa ensimmäisenä olevaan vasanaarassuhteen päävaikutusmalliin.

Jos parhaiksi arvioiduista malleista halutaan vielä valita kaikista paras, valinta tehdään vasanaarassuhteen päävaikutusmallin ja kahden selittäjän sekamallin välillä. Aiemmin luvussa 6.3 selvitimme, kumpi yllä olevan listan päävaikutus- ja yhdysvaikutusmalleista on parempi ja luvussa 7.4 selvisi paras sekamalli. Vasanaarassuhteen päävaikutusmallin ja kahden selittäjän sekamallin välisen paremmuusjärjestyksen ratkaisemiseksi vertailemme niiden AIC-lukuja. Niiden perusteella vasanaarassuhteen päävaikutusmalli on kaikista paras malli ($AIC_{\text{main}} = -309.61$, $AIC_{\text{mixed}} = -299.43$).

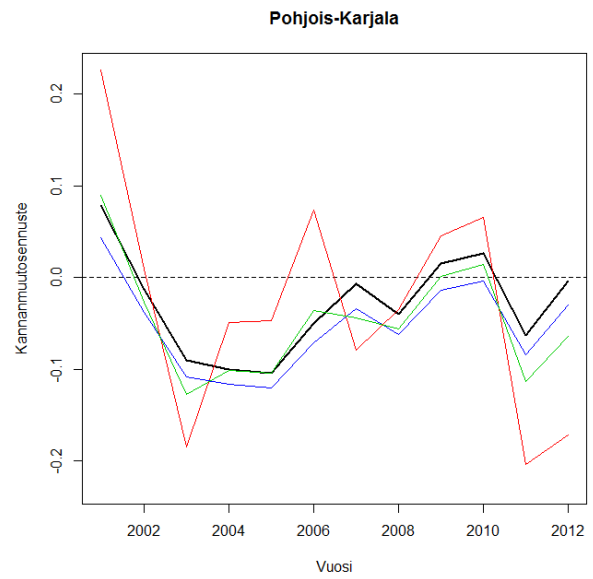
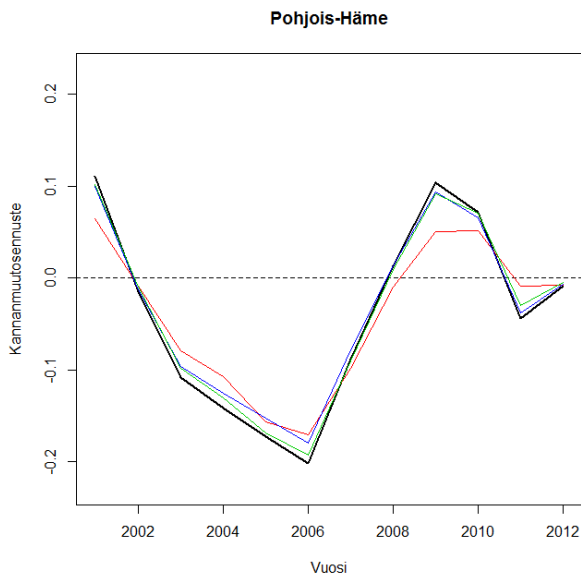
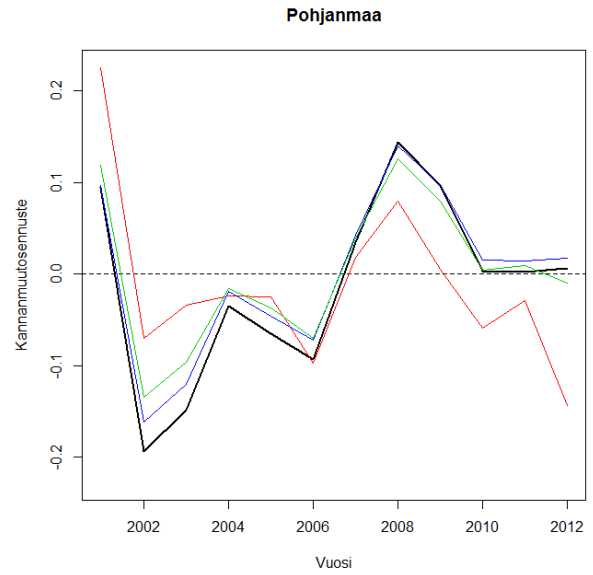
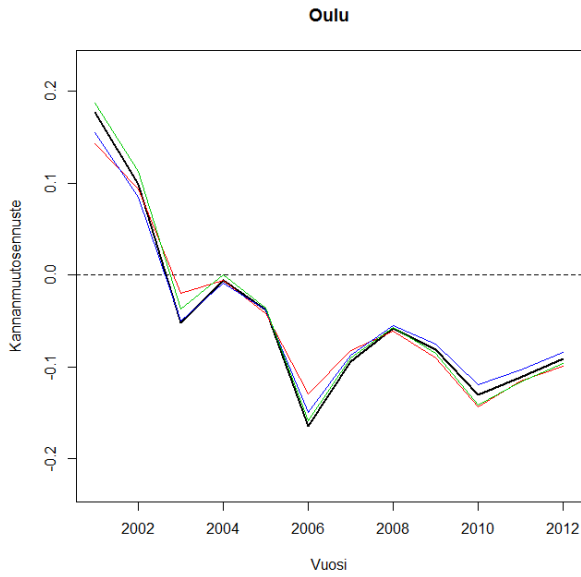
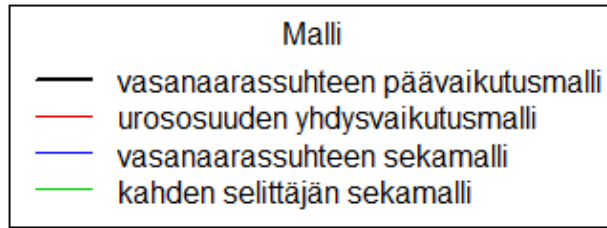
Seuraavaksi luvussa 8.2 vertailemme neljän parhaiksi arvioidun mallin antamia kannanmuutosennusteita. Sen jälkeen luvussa 8.3 esittelemme parhaan (kahden selittäjän) sekamallin kannanmuutosennusteista piirretyt tasa-arvokäyrät. Luvussa 8.4 käymme läpi kahden parhaan mallin, vasanaarassuhteen päävaikutusmallin ja parhaan (kahden selittäjän) sekamallin, diagnostiikkaa.

8.2 Kannanmuutoksen ennustaminen

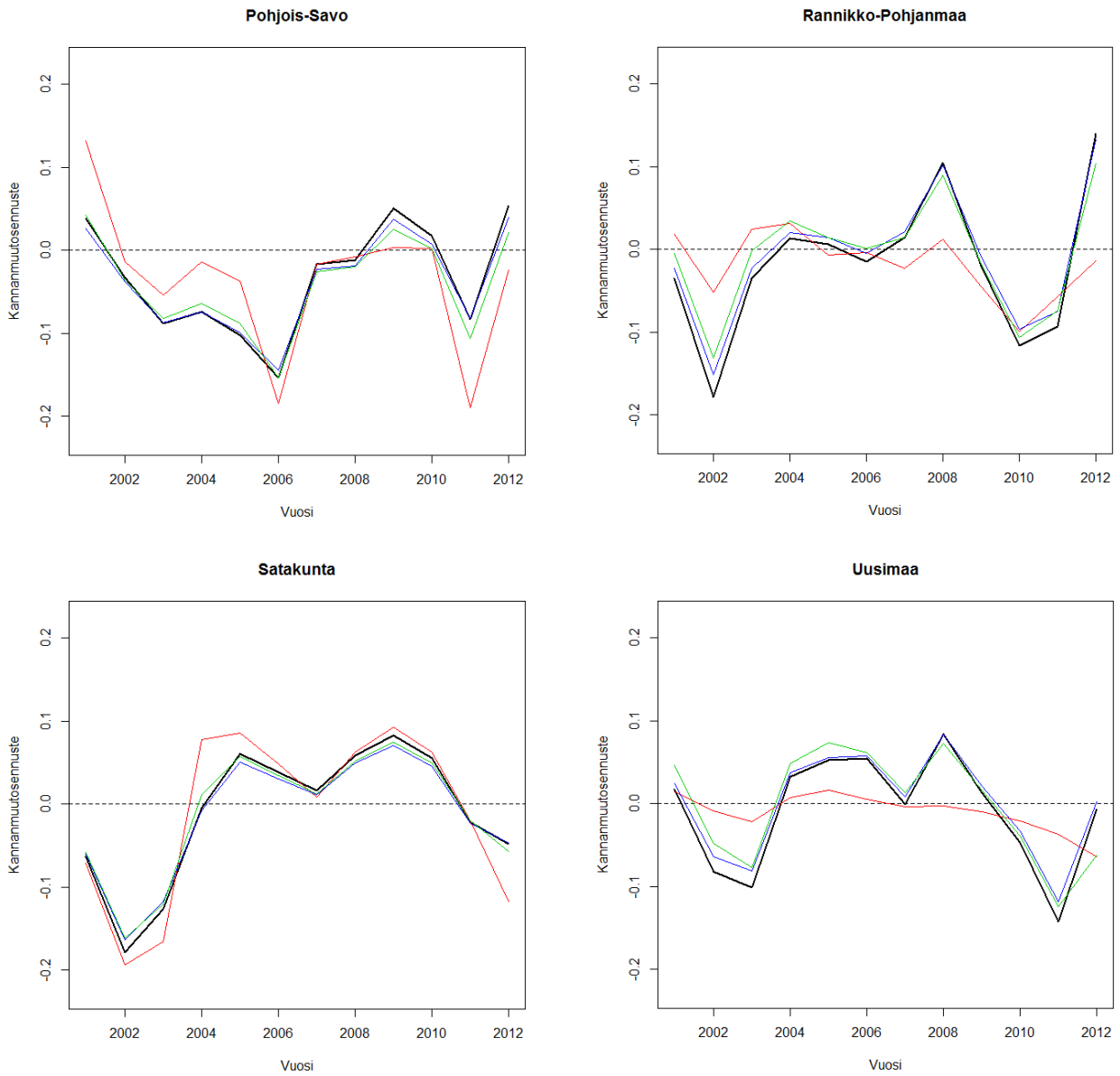
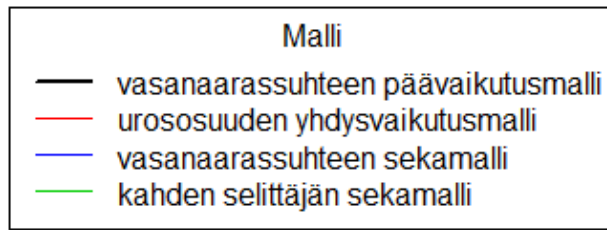
Olemme piirtäneet jokaiselle 13 piirille erikseen kuvan (ks. Kuvat 6a-d), jossa näkyy havaituista kannanmuutosluvuista laskettujen ennusteiden vaihtelu vuodesta toiseen luvussa 8.1 esiteltyjen neljän parhaaksi arvioidun mallin tapauksessa. Pääsääntöisesti kaikki neljä mallia näyttävät olevan melko samoilla linjoilla kannanmuutosennusteiden vuosivaihtelun suhteen eri piireissä. Parhaiksi arvioitujen mallien listasta vasanaarassuhteen päävaikutusmalli, vasanaarassuhteen sekamalli ja kahden selittäjän sekamalli (samalla paras sekamalli) ovat eniten samaa mieltä kannanmuutosennusteiden vaihtelusta vuosittain. Poikkeavimmat käyrät ovat urososuuden yhdysvaikutusmallilla. Kaikkein erikoisimmat kannanmuutosennustekäyrät sillä on muihin malleihin verrattuna Pohjois-Karjalassa ja Uudellamaalla. Näin ollen ainoa malli, jossa urososuus on selittämässä kannanmuutosta ilman vasanaarassuhdetta, näyttää antavan kaikkein poikkeavimpia vuosittaisia kannanmuutosennusteita.



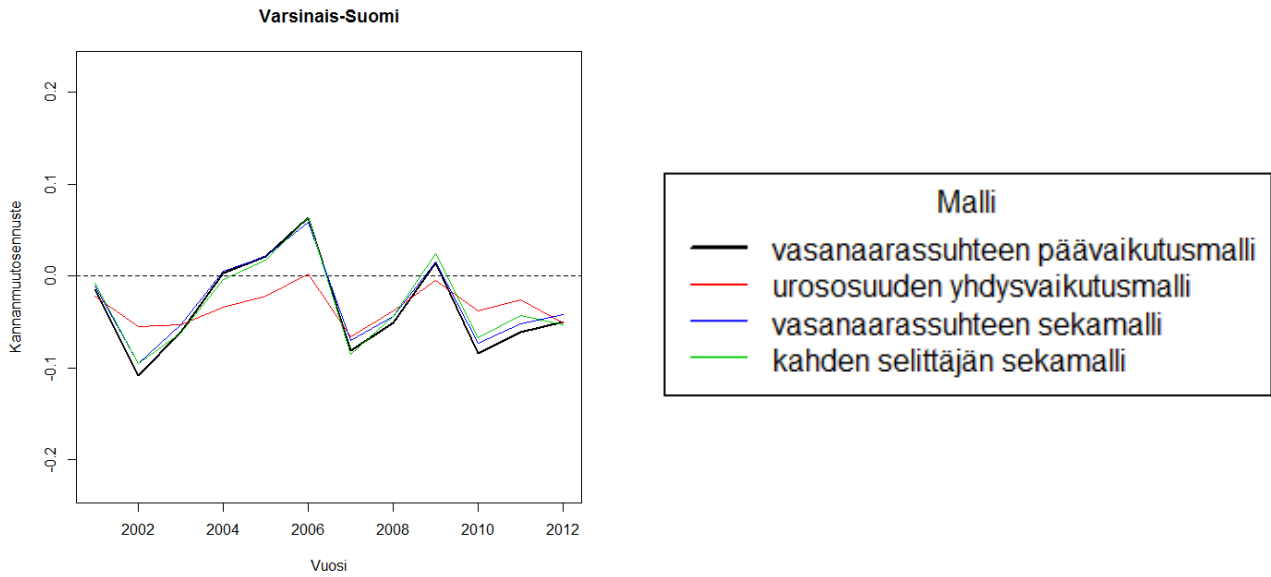
Kuva 6a Havaituista kannanmuutosluvuista laskettujen ennusteiden vaihtelu vuodesta toiseen neljän parhaaksi valitun mallin tapauksessa. Kaikista parhaimmaksi valittu malli, eli vasanaarassuhteen päävaikutusmalli, on erotettu hieman paksummalla mustalla. Etelä-Häme, Etelä-Savo, Kaakkois-Suomi ja Keski-Suomi (ks. myös Kuva 6b, Kuva 6c ja Kuva 6d).



Kuva 6b Havaituista kannanmuutosluvuista laskettujen ennusteiden vaihtelu vuodesta toiseen neljän parhaaksi valitun mallin tapauksessa. Kaikista parhaimmaksi valittu malli, eli vasanaarassuhteen päävaikutusmalli, on erotettu hieman paksummalla mustalla. Oulu, Pohjanmaa, Pohjois-Häme ja Pohjois-Karjala (ks. myös Kuva 6a, Kuva 6c ja Kuva 6d).



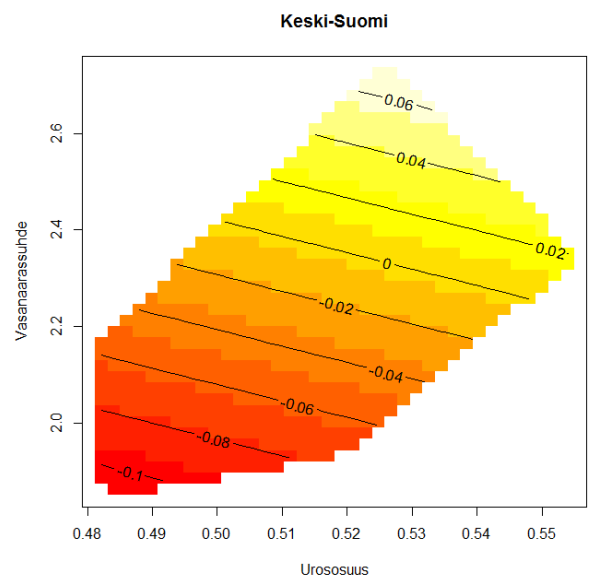
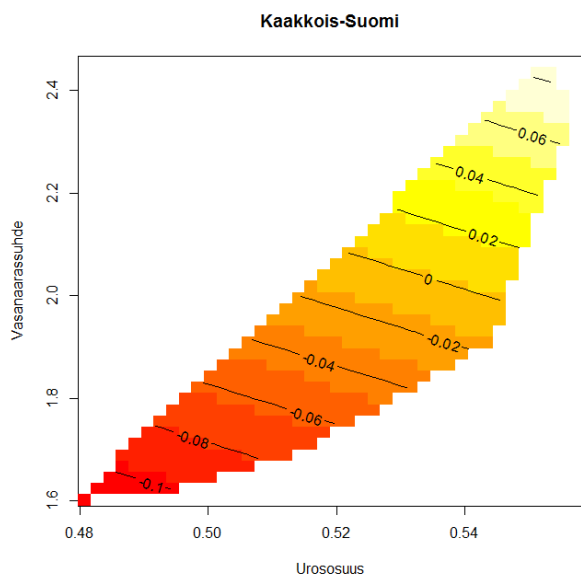
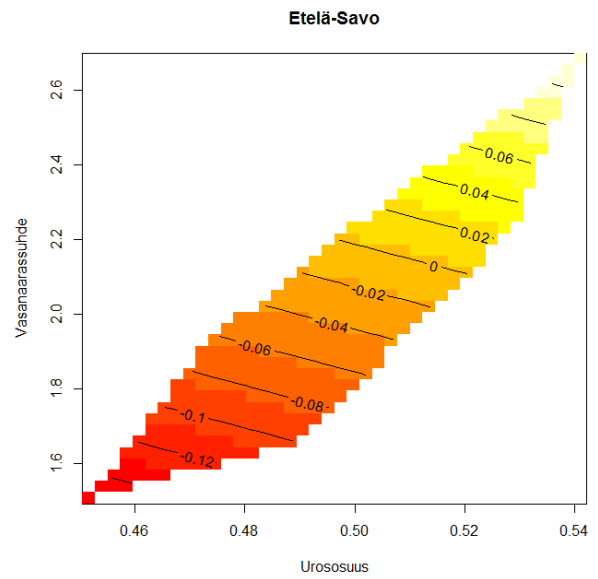
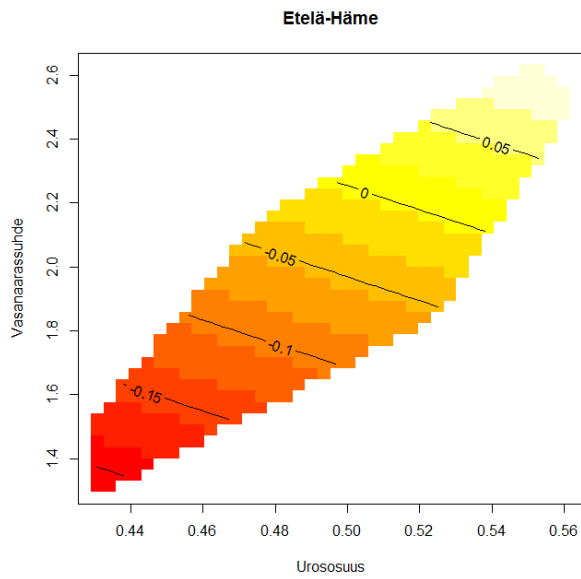
Kuva 6c Havaituista kannanmuutosluvuista laskettujen ennusteiden vaihtelu vuodesta toiseen neljän parhaaksi valitun mallin tapauksessa. Kaikista parhaimmaksi valittu malli, eli vasanaarassuhteen päävaikutusmalli, on erotettu hieman paksummalla mustalla. Pohjois-Savo, Rannikko-Pohjanmaa, Satakunta ja Uusimaa (ks. myös Kuva 6a, Kuva 6b ja Kuva 6d).



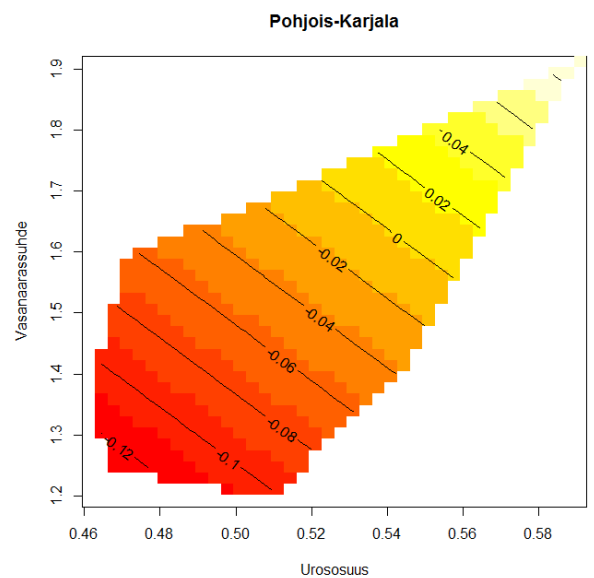
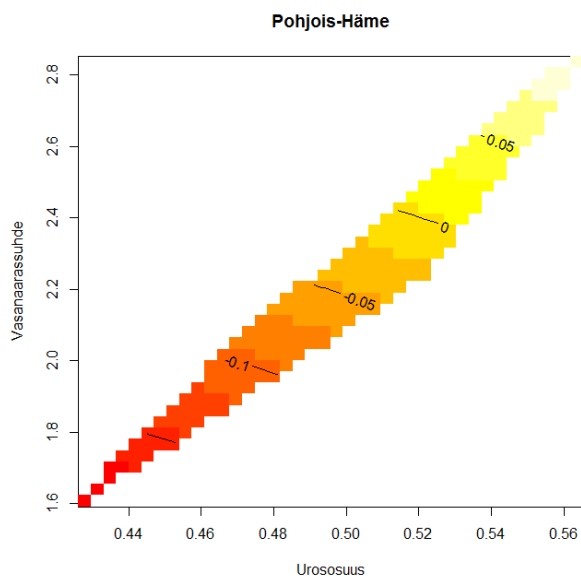
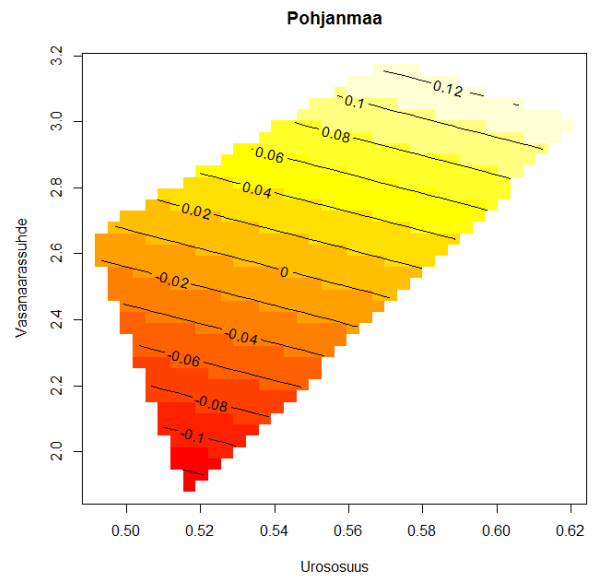
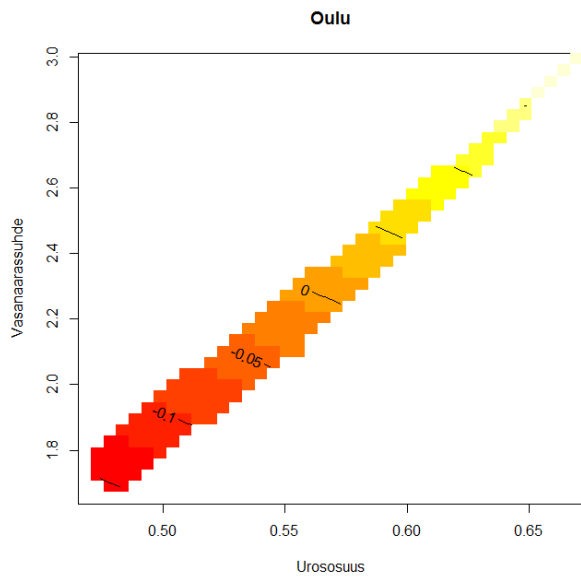
Kuva 6d Havaituista kannanmuutosluvuista laskettujen ennusteiden vaihtelu vuodesta toiseen neljän parhaaksi valitun mallin tapauksessa. Kaikista parhaimmaksi valittu malli, eli vasanaarassuhteen päävaikutusmalli, on erotettu hieman paksummalla mustalla. Varsinais-Suomi (ks. myös Kuva 6a, Kuva 6b ja Kuva 6c).

8.3 Tasa-arvokäyrät

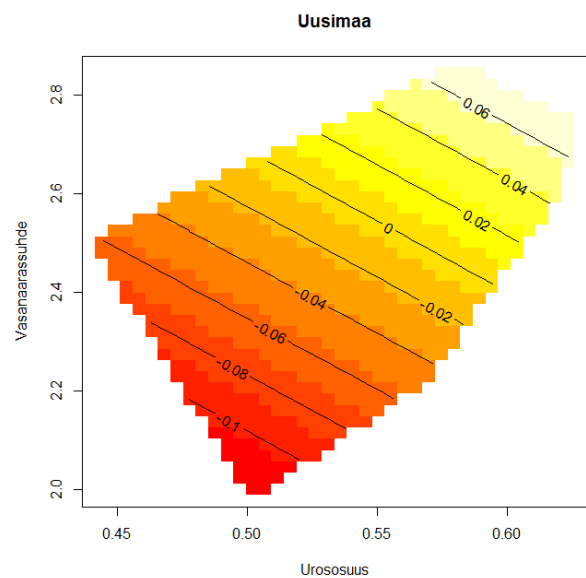
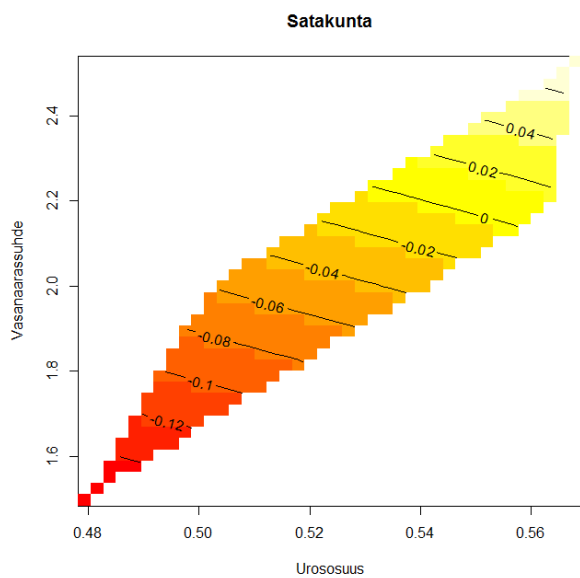
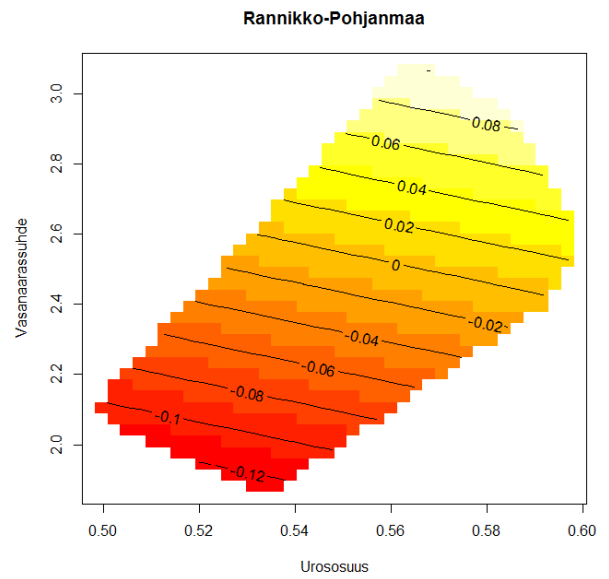
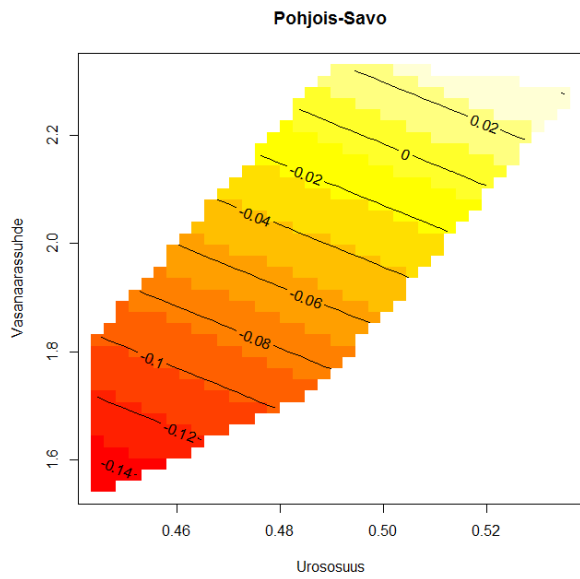
Kannanmuutosennusteista ja selittäjämuuttujista (vasanaarassuhde ja urososuus) voi piirtää myös niin sanotut tasa-arvokäyrät. Olemme piirtäneet tällaiset kuvat jokaiselle piirille käyttäen kahden selittäjän sekamallin (samalla paras sekamalli) kannanmuutosennusteita, sillä se on parhaiksi valituista malleista ainoa, jossa selittäjänä esiintyy sekä vasanaarassuhde että urososuus. Tasa-arvokäyristä saadaan selville kannanmuutosennusteet piireittäin, kun tiedetään vasanaarassuhde ja urososuus. Tulkitsemisen helpottamiseksi x- ja y-akseleiden skaalaukset ovat piirikohtaisia, sillä muutoin kuvien keskellä olevat kannanmuutosennusteluvut menevät joidenkin piirien tapauksessa lukukelvottomiksi.



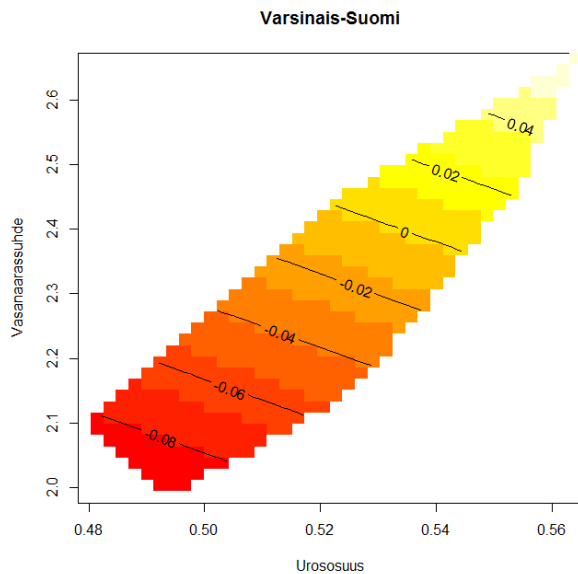
Kuva 7a Tasa-arvokäyrät, joista saadaan selville kannamuutosennusteet, kun tiedetään vasanaarassuhde ja urososuus. Ku-
vissa keskellä olevat kannamuutosennusteet on laskettu kahden selittäjän sekamallista (samalla paras sekamalli). Akseleiden
skaalaukset ovat tulkinnallisista syistä piirikohtaisia. Etelä-Häme, Etelä-Savo, Kaakkois-Suomi ja Keski-Suomi (ks. myös Kuva
7b, Kuva 7c ja Kuva 7d).



**Kuva 7b Tasa-arvokäyrät, joista saadaan selville kannanmuutosennusteet, kun tiedetään vasanaarassuhde ja urososuus. Ku-
vissa keskellä olevat kannanmuutosennusteet on laskettu kahden selittäjän sekamallista (samalla paras sekamalli). Akseleiden
skaalaukset ovat tulkinnallisista syistä piirikohtaisia. Oulu, Pohjanmaa, Pohjois-Häme ja Pohjois-Karjala (ks. myös Kuva 7a,
Kuva 7c ja Kuva 7d).**



Kuva 7c Tasa-arvokäyrät, joista saadaan selville kannanmuutosennusteet, kun tiedetään vasanaarassuhde ja urososuus. Kuvis-
sa keskellä olevat kannanmuutosennusteet on laskettu kahden selittäjän sekamallista (samalla paras sekamalli). Akseleiden
skaalaukset ovat tulkinnallisista syistä piirikohtaisia. Pohjois-Savo, Rannikko-Pohjanmaa, Satakunta ja Uusimaa (ks. myös
Kuva 7a, Kuva 7b ja Kuva 7d).



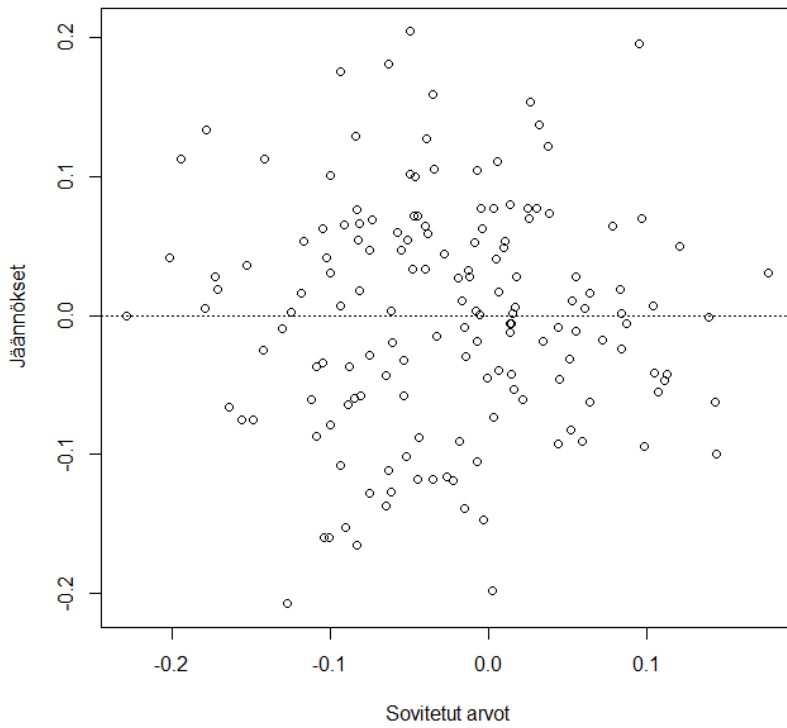
Kuva 7d Tasa-arvokäyrät, joista saadaan selville kannanmuutosennusteet, kun tiedetään vasanaarassuhde ja urososuus. Kuvissa keskellä olevat kannanmuutosennusteet on laskettu kahden selittäjän sekamallista (samalla paras sekamalli). Akseleiden skaalaukset ovat tulkinnallisista syistä piirikohtaisia. Varsinais-Suomi (ks. myös Kuva 7a, Kuva 7b ja Kuva 7c).

8.4 Mallien diagnostiikkaa

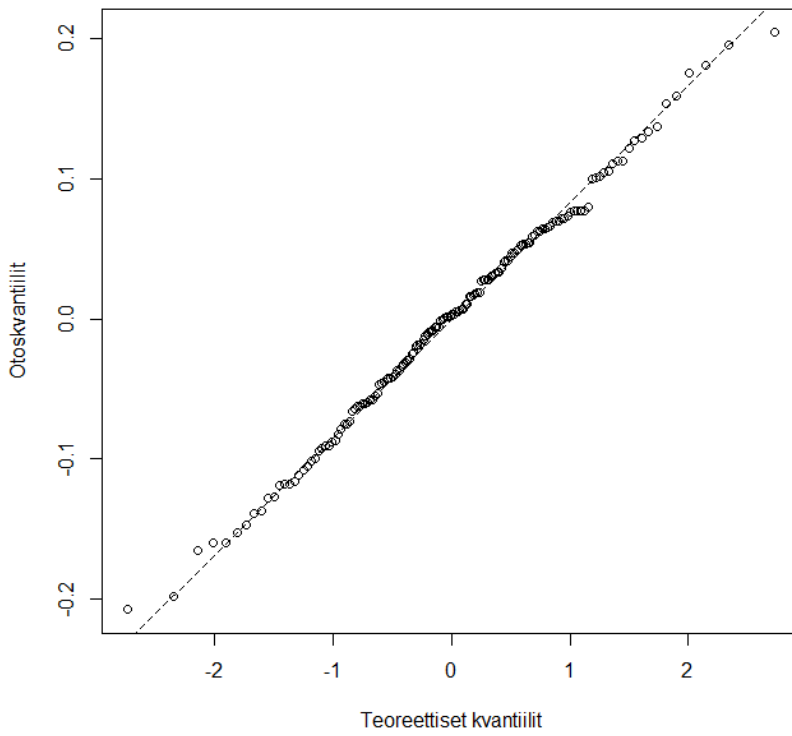
Seuraavaksi tarkastelemme hieman tarkemmin parasta regressiomallia eli vasanaarassuhteen päävaikutusmallia ja parasta sekamallia eli kahden selittäjän sekamallia. Tutkimme niiden jäännöksiä, mallien soveltuvuutta havaintoihin ja sekamallin osalta satunnaisvaikutuksia.

Vasanaarassuhteen päävaikutusmalli

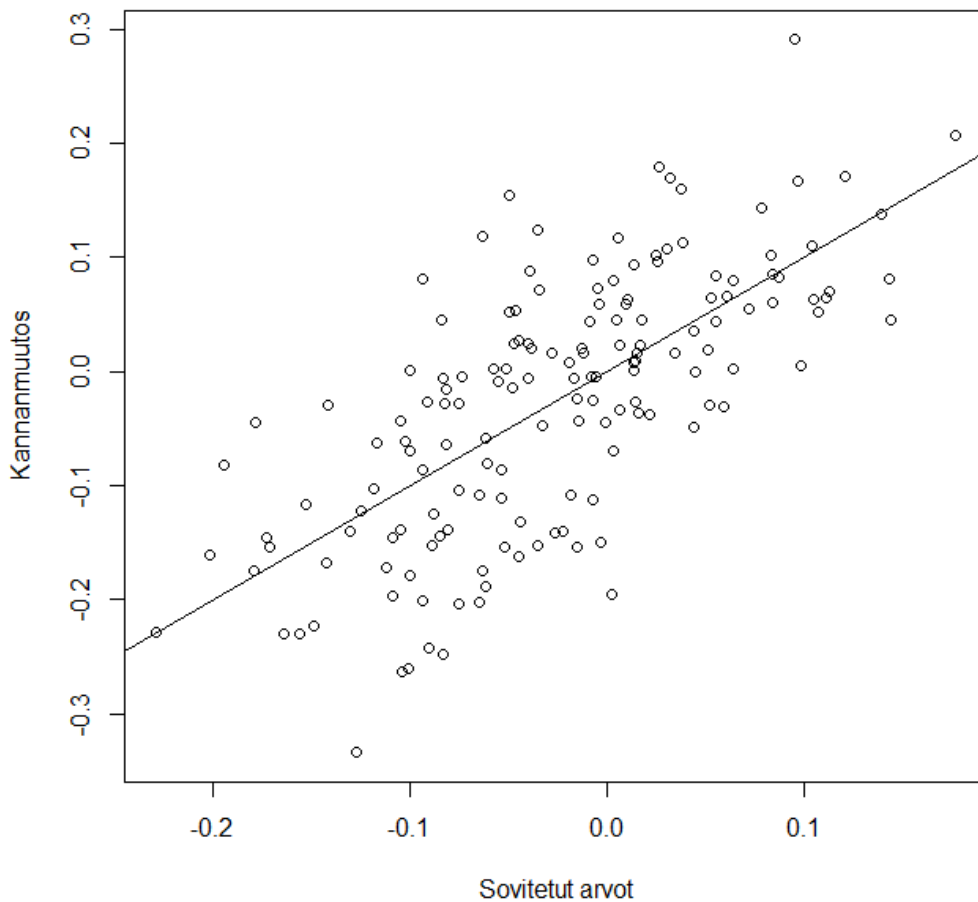
Kuvasta 8 näemme, että kannanmuutoksen jäännösten ja sovitettujen arvojen välillä ei ole minikäänlaista riippuvuutta. Pisteet näyttävät olevan satunnaisesti jakautuneita keskiviivan ympärille. Kuvan 9 perusteella jäännökset noudattavat oikein hyvin normaalijakaumaa. Vasanaarassuhteen päävaikutusmalli sopii hyvin havaintoihin (Kuva 10), sillä alkuperäisten kannanmuutoshavaintojen ja sovitettujen arvojen yhteisjakauma näyttää hyvältä muutamaa poikkeavaa havaintoa lukuun ottamatta.



Kuva 8 Kannanmuutoksen jäännösten ja sovitettujen arvojen yhteisjakauma.



Kuva 9 Kannanmuutoksen jäännösten normalisuus.

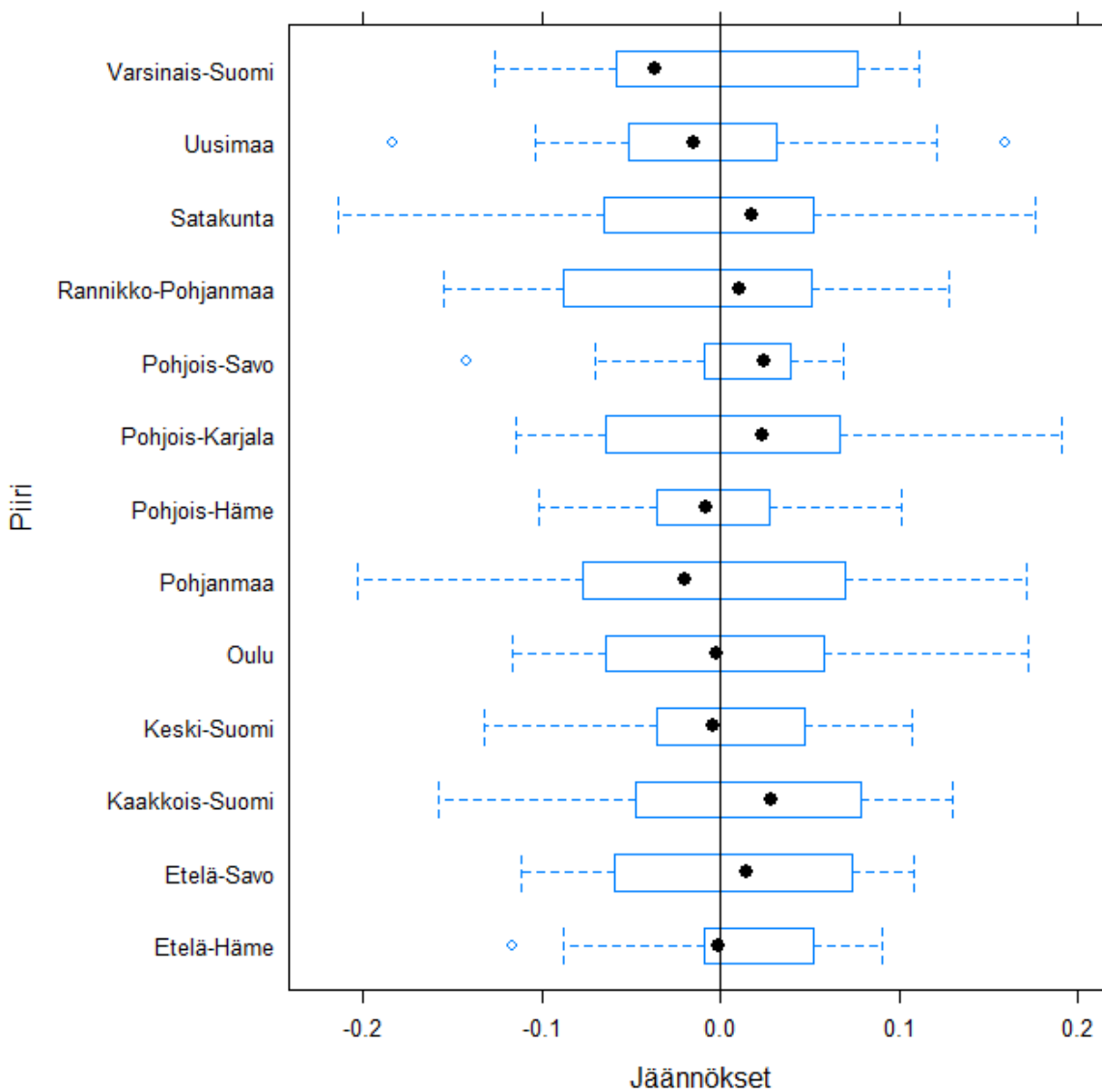


Kuva 10 Alkuperäisten kannanmuutoshavaintojen ja sovitettujen arvojen yhteisjakauma.

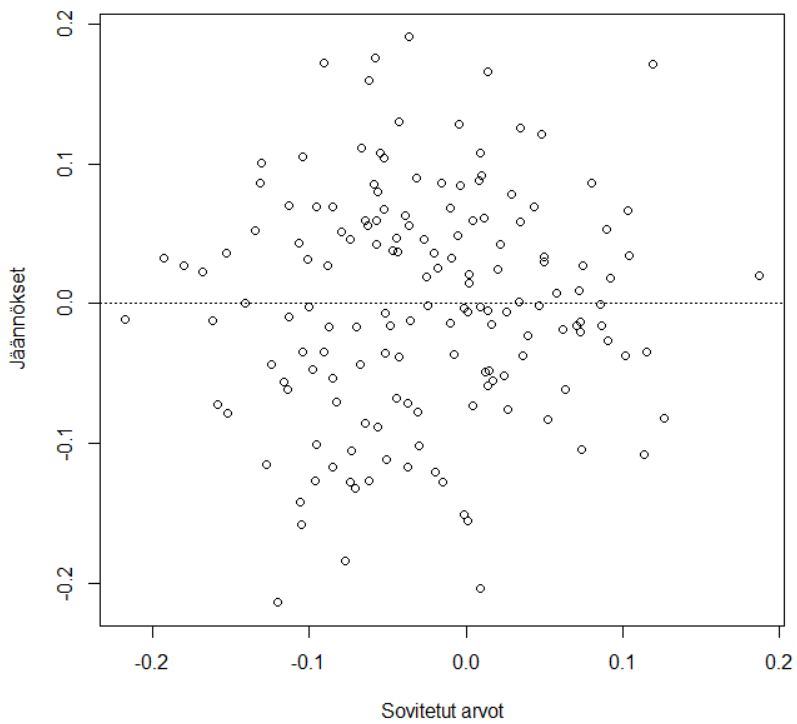
Kahden selittäjän sekamalli

Kuvasta 11 huomaamme, että kannanmuutoksen jäännösten vaihtelu on suurinta Pohjanmaalla ja Satakunnassa, ja poikkeavia jäännöshavaintoja on Etelä-Hämeessä, Pohjois-Savossa ja Uudella- maalla. Lisäksi Pohjois-Savon jäännösjakauma on hieman vasemmalle vino. Kannanmuutoksen jäännösten ja sovitettujen arvojen välillä ei ole lainkaan riippuvuutta, sillä pisteet näyttävät olevan satunnaisesti jakautuneita keskiviivan ympärille (Kuva 12). Kuvan 13 perusteella kannanmuutoksen jäännökset näyttävät noudattavan hyvin normaalijakaumaa. Urososuuden satunnaisvaikutusten ennusteet taas eivät noudata normaalijakaumaa kovin hyvin (Kuva 14). Taulukosta 9 näemme, että

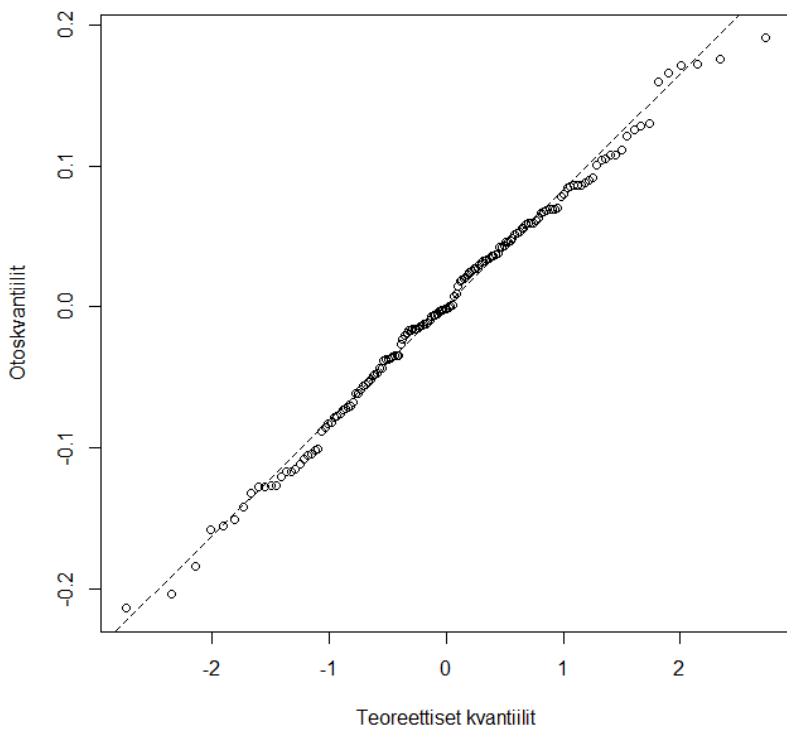
selvästi pienimmät urososuuden kulmakertoimen satunnaisvaikutuksen ennusteet ja lopulliset kulmakertoimet ovat Rannikko-Pohjanmaalla, Uudellamaalla ja Pohjanmaalla. Silmiinpistäväntä on kuitenkin urososuuden kulmakertoimen satunnaisvaikutuksen ennusteen ja lopullisen kulmakertoimen suuruus Pohjois-Karjalan tapauksessa. Alkuperäisten havaintojen ja sovitettujen arvojen yhteisjakauma näyttää todella hyvältä (Kuva 15), mikä tarkoittaa sitä, että lineaarinen sekamalli sopii hyvin havaintoihin. Muutamia poikkeavia havaintoja on Oulussa, Pohjanmaalla, Pohjois-Karjalassa, Satakunnassa ja Uudellamaalla.



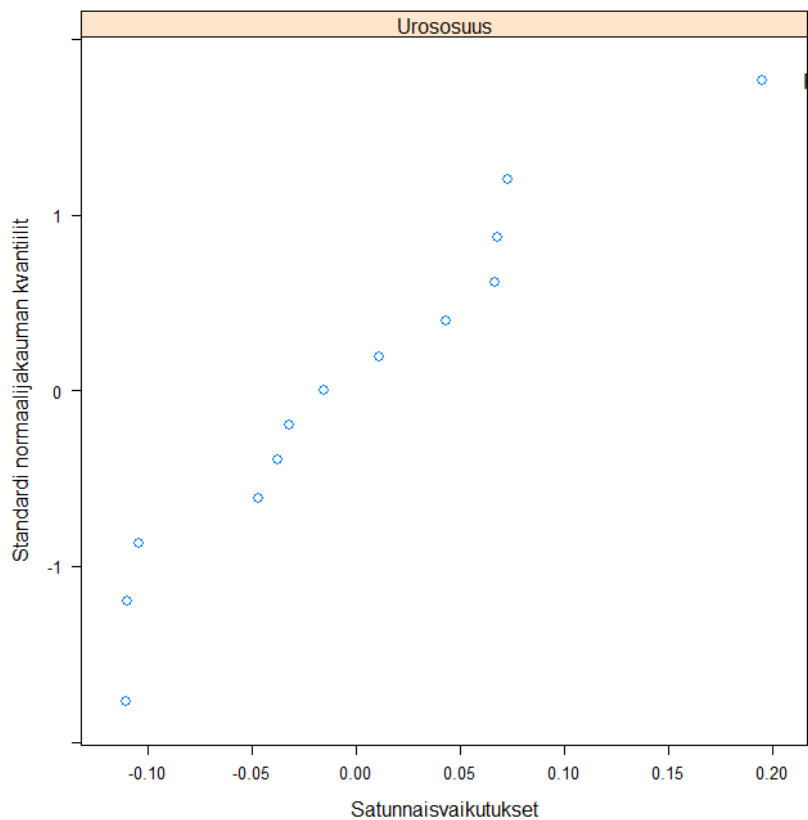
Kuva 11 Kannanmuutoksen jäännökset piireittäin.



Kuva 12 Kannanmuutoksen jäännösten ja sovitettujen arvojen yhteisjakauma.



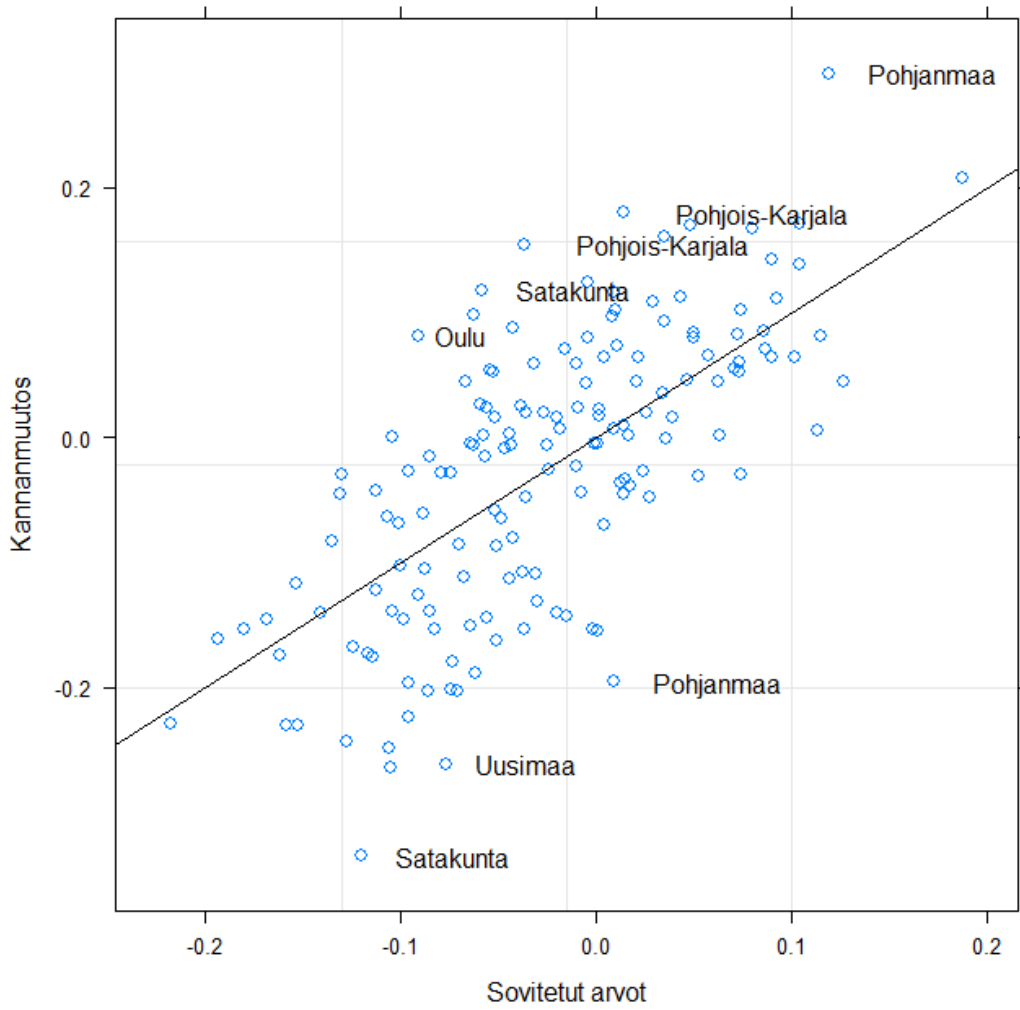
Kuva 13 Kannanmuutoksen jäännösten normalisuus.



Kuva 14 Urososuuden satunnaisvaikutusten ennusteiden normalisuus.

Taulukko 9 Urososuuden kulmakertoimen satunnaisvaikutuksen piirikohtaiset ennusteet ja lopulliset kulmakertoimet.

	\hat{u}_{2i}	$\hat{\beta}_2 + \hat{u}_{2i}$
Etelä-Häme	0.043	0.701
Etelä-Savo	0.066	0.747
Kaakkois-Suomi	0.073	0.760
Keski-Suomi	-0.016	0.583
Oulu	-0.038	0.538
Pohjanmaa	-0.104	0.405
Pohjois-Häme	-0.032	0.549
Pohjois-Karjala	0.195	1.004
Pohjois-Savo	0.068	0.749
Rannikko-Pohjanmaa	-0.110	0.393
Satakunta	0.011	0.636
Uusimaa	-0.109	0.395
Varsinais-Suomi	-0.047	0.521



Kuva 15 Alkuperäisten kannanmuutoshavaintojen ja sovitettujen arvojen yhteisjakauma. Poikkeavimpien havaintojen piirit on merkitty näkyviin.

9 POHDINTAA

Tutkimuksemme tarkoituksena oli selvittää, voidaanko saalisrakenneindekseillä ennustaa hirvikannan muutosta vuodesta toiseen. Lisäksi halusimme nähdä, vaihtelee kannanmuutoksen ja saalisrakenneindeksien välinen yhteys alueittain. Saimme sekä odotettuja että odottamattomia tuloksia. Saalisrakenneindeksit näyttäsivät joka tapauksessa soveltuvan hyvin kannanmuutoksen ennustamiseen, sillä kehitimme monta erilaista toimivaa mallia.

Huomasimme, että valtakunnallisella tasolla ilman piirierottelua varsinkin vasanaarassuhde sekä vasanaarassuhde ja urososuus yhdessä toimivat erittäin hyvin kannanmuutoksen selittäjinä. Myös urososuuden tapauksessa regressiomalli oli toimiva, mutta sen selitysaste jäi paljon alhaisemmaksi. Saimme myös oletetun tuloksen: sekä vasanaarassuhde että urososuus korreloivat kannanmuutoksen kanssa positiivisesti. Piiritasolla kannanmuutoksen ja vasanaarassuhteen välinen regressiomalli toimi suuremmalle määrälle piirejä kuin urososuuden ja kahden selittäjän (selittäjinä vasanaarassuhde ja urososuus) vastaavat mallit. Piirien välillä näytti olevan enemmän eroa urososuuden kuin vasanaarassuhteen selittäessä kannanmuutosta. Lisäksi tutkiessamme saalisrakenneindeksien ja piirifaktorin välisiä yhdysvaikutuksia huomasimme, että vasanaarassuhteen ja piirifaktorin välillä ei ole yhdysvaikutusta, mutta urososuuden ja piirifaktorin välillä taas on yhdysvaikutus. Näin ollen ajattelimme, että varsinkin urososuuden ja kannanmuutoksen väliseen regressiomalliin voisi olla hyvä lisätä piirien välistä vaihtelua kuvaava satunnaisvaikutus. Sekamalli ei yllättäen osoittautunutkaan tarpeelliseksi urososuuden tapauksessa, vaan regressiomalli oli riittävä. Totesimme, että selitys tälle saattaa olla kiinteiden vaikutusten lineaaristen regressiomallien ja lineaaristen sekamallien erilaisissa riippuvuusoletuksissa. Vasanaarassuhteelle ja kahdelle selittäjälle sekamalli sen sijaan oli tarpeellinen. Näin ollen sekamallien käytöstä oli hyötyä, vaikka emme tarvinneetkaan niitä juuri siinä tapauksessa kuin regressiotulosten perusteella arvelimme.

Regressio- ja sekamallivertailut tuottivat neljä parhaaksi arvioitua mallia: vasanaarassuhteen päävaikutusmalli, urososuuden yhdysvaikutusmalli, vasanaarassuhteen sekamalli ja kahden selittäjän sekamalli (samalla paras sekamalli). Näistä malleista kaikkien muiden paitsi urososuuden yhdysvaikutusmallin antamat piirikohtaiset kannanmuutosennusteet vuosille 2001–2012 olivat yllättäen hyvin samanlaiset. Kahden selittäjän sekamallin (paras sekamalli) kannanmuutosennusteista ja

selittäjämuuttujista piirsimme lisäksi jokaiselle piirille tasa-arvokäyrät, joiden tulkinta on pitkälti piirikohtaista akseleiden skaalauksen vuoksi.

Parhaan regressiomallin ja parhaan sekamallin diagnostiikkatarkastelut osoittivat mallien valinnan olleen pääosin onnistunut. Mallit sopivat hyvin havaintoihin ja jäännökset käyttäytyivät hyvin. Oikeastaan ainoaksi ongelmaksi osoittautui se, että kahden selittäjän sekamallin urososuuden satunnaisvaikutusten ennusteet eivät noudattaneet kovin hyvin normaalijakaumaa.

Yhteenvetona voimme todeta, että saalisrakennepäävaikutuksesta vasanaarassuhde näyttäisi toimivan oikein hyvin kannanmuutoksen ennustajana. Se toimi urososuutta paremmin sekä valtakunnallisesti että piiritasolla. Lisäksi löysimme sille hyvän kiinteiden vaikutusten regressiomallin ja sekamallin. Näistä malleista vasanaarassuhteen päävaikutusmallin tarkemmat diagnostiikkatarkastelut osoittivat, että malli oli erittäin onnistunut. Urososuus ei sopinut ennustamaan kannanmuutosta aivan yhtä hyvin, sillä sille regressiomalli todettiin riittäväksi, vaikka sekamallille näytti olevan tarvetta. Vasanaarassuhde ja urososuus yhdessä taas näyttivät olevan toimiva kokonaisuus kannanmuutosta ennustettaessa. Regressiomalli toimi hyvin valtakunnallisesti ja sekamallikin saatiin aikaiseksi. Kahden selittäjän sekamallin lähemmät tarkastelut osoittivat mallin olevan hyvä lukuun ottamatta urososuuden satunnaisvaikutusten ennusteiden epänormaalisuutta.

Erityisesti parhaan regressiomallin, vasanaarassuhteen päävaikutusmallin, osalta olisi hyvä tarkastella vielä ennuste- ja luottamusvälejä. Sekamalleissa satunnaisvaikutusten kovarianssimatriisille voisi kokeilla esimerkiksi diagonaalista rakennetta. Tässä työssä saatuja malleja voisi soveltaa myös esimerkiksi muiden hirvieläinten kannanmuutoksen ennustamiseen siltä osin kuin niiden saalis- ja jäävän kannan tietoja on saatavilla.

KIITOKSET

Haluan kiittää lehtori Harri Högmanderia, FM Tuomas Kukkoa, tutkija Jyrki Puseniusta (Riista- ja kalatalouden tutkimuslaitos, Joensuu) ja riistapäällikkö Jukka Purhosta (Suomen riistakeskus, Jyväskylä), joiden yhteispohdinnat tuottivat minulle erittäin mielenkiintoisen pro gradu -tutkielman aiheen. Tuomas Kukolle haluan osoittaa kiitokset myös työn alkuvaiheen aineisto- ja asiantunte-
musavusta.

Suurimmat kiitokset esitän tutkielmani ohjaajille Harri Högmanderille ja FT Salme Kärkkäiselle todella kannustavasta ja paneutuvasta ohjauksesta. Jokainen tapaaminen heidän kanssaan sai minut hyvälle tuulelle ja innosti eteenpäin. Lisäksi he auttoivat teoriapuolen sisäistämisessä ja R-ohjelmiston käytössä.

Lisäksi kiitän dosentti Sara Taskista työn loppuvaiheessa annetusta palautteesta.

Jyväskylässä 27.6.2014

Viivi Nuojua

LÄHTEET

Brown, H. & Prescott, R. (2006). *Applied Mixed Models in Medicine*. 2. painos. Wiley, New York.

McCulloch, C. E. & Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. Wiley, New York.

Morrison, D. F. (1990). *Multivariate Statistical Methods*. 3. painos. McGraw-Hill, New York.

Nygrén, T. (2010). *Hirvikannan säätely: Perusbiologiaa*. Riista- ja kalatalouden tutkimuslaitos (julkaisematon).

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer-Verlag, New York.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

<http://www.riistaweb.riista.fi> (viitattu 2.6.2014)

<http://www.metsastajaliitto.fi> (viitattu 2.6.2014)

<http://www.riista.fi> (viitattu 2.6.2014)