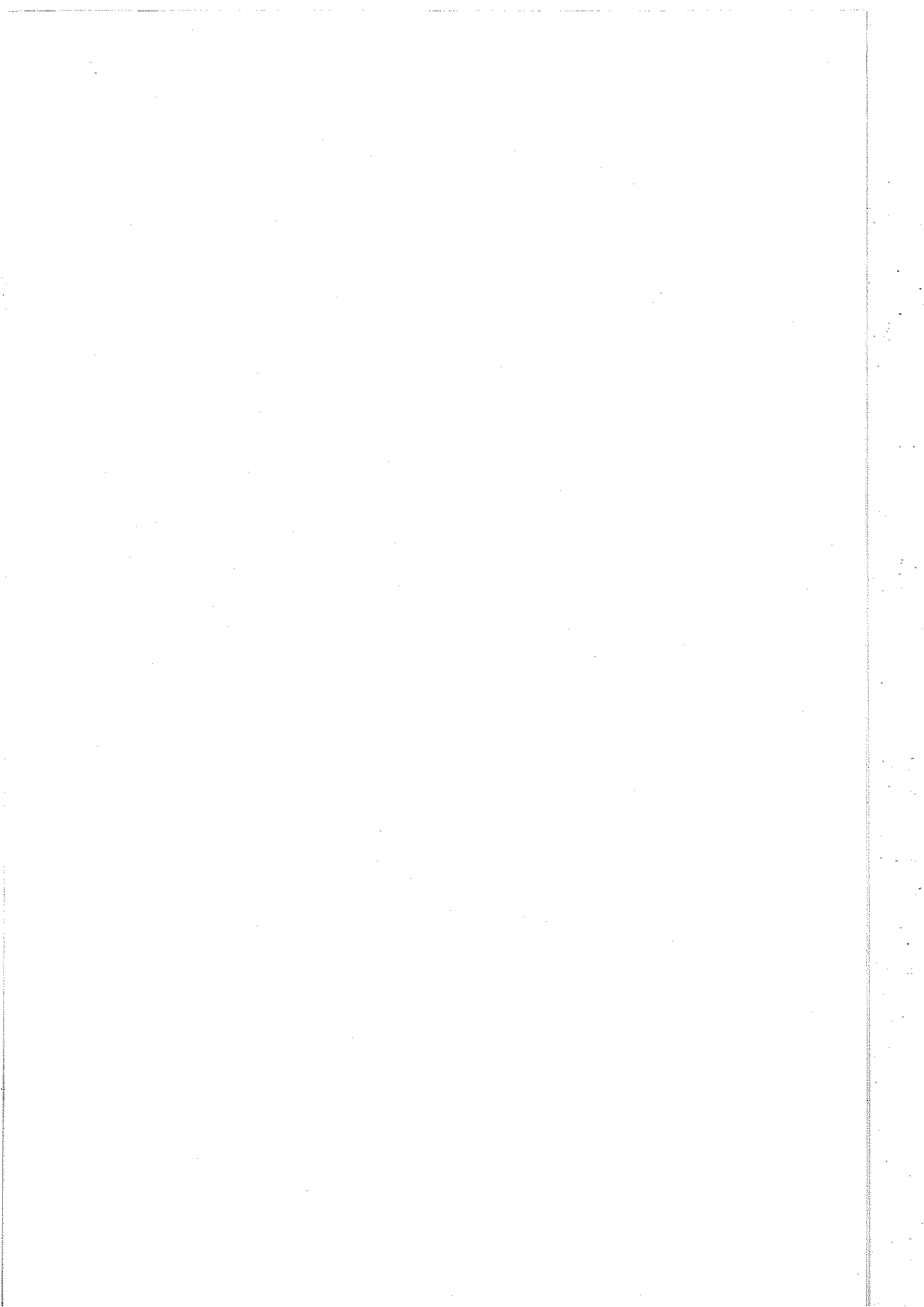


5.

Moottoriajoneuvovakuutusten hinnoittelua yleistetyllä lineaarisella mallilla.

SHV-harjoitustyö
Heikki Arminen
13.12.2004



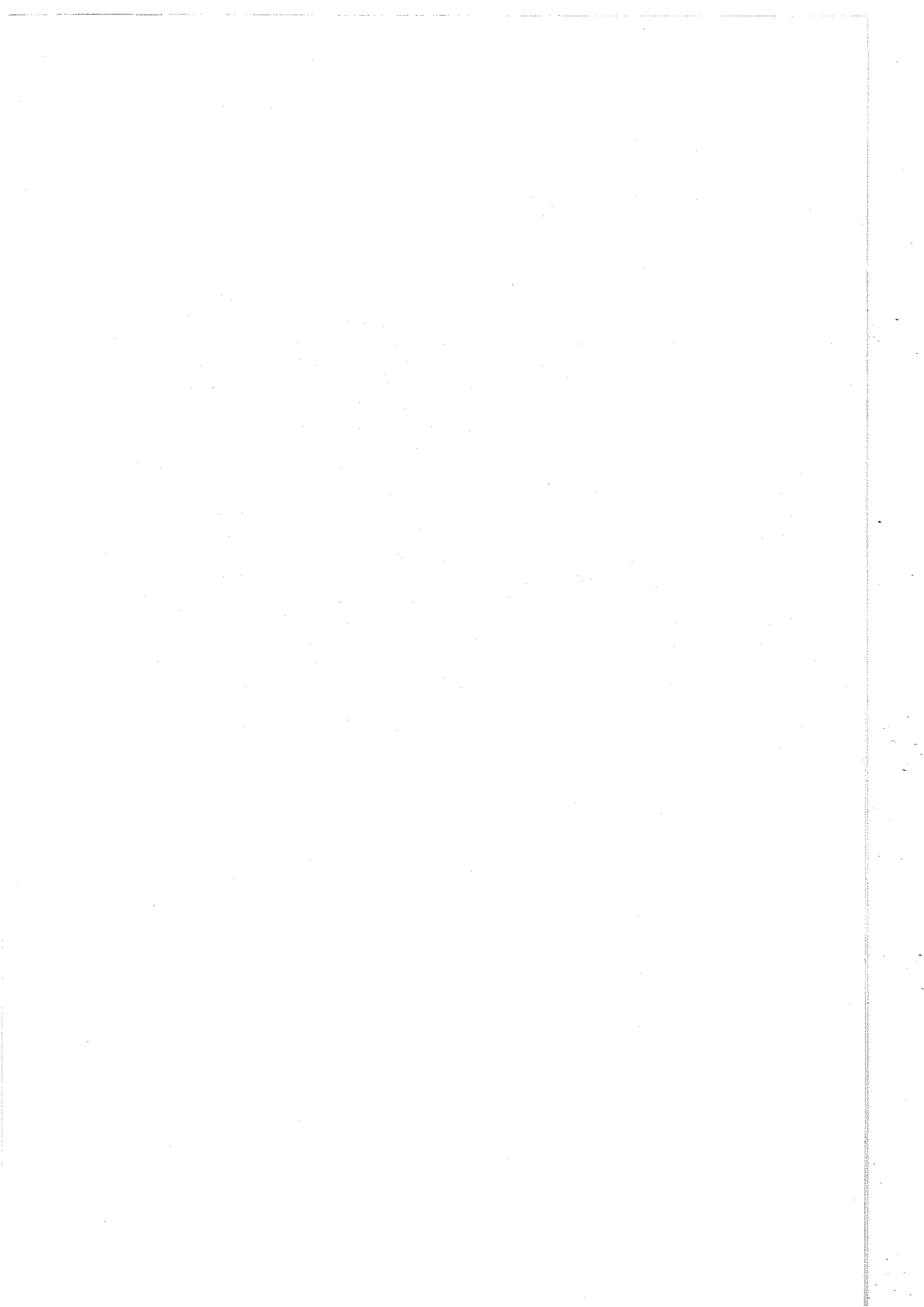
Motor insurance rating by generalized linear model

Abstract

Generalized linear models are nowadays used as a standard method for pricing motor and other personal lines insurance in the European Union and in many other markets. Models show the true statistical effect of rating factors upon the expected claims cost taking into account correlations and interactions between rating factors and the nature of random process. Claim frequency and severity are often modelled separately for different claim types. The technique of iterative weighted linear least-squares can be used to obtain maximum likelihood estimates of the parameters with observations distributed according to a distribution of the exponential family.

The structure and components of a generalized linear model, the measures of a model fit and the method of solving the parameters of a model are presented. The treatment of a generalized linear model using quasi-likelihood is introduced. In the 1960s actuaries developed ratemaking techniques, which are known as minimum bias procedures. The correspondence between minimum bias procedures and generalized linear models is defined. Some but not all minimum bias procedures have a corresponding generalized linear model and vice versa.

The effect of the selection of a distribution and a link function is investigated with the aid of an example, in which models has been fitted to British motor vehicle insurance data. The data gives average claim severities for private passenger car collisions. Claim frequency and severity models are fitted to motor third party liability data of Pohjola Insurance Company. The data consists of the material damage and personal injury claims of the passenger car policies of private persons. It turns out, that in addition to the region of use, the engine cubic capacity and the make and model of a car also the age of a driver and the age of a car influences to material damage and personal injury risk.



SISÄLLYSLUETTELO

1. JOHDANTO	3
2. YLEISTETTY LINEAARINEN MALLI	7
2.1 Yleinen lineaarinen malli	7
2.2 Yleistetyn lineaarisen mallin rakenne	8
2.2.1 Yleistetyn lineaarisen mallin uskottavuusfunktiot	9
2.2.2 Linkkifunktiot	11
2.3 Mallin hyvyyden mittaaminen	12
2.3.1 Devianssi	12
2.3.2 Residuaalit	15
2.4 Yleistetyn lineaarisen mallin kertoimien ratkaisu	16
2.5 Kvasiuskottavuus	21
3. TARIFFIN PARAMETRIEN ESTIMOINTI	29
3.1 Taulukoidut tiedot	29
3.1.1 Bailey-Simon:in menetelmä	30
3.1.2 Reunasummamenetelmä	32
3.1.3 Pienimmän neliösumman estimaatit	33
3.2 Estimointi yleistetyllä lineaarisella mallilla	34
4. AUTOVAKUUTUKSEN KESKIVAHINKO	42
5. HENKILÖAUTOJEN LIIKENNEVAKUUTUKSEN HINNOITTELUA	48
5.1 Omaisuusvahingon vahinkotiheys	48
5.2 Omaisuusvahingon keskivahinko	52
5.3 Henkilövahingon vahinkotiheys	53
5.4 Henkilövahingon keskivahinko	53
6. YHTEENVETO	54

LÄHDEVIITTEET

LIITTEET 1 - 7



1. JOHDANTO

Nelder ja Wedderburn (1972) esittivät, miten lineaaristen mallien yhteydessä käytettävät menetelmät voidaan laajentaa toimimaan yleisemmin. Selitettävän muuttujan ei tarvitse olla normaalisti jakautunut ja jatkuva, ja sen suhteen selittäviin muuttujiin ei tarvitse olla lineaarinen. Normaalijakauman ja lineaaristen mallien hyvät ominaisuudet saadaan käyttöön eksponenttijakaumien perheessä, joka sisältää mm. vakuutusosalalla käyttökelpoiset Poisson-, binomi- ja gammajakaumat. He kutsuivat malliaan yleistetyksi lineaariseksi malliksi (Generalized Linear Model). He kehittivät myös jo vuonna 1972 mallille ratkaisumenetelmän, jolla saadaan mallin parametrien suurimman uskottavuuden estimaatit. Vielä nykyäänkin yleistettyjen lineaaristen mallien ratkaisemisessa käytettävät ohjelmistot perustuvat tähän vuonna 1972 kehitettyyn menetelmään.

Wedderburn laajensi yleistettyjen lineaaristen mallien sovelluttavuutta vuonna 1974, kun hän esitti kvasiuskottavuuden. Hän osoitti, että usein ei ole tarpeellista tehdä satunnaisvaihtelua koskevia yksityiskohtaisia oletuksia. Sen sijaan parametriestimaattien monet hyödylliset alun perin suurimman uskottavuuden teorian perusteella johdetut ominaisuudet voidaan perustella heikompien muuttujien riippumattomuutta ja toisia momenteja koskevien oletusten avulla. On tarpeen tietää, miten jokaisen havainnon varianssi muuttuu havainnon odotusarvon muuttuessa, mutta ei ole tarpeen määrittellä koko jakaumaa. Käytännön ongelmat ratkaistaan nykyään useimmiten soveltamalla malleja kvasiuskottavuuden mukaisesti.

Vakuutusosalalla yleistettyjä lineaarisia malleja on käytetty 1980-luvun alusta alkaen. Yksi tärkeimpiä sovellusalueita on vahinkovakuutusten hinnoittelu, mutta GLM-malleja voidaan käyttää myös monien muiden vakuutusalan perusasioiden ratkaisuvälineenä. Haberman ja Renshaw (1996) mainitsevat ja esittävät esimerkkejä vahinkovakuutuksen hinnoittelun lisäksi seuraavista sovelluksista :

- (a) eliniän mallintaminen henkivakuutuksessa
- (b) sairausvakuutuksen tilamallit
- (c) vahingon suuruusjakauman sovittaminen vahinkoaineistoon
- (d) riskien luokittelu ; vahingoittuneiden henkilöiden kuolevuuden mallintaminen; tupakoitsijoiden lisäkuolevuuden mallintaminen; henkivakuutusten ennaikaisen päättymisen mallintaminen
- (e) vahinkovakuutuksen vahinkojen varaaminen.

Vahinkovakuutuksen hinnoittelussa GLM-mallien käyttö yleistyi 1990-luvulla, ja nykyään myös monet pohjoismaiset vakuutusyhtiöt käyttävät niitä riskimaksujen laskentavälineenä erityisesti moottoriajoneuvojen vakuutuksissa ja kotivakuutuksissa. Malleilla voidaan testata eri tekijöiden vaikutusta vahinkojen lukumäärään ja vahingon suuruuteen, ja valita maksun määräämisessä käytettävät tariffitekijät. Mallien kertoimien avulla voidaan laskea vahinkotiheys ja yksittäisen vahingon suuruus, joiden avulla saadaan riskimaksu. Murphy, Brockman ja Lee (2000) esittävät, miten GLM-malleja voidaan riskimaksujen laskemisen lisäksi käyttää vakuutusten uudistamisen, uusien vakuutusten alkamisen ja asiakkaan hintajoustopuolellisuuden mallintamiseen.

Ulrich Meyer (2002) teki tutkimuksen Euroopan Unionin jäsenvaltioissa toimivien vakuutusyhtiöiden yksityiskäyttöisten henkilöautojen moottoriajoneuvon vastuuvakuutuksessa

käyttämistä hinnoittelutekijöistä. Käytössä olevien tekijöiden lisäksi selvitettiin, miten paljon eri vakuutusentottajaryhmien maksamat vakuutusmaksut poikkeavat toisistaan. Tutkimukseen osallistuneita maita koskevat tiedot ovat vuodelta 1998. EU:n alueella yhdenmuikaistettiin vakuutusvalvontaa 1990-luvun puolivälissä, mikä johti useassa jäsenvaltiossa valvonnan lieventymiseen. Vakuutusyhtiöt voivat itse määrätä vakuutusmaksut, ja yhtiön käyttämät hinnoittelutekijät ovat suuressa määrin yhtiön valittavissa. Uusi tilanne johti useassa maassa kilpailun lisääntymiseen, mikä ilmeni maksutason alenemisena, yhtiöiden tulosten heikkenemisenä ja vakuutusmaksujen suuruuserojen kasvuna. Alkoi kilpailu ”hyvistä riskeistä”, joita yhtiöt yrittivät saada asiakkaikseen tarjoamalla matalia vakuutusmaksuja. Maksujen laskennassa sovellettiin uusia vakuutettavaan riskiin yhteydessä olevia tekijöitä. Kun jollakin yhtiöllä on vähemmän riskin luokittelutekijöitä kuin kilpailijoilla, siihen kohdistuu epäedullinen riskin valinta, ja sen osuus riskimielessä huonommista vakuutusentottajista kasvaa, ja yhtiön kannattavuus heikkenee.

Saksassa vakuutusyhtiöiden on lainsäädännön perusteella toimitettava vuosittain tilastoaineisto vakuutusyhtiöiden yhdistykselle (Gesamtverband der Deutschen Versicherungswirtschaft), joka määrittää yhteisestä tilastoaineistosta suosituksen riskimaksuista, joissa on käytetty yleisimpiä tariffitekijöitä. Yksittäiset yhtiöt voivat laskea omat maksunsa näiden GDV:n riskimaksujen perusteella. GDV:n riskimaksuissa on 5 tariffitekijää, jotka ovat auton tyyppi ja malli, alue, kuljettajan ammatti, vuosittain ajettava km-määrä ja autotalli. Auton tyyppi ja malli on tärkein tariffitekijä. Henkilöautot on luokiteltu 16:ta luokkaan autojen aiheuttamien vahinkojen perusteella. Moottoriajoneuvon vastuussa, laajassa moottoriajoneuvovakuutuksessa ja suppeammassa moottoriajoneuvovakuutuksessa on omat luokittelunsa. Henkilöautomalleja on 11 000. Luokittelussa käytetään automallin vahinkotilaston lisäksi mallin moottorin sylinteritilavuutta, moottorin tehoa, auton painoa, huippunopeutta ja polttoaineen lajia. Ylimmän ja alimman luokan suhde on 5:n ja 6:n välillä. GDV:n tariffitekijöiden lisäksi yhtiöt käyttävät muita tekijöitä. Auton kuljettajan luokitteluun käytetään seuraavia tekijöitä: ikä, sukupuoli, ammatti, siviilisääty ja nuoret lapset, asumismuoto omakotitalo, ajokortin ikä, vakuutusten lukumäärä, julkisen liikenteen käyttäjä, turvallisen ajon harjoituskurssin suoritus, asiakkuuden kesto ja autokerhon jäsenyys. Auton käytön luokittelutekijöitä ovat auton rekisteröintipaikan tai sijaintipaikan mukainen alue, yksityistai yrityskäyttö, vuosittain ajettava km-määrä, autotalli ja käyttö ensimmäisenä tai toisena autonä. Auto luokitellaan auton tyyppiin ja mallin, auton iän, polttoaineen kulutuksen ja autoon tehtyjen rakennemuutosten perusteella. Meyerin mukaan näiden ensisijaisten tariffitekijöiden perusteella muodostuvissa luokissa alimman ja ylimmän luokan suhde on 1:30, joka on hyvin korkea tutkittujen maiden joukossa. Yhtiöt soveltavat tekijöiden yhteisvaikutukselle ”kattoa”, jolla yhteisvaikutusta katkaistaan. Maksujen vaihtelua lisää GDV:n suosittama bonusasteikko, jossa on 29 luokkaa, ja jossa alimman ja ylimmän luokan suhde on 1:8.2. Vakuutusyhtiöt noudattavat joko tätä asteikkoa sellaisenaan tai ovat tehneet siihen pieniä muutoksia.

Isossa-Britanniassa valvova viranomainen ei ole koskaan vahvistanut tariffeja. Vakuutusyhtiöiden yhdistykselläkin on pienempi merkitys kuin useimmissa Euroopan maissa. Maksutaso on korvauksiin verrattuna alhainen ja yhtiöt muuttavat usein tariffirakenteita ja hintoja. Vakuutussopimus ei jatku automaattisesti seuraavalle vakuutuskaudelle toisin kuin muissa Euroopan maissa, joissa vakuutus jatkuu, ellei sitä ole irtisanottu. Tästä johtuen yli 50 % kuluttajista vaihtaa yhtiötä joka vuosi. Vuonna 1998 yhtiöt tekivät siihen mennessä kaikkien aikojen suurimmat tappiot. Tariffitekijöitä on käytössä vielä enemmän kuin Saksassa. Kuljettajan luokitteluun on käytössä muuten samat tekijät kuin Saksassa, mutta va-

kuutusten lukumäärää ei käytetä ja asumisaika Isossa-Britanniassa, sakko vakavasta liikenerikkomuksesta ja terveydentila ovat sellaisia tekijöitä, joita Saksassa ei käytetä. Auton käyttöä luokitellaan Saksassa käytettävien tekijöiden lisäksi käyttäjäryhmällä ja käyttöalueella. Auton luokittelussa on käytössä auton tyyppi ja malli, auton ikä, ja autoon tehdyt rakennemuutokset kuten Saksassakin. Uusi tai käytetty auto, ostohinta, moottorin sylinteritilavuus ja teho ja turvallisuuslaitteet (esimerkiksi hyväksyty varkaudenestolaite) ovat sellaisia tekijöitä, joita Saksassa ei käytetä. Myyntikanava ja vasemman puoleinen ohjaus ovat vielä lisätekijöinä. Halvimman luokan hinnan suhde kalleimman luokan hintaan on 1:31. Bonusasteikkoja on käytössä useita erilaisia. Niissä on vähän luokkia ja malusluokkia (maksu > 100 %) ei käytetä. Erilaisten asteikkojen alimman ja korkeimman maksun suhde on keskimäärin 1:3.

Suomessa kehitys on ollut erilainen kuin Saksassa ja Isossa-Britanniassa. Tariffitekijöitä henkilöautojen hinnoittelussa on käytössä selvästi vähemmän. Liikennevakuutuksessa on käytössä auton maksuluokka, alue ja joillakin yhtiöillä auton haltijan ikä ja sukupuoli sekä yksityis- tai yrityskäyttö. Useimmat yhtiöt määräävät maksuluokan sylinteritilavuuden ja merkki- ja mallikohtaisen tilaston perusteella. Autovakuutuksessa moni yhtiö käyttää lisäksi auton ikää tariffitekijänä. Lisäksi useimmat yhtiöt antavat keskittämisalennusta, jos vakuutuksenottajalla on usean eri vakuutuslajin vakuutuksia. Bonusasteikossa on 17 luokkaa, ja asteikon alimman maksun ja korkeimman maksun suhde on 1:3.3. Kaikki yhtiöt käyttävät henkilöautoille yhteistä bonusjärjestelmää vielä vuonna 2004, vaikka yhtiöt voivat nykyään valita oman bonusjärjestelmän.

Meyer löysi tutkimuksessaan 43 hinnoittelutekijää, joita käytettiin henkilöautojen moottoriajoneuvon vastuuvakuutuksen hinnoittelussa EU:n valtioissa vuonna 1998. Hän vertasi eri maiden maksuja viiden erilaisen asiakkaan maksun perusteella. Suomen maksu oli tässä vertailussa halvin. Hollannin vakuutusmaksu oli 4.5 kertainen Suomen maksuun verrattuna. Jokaisen viiden esimerkkiasiakkaan maksun vaihtelu eri maiden välillä oli huomattava. Belgiassa ensimmäisen asiakkaan maksuun suhde neljännen asiakkaan maksuun oli 1.0, kun Saksassa vastaava suhde oli 3.6. Meyer'in mukaan useassa maassa kilpailutilanne on edellä kuvatun kehityksen seurauksena häiriintynyt, koska eri yhtiöiden vakuutusmaksujen vertailu on vaikeaa, ja eri tekijöiden vaikutus maksuun jää kuluttajalle epäselväksi. Hän ehdottaa maksujen läpinäkyvyyden lisäämistä asiakkaalle annettavien tarkempien maksua koskevien tietojen avulla, korostaa maksujen oikeudenmukaisuutta ja maksun laskennan riskivastaavuutta.

Brockman ja Wright (1992) korostavat tilastollisen mallintamisen ja tilastoaineiston analyysin tärkeyttä riskimaksujen määrittämisessä. Käytettäessä yhden tariffitekijän mukaan tehtyjä tilastoja, tariffitekijän eri tasojen suhteellisesta kannattavuudesta voidaan saada harhaanjohtava käsitys. Tariffitekijöiden välillä voi olla tärkeitä keskinäisiä vaikutuksia (interaktioita), jotka yhden tekijän mukaisissa tilastoissa jäävät huomaamatta. Esimerkiksi miesten ja naisten suhteellinen vahinkokehitys vaihtelee henkilön iän mukaan. Vakuutuskannan jakautuminen eri tariffitekijöiden eri tasoille voi olla epätasainen. Esimerkiksi nuorten naisten kannattavuus voi tilastossa olla parempi kuin nuorten miesten. Aineiston tarkempi tutkiminen voi kuitenkin osoittaa, että nuoret naiset ajavat pienempiä autoja kuin nuoret miehet. Saman tyyppistä autoa ajavien nuorten kannattavuusero sukupuolten välillä voi olla huomattavasti pienempi, kuin pelkästään sukupuolen ja iän mukaisesta tilastosta voi päätellä. Sovittamalla malli tilastoaineistoon eri tekijät vaikuttavat samanaikaisesti, ja tällöin edellä kuvatut interaktiosta ja korrelaatiosta johtuvat ongelmatilanteet voidaan hallita. Tilastoaineisto luokitellaan tariffitekijöiden eri tasojen yhdistelmien mukaisesti. Jokaiselle yhdistelmien mukaiselle solulle lasketaan vahinko-

tiheys ja keskivahinko, joiden avulla saadaan solun riskimaksu. Jos esimerkiksi mallinnetaan henkilöautojen autovakuutuksen tariffia, jossa on laaja ja suppea turva, 7 maksuluokkaa, auton iällä 5 luokkaa, vakuutuksenottajan iällä 8 luokkaa, 6 alueluokkaa, 7 bonusluokkaa ja auton käytöllä 3 luokkaa, mallissa on yli 60 000 solua. Brockman ja Wright kehoittavat mallintamaan erikseen jokaisen vahinkotyyppin vahinkotiheyttä ja keskivahinkoa. Esimerkiksi moottoriajoneuvovakuutuksessa ajoneuvon vaunu-, varkaus- ja palovahingolle ja liikennevakuutuksen omaisuus- ja henkilövahingoille tehdään omat mallit. Brockman ja Wright sovittavat brittiläiseen henkilöautojen vaunuvahinkoaineistoon vahinkotiheyksille Poisson mallin ja keskivahingoille gamma mallin GLIM-ohjelmistoa käyttäen. Tariffitekijöinä ovat alue, vakuutuksenottajan ikä, auton ikä ja auton tyyppi. He osoittavat myös miten riskimaksuihin voidaan lisätä kuorimitukset, jotta saadaan maksutauluston vakuutusmaksu, ja antavat ohjeita, miten mallintamista voidaan hyödyntää markkinoiden kilpailutilanteessa.

Tämän harjoitustyön tarkoituksen on osoittaa, miten yleistettyjä lineaarisia malleja voidaan käyttää moottoriajoneuvovakuutusten hinnoittelussa. Luvussa 2 esitetään yleistettyjen lineaaristen mallien teoria. Luku 2 perustuu pääosin McCullagh'in ja Nelder'in (1989) ja Dobson'in (2001) kirjoihin sekä Mildenhall'in (1999) artikkeliin. Luvussa 3 tarkastellaan tariffin parametrien estimointia ensin usein käytetyillä 1960-luvulla kehitetyillä menetelmillä (linear bias methods) ja sen jälkeen yleistetyllä lineaarisella mallilla. Lisäksi osoitetaan edellä mainittujen menetelmien yhteys (Mildenhall 1999). Luvussa 4 esitetään autovakuutuksen hinnoittelusta esimerkkinä Baxter'in, Coutts'in ja Ross'in (1980) tekemä autovakuutuksen mallinnus, jota myös McCullagh ja Nelder (1989) ja Mildenhall'in (1999) tarkastelevat. Luvussa 5 sovitetaan yleistettyjä lineaarisia malleja Vahinkovakuutusyhtiö Pohjolan liikennevakuusaineistoon. Tietokonetyöt on tehty Watson Wyattin Pretium ohjelmistolla, joka on SAS'in genmod-ohjelmasta kehitetty yleistettyjen lineaaristen mallien teoriaan perustuva hinnoitteluohjelmisto.

Satunnaismuuttujia merkitään isoilla kirjaimilla, ja niiden toteutuneita arvoja pienillä kirjaimilla. Vektoreita merkitään pienillä lihavoiduilla kirjaimilla, ja matriiseja merkitään isoilla lihavoiduilla kirjaimilla. Kreikkalaiset kirjaimet viittaavat mallin parametreihin tai niiden estimaatteihin.

2. YLEISTETTY LINEAARINEN MALLI

2.1 Yleinen lineaarinen malli

Yleistetty lineaarinen malli on yleisen lineaarisen mallin laajennus. Tarkastellaan ensin yleistä lineaarista mallia

$$\mathbf{Y} = E(\mathbf{Y}) + \mathbf{e} \quad , \quad E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} \quad (2.1)$$

missä

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad , \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1^t \\ \mathbf{x}_2^t \\ \vdots \\ \mathbf{x}_n^t \end{bmatrix} \quad , \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \quad , \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \quad .$$

Tässä e_i :t ovat riippumattomia, samoin jakautuneita satunnaismuuttujia ja $e_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$. \mathbf{Y} :n komponentit ovat riippumattomia ja normaalisti jakautuneita. Y_i :tten odotusarvot μ_i vaihtelevat, mutta niillä on yhteinen varianssi σ^2 , joten

$$E(Y_i) = \mu_i = \mathbf{x}_i \boldsymbol{\beta} \quad , \quad Y_i \sim N(\mu_i, \sigma^2) \quad . \quad (2.2)$$

β_i :t ovat yleensä tuntemattomia parametreja, joiden arvot on estimoitava aineistosta. Kun havaintojen indeksi on i , mallin systemaattinen osa on

$$E(Y_i) = \mu_i = \sum_{j=1}^p x_{ij} \beta_j \quad ; \quad i = 1, \dots, n \quad , \quad (2.3)$$

missä x_{ij} on selittäjän j havainnon i arvo. Matriisimuodossa ($\boldsymbol{\mu}$ on $n \times 1$, \mathbf{X} on $n \times p$ ja $\boldsymbol{\beta}$ on $p \times 1$) saadaan

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \quad ,$$

missä \mathbf{X} on kerroinmatriisi ja $\boldsymbol{\beta}$ on parametrivektori. Lineaariset regressio-, varianssianalyysi- ja kovarianssianalyysimallit voidaan esittää tässä muodossa. Regressioanalyysissä selittävät muuttujat ovat jatkuvia, kerroinmatriisissa on 1:sistä muodostettu sarake, joka vastaa β :n regressiovakiokomponenttia, ja kerroinmatriisin muiden sarakkeiden komponentit ovat selittävien muuttujien havaittuja arvoja. Varianssianalyysissä verrataan selitettävän muuttujan keskiarvoja selittävien muuttujien arvotasojen avulla määritellyissä ryhmissä. Selittävät muuttujat eivät ole jatkuvia, ja ne voivat olla myös kvalitatiivisia. Kerroinmatriisin sa-

rakkeiden komponentit ovat yleensä 0:ia ja 1:iä, ja niiden avulla määritellään koesuunnitelun rakenne. Kovarianssianalyysissä osa selittävästä muuttujista on luokittelumuuttujia, ja loput ovat jatkuvia muuttujia. Regressioanalyysillä voitaisiin mallintaa esimerkiksi autovakuutuksen henkilöautojen kolarivahingon suuruutta. Selitettävä muuttuja on vahingosta korvattu euro-määrä, ja selittävät muuttujat ovat henkilöauton moottorin sylinteritilavuus ja moottorin teho. Mallinnettaessa vahinkovakuutuksen vahinkojen lukumääriä, vahingon suuruutta tai vahinkomenoa, normaalijakauma ja vakio varianssi eivät yleensä ole realistisia oletuksia. Vahinkojen lukumäärä ja suuruus sekä vahinkomeno ovat positiivisia arvoja saavia muuttujia ja kahden jälkimmäisen jakaumat ovat selvästi oikealle vinoja. Jos joidenkin Y_i -muuttujien odotusarvo on lähellä 0:aa, näiden varianssin voisi olettaa pienemmäksi kuin sellaisten Y_i -tten, joiden odotusarvo on selvästi suurempi. Lisäksi selittävien tekijöiden vaikutus on vakuutussovelluksissa usein perusteltavissa paremmin multiplikatiivisena kuin additiivisena. Edellisen kappaleen henkilöautoesimerkissä mallinnus voitaisiin tehdä korvaamalla Y_i -tten arvot $\ln Y_i$ -illä, jolloin selitettävän muuttujan jakauma olisi lähempänä normaalijakaumaa. Malli jäisi kuitenkin additiiviseksi.

2.2 Yleistetyn lineaarisen mallin rakenne

Yleinen lineaarinen malli on yleistetyn lineaarisen mallin erikoistapaus. Yleinen lineaarinen malli voidaan kuvata seuraavasti yleistettynä lineaarisena mallina :

1. Satunnainen osa: Y :n komponentit ovat riippumattomia ja normaalisti (μ_i, σ^2) -jakautuneita.
2. Systemaattinen osa: Selittäjät x_1, x_2, \dots, x_p muodostavat lineaarisen prediktorin η , joka on

$$\eta = \sum_{j=1}^p x_j \beta_j .$$

3. Yhteys satunnaisen ja systemaattisen osan välillä:

$$\mu = \eta$$

Yleistetyssä lineaarisessa mallissa määritellään

$$g(\mu_i) = \eta_i$$

missä $g(\cdot)$ on linkkifunktio. Yleisessä lineaarisessa mallissa on siis normaalijakauma osassa 1 ja identiteettifunktio kohdassa 3. Yleistetyssä lineaarisessa mallissa tehdään kaksi laajennusta: osan 1 jakauma kuuluu eksponentiaaliseen jakaumaperheeseen, ja kohdan 3 linkkifunktio on monotooninen ja differentioituva funktio. Linkkifunktiolla on tällöin olemassa käänteisfunktio. Normaalijakauma- ja vakiovariانسsioletuksesta on päästy eroon ja selittävien muuttujien vaikutus on nyt additiivinen muunnetulla asteikolla.

2.2.1. Yleistetyt lineaarisen mallin uskottavuusfunktiot

Oletetaan, että jokaisen Y :n komponentin jakauman tiheysfunktio on muotoa

$$f_Y(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}, \quad (2.4)$$

missä $a(\cdot)$, $b(\cdot)$ ja $c(\cdot)$ ovat tiettyjä funktioita. Jos ϕ on tunnettu, tämä on eksponenttijakaumaperheen jakauma, jonka kanoninen parametri on θ . Jos ϕ on tuntematon, kyseessä voi olla kaksiparametrinen eksponenttiperheen jakauma. Normaalijakaumalla on

$$\begin{aligned} f_Y &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-(y - \mu)^2 / 2\sigma^2\} \\ &= \exp\{(y\mu - \mu^2 / 2) / \sigma^2 - \frac{1}{2}(y^2 / \sigma^2 + \log(2\pi\sigma^2))\}, \end{aligned}$$

joten $\theta = \mu$, $\phi = \sigma^2$, ja

$$a(\phi) = \phi, \quad b(\theta) = \theta^2 / 2, \quad c(y, \phi) = -\frac{1}{2}\{y^2 / \sigma^2 + \log(2\pi\sigma^2)\}.$$

θ :n ja ϕ :n uskottavuusfunktio on $L(\theta, \phi; y)$. Uskottavuusfunktio $L(\theta, \phi; y)$:n algebrallinen lauseke on sama kuin tiheysfunktion $f_Y(y; \theta, \phi)$ lauseke. Tiheysfunktio $f_Y(y; \theta, \phi)$ on y :n funktio, kun parametrit θ ja ϕ on kiinnitetty, mutta uskottavuusfunktio $L(\theta, \phi; y)$ tulkitaan θ :n ja ϕ :n funktioksi, kun tietty havainto y on havaittu. Logaritminen uskottavuusfunktio on $l(\theta, \phi; y) = \log L(\theta, \phi; y)$. Logaritmisen uskottavuusfunktion maksimi saadaan samoilla parametrien arvoilla kuin uskottavuusfunktion maksimi, koska logaritmfunktio on monotooninen funktio. Usein on helpompi käyttää logaritmista uskottavuusfunktiota kuin itse uskottavuusfunktiota.

Y :n odotusarvo ja varianssi saadaan tunnetuista tuloksista (esim. Dobson 2001, sivu 48)

$$E\left(\frac{\partial l}{\partial \theta}\right) = 0 \quad (2.5)$$

$$E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left(\frac{\partial l}{\partial \theta}\right)^2 = 0. \quad (2.6)$$

(2.4):stä saadaan

$$\frac{\partial l(\theta; y)}{\partial \theta} = \{y - b'(\theta)\} / a(\phi) \quad (2.7)$$

ja

$$\frac{\partial^2 l}{\partial \theta^2} = -b''(\theta) / a(\phi) \quad (2.8)$$

(2.5):n ja (2.7):n perusteella

$$0 = E\left(\frac{\partial l}{\partial \theta}\right) = \{\mu - b'(\theta)\} / a(\phi) ,$$

joten

$$E(Y) = \mu = b'(\theta) .$$

Vastaavasti (2.6):sta, (2.7):sta ja (2.8):sta seuraa, että

$$0 = -\frac{b''(\theta)}{a(\phi)} + \frac{\text{var}(Y)}{a^2(\phi)} ,$$

joten

$$\text{Var}(Y) = b''(\theta)a(\phi) . \quad (2.9)$$

Y:n varianssi on siis kahden funktion tulo. Näistä $b''(\theta)$ on kanonisen parametrin ja täten myös odotusarvon funktio, ja sitä kutsutaan varianssifunktioksi. $a(\phi)$ on θ :sta riippumaton ja riippuu ainoastaan ϕ :stä. Varianssifunktiota pidetään μ :n funktiona, ja sitä merkitään $V(\mu)$:llä.

Funktio $a(\phi)$ on yleensä muotoa

$$a(\phi) = \phi / w ,$$

missä ϕ , jota merkitään myös σ^2 :lla ja kutsutaan dispersioparametriksi, on vakio kaikkien havaintojen suhteen. w on paino, joka vaihtelee havainnosta havaintoon. ϕ skaalaa havainnon i varianssin. Poisson jakaumalle $\phi = 1$. Muille jakaumille se lasketaan mallia sovitettaessa kaikkien havaintojen ja mallin antamien arvojen poikkeamien perusteella. Paino w pienentää sellaisen havainnon varianssia, jolla on suuri paino. Painot valitaan yleensä siten, että ne kasvavat havaintoon liittyvän aineiston kasvaessa. Normaalijakaumamallille, missä jokainen havainto on m :n riippumattoman luvun keskiarvo, saadaan

$$a(\phi) = \phi / m .$$

Yhteenveto tärkeimpien muotoa (2.4) olevien jakaumien ominaisuuksista on taulukossa 2.1.

Taulukko 2.1. Eksponenttijakaumaperheen jakaumien ominaisuuksia.

	Normaali	Poisson	Binomi	Gamma	Käänteinen Gaussinen
Merkintä	$N(\mu, \sigma^2)$	$P(\mu)$	$B(m, \pi)/m$	$G(\mu, \nu)$	$IG(\mu, \sigma^2)$
y:n määrittelyalue	$(-\infty, \infty)$	$0(1)\infty$	$\frac{0(1)m}{m}$	$(0, \infty)$	$(0, \infty)$
Dispersioparametri ϕ	σ^2	1	$\frac{1}{m}$	ν^{-1}	σ^2
Funktio $b(\theta)$	$\frac{\theta^2}{2}$	$\exp(\theta)$	$\log(1 + e^\theta)$	$-\log(-\theta)$	$-\frac{1}{(2\theta)^2}$
$c(y; \phi)$	$-\frac{1}{2} \left[\frac{y^2}{\phi} + \log(2\pi\phi) \right]$	$-\log y!$	$\log \binom{m}{my}$	$\nu \log(\nu y) - \log y - \log \Gamma(\nu)$	$-\frac{1}{2} \left[\log(2\pi\phi y^3) + \frac{1}{\phi y} \right]$
$\mu(\theta) = E(Y; \theta)$	θ	$\exp(\theta)$	$\frac{e^\theta}{(1 + e^\theta)}$	$-\frac{1}{\theta}$	$(-2\theta)^{-\frac{1}{2}}$
Kanoninen linkki $\theta(\mu)$	Identiteetti	log	logit	käänteisfunktio	$\frac{1}{\mu^2}$
Varianssifunktio $V(\mu)$	1	μ	$\mu(1 - \mu)$	μ^2	μ^3

Odotusarvoparametria on merkitty μ :llä muulloin, paitsi binomijakaumalle π :llä.

Gammajakauma on parametrisoitu siten, että sen varianssi on $\frac{\mu^2}{\nu}$.

2.2.2 Linkkifunktiot

Linkkifunktio yhdistää lineaarisen prediktorin η ja havainnon y odotusarvon μ . Yleisessä lineaarisessa mallissa lineaarinen prediktori ja odotusarvo ovat yhtä suuret, ja molemmat voivat saada minkä tahansa reaalilukuarvon. Linkkifunktio on tässä tapauksessa identiteettifunktio. Kun on kyseessä lukumäärät, ja jakauma on Poisson-jakauma, niin on oltava $\mu > 0$, joten identiteettifunktio ei tule kyseeseen. Kun käytetään logaritmfunktiota linkkifunktiona, prediktorin muuttujien vaikutus on multiplikatiivinen, sillä tällöin

$$\mu_i = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) = \exp(\beta_1 x_{i1}) \cdot \exp(\beta_2 x_{i2}) \cdot \dots \cdot \exp(\beta_p x_{ip})$$

Binomijakauman tapauksessa $0 < \mu < 1$, ja linkin pitää kuvata väli $(0,1)$ koko reaalisuoralle. Yleisimmät linkkifunktiot ovat:

1. logit

$$\eta = \log\left\{\frac{\mu}{1-\mu}\right\};$$

$$2. \text{ probit} \quad \eta = \Phi^{-1}(\mu) \quad ;$$

missä $\Phi(\cdot)$ on normaalijakauman kertymäfunktio.

$$3. \text{ komplementaarinen log-log} \quad \eta = \log\{-\log(1-\mu)\};$$

$$4. \text{ potenssiperheen linkit} \quad \eta = (\mu^\lambda - 1) / \lambda \quad ,$$

missä $\eta = \log \mu$, kun $\lambda \rightarrow 0$,

tai

$$\eta = \begin{cases} \mu^\lambda & , \lambda \neq 0 \quad , \\ \log \mu & , \lambda = 0 \quad . \end{cases}$$

Molemmissa tapauksissa laskettaessa funktion arvoa λ :n arvolla 0 tarvitaan erikoistoimenpiteitä.

Jokaisella taulukon 2.1 jakaumalla on tietty linkkifunktio, jolle on olemassa tyhjentävä tunnusluku (katso esim. Lokki Olli 1979) , jonka dimensio on sama kuin β :n. Tällöin $\theta = \eta$, ja kyseessä on kanoninen linkki. Edellä (2.4):ssä määritelty θ on kanoninen parametri. Taulukon 2.1. jakaumien kanoniset linkit ovat

Normaali	$\eta = \mu$,
Poisson	$\eta = \log \mu$,
Binomi	$\eta = \log\{\pi/(1-\pi)\}$,
Gamma	$\eta = \mu^{-1}$,
Käänteinen normaali	$\eta = \mu^{-2}$.

Kanonisen linkin tyhjentävä tunnusluku vektorimuodossa on $\mathbf{X}^T \mathbf{Y}$, jonka alkiot ovat $\sum_i x_{ij} Y_i$, $j = 1, \dots, p$. Kanonisella linkillä saadaan mallille halutut tilastolliset ominaisuudet, mutta linkin ei tarvitse olla kanoninen.

2.3 Mallin hyvyyden mittaaminen

2.3.1 Devianssi

Mallin sovituksen hyvyyttä voidaan arvioida vertaamalla mallia ns. täydelliseen eli kyllästettyyn malliin. Täydellisessä mallissa on yhtä monta parametria kuin havaintoja on. Olkoon β_{\max} täydellisen mallin parametrivektori, ja \mathbf{b}_{\max} β_{\max} :n suurimman uskottavuuden estimaatti. Täydellisen mallin uskottavuusfunktion arvo arvo laskettuna \mathbf{b}_{\max} :ssa on $L(\mathbf{b}_{\max}; \mathbf{y})$. $L(\mathbf{b}_{\max}; \mathbf{y})$ on suurempi kuin mikä muu tahansa näille havainnoille samalla

jakaumalla ja linkkifunktiolla laskettu uskottavuusfunktion arvo, sillä sitä vastaava malli kuvaa aineistoa parhaiten. Olkoon $L(\mathbf{b}; \mathbf{y})$ tarkasteltavan mallin uskottavuusfunktion suurin arvo. Uskottavuussuhteen

$$\lambda = \frac{L(\mathbf{b}_{\max}; \mathbf{y})}{L(\mathbf{b}; \mathbf{y})}$$

avulla voidaan arvioida tarkasteltavan mallin hyvyyttä. Käytännössä lasketaan

$$2 \log \lambda = 2 [l(\mathbf{b}_{\max}; \mathbf{y}) - l(\mathbf{b}; \mathbf{y})] .$$

Suuri arvo merkitsee, että kyseessä oleva malli ei kuvaa aineistoa hyvin. $2 \log \lambda$:n jakaumaa voidaan approksimoida $\chi^2(n-p, \nu)$ jakaumalla, missä n on havaintojen lukumäärä, p on estimoitavien parametrien lukumäärä ja ν on χ^2 -jakauman epäkeskisyyssparametri. (Dobson A. J. 2001 kappale 5.6). Normaalijakaumamalleilla $2 \log \lambda$ on tarkasti χ^2 -jakautunut. Jos malli kuvaa aineistoa hyvin, parametri ν on lähellä 0:aa.

Devianssi määritellään yleensä käyttämällä β :n sijasta β :ta vastaavaa μ :tä. Skaalattu devianssi on

$$D^*(\mathbf{y}; \mu) = 2 [l(\mathbf{y}; \mathbf{y}) - l(\mu; \mathbf{y})] . \quad (2.10)$$

Normaalisti jakautuneelle havainnolle, jonka varianssi on σ^2 , saadaan

$$\begin{aligned} l(\mu; \mathbf{y}) &= -\frac{1}{2} \log(2\pi\sigma^2) - (y - \mu)^2 / (2\sigma^2) , \\ l(\mathbf{y}; \mathbf{y}) &= -\frac{1}{2} \log(2\pi\sigma^2) , \\ D^*(\mathbf{y}; \mu) &= (y - \mu)^2 / \sigma^2 , \end{aligned}$$

joten devianssin minimi on sama kuin pienin neliösumma.

Oletetaan, että hajontaparametrilla ϕ on kiinteä arvo, ja että $a_i(\phi) = \phi / w_i$. Olkoon $l(\hat{\mu}; \mathbf{y})$ β :n suhteen maksimoitu uskottavuus, $l(\mathbf{y}; \mathbf{y})$ n :n parametrin täydellä mallilla saavutettavissa oleva suurin uskottavuus ja $\hat{\theta} = \theta(\hat{\mu})$ ja $\tilde{\theta} = \theta(\mathbf{y})$ kyseessä olevien kahden mallin kanoniset parametrit. Nyt saadaan

$$D^*(\mathbf{y}; \hat{\mu}) = \sum 2w_i \{y_i (\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\} / \phi = D(\mathbf{y}; \hat{\mu}) / \phi , \quad (2.11)$$

missä $D(\mathbf{y}; \hat{\mu})$ on tarkasteltavana olevan mallin devianssi.

Taulukon 2.1 jakaumien devianssit ovat (summaus $i = 1, \dots, n$)

Normaali	$\sum (y - \hat{\mu})^2$
Poisson	$2 \sum \{y \log(y / \hat{\mu}) - (y - \hat{\mu})\}$
Binomi	$2 \sum \{y \log(y / \hat{\mu}) + (m - y) \log[(m - y) / (m - \hat{\mu})]\}$
Gamma	$2 \sum \{-\log(y / \hat{\mu}) + (y - \hat{\mu}) / \hat{\mu}\}$
Käänteinen normaali	$\sum (y - \hat{\mu})^2 / (\hat{\mu}^2 y)$

Poisson- ja gammajakauman lausekkeen toinen termi jätetään usein pois. Jos sovitetussa mallissa on vakiotermi, summattaessa toisen termin lauseke saadaan 0, minkä vuoksi termin jättäminen pois on perusteltua.

Mallin hyvyttä voidaan mitata myös yleistetyllä Pearson X^2 statistikalla, joka on

$$X^2 = \sum (y - \hat{\mu})^2 / V(\hat{\mu}) \quad , \quad (2.12)$$

missä $V(\hat{\mu})$ on kyseessä olevan jakauman varianssifunktio. Normaalijakauman malleilla yleistetty Pearson X^2 on χ^2 -jakautunut.

Tarkastellaan kahta sisäkkäistä mallia. Malleilla on sama jakauma- ja linkkifunktio, mutta yksinkertaisen mallin M_0 lineaarinen prediktori on yleisemmän mallin M_1 prediktorin erikoistapaus. Nollahypoteesi

$$H_0 : \beta = \beta_0 = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{bmatrix}$$

vastaa mallia M_0 , ja yleisempi hypoteesi

$$H_1 : \beta = \beta_1 = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

vastaa mallia M_1 , ja $q < p < n$. Voimme testata H_0 :aa ja H_1 :tä käyttäen skaalattujen devianssien erotusta

$$\begin{aligned}\Delta D^* &= D_0^* - D_1^* = 2 [l(\mathbf{y}; \mathbf{y}) - l(\mu_0; \mathbf{y})] - 2 [l(\mathbf{y}; \mathbf{y}) - l(\mu_1; \mathbf{y})] \\ &= 2 [l(\mu_1; \mathbf{y}) - l(\mu_0; \mathbf{y})] .\end{aligned}$$

Jos molemmat mallit kuvaavat aineistoa hyvin, niin $D_0^* \sim \chi^2(n-q)$, $D_1^* \sim \chi^2(n-p)$ ja $\Delta D^* \sim \chi^2(p-q)$ olettaen, että tietyt riippumattomuusehdot ovat voimassa. Jos ΔD^* :n arvo on χ^2 -jakauman mukainen, valitaan H_0 :aa vastaava malli M_0 , koska M_0 on yksinkertaisempi. ΔD^* :n jakaumaa voidaan approksimoida paremmin χ^2 -jakaumalla kuin pelkän devianssin jakaumaa.

2.3.2 Residuaalit

Residuaaleilla voidaan tutkia linkkifunktion, varianssifunktion ja lineaarisen prediktorin selittäjien valintojen onnistumista. Residuaalien avulla voidaan myös löytää muista havainnoista poikkeavat havainnot, jotka tarvitsevat lisätarkastelua. Useimmiten käytetyt residuaalit ovat Pearson, Anscombe ja devianssi residuaalit. Pearson residuaali on

$$r_p = \frac{y - \mu}{\sqrt{V(\mu)}} .$$

Tässä y :n ja μ :n erotus on skaalattu jakamalla se Y :n keskihajonnalla. Nyt saadaan

$$\sum r_p^2 = X^2 ,$$

missä X^2 on määritelty kohdassa (2.12).

Pearson residuaalin heikkous on se, että r_p :n jakauma on usein huomattavan vino. Anscombe residuaalissa käytetään y :n sijasta funktiota $A(y)$, missä $A(\cdot)$ on valittu siten, että $A(Y)$:n jakauma on ”niin normaali kuin mahdollista”. Yleistetyn lineaarisen mallin tapauksessa voidaan valita

$$A(\cdot) = \int \frac{d\mu}{V^{\frac{1}{3}}(\mu)} .$$

Poisson jakaumalle saadaan

$$\int \frac{d\mu}{\mu^{\frac{1}{3}}} = \frac{3}{2} \mu^{\frac{2}{3}} ,$$

joten residuaalin voi perustua lausekkeeseen $y^{\frac{2}{3}} - \mu^{\frac{2}{3}}$. Kun edellä oleva lauseke vielä jaetaan $A(Y)$:n hajonnalla, Poisson jakauman Anscombe residuaaliksi saadaan

$$r_A = \frac{\frac{3}{2}(y^{\frac{2}{3}} - \mu^{\frac{2}{3}})}{\mu^{\frac{1}{6}}} .$$

Gammajakaumalle ja käänteiselle normaalijakaumalle saadaan vastaavasti

$$r_A = \frac{3(y^{\frac{1}{3}} - \mu^{\frac{1}{3}})}{\mu^{\frac{1}{3}}} .$$

ja

$$r_A = \frac{(\log y - \log \mu)}{\mu^{\frac{1}{2}}} .$$

Käytettäessä devianssia mallin hyvyyden mittana, jokainen havainto vaikuttaa d_i :n verran tähän mittaan, ja saadaan $\sum d_i = D$. Kun määritellään

$$r_D = \text{sign}(y - \mu)\sqrt{d_i} ,$$

niin r_D kasvaa $(y_i - \mu_i)$:n mukaisesti, ja $\sum r_D^2 = D$. Poisson jakaumalle saadaan

$$r_D = \text{sign}(y - \mu)\{2(y \log \frac{y}{\mu}) - y + \mu\}^{\frac{1}{2}} .$$

Vaikka muiden paitsi normaalijakauman Anscombe ja devianssi residuaalien lausekkeet näyttävät hyvin erilaisilta, niin niiden saamat arvot suurella aineistolla ovat lähellä toisiaan. (katso P. McCullagh and J.A. Nelder luku 2.4.2).

2.4 Yleistetyn lineaarisen mallin kertoimien ratkaisu

Kun Y_1, \dots, Y_n ovat riippumattomia eksponenttijakaumaperheen satunnaismuuttujia, niin niiden logaritminen uskottavuus on

$$l = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i) \quad (2.13)$$

Haluamme estimoida parametrit β , kun $E(Y_i) = \mu_i$ ja $g(\mu_i) = \sum_{r=1}^p x_{ir} \beta_r$, $i = 1, \dots, n$.

Estimointi tehdään suurimman uskottavuuden menetelmällä. Välttämättömät ehdot sille, että \mathbf{b} on β :n suurimman uskottavuuden estimaatti, ovat

$$\left(\frac{\partial l}{\partial \beta} \right)_{\beta=\mathbf{b}} = \mathbf{0} \quad (2.14)$$

ja

$$\left(\frac{\partial^2 l}{\partial \beta \partial \beta^t} \right)_{\beta=\mathbf{b}} \text{ on negatiividefiniitti matriisi.}$$

Jos on useita β :n arvoja, jotka toteuttavat yllä olevat ehdot, valitaan se β , jolla logaritmissen uskottavuusfunktio l :n arvo on suurin. Jos $\left(\frac{\partial^2 l}{\partial \beta \partial \beta^t} \right)$ on negatiividefiniitti kaikilla β :n

arvoilla, niin silloin logaritminen uskottavuusfunktio l on konkaavi, ja sen suurin arvo on yksikäsitteinen. Useimmilla yleistetyillä lineaarisilla malleilla on yksikäsitteinen maksimi-arvo, ja se saadaan yhtälöiden (2.14) ratkaisuna.

β :n arvot voitaisiin ratkaista myös pienimmän neliösumman menetelmällä. Suurimman uskottavuuden ja pienimmän neliösumman menetelmällä ei kuitenkaan saada samoja β :n estimaatteja muille kuin normaalijakaumamallille, jossa linkkifunktio on identiteettifunktio. Kumpikaan menetelmä ei johda analyttisesti ratkaistavissa oleviin yhtälöihin muille kuin normaali-identiteetti-mallille, vaan yhtälöt ratkaistaan numeerisilla menetelmillä. Suurimman uskottavuuden avulla saadaan β :n arvojen testaamiseksi, ja luottamusvälien muodostamiseksi yhtenäinen teoria, ja uskottavuusyhtälöiden ratkaisemiseksi on olemassa hyvin toimiva menetelmä.

Suurimman uskottavuuden mukaiset yhtälöt ovat

$$\frac{\partial l}{\partial \beta_j} = u_j = \sum_{i=1}^n \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j} = 0 \quad (2.15)$$

Nyt

$$\frac{\partial l_i}{\partial \theta_i} = \frac{[y_i - b'(\theta_i)]}{a(\phi_i)} = \frac{[y_i - \mu_i]}{a(\phi_i)},$$

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{\frac{\partial \mu_i}{\partial \theta_i}} = \frac{1}{b''(\theta_i)} = \frac{a(\phi_i)}{\text{Var}(Y_i)},$$

$$\frac{\partial \mu_i}{\partial \beta_i} = \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_i} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij}.$$

Saadaan

$$u_j = \sum_{i=1}^n \left[\frac{(y_i - \mu_i)}{\text{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right]. \quad (2.16)$$

Nyt u_j :den kovarianssimatriisilla on alkiot

$$I_{jk} = E[u_j u_k],$$

jotka muodostavat informaatiomatriisin I . Käyttämällä (2.16):ta saadaan

$$\begin{aligned} I_{jk} &= E \left\{ \sum_{i=1}^n \left[\frac{(Y_i - \mu_i)}{\text{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right] \sum_{l=1}^n \left[\frac{(Y_l - \mu_l)}{\text{var}(Y_l)} x_{lk} \left(\frac{\partial \mu_l}{\partial \eta_l} \right) \right] \right\} \\ &= \sum_{i=1}^n \frac{E[(Y_i - \mu_i)^2] x_{ij} x_{ik} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2}{[\text{var}(Y_i)]^2}, \end{aligned}$$

koska $E[(Y_i - \mu_i)(Y_l - \mu_l)] = 0$, kun $i \neq l$, sillä Y_i :t ovat riippumattomia. Siis

$$I_{jk} = \sum_{i=1}^n \frac{x_{ij} x_{ik}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2. \quad (2.17)$$

Yhtälöiden (2.15) mukaisesti ratkaisemme vektoryhtälöä $\mathbf{u}(\beta) = \mathbf{0}$, missä

$$\mathbf{u}(\beta) = \mathbf{u} = \left(\frac{\partial l}{\partial \beta_1}, \dots, \frac{\partial l}{\partial \beta_p} \right)^t.$$

Käyttämällä Newton-Raphson menetelmää saadaan

$$\beta_m = \beta_{m-1} - \left(\frac{\partial \mathbf{u}}{\partial \beta} \right)^{-1} \mathbf{u} .$$

Termi $\frac{\partial \mathbf{u}}{\partial \beta}$ on Hessin matriisi. Se on yleensä satunnainen suure. Fisher'in scoring menetelmässä käytetään Hessin matriisin sijalla Hessin matriisin odotusarvoa. Varsin lievien säännöllisyysoletusten ollessa voimassa on

$$E \left(\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right) = -E \left(\frac{\partial l}{\partial \beta_j \partial \beta_k} \right) = -I_{jk} , \quad j, k = 1, \dots, p . \quad (2.18)$$

(katso esim Greene 2003 sivut 474-475).

Ratkaistaan siis yhtälö (2.15) iteratiivisesti :

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} + [\mathbf{I}^{(m-1)}]^{-1} \mathbf{u}^{(m-1)} , \quad (2.19)$$

missä $\mathbf{b}^{(m)}$ on parametrien β_1, \dots, β_p estimaattivektori m :ssä iteraatiossa . Yhtälössä (2.19) $[\mathbf{I}^{(m-1)}]^{-1}$ on informaatiomatriisin käänteismatriisi, ja $\mathbf{u}^{(m-1)}$ on vektori, jonka alkiot ovat (2.16):ta mukaisia. Molempien arvot on laskettu $\mathbf{b}^{(m-1)}$:ssä. Kertomalla (2.19) $[\mathbf{I}^{(m-1)}]$:lla saadaan

$$\mathbf{I}^{(m-1)} \mathbf{b}^{(m)} = \mathbf{I}^{(m-1)} \mathbf{b}^{(m-1)} + \mathbf{u}^{(m-1)} . \quad (2.20)$$

(2.17):sta seuraa, että \mathbf{I} :n voi esittää muodossa

$$\mathbf{I} = \mathbf{X}^t \tilde{\mathbf{W}} \mathbf{X} ,$$

missä $\tilde{\mathbf{W}}$ on $n \times n$ diagonaalimatriisi, jonka alkiot ovat

$$\tilde{w}_{ii} = \frac{1}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 .$$

(2.20):n oikea puoli voidaan (2.16) ja (2.17) avulla esittää vektorina, jonka alkiot ovat

$$\sum_{k=1}^p \sum_{i=1}^n \frac{x_{ij} x_{ik}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 b_k^{(m-1)} + \sum_{i=1}^n \frac{(y_i - \mu_i) x_{ij}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)$$

laskettuna $\mathbf{b}^{(m-1)}$:ssä. (2.20) oikea puoli voidaan täten kirjoittaa muodossa

$$\mathbf{X}^t \tilde{\mathbf{W}} \mathbf{z},$$

missä \mathbf{z} :lla on alkiot

$$z_i = \sum_{k=1}^p x_{ik} b_k^{(m-1)} + (y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right),$$

missä μ_i ja $\frac{\partial \eta_i}{\partial \mu_i}$ on laskettu $\mathbf{b}^{(m-1)}$:ssä.

Iteratiivinen yhtälö (2.20) on voidaan siis esittää muodossa

$$\mathbf{X}^t \tilde{\mathbf{W}} \mathbf{X} \mathbf{b}^{(m)} = \mathbf{X}^t \tilde{\mathbf{W}} \mathbf{z} . \quad (2.21)$$

Tämä on saman muotoinen kuin painotetulla pienimmän neliösumman menetelmällä saadut lineaarisen mallin normaaliyhtälöt. Tämä on kuitenkin ratkaistava iteratiivisesti, koska yleensä \mathbf{z} ja $\tilde{\mathbf{W}}$ riippuvat \mathbf{b} :stä. Yleistetyn lineaarisen mallin suurimman uskottavuuden estimaattorit saadaan siis painotetulla pienimmän neliösumman menetelmällä. Yleistetyn lineaarisen mallin sisältävillä tilastollisilla ohjelmistopaketeilla on tehokas (2.21):een perustuva algoritmi. Ne aloittavat käyttämällä alkuarvoa $\mathbf{b}^{(0)}$, jonka avulla lasketaan \mathbf{z} ja $\tilde{\mathbf{W}}$. Sen jälkeen ratkaistaan (2.21), ja saadaan $\mathbf{b}^{(1)}$, jota käytetään edelleen \mathbf{z} :n ja $\tilde{\mathbf{W}}$:n uusien arvojen laskentaan. Kun $\mathbf{b}^{(m-1)}$:n ja $\mathbf{b}^{(m)}$:n etäisyys on riittävän pieni, $\mathbf{b}^{(m)}$ hyväksytään suurimman uskottavuuden estimaatiksi.

Suurimman uskottavuuden mukaiset estimaatit ovat asymptoottisesti normaalisti jakautuneita ja asymptoottisesti tehokkaita, joten

$$\mathbf{b} \sim \mathbf{N}(\beta, \mathbf{I}^{-1}) . \quad (2.22)$$

β :n komponenteille voi laskea luottamusvälit (2.22):n avulla. Tulos (2.22) on kuitenkin voimassa vain asymptoottisesti, ja luottamusvälejä muodostettaessa joudutaan β korvaamaan estimaatillaan \mathbf{b} .

Luottamusvälit voidaan muodostaa myös uskottavuussuhteen avulla. Oletetaan, että β :n komponentille β_j halutaan luottamusvälit. β_j :n profiiliuskottavuusfunktio on

$$l^*(\beta_j) = \max_{\hat{\beta}} l(\beta) .$$

missä $\hat{\beta}$:n j :s komponentti on kiinnitetty β_j :ksi. Kun β :n logaritmisien uskottavuusfunktion arvo on $l(\mathbf{b})$, missä \mathbf{b} on suurimman uskottavuuden estimaatti, $2[l(\mathbf{b}) - l^*(\beta_j)]$ on χ^2 -jakautunut yhdellä vapausasteella. β_j :n $(1 - \alpha) \cdot 100$ %:n luottamusvälit ovat

$$\{ \beta_j : l^*(\beta_j) = l(\mathbf{b}) - 0.5 \cdot \chi_{1-\alpha,1}^2 \} ,$$

missä $\chi_{1-\alpha,1}^2$ on $\chi^2(1)$ -jakauman $(1-\alpha) \cdot 100\%$:n fraktiili.

2.5 Kvasiuskottavuus

Kun halutaan varmistua, että havainnot noudattavat (2.4):n mukaista jakaumaa, on tunnettava havainnot synnyttävä satunnaismekanismi yksityiskohtaisesti. Usein kuitenkin on helpompi saada selville havainnoista esimerkiksi

1. miten eri tekijät vaikuttavat vasteen keskiarvoon tai mediaaniin
2. miten vasteen vaihtelu riippuu vasteen keskiarvosta
3. ovatko havainnot tilastollisesti riippumattomia
4. onko vasteen jakauma positiivisesti tai negatiivisesti vino tai symmetrinen.

Luopumalla (2.13):ta mukaisesta uskottavuusoletuksesta ja sen alkuperänä olevasta jakaumaoletuksesta (2.4), yhtälöiden (2.15) voimassaoloaluetta voidaan laajentaa.

Määritellään ensin kvasiuskottavuusfunktio ja sen avulla kvasidevianssifunktio. Oletetaan, että vastevektorin \mathbf{Y} alkiot ovat riippumattomia, ja μ on \mathbf{Y} :n odotusarvo. μ :n havaintoon \mathbf{y} perustuva logaritminen kvasiuskottavuus on

$$q(\mathbf{y}; \mu) = \sum_{i=1}^n q_i = \sum_{i=1}^n w_i \int_{y_i}^{\mu_i} \frac{y_i - s}{\phi V(s)} ds . \quad (2.23)$$

Varianssifunktio $V(s)$ aidosti positiivinen, ja w_i ja ϕ ovat havainnon i paino ja dispersioparametri kuten luvussa 2.2.1. $V(s)$ ja ϕ määräävät havaintojen varianssin. Nyt μ :n ja \mathbf{y} :n kvasidevianssi on

$$D(\mathbf{y}; \mu) = \sum_{i=1}^n d_i = \sum_{i=1}^n 2 w_i \int_{\mu_i}^{y_i} \frac{y_i - s}{V(s)} ds = -2\phi q(\mathbf{y}; \mu) . \quad (2.24)$$

Tämä on aidosti positiivinen funktio, paitsi jos $\mathbf{y} = \mu$. Merkitään linkkifunktion g käänteisfunktioita h :lla. Nyt $\mu_i = h(\mathbf{x}_i; \beta)$, missä $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$. Devianssin minimi vektorin β suhteen saadaan ratkaisemalla p :n yhtälön ryhmä

$$\frac{\partial D}{\partial \beta_j} = 0 \quad , \quad j = 1, \dots, p . \quad (2.25)$$

Saadaan

$$\frac{\partial D}{\partial \beta_j} = \sum_i \frac{\partial d_i}{\partial \beta_j} = \sum_i \frac{\partial d_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j} = -2 \sum_i \frac{w_i (y_i - \mu_i)}{V(\mu_i)} h'(\mathbf{x}_i \beta) x_{ij} ,$$

joten yhtälöryhmä (2.25) on

$$\sum_i \frac{w_i (y_i - \mu_i)}{V(\mu_i)} h'(\mathbf{x}_i \beta) x_{ij} = 0 \quad , j = 1, \dots, p . \quad (2.26)$$

Nämä ovat samat yhtälöt kuin (2.15). Tästä saatavilla kvasiuskottavuusparametreilla on samat asymptoottiset ominaisuudet kuin (2.4):n määrittelemien jakaumien suurimman uskottavuuden parametreilla. Olennaista on, että kvasiuskottavuuden lausekkeessa (2.23) ja yhtälössä (2.26) riittää tuntea Y_i :den jakaumien ensimmäinen ja toinen momentti. Uskottavuusfunktio-oletuksesta (2.13) voidaan luopua, ja laajentaa käytettävissä olevien jakaumien joukkoa.

Määritellään eksponentiaaliset jakaumat kvasidevianssin (2.24) avulla. Eksponentiaalisen jakaumaperheen jakaumien tiheysfunktio on

$$f(y; \mu, \phi) = c(y, \phi) \exp\left(-\frac{1}{2\phi} d(y; \mu)\right) , \quad (2.27)$$

missä siis

$$d(y; \mu) = 2w \int_{\mu}^y \frac{(y-t)}{V(t)} dt .$$

Yhtälöitä (2.5) ja (2.6) vastaavien yhtälöiden avulla voidaan osoittaa, että

$$E(Y) = \mu \quad (2.28)$$

ja

$$\text{Var}(Y) = \frac{\phi}{w} V(\mu) . \quad (2.29)$$

Nämä ovat samat kuin luvun (2.2.1) odotusarvo ja varianssi (2.4):n mukaisille jakaumille. Luvussa (2.2.1) $b''(\theta)$:ta kutsuttiin varianssifunktioksi, ja sitä merkittiin $V(\mu)$:lla. Verrataan tiheysfunktion määritelmiä (2.4) ja (2.27). Oletetaan, että $a(\phi) = \phi/w$. Jotta (2.4) ja (2.27) määrittelisivät saman jakaumaperheen, on oltava voimassa

$$y\theta - b(\theta) = -\frac{1}{2} d(y; \mu) = \int_y^\mu \frac{(y-t)}{V(t)} dt$$

Derivoimalla θ :n suhteen saadaan

$$y - b'(\theta) = \frac{y - \mu}{V(\mu)} \frac{d\mu}{d\theta}$$

Ottamalla odotusarvo saadaan $\mu = b'(\theta)$. Sijoittamalla tämä yllä olevaan yhtälöön saadaan

$$V(\mu) = b''(\theta).$$

Funktio b toteuttaa siis differentiaaliyhtälön

$$V(b'(\theta)) = b''(\theta) \quad (2.30)$$

Yhtälön (2.30) avulla pystytään määrittämään b , kun V tunnetaan. Johdetaan varianssi-funktiota $V(\mu) = \mu^2$ vastaava jakauma. Yhtälöstä (2.30) saadaan $(b'(\theta))^2 = b''(\theta)$, joten

$$\mu = b'(\theta) = -\frac{1}{\theta},$$

ja

$$b(\theta) = -\log(-\theta).$$

Sijoitetaan tämä (2.4):een. Saadaan

$$f(y; \theta, \phi) = \exp\{[y\theta + \log(-\theta)]w/\phi + c(y, \phi)\}.$$

Merkitään $\nu = \frac{w}{\phi}$, valitaan $c(y, \phi) = \nu \log \nu y - \log y - \log \Gamma(\nu)$ ja sijoitetaan $\theta = -\frac{1}{\mu}$ yllä olevaan, jolloin saadaan

$$\begin{aligned} f(y; \mu, \nu) &= \exp\{[y(-1/\mu) - \log \mu] \nu + \nu \log \nu y - \log y - \log \Gamma(\nu)\} \\ &= \exp[-\log \Gamma(\nu) + \nu(\log \nu y - \log \mu) - (\nu y / \mu) - \log y] \\ &= \exp\{\log[1/\Gamma(\nu)] + \nu \log(\nu y / \mu) - (\nu y / \mu) + \log(1/y)\} \\ &= \left(\frac{\nu}{\mu}\right)^\nu \frac{1}{\Gamma(\nu)} y^{\nu-1} \exp\left(-\frac{\nu y}{\mu}\right), \end{aligned}$$

joka on gammajakauman tiheysfunktio. Muokkaamalla tiheysfunktioita saadaan

$$\begin{aligned}
 f(y; \mu, \nu) &= \frac{\nu^\nu y^{-1}}{\Gamma(\nu)} \exp[\nu \log(y/\mu) - \nu y/\mu] \\
 &= \frac{\nu^\nu y^{-1} e^{-\nu}}{\Gamma(\nu)} \exp\left\{-\frac{\nu}{2} 2 \left[\left(\frac{y-\mu}{\mu} \right) - \log\left(\frac{y}{\mu} \right) \right]\right\} \\
 &= \frac{\nu^\nu y^{-1} e^{-\nu}}{\Gamma(\nu)} \exp\left\{\left\{-\frac{\nu}{2} 2 \int_{\mu}^y \frac{y-t}{t^2} dt\right\}\right\}
 \end{aligned}$$

Tämä on muotoa (2.27) oleva tiheysfunktio, missä varianssifunktio on $V(\mu) = \mu^2$. Siis varianssifunktiota $V(\mu) = \mu^2$ vastaa gammajakauma. Tavallisimpia varianssifunktioita vastaavat kvasiuskottavuudet ja jakaumat on esitetty seuraavassa taulukossa.

Taulukko 2.2. Eri varianssifunktioita vastaavat kvasiuskottavuudet ja jakaumat.

Varianssifunktio $V(\mu)$	Kvasiuskottavuus $q(\mu; y)$	Kanoninen parametri θ	Jakauma	Rajoitukset
1	$\frac{-(y-\mu)^2}{2}$	μ	Normaali	
μ	$y \log \mu - \mu$	$\log \mu$	Poisson	$\mu > 0, y \geq 0$
μ^2	$-\frac{y}{\mu} - \log \mu$	$-\frac{1}{\mu}$	Gamma	$\mu > 0, y > 0$
μ^3	$-\frac{y}{2\mu^2} + \frac{1}{\mu}$	$-\frac{1}{2\mu^2}$	Käänteinen normaali	$\mu > 0, y > 0$
μ^ζ	$\mu^{-\zeta} \left(\frac{\mu y}{1-\zeta} - \frac{\mu^2}{2-\zeta} \right)$	$\frac{1}{(1-\zeta) \mu^{\zeta-1}}$	Useita jakaumia mm Tweedie	$\mu > 0, \zeta \neq 0, 1, 2$
$\mu(1-\mu)$	$y \log\left(\frac{\mu}{1-\mu}\right) + \log(1-\mu)$	$\log\left(\frac{\mu}{1-\mu}\right)$	Binomi / m	$0 < \mu < 1, 0 \leq y \leq 1$
$\mu^2(1-\mu)^2$	$(2y-1) \log\left(\frac{\mu}{1-\mu}\right) - \frac{y}{\mu} - \frac{1-y}{1-\mu}$	—	—	$0 < \mu < 1, 0 < y < 1$
$\mu + \frac{\mu^2}{\nu}$	$y \log\left(\frac{\mu}{\nu+\mu}\right) + \nu \log\left(\frac{\nu}{\nu+\mu}\right)$	$\log\left(\frac{\mu}{\nu+\mu}\right)$	Negatiivinen binomi	$\mu > 0, y \geq 0$

Taulukossa 2.2 viidennen rivin varianssifunktiota μ^ζ vastaa käänteinen normaalijakauma ζ :n arvolla 3, kun $y > 0$. Kuudennen rivin binomijakaumalla $0 < y < 1$, $y = \frac{k}{m}$, $k = 0, 1, 2, \dots, m$ ja tiheysfunktio on muodossa

$$f(y) = \binom{m}{k} \mu^k (1-\mu)^{m-k}$$

Seitsemännen rivin varianssifunktiota ei vastaa mikään tunnettu jakauma.

Taulukon 2.2. varianssifunktioille μ^ζ , $\zeta \neq 0, 1, 2$, on olemassa vastaavia jakaumia seuraavasti (Jørgensen (1987)):

1. $V(\mu) = \mu^\zeta$, $1 < \zeta < 2$ vastaa Tweedien jakaumaa, joka on Poisson – Gamma yhdistetty jakauma. Se soveltuu vahinkovakuutuksen kokonaisvahinkomenon mallintamiseen.
2. $V(\mu) = \mu^\zeta$, $\zeta < 0$ vastaa hyvin stabiilia¹ jakaumaa. Ei-normaalit stabiilit jakaumat ovat ”paksuhäntäisiä” jakaumia, jotka saattavat soveltua vahinkomenon mallintamiseen.
3. $V(\mu) = \mu^\zeta$, $2 < \zeta < \infty$, $\zeta \neq 3$ vastaa positiivista stabiilia jakaumaa.
4. $V(\mu) = \mu^\zeta$, $0 < \zeta < 1$:lle ei löydy vastaavaa eksponenttijakaumaperheen jakaumaa.

Lisäksi varianssifunktiota $V(\mu) = \mu(1 + \tau\mu^2)$ vastaa yhdistetty Poisson-käänteinen Gamma jakauma. Renshaw (1994) esittää tämän jakauman ja negatiivisen binomijakauman devianssin lausekkeen.

Taulukossa 2.2 kvasiuskuottavuuden lausekkeista puuttuu ϕ ja pelkästään y :tä sisältävät termit, koska he häviävät ratkaistaessa yhtälöitä (2.25). Parametrit β ja mallin arvot $\hat{\mu}$ ovat ϕ :stä riippumattomia, mutta ϕ :tä tarvitaan varianssien laskentaan. ϕ voidaan estimoida momenttiestimaattorilla

$$\hat{\phi} = \frac{1}{n-p} \sum_i w_i \frac{(y_i - \mu_i)^2}{V(\mu_i)},$$

tai devianssin avulla, jolloin

¹ Stabiililla jakaumalla, jonka indeksi on α , tarkoitetaan tässä jakaumaa P, joka toteuttaa seuraavan ehdon:

$(x_1 + x_2 + \dots + x_n)$:llä ja $n^\alpha x_1$:llä on identtinen jakauma, jos x_1, x_2, \dots, x_n ovat riippumattomia, ja niiden kaikkien jakauma on P (Jørgensen (1987) sivu 133).

$$\hat{\phi} = \frac{D}{n-p} = \frac{1}{n-p} \sum_i d(y_i; \mu_i) \quad .$$

Luvussa (2.3.1) skaalatuksi devianssiksi saatiin (2.11):ta mukaisesti

$$D^*(\mathbf{y}; \hat{\mu}) = \sum 2w_i \{ y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i) \} / \phi = D(\mathbf{y}; \hat{\mu}) / \phi \quad . \quad (2.31)$$

Käyttämällä yhtälöä

$$\int_{\tilde{\theta}_i}^{\theta_i} (y_i - b'(t)) dt = \int_{y_i}^{\hat{\mu}_i} (y_i - b'(t)) dt$$

ja differentiaaliyhtälöä (2.30) saadaan lähtemällä (2.31):stä

$$D(\mathbf{y}; \hat{\mu}) = \sum_{i=1}^n d_i = \sum_{i=1}^n 2w_i \int_{\hat{\mu}_i}^{y_i} \frac{y_i - s}{V(s)} ds = -2\phi q(\mathbf{y}; \hat{\mu}) \quad ,$$

mikä on sama kuin tässä luvussa (2.23):ssa kvasiuskottavuuden avulla määritelty devianssi. Devianssin laskemiseen riittää siis Y_i :den jakauman ensimmäisen ja toisen momentin tunteminen.

Tarkastellaan mallin parametrien β ratkaisemista lähtien liikkeelle (2.27):n mukaisesta tiheysfunktioista. Logaritmiseksi uskottavuudeksi saadaan

$$l = l(\beta; \mathbf{y}) = \sum_{i=1}^n -\frac{1}{2\phi} d(y_i; \mu_i) + \log(c(y_i, \phi)) \quad .$$

Suurimman uskottavuuden parametrit $\hat{\beta}$ löydetään ratkaisemalla p :n yhtälön ryhmä

$$\frac{\partial l}{\partial \beta_j} = 0 \quad j = 1, \dots, p \quad .$$

Saadaan

$$\frac{\partial l}{\partial \beta_j} = -\sum_i \frac{1}{2\phi} \frac{\partial d(y_i; \mu_i)}{\partial \beta_j}$$

$$\begin{aligned}
&= -\sum_i \frac{w_i}{2\phi} \frac{\partial}{\partial \mu_i} \left(2 \int_{\mu_i}^{y_i} \frac{y_i - t}{V(t)} dt \right) \frac{\partial \mu_i}{\partial \beta_j} \\
&= \sum_i \frac{w_i}{\phi} \frac{y_i - \mu_i}{V(\mu_i)} \frac{\partial h(\mathbf{x}_i \boldsymbol{\beta})}{\partial \beta_j} \\
&= \sum_i \frac{w_i}{\phi} \frac{y_i - \mu_i}{V(\mu_i)} h'(\mathbf{x}_i \boldsymbol{\beta}) \mathbf{x}_{ij} .
\end{aligned}$$

β_j :n suurimman uskottavuuden yhtälö on siis

$$\sum_{i=1}^n \hat{w}_i (y_i - \mu_i) \mathbf{x}_{ij} = 0 \quad , \quad (2.32)$$

missä muunnettu paino on

$$\hat{w}_i = \frac{w_i h'(\mathbf{x}_i \boldsymbol{\beta})}{V(\mu_i)} . \quad (2.33)$$

Tämä on sama yhtälö kuin (2.26), joka johdettiin minimoimalla devianssi. Olkoon \mathbf{W} muunnettujen painojen \hat{w}_i muodostama $n \times n$ diagonaalimatriisi. Yhtälö (2.32) matriisimuodossa on

$$\mathbf{X}^t \mathbf{W}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0} . \quad (2.34)$$

Yhtälöstä (2.33) nähdään kanonisen linkin merkitys. Jos $\hat{w}_i = w_i$, niin h on varianssifunktiota V vastaavan kanonisen linkin käänteisfunktio. Tällöin h toteuttaa differentiaaliyhtälön $V(h(\eta)) = h'(\eta)$. Jos esimerkiksi $V(\mu) = \mu$, niin $h(\eta) = e^\eta$, ja $g(\mu) = \log(\mu)$ on kanoninen linkki. Kun käytetään kanonista linkkiä, painomatriisi \mathbf{W} ei riipu mallin antamista arvoista, ja laskenta helpottuu.

Luvussa (2.4) esitettiin ratkaisusualgoritmi, kun lähtökohtana oli (2.4):n mukaiset uskottavuusyhtälöt. Kvasiuskottavuuden mukaiset yhtälöt ratkaistaan samalla Fisher'in scoring menetelmällä. Painomatriissa $\tilde{\mathbf{W}}$ ei tarvitse tuntea Y_i :den varianssia, vaan käyttää $V(\mu_i)$:tä. Merkitään negatiivista Hessin matriisia \mathbf{H} :lla. Siis

$$\mathbf{H} = -E \left(\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right) = -E \left(\frac{\partial \mathbf{u}}{\partial \boldsymbol{\beta}} \right)$$

Yhtälön (2.34) mukaisesti $\mathbf{u} = \mathbf{X}^t \mathbf{W}(\mathbf{y} - \boldsymbol{\mu})$, ja

$$\begin{aligned}
\mathbf{H} &= -\mathbf{E}\left(\frac{\partial \mathbf{u}}{\partial \boldsymbol{\beta}}\right) \\
&= -\mathbf{E}\left(\frac{\partial}{\partial \boldsymbol{\beta}} \mathbf{X}^t \mathbf{W} (\mathbf{y} - \boldsymbol{\mu})\right) \\
&= -\mathbf{E}\left(\frac{\partial(\mathbf{X}^t \mathbf{W})}{\partial \boldsymbol{\beta}} (\mathbf{y} - \boldsymbol{\mu}) + (\mathbf{X}^t \mathbf{W}) \frac{\partial}{\partial \boldsymbol{\beta}} (\mathbf{y} - \boldsymbol{\mu})\right) \\
&= \mathbf{E}\left(\mathbf{X}^t \mathbf{W} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}\right) \\
&= \mathbf{E}\left(\mathbf{X}^t \mathbf{W} \frac{d\boldsymbol{\mu}}{d\boldsymbol{\eta}} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}}\right) \\
&= \mathbf{X}^t \tilde{\mathbf{W}} \mathbf{X} ,
\end{aligned}$$

missä painomatriisi on

$$\tilde{\mathbf{W}} = \text{diag}\left(\frac{w_i (h'(\mathbf{x}_i \boldsymbol{\beta}))^2}{V(\mu_i)}\right) .$$

Merkitään $\mathbf{a} = \mathbf{H}^{-1} \mathbf{u}$, ja käytetään Newton-Raphson menetelmää kuten luvussa 2.4. Nyt $\boldsymbol{\beta}_{n+1} = \boldsymbol{\beta}_n + \mathbf{a}$, $\mathbf{H}\boldsymbol{\beta}_{n+1} = \mathbf{H}\boldsymbol{\beta}_n + \mathbf{H}\mathbf{a} = \mathbf{H}\boldsymbol{\beta}_n + \mathbf{u}$, joten

$$\begin{aligned}
\mathbf{X}^t \tilde{\mathbf{W}} \mathbf{X} \boldsymbol{\beta}_{n+1} &= \mathbf{X}^t \tilde{\mathbf{W}} \mathbf{X} \boldsymbol{\beta}_n + \mathbf{u} \\
&= \mathbf{X}^t \tilde{\mathbf{W}} \boldsymbol{\eta}_n + \mathbf{X}^t \tilde{\mathbf{W}} \frac{d\boldsymbol{\eta}_n}{d\boldsymbol{\mu}} (\mathbf{y} - \boldsymbol{\mu}) \\
&= \mathbf{X}^t \tilde{\mathbf{W}} \left(\boldsymbol{\eta}_n + \frac{d\boldsymbol{\eta}_n}{d\boldsymbol{\mu}} (\mathbf{y} - \boldsymbol{\mu}) \right) . \tag{2.35}
\end{aligned}$$

Yhtälö (2.35) on siis painotetun lineaarisen mallin normaaliyhtälö, kun havainnot ovat

$\boldsymbol{\eta}_n + \frac{d\boldsymbol{\eta}_n}{d\boldsymbol{\mu}} (\mathbf{y} - \boldsymbol{\mu})$, ja mallin matriisi on \mathbf{X} ja painot ovat $\tilde{\mathbf{W}}$. Nyt

$$g(\boldsymbol{\mu}) + (\mathbf{y} - \boldsymbol{\mu}) g'(\boldsymbol{\mu}) = \boldsymbol{\eta} + (\mathbf{y} - \boldsymbol{\mu}) \frac{d\boldsymbol{\eta}}{d\boldsymbol{\mu}} ,$$

on $g(\mathbf{y})$:n lineaarinen approksimaatio, ja

$$\text{Var}[g(\mu) + (y - \mu)g'(\mu)] = V(\mu) \left(\frac{d\eta}{d\mu} \right)^2 = \tilde{W}^{-1},$$

kun ϕ :tä ei huomioida.

Ratkaistaessa β :tä tällä iteratiivisella ja uudelleen painotettavalla menetelmällä voidaan aloittaa asettamalla $\mu = y$.

3. TARIFFIN PARAMETRIEN ESTIMOINTI

3.1 Taulukoidut tiedot

Oletetaan, että vakuutukset on luokiteltu tariffimuuttujien A_1, \dots, A_k perusteella, missä A_i on jaettu n_i luokkaan. Solujen lukumäärä on tällöin $n_1 \cdot n_2 \cdot \dots \cdot n_k$. Havainnollisuuden vuoksi tarkastellaan tilannetta, missä $k = 2$. Tällöin on kyseessä kuvan 3.1 mukainen tilanne.

Kuva 3.1

A_1 / A_2	1	2	...	j	...	n_2
1						
2						
...						
i				(i, j)		
...						
n_1						

Olkoon

$$y_{ij} = \frac{r_{ij}}{w_{ij}} \quad i = 1, \dots, n_1, \quad j = 1, \dots, n_2 \quad (3.1)$$

y_{ij} :llä on 3 mahdollista tulkintaa :

1. y_{ij} on vahinkotiheys, jolloin r_{ij} on solun (i,j) vahinkojen lukumäärä, ja w_{ij} on solun (i,j) vakuutusosan suuruus (vakuutusvuosien lukumäärä, vakuutettujen lukumäärä).
2. y_{ij} on keskivahinko, jolloin r_{ij} on solun (i,j) vahinkomeno, ja w_{ij} on solun (i,j) vahinkojen lukumäärä.

3. y_{ij} on havaittu riskimaksu, jolloin r_{ij} on solun (i,j) vahinkomeno, ja w_{ij} on solun (i,j) vakuutuskannan suuruus (vakuutusvuosien lukumäärä, vakuutettujen lukumäärä).

y_{ij} :t ja r_{ij} :t ovat satunnaismuuttujien havaittuja arvoja, ja w_{ij} :t ovat annettuja vakioita. Tarkastellaan seuraavia malleja:

- multiplikatiivinen malli

$$E(Y_{ij}) = a_i \cdot b_j \quad i = 1, \dots, n_1, \quad j = 1, \dots, n_2 \quad (3.2)$$

- additiivinen malli

$$E(Y_{ij}) = a_i + b_j \quad i = 1, \dots, n_1, \quad j = 1, \dots, n_2 \quad (3.3)$$

a_i :t ja b_j :t ovat mallin parametrit, jotka on estimoitava. (3.2) ja (3.3) eivät määrää mallien parametreja yksikäsitteisesti. Asettamalla parametreille reunaehto, päästään tilanteeseen, jossa on estimoitava $n_1 + n_2 - 1$ parametria $n_1 \cdot n_2$:stä havainnosta.

3.1.1 Bailey - Simon'in menetelmä

Multiplikatiivisessa mallissa a_i :n ja b_j :n estimaatit saadaan minimoimalla

$$Q = \sum_{i,j} \frac{w_{ij}}{a_i b_j} (y_{ij} - a_i b_j)^2 \quad (3.4)$$

a_i :n ja b_j :n suhteen. Laskemalla osittaisderivaatat a_i :n ja b_j :n suhteen, ja laittamalla ne 0:ksi, saadaan

$$a_i = \left(\frac{\sum_j \frac{w_{ij} y_{ij}^2}{b_j}}{\sum_j w_{ij} b_j} \right)^{\frac{1}{2}} \quad i = 1, \dots, n_1 \quad (3.5)$$

$$b_j = \left(\frac{\sum_i \frac{w_{ij} y_{ij}^2}{a_i}}{\sum_i w_{ij} a_i} \right)^{\frac{1}{2}} \quad j=1, \dots, n_2 \quad (3.6)$$

Näiden yhtälöiden ratkaisu saadaan iteratiivisesti. Alkuarvoksi voidaan valita esimerkiksi $b_j = 1$, $j=1, \dots, n_2$. Konvergointi tapahtuu nopeasti. Valitsemalla eri alkuarvot saadaan ratkaisut $\{a_i', b_j'\}$, joille on voimassa

$$\begin{aligned} a_i' &= ca_i & i &= 1, \dots, n_1 \\ b_j' &= \frac{1}{c} b_j & j &= 1, \dots, n_2 \end{aligned}$$

$\{a_i', b_j'\}$:lla saadaan siis sama arvo $E(Y_{ij})$:lle kuin $\{a_i, b_j\}$:lla.

Additiivista mallia käytettäessä minimoidaan

$$Q = \sum_{i,j} \frac{w_{ij}}{a_i + b_j} [y_{ij} - (a_i + b_j)]^2 \quad (3.7)$$

Laittamalla (3.7):n a_i :n ja b_j :n suhteen lasketut osittaisderivaatat 0:ksi, saadaan yhtälöt

$$\begin{aligned} \sum_j \frac{w_{ij} y_{ij}^2}{(a_i + b_j)^2} &= \sum_j w_{ij} & i &= 1, \dots, n_1 \\ \sum_i \frac{w_{ij} y_{ij}^2}{(a_i + b_j)^2} &= \sum_i w_{ij} & j &= 1, \dots, n_2 \end{aligned}$$

Näiden yhtälöiden ratkaisu löydetään iteratiivisella Newton-Raphson menetelmällä, jolloin määritellään rekursiivisesti

$$a_i^{(k+1)} = a_i^{(k)} + \frac{\sum_j w_{ij} \left(\frac{y_{ij}}{a_i^{(k)} + b_j^{(k)}} \right)^2 - \sum_j w_{ij}}{2 \sum_j w_{ij} \left(\frac{y_{ij}}{a_i^{(k)} + b_j^{(k)}} \right)^2 \left(\frac{1}{a_i^{(k)} + b_j^{(k)}} \right)}$$

$$b_j^{(k+1)} = b_j^{(k)} + \frac{\sum_i w_{ij} \left(\frac{y_{ij}}{a_i^{(k)} + b_j^{(k)}} \right)^2 - \sum_i w_{ij}}{2 \sum_i w_{ij} \left(\frac{y_{ij}}{a_i^{(k)} + b_j^{(k)}} \right)^2 \left(\frac{1}{a_i^{(k)} + b_j^{(k)}} \right)}$$

Ratkaisu riippuu valituista alkuarvoista, mutta kaikilla ratkaisuilla saadaan samat arvot $a_i + b_j$:lle.

3.1.2 Reunasummamenetelmä

Tämä menetelmä tunnetaan myös Bailey:n menetelmän ja Jung'in menetelmän nimellä (Bailey 1963, Jung 1968). Välttämätön ehto, joka hinnoittelumenetelmän on toteutettava, on se, että suurelle määrälle vakuutuksia menetelmällä saatu tulos on yhtä suuri kuin havaitut arvot.

Multiplikatiivisessa mallissa $\{a_i, b_j\}$:n on toteutettava

$$\begin{aligned} \sum_j w_{ij} a_i b_j &= \sum_j w_{ij} y_{ij} & i = 1, \dots, n_1 \\ \sum_i w_{ij} a_i b_j &= \sum_i w_{ij} y_{ij} & j = 1, \dots, n_2 \end{aligned} \quad (3.8)$$

Tämä on $n_1 + n_2$:n yhtälön ryhmä, jossa on $n_1 + n_2$ tuntematonta, ja yksi yhtälö on muista lineaarisesti riippuva. Ratkaisu saadaan laskemalla iteratiivisesti

$$\begin{aligned} a_i &= \frac{\sum_j w_{ij} y_{ij}}{\sum_j w_{ij} b_j} & i = 1, \dots, n_1 \\ b_j &= \frac{\sum_i w_{ij} y_{ij}}{\sum_i w_{ij} a_i} & j = 1, \dots, n_2 \end{aligned} \quad (3.9)$$

b_j :lle valitaan 0:sta poikkeavat alkuarvot. Menetelmä konvergoi nopeasti.

Additiivinen malli on vastaavasti

$$\begin{aligned} \sum_j w_{ij}(a_i + b_j) &= \sum_j w_{ij}y_{ij} & i = 1, \dots, n_1 \\ \sum_i w_{ij}(a_i + b_j) &= \sum_i w_{ij}y_{ij} & j = 1, \dots, n_2 \end{aligned} \quad (3.10)$$

Nämä yhtälöt voidaan ratkaista iteratiivisesti samaan tapaan kuin multiplikatiivinen malli.

3.1.3 Pienimmän neliösumman estimaatit

Multiplikatiivisessa menetelmässä valitaan sellaiset a_i :t ja b_j :t, jotka minimoivat

$$SS = \sum_{i,j} w_{ij} (y_{ij} - a_i b_j)^2 .$$

Laittamalla SS:n osittaisderivaatat a_i :n ja b_j :n suhteen 0:ksi, saadaan

$$\begin{aligned} a_i &= \frac{\sum_j w_{ij} y_{ij} b_j}{\sum_j w_{ij} b_j^2} & i = 1, \dots, n_1 \\ b_j &= \frac{\sum_i w_{ij} y_{ij} a_i}{\sum_i w_{ij} a_i^2} & j = 1, \dots, n_2 \end{aligned} \quad (3.11)$$

jotka voidaan ratkaista iteratiivisesti.

Additiivisessa mallissa minimoitava neliösumma on

$$SS = \sum_{i,j} w_{ij} [y_{ij} - (a_i + b_j)]^2 , \quad (3.12)$$

mikä on sama ehto kuin kaksisuuntaisessa varianssianalyysissä. Derivoimalla SS a_i :n ja b_j :n suhteen, ja laittamalla osittaisderivaatat 0:ksi, saadaan yhtälöt (3.10). a_i :n ja b_j :n estimaatit ovat siis samat kuin saadaan reunasummamenetelmän additiivisella mallilla.

3.2 Estimointi yleistetyllä lineaarisella mallilla.

Oletetaan, että havainnon Y_i tiheysfunktio on

$$f(y_i; \theta, \phi) = \exp\left\{\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right\}. \quad (3.13)$$

Vahinkotiheyttä mallinnettaessa y_i on i :nnen vakuutetun vahinkojen lukumäärä, ja vahingon suuruutta mallinnettaessa y_i on i :nnen vahingon suuruus. Täten jokaisella havainnolla on oma kanoninen parametri θ_i , mutta dispersioparametri ϕ ja funktiot $a(\phi)$, $b(\theta)$ ja $c(y, \phi)$ ovat samat jokaiselle i :lle. Jokainen havainto tulee siis samasta eksponenttijakaumaperheen jakaumasta, mutta jokaisen havainnon odotusarvo voi vaihdella, koska θ vaihtelee havaintojen välillä. Vakuutusyhtiön kantaa mallinnettaessa havaintojen lukumäärä on usein satoja tuhansia.

Y_i :n odotusarvo- ja varianssi-informaatio on parametreissa θ_i ja ϕ . Luvun 2.2.1 mukaisesti

$$\mu_i = E(Y_i) = b'(\theta_i) \quad (3.14)$$

$$\text{Var}(Y_i) = b''(\theta_i)a(\phi). \quad (3.15)$$

Ylemmästä yhtälöstä saadaan

$$\theta_i = (b')^{-1}(\mu_i).$$

Yleistetyn lineaarisen mallin mukaisesti

$$\mu_i = h(\beta_1 x_{i1} + \dots + \beta_p x_{ip}),$$

joten

$$\theta_i = (b')^{-1}[h(\beta_1 x_{i1} + \dots + \beta_p x_{ip})].$$

Yllä olevasta ilmenee Y_i :n jakauman ja mallin parametrien β_1, \dots, β_p välinen yhteys.

Koska θ_i on odotusarvon μ_i funktio, niin (3.15):ta mukaisesti Y_i :n varianssi on Y_i :n odotusarvon μ_i :n funktio kerrottuna skaalaustermillä $a(\phi)$. Täten dispersioparametri ϕ on jakauman odotusarvon ja varianssin funktio.

Jos

$$a(\phi) = \frac{\phi}{w_i}, \quad (3.16)$$

ja käytetään varianssifunktiota, niin

$$\text{Var}(Y_i) = \phi \frac{V(\mu_i)}{w_i} \quad , \quad (3.17)$$

josta ilmenee, miten Y_i :n varianssi on Y_i :n odotusarvon funktio. Tarkasteltaessa vakuutus-kantaa vakuutus- tai vahinkokohtaisesti, painot $w_i = 1$. Aineisto voidaan myös luokitella luvun 3.1 mukaisesti, jolloin indeksin i arvot tarkoittavat aineiston eri soluja.

Tariffimuuttujien arvot saadaan ratkaisemalla parametrien β_1, \dots, β_p arvot. Luvun 2.4 mukaisesti aineiston logaritminen uskottavuus on

$$l = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \quad .$$

Uskottavuusyhtälöt (2.15) ratkaistaan luvun 2.4 mukaisesti Newton-Raphson menetelmällä, jolloin saadaan (2.21) mukaiset yhtälöt

$$\mathbf{X}^t \tilde{\mathbf{W}} \mathbf{X} \beta^{(m)} = \mathbf{X}^t \tilde{\mathbf{W}} \mathbf{z} \quad , \quad (3.18)$$

jotka ratkaistaan iteratiivisesti.

Aineistoa voidaan myös tarkastella kvasiuskottavuutta käyttäen. Logaritminen kvasiuskottavuus on luvun 2.5 yhtälön (2.23) mukaisesti

$$q(\mathbf{y}; \mu) = \sum_{i=1}^n q_i = \sum_{i=1}^n w_i \int_{y_i}^{\mu_i} \frac{y_i - s}{\phi V(s)} ds \quad .$$

Minimoimalla kvasiuskottavuuden avulla määritelty kvasidevianssi (2.24) saadaan luvun 2.5 mukaisesti yhtälöt (3.18).

Tarkastellaan luvussa 3.1 esitettyjen menetelmien ja yleistetyn lineaarisen mallin yhteyttä. Linearisissa malleissa havainnot on indeksoitu yhdellä indeksillä, mutta luvun 3.1 menetelmissä indeksejä on yhtä monta kuin luokittelumuuttujia. Kun luokittelumuuttujia on 2, voidaan muunnosta kuvata seuraavalla $n_1 n_2 \times 1$ vektorien välisellä muunnoksella. Vasemmalla on luvun 3.1 menetelmiä vastaava merkintä, ja oikealla lineaarista mallia vastaava merkintä. Vasemmalla on indeksien väliin lisätty pilkut selvyysden vuoksi.

$$\begin{pmatrix} y_{1,1} \\ y_{1,2} \\ \vdots \\ y_{1,n_2} \\ y_{2,1} \\ \vdots \\ y_{n_1,1} \\ \vdots \\ y_{n_1,n_2} \end{pmatrix} \leftrightarrow \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_2} \\ y_{n_2+1} \\ \vdots \\ y_{(n_1-1)n_2+1} \\ \vdots \\ y_{n_1n_2} \end{pmatrix} \quad \text{ja} \quad \begin{pmatrix} w_{1,1} \\ w_{1,2} \\ \vdots \\ w_{1,n_2} \\ w_{2,1} \\ \vdots \\ w_{n_1,1} \\ \vdots \\ w_{n_1,n_2} \end{pmatrix} \leftrightarrow \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{n_2} \\ w_{n_2+1} \\ \vdots \\ w_{(n_1-1)n_2+1} \\ \vdots \\ w_{n_1n_2} \end{pmatrix} .$$

Vastaavasti on

$$\begin{pmatrix} a_1 \\ \vdots \\ a_{n_1} \\ b_1 \\ \vdots \\ b_{n_2} \end{pmatrix} \leftrightarrow \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{n_1} \\ \beta_{n_1+1} \\ \vdots \\ \beta_{n_1+n_2} \end{pmatrix} \quad (3.19)$$

Havaintojen lukumäärä on $n = n_1n_2$, ja parametrien lukumäärä on $p = n_1 + n_2$. Kahden muuttujan additiivisen lineaarisen mallin matriisi on $n \times p$ matriisi

$$\mathbf{X} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{I} \\ \vdots & \vdots \\ \mathbf{A}_{n_1} & \mathbf{I} \end{pmatrix},$$

missä

$$\mathbf{A}_i = \begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \end{pmatrix},$$

jonka kertaluku on $n_1 \times n_2$ ja jonka alkiot ovat muuten 0 paitsi i :nnessä sarakkeessa on 1:iä. \mathbf{I} on $n_2 \times n_2$ yksikkömatriisi. Muunnoksen (3.19) mukaisesti

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{X} \begin{pmatrix} a_1 \\ \vdots \\ a_{n_1} \\ b_1 \\ \vdots \\ b_{n_2} \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_1 + b_{n_2} \\ a_2 + b_1 \\ \vdots \\ a_2 + b_{n_2} \\ \vdots \\ a_{n_1} + b_1 \\ \vdots \\ a_{n_1} + b_1 \\ \vdots \\ a_{n_1} + b_{n_2} \end{pmatrix}$$

missä kertaluvut ovat $(n \times p)(p \times 1) = (n \times 1)$. Tämä kuvaa muunnosta luvussa 3.1 käytettyjen merkintöjen ja lineaarisen mallin merkintöjen välillä.

Lineaarisen mallin suurimman uskottavuuden estimaattori $\hat{\boldsymbol{\beta}}$ toteuttaa normaaliyhtälöt

$$\mathbf{X}^t \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^t \mathbf{y} .$$

Jos havainnon i paino on w_i , ratkaisu toteuttaa yhtälön

$$\mathbf{X}^t \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^t \mathbf{W} \mathbf{y} , \quad (3.20)$$

missä $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$ on painojen lävistämatriisi. Lasketaan yhtälö (3.20) kahden muuttujan additiiviselle mallille. Painomatriisi $(n \times n)$ on

$$\mathbf{W} = \begin{pmatrix} w_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_{n_1} \end{pmatrix} ,$$

missä

$$\mathbf{W}_i = \begin{pmatrix} w_{i1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_{in_2} \end{pmatrix} \quad (n_2 \times n_2) .$$

Saadaan

$$\mathbf{X}^t \mathbf{W} \mathbf{X} = \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}^t & \mathbf{D} \end{pmatrix} \quad (p \times p) \quad ,$$

missä \mathbf{B} , \mathbf{C} ja \mathbf{D} ovat

$$\mathbf{B} = \begin{pmatrix} \sum_j w_{1j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_j w_{n_1 j} \end{pmatrix} \quad (n_1 \times n_1) \quad ,$$

$$\mathbf{C} = \begin{pmatrix} w_{11} & \cdots & w_{1n_2} \\ \vdots & \ddots & \vdots \\ w_{n_1 1} & \cdots & w_{n_1 n_2} \end{pmatrix} \quad (n_1 \times n_2) \quad ,$$

ja

$$\mathbf{D} = \begin{pmatrix} \sum_i w_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_i w_{in_2} \end{pmatrix} \quad (n_2 \times n_2) \quad .$$

Nyt

$$\mathbf{X}^t \mathbf{W} \mathbf{X} \beta = \mathbf{X}^t \mathbf{W} \mathbf{X} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}^t & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{B}\mathbf{a} + \mathbf{C}\mathbf{b} \\ \mathbf{C}^t \mathbf{a} + \mathbf{D}\mathbf{b} \end{pmatrix}$$

$$= \begin{pmatrix} a_1 \sum_j w_{1j} + \sum_j w_{1j} b_j \\ \vdots \\ a_{n_1} \sum_j w_{n_1 j} + \sum_j w_{n_1 j} b_j \\ \sum_i w_{i1} a_i + b_1 \sum_i w_{i1} \\ \vdots \\ \sum_i w_{in_2} a_i + b_{n_2} \sum_i w_{in_2} \end{pmatrix} \quad (p \times 1) \quad . \quad (3.21)$$

Toisaalta

$$\mathbf{X}'\mathbf{W}\mathbf{y} = \begin{pmatrix} \sum_j w_{1j} y_{1j} \\ \vdots \\ \sum_j w_{n_1 j} y_{n_1 j} \\ \vdots \\ \sum_i w_{i1} y_{i1} \\ \vdots \\ \sum_i w_{in_2} y_{in_2} \end{pmatrix} \quad (\mathbf{p} \times 1) \quad (3.22)$$

Yhtälöistä (3.21) ja (3.22) saadaan luvussa 3.1.2 esitetetyt Bailey:n additiivisen mallin yhtälöt (3.10), joten kahden muuttujan additiivinen yleinen lineaarinen malli on sama kuin Bailey:n additiivinen menetelmä. Normaaliyhtälöiden ratkaiseminen on tehokkaampaa kuin iteroida luvun 3.1.2 menetelmällä. Lisäksi voidaan hyödyntää lineaarisen mallin analysointiin kehitettyjä menetelmiä. Voidaan esimerkiksi tutkia residuaaleja ja poikkeavien havaintojen vaikutusta.

Yleistetyllä lineaarisella mallilla havainnon i paino on (2.33):n mukaisesti

$$\hat{w}_i = \frac{w_i h'(x_i \beta)}{V(\mu_i)}$$

Merkitään \mathbf{W} :llä \hat{w}_i :n muodostamaa $n \times n$ diagonaalimatriisia. Yhtälö (2.34) on

$$\mathbf{0} = \mathbf{X}'\mathbf{W}(\mathbf{y} - \boldsymbol{\mu}) = \begin{pmatrix} \sum_j \hat{w}_{1j} (y_{1j} - \mu_{1j}) \\ \vdots \\ \sum_j \hat{w}_{n_1 j} (y_{n_1 j} - \mu_{n_1 j}) \\ \vdots \\ \sum_i \hat{w}_{i1} (y_{i1} - \mu_{i1}) \\ \vdots \\ \sum_i \hat{w}_{in_2} (y_{in_2} - \mu_{in_2}) \end{pmatrix} \quad (\mathbf{p} \times 1),$$

josta saadaan

$$\sum_{j=1}^{n_2} \hat{w}_{ij} (y_{ij} - \mu_{ij}) = 0 \quad i = 1, \dots, n_1 \quad (3.23)$$

$$\sum_{i=1}^{n_1} \hat{w}_{ij} (y_{ij} - \mu_{ij}) = 0 \quad j = 1, \dots, n_2 \quad (3.24)$$

Kun valitaan linkkifunktioksi identiteettifunktio, niin $\eta_{ij} = \mu_{ij}$ ja $\frac{d\eta}{d\mu} = 1$, joten

$$\hat{w}_{ij} = \frac{w_{ij}}{V(\mu_{ij})}$$

Additiivisella mallilla on $\eta_{ij} = a_i + b_j$. Yhtälöistä (3.23) ja (3.24) saadaan

$$a_i = \frac{\sum_{j=1}^{n_2} \frac{w_{ij} (y_{ij} - b_j)}{V(\mu_{ij})}}{\sum_{j=1}^{n_2} \frac{w_{ij}}{V(\mu_{ij})}} \quad i = 1, \dots, n_1 \quad (3.25)$$

$$b_j = \frac{\sum_{i=1}^{n_1} \frac{w_{ij} (y_{ij} - a_i)}{V(\mu_{ij})}}{\sum_{i=1}^{n_1} \frac{w_{ij}}{V(\mu_{ij})}} \quad j = 1, \dots, n_2 \quad (3.26)$$

Normaalijakaumalla $V(\mu) = 1$, ja tällöin saadaan yllä olevasta yhtälöstä (3.25)

$$a_i = \frac{\sum_{j=1}^{n_2} w_{ij} (y_{ij} - b_j)}{\sum_{j=1}^{n_2} w_{ij}} \quad i = 1, \dots, n_1,$$

joka on luvussa 3.1.2 esitetty Bailey:n additiivinen malli.

Poisson-, gamma- ja käänteiselle normaalijakaumalle saadaan (3.25):stä

$$a_i = \frac{\sum_{j=1}^{n_2} \frac{w_{ij}(y_{ij} - b_j)}{\mu_{ij}}}{\sum_{j=1}^{n_2} \frac{w_{ij}}{\mu_{ij}}} \quad i = 1, \dots, n_1, \quad a_i = \frac{\sum_{j=1}^{n_2} \frac{w_{ij}(y_{ij} - b_j)}{\mu_{ij}^2}}{\sum_{j=1}^{n_2} \frac{w_{ij}}{\mu_{ij}^2}} \quad i = 1, \dots, n_1,$$

ja

$$a_i = \frac{\sum_{j=1}^{n_2} \frac{w_{ij}(y_{ij} - b_j)}{\mu_{ij}^3}}{\sum_{j=1}^{n_2} \frac{w_{ij}}{\mu_{ij}^3}} \quad i = 1, \dots, n_1.$$

Nämä ovat uusia menetelmiä. Kun linkkifunktio on logaritmfunktio, niin $\eta = \log \mu$, ja siis

$$\frac{d\eta}{d\mu} = \frac{1}{\mu}, \text{ jolloin}$$

$$\hat{w}_{ij} = \frac{w_{ij} \mu_{ij}}{V(\mu_{ij})},$$

Tässä tapauksessa $\mu_{ij} = \exp \eta_{ij} = \exp(a_i + b_j) =: c_i d_j$. Logaritminen linkki siis muuntaa additiivisen mallin multiplikatiiviseksi. Yhtälöstä (3.23) saadaan

$$c_i = \frac{\sum_{j=1}^{n_2} \frac{w_{ij} y_{ij} d_j}{V(\mu_{ij})}}{\sum_{j=1}^{n_2} \frac{w_{ij} d_j^2}{V(\mu_{ij})}} \quad i = 1, \dots, n_1,$$

ja vastaavasti d_j :lle. Sijoittamalla normaali-, Poisson-, gamma- ja käänteisen normaalijauman V :t saadaan

$$c_i = \frac{\sum_{j=1}^{n_2} w_{ij} y_{ij} d_j}{\sum_{j=1}^{n_2} w_{ij} d_j^2} \quad (\text{Normaali}),$$

$$c_i = \frac{\sum_{j=1}^{n_2} w_{ij} y_{ij}}{\sum_{j=1}^{n_2} w_{ij} d_j} \quad (\text{Poisson}) ,$$

$$c_i = \frac{\sum_{j=1}^{n_2} \frac{w_{ij} y_{ij}}{d_j}}{\sum_{j=1}^{n_2} w_{ij}} \quad (\text{Gamma}) ,$$

$$c_i = \frac{\sum_{j=1}^{n_2} \frac{w_{ij} y_{ij}}{d_j^2}}{\sum_{j=1}^{n_2} \frac{w_{ij}}{d_j}} \quad (\text{Käänteinen normaali}) .$$

Näistä ensimmäinen on luvun (3.1.3) pienimmän neliösumman multiplikatiivinen malli, toinen on Baileyn multiplikatiivinen malli, kolmas on ns. Brownin eksponentiaalinen malli ja neljäs on uusi menetelmä.

Käyttämällä muita linkkifunktioita kuin identiteetti- ja logaritmfunktio, voidaan johtaa lisää saman tyyllisiä menetelmiä. Luvussa 3.1.1 esitetty Bailey-Simon'in menetelmä perustuu χ^2 -statistikan minimointiin eikä uskottavuusfunktion maksimointiin, joten Bailey-Simon'in menetelmää ei voida johtaa yleistetystä lineaarisesta mallista. Useilla mutta ei kaikilla 1960-luvulla kehitetyillä ”minimum bias hinnoittelumenetelmillä” on siis vastaava yleistetty lineaarinen malli. Vastaavasti monille GLM-malleille ei ole vastaavaa ”minimum bias menetelmää”.

Kappaleessa 3 on tarkasteltu 2:n tariffimuuttujan tilannetta. Kappaleen tulokset on yleistettävissä useamman muuttujan perusteella tehtävään tariffointiin.

4. AUTOVAKUUTUKSEN KESKIVAHINKO

Baxter, Coutts ja Ross (1980) analysoivat brittiläisen vakuutusyhtiön yksityishenkilöiden yksityiskäyttöisten henkilöautojen laajan moottoriajoneuvovakuutuksen oman ajoneuvon vaunuvahinkojen keskivahinkoaineistoa lineaarisella ortogonaalisella painotetulla pienimmän neliösumman mallilla (OWLS) ja GLIM-ohjelmistoa käyttäen yleistetyllä lineaarisella mallilla. McCullagh ja Nelder (1989) ja Mildenhall'in (1999) tarkastelivat samaa aineistoa. Aineistossa on 3 tariffitekijää: vakuutuksenottajan ikä (8 tasoa), auton ikä (4 tasoa) ja auton luokka (4 tasoa). Aineisto on liitteen 1 sarakkeissa 1 – 5 .

Baxter, Coutts ja Ross olettavat kummassakin mallissa, että solun (i,j,k) keskivahinko $Y_{ijk} \sim N(\theta_{ijk}, \sigma^2 / n_{ijk})$, missä n_{ijk} on vahinkojen lukumäärä, σ^2 on tuntematon vakio ja

$$\theta_{ijk} = \xi + \delta_i + \kappa_j + \tau_k + [I_n] .$$

Tässä ξ on vakio, δ_i , κ_j ja τ_k ovat tariffitekijöiden päävaikutukset ja $[I_n]$:llä on merkitty interaktiovaikutuksia. Linearisessa mallissa ratkaistaan normaaliyhtälöistä

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

$\boldsymbol{\beta}$:n pienimmän neliösumman estimaattori $\hat{\boldsymbol{\beta}}$. Yleistetyssä lineaarisessa mallissa on normaalijakauma ja linkkifunktiona identiteettifunktio. Mallien parametrien estimaatit on taulukossa 4.1.

Taulukko 4.1 Mallien parametrien estimaatit (Baxter, Coutts, Ross (1980))

Parametri		GLIM		OWLS	
		Estimaatti	Keski-hajonta	Estimaatti	Keski-hajonta
Vakio		210.6	20.0	217.5	18.1
Vakuutusikä	17-20	76.3	31.0	63.1	31.6
	21-24	68.4	17.2	57.5	17.1
	25-29	40.6	12.4	38.5	12.5
	30-34	24.9	11.7	29.9	12.0
	35-39	-23.1	11.7	-24.8	11.7
	40-49	- 3.7	9.9	- 3.8	10.2
	50-59	-15.9	10.5	- 0.2	10.6
	60+	0	-	0	-
Auton luokka	A	-102.8	12.5	-108.7	12.2
	B	-108.4	11.5	-111.1	10.1
	C	- 72.0	12.1	- 74.5	10.4
	D	0	-	0	-
Auton ikä	0-3	131.3	16.3	126.6	14.7
	4-7	107.0	16.4	102.4	14.7
	8-9	52.5	18.9	49.8	17.1
	10+	0	-	0	-

Baxter:in, Coutts:in ja Ross:in mukaan eri menetelmillä lasketut parametrien estimaatit vastaavat hyvin toisiaan. GLIM-mallissa tariffitekijöiden päävaikutukset selittävät 76 % solujen välisestä vaihtelusta, ja OWLS-mallissa vastaavasti 72 %. Sen vuoksi aineisto voidaan esittää tyydyttävästi ainoastaan päävaikutukset sisältävällä mallilla. Baxter, Coutts ja Ross tutkivat varianssioletusta residuaalien

$$\frac{x_{ijk} - \hat{\theta}_{ijk}}{S / \sqrt{n_{ijk}}}$$

avulla, missä x_{ijk} on solun (i,j,k) havaittu keskivahinko, ja $\hat{\theta}_{ijk}$ on GLIM mallilla saatu solun (i,j,k) keskivahingon odotusarvon suurimman uskottavuuden estimaatti. S^2 on varianssin σ^2 estimaatti

$$S^2 = D^{-1} \sum_{i=1}^8 \sum_{j=1}^4 \sum_{k=1}^4 n_{ijk} (x_{ijk} - \hat{\theta}_{ijk})^2 .$$

D on vapausasteiden lukumäärä. Baxter:in, Coutts:in ja Ross:in mukaan ei ole mahdotonta, että varianssi vaihtelisi solun odotusarvon mukaisesti. Heidän mukaansa residuaalit eivät kuitenkaan poikkea merkittävästi normaaleista, ja residuaaleissa on ainoastaan pientä trendiä. (Baxter, Coutts ja Ross (1980) sivu 27) Sen vuoksi heidän mielestään jakaumaksi voidaan olettaa $N(\theta_{ijk}, \sigma^2 / n_{ijk})$. Lisäksi heidän mielestään vaikka OWLS-malli on ainoastaan GLIM-mallin approksimaatio, sillä on GLIM-malliin verrattuna laskennallisia etuja.

McCullagh ja Nelder (1989) olettivat, että solujen variaatiokerroin on vakio ja käyttivät gammajakaumamallia ja linkkifunktiona käänteisfunktiota. Tariffitekijöiden vaikutukset ovat siis lineaarisia käänteisellä asteikolla. Päävaikutukset sisältävä malli on

$$\mu_{ijk} = E(Y_{ijk}) = (\mu_0 + \alpha_i + \beta_j + \gamma_k)^{-1}$$

$$\text{Var}(Y_{ijk}) = \frac{\sigma^2 \mu_{ijk}^2}{m_{ijk}} ,$$

missä α_i , β_j ja γ_k ovat vakuutusnottajan ikää (VI), auton luokkaa (AL), auton ikää (AI) vastaavat parametrit ja m_{ijk} on vahinkojen lukumäärä. Tässä $\eta_{ijk} = 1/\mu_{ijk}$ ”voidaan ajatella nopeutena, millä £1 :n lisäykset pitää maksaa, kun palvellaan solun keskivahinkoa aikayksikön ajan”. Taukossa 4.2 on mallin sovitusten yhteensopivuustilasto ja taulukossa 4.3 parametrien estimaatit ja estimaattien hajonnat.

Taulukko 4.2 Mallien yhteensopivuustilastot (McCullagh ja Nelder (1989))

Malli	Devianssi	Erotus	Vapausasteet	Keski-devianssi
1	649.9			
VI	567.7	82.2	7	11.7
VI + AL	339.4	228.3	3	76.1
VI + AL + AI	124.8	214.7	3	71.6
+ VI · AL	90.7	34.0	21	1.7
+ VI · AI	71.0	19.7	21	0.9
+ AL · AI	65.6	5.4	9	0.6
Täysi	0	65.6	58	1.1

Taulukko 4.3 Päävaikutusmallin parametrien estimaatit ja hajonnat ($\times 10^6$) (McCullagh, Nelder (1989))

Taso	Vakuutusnottajan ikä (VI)	Auton luokka (AL)	Auton ikä (AI)
1	0	0	0
2	101 (436)	38 (169)	336 (101)
3	350 (412)	-614 (170)	1651 (227)
4	462 (410)	-1421 (181)	4154 (442)
5	1370 (419)		
6	970 (405)		
7	916 (408)		
8	920 (416)		

Kun oletetaan, että taulukon 4.2 erotukset noudattavat χ^2 -jakaumaa, niin havaitaan, että päävaikutukset sisältävä malli on riittävä, ja interaktiotermi ei lisää mallin selitysvoimaa. Päävaikutusmallissa residuaaleihin perustuva σ^2 :n estimaatti on

$$\tilde{\sigma}^2 = \frac{1}{109} \sum \frac{m(y - \hat{\mu})^2}{\hat{\mu}^2} = 1.21,$$

joten yksittäisen vahingon variaatiokertoimen estimaatti on 1.1. Standardisoidut residuaalit voi laskea kaavalla $\sqrt{m}(y - \hat{\mu})/\tilde{\sigma}\hat{\mu}$. Suurimmat residuaalit vastaavat havaintoja (2,2,1), (3,2,4), (3,4,4), (5,1,2), (5,4,1) ja (7,2,1), ja ne ovat 3.4, 3.2, 2.8, -2.6, -2.2 ja -2.5. Taulukon 4.3 keskihajonnat perustuvat estimaattiin $\tilde{\sigma} = 1.1$ ja μ_0 :n estimaatti on $3410 \cdot 10^{-6}$.

McCullagh ja Nelder päättelivät taulukon 4.3 tulosten perusteella, että neljän nuorimman ikäryhmän vakuutusnottajilla keskivahinko on suurin, 35-39 vuotiaille on pienin keskivahinko ja yli 40 vuotiaat ovat keskitasolla. Vahingon koko pienenee auton vanhetessa, mutta

pieneneminen ei ole lineaarista. Auton luokkien vahingoissa on selvät erot. Luokan D autot ovat kalleimmat, C on keskitasoa ja halvimpien luokkien A ja B autojen välillä ei ole merkittävää eroa.

McCullagh ja Nelder tutkivat devianssin muuttumista λ :n funktiona, kun $\eta = \mu^\lambda$ ja varianssifunktio $V(\mu) \propto \mu^2$. Devianssin minimi 124.51 saadaan lähellä λ :n arvoa -1 , joka vastaa heidän käyttämää käänteisfunktiota. λ :n arvoa 0 vastaava logaritmifunktio on 95 %:n varmuusvälin sisäpuolella. He tarkastelivat myös λ :n ja ζ :n yhtäaikaista muuttumista laajennetun kvasidevianssin

$$\sum_i \log(\hat{\phi}V(y_i; \zeta))$$

avulla, kun linkkifunktio on muotoa $g(\mu; \lambda) = \mu^\lambda$ ja vastaavasti varianssifunktio kuuluu samaan potenssifunktioiluokkaan $V(\mu; \zeta) = \mu^\zeta$. Minimi saavutetaan, kun $\zeta = 2.4$ ja $\lambda = 0.75$. Alkuperäinen valinta $(2, -1)$ on 95 %:n alueen sisäpuolella kuten myös logaritmifunktiota vastaava $(2, 0)$.

Mildenhall (1999) käytti myös Baxter:in, Coutts:in ja Ross:in aineistoa. Hän tarkasteli aineistoa kuitenkin käyttäen tariffitekijöinä ainoastaan vakuutuksenottajien ikää ja auton luokkaa. Mildenhall sovitti aineistoon 10 mallia, joissa jakaumina olivat normaali-, gamma- ja käännteinen normaalijakauma. Linkkifunktioina oli identiteetti-, logaritmi-, käänteis- ja käännteinen neliöfunktio, joista viimeistä käytettiin ainoastaan käänteiselle normaalijakaumalle.

Kaikissa 10:ssä mallissa kaikki muut parametrit olivat tilastollisesti merkitseviä χ^2 -jakauman perusteella, mutta auton luokkien A ja B välillä ei ollut eroa (Mildenhall 1999 liitteet 7.1 – 7.10). Logaritmisella uskottavuudella voidaan verrata mallien sovituksia. Logaritmisten uskottavuuksien arvot on taulukossa 4.4.

Taulukko 4.4. Mallien logaritmiset uskottavuudet (Mildenhall (1999)).

Malli	Jakauma	Linkki	Logaritminen uskottavuus
1	Normaali	Identiteetti	- 144.30
2	Normaali	Log	- 144.44
3	Normaali	Käänteis	- 145.79
4	Gamma	Identiteetti	- 140.75
5	Gamma	Log	- 141.06
6	Gamma	Käänteis	- 143.27
7	Käännteinen normaali	Identiteetti	- 141.08
8	Käännteinen normaali	Log	- 141.35
9	Käännteinen normaali	Käänteis	- 143.34
10	Käännteinen normaali	Käännteinen neliö	- 147.22

Gammajakaumalla identiteetti- tai logaritmifunktio linkkifunktiona on suurin logaritminen uskottavuus. Gamma-identiteetti mallin devianssanalyysin tulokset on taulukossa 4.5.

Taulukko 4.5 Mallien yhteensopivuustilastot (Mildenhall (1999))

Malli	Devianssi	Erotus	Vapausasteet	Keski-devianssi
1	347.0			
VI	264.9	82.1	7	11.7
VI + AL	31.2	233.7	3	77.9
Täysi	0	31.2	21	1.5

Vakuutuksenottajan ikä ja auton luokka ovat tilastollisesti merkitseviä ja pienentävät selvästi devianssia. Devianssit ovat pienempiä kuin McCullagh:in ja Nelder:in malleissa, koska auton ikä ei ole mukana aineistossa. Vakuutuksenottajan iän ja auton luokan lisäämisen aikaan sama devianssin pienennys on samaa tasoa kuin McCullagh:lla ja Nelder:llä.

Myös Mildenhall vertasi linkkifunktioita laskemalla devianssit potenssiperheen

$$\eta = \begin{cases} \mu^\lambda, & \lambda \neq 0, \\ \log \mu, & \lambda = 0. \end{cases}$$

funktioille. Taulukossa 4.6 on devianssit eri λ :n arvoilla, kun jakaumana on gamma-jakauma.

Taulukko 4.6. Devianssi linkkifunktion potenssin λ arvoilla (Mildenhall (1999)).

λ	Devianssi
-1.80	43.83
-1.30	38.97
-0.80	35.19
-0.30	32.72
0.20	31.46
0.70	31.13
1.20	31.42
1.45	31.72

Devianssi on melko tasainen alueella $0.20 \leq \lambda \leq 1.2$, joka sisältää identiteettifunktion. Käänteisfunktioita ($\lambda = -1$) vastaavan linkin devianssi on selvästi suurempi kuin devianssi tällä alueella. McCullagh:lla ja Nelder :llä devianssin minimi saatiin lähellä λ :n arvoa -1 .

Mildenhall vertasi varianssifunktioita $V(\mu; \zeta) = \mu^\zeta$ käyttäen laajennettua devianssia muodossa

$$d(y_i; \mu) = 2 \frac{w_i}{\phi} \int_{\mu}^{y_i} \frac{y_i - t}{V(t)} dt + \log(\phi V(y_i)) .$$

Sivun 44 taulukossa 4.7 on laajennetut devianssit eri ζ :n arvoilla. Linkkifunktiona on identiteettifunktio. Laajennettu devianssi on pieni alueella $1.95 \leq \zeta \leq 3.45$, joka sisältää sekä gammajakauman ($\zeta = 2$) että käänteisen normaalijakauman ($\zeta = 3$). Yhdistämällä taulukojen 4.6 ja 4.7 tulokset havaitaan, että gammajakauma tai käänteinen normaalijakauma ja identiteettifunktio linkkifunktiona on hyvä valinta, vaikka voi valita linkkifunktion ja varianssifunktion potenssifunktioperheestä. Nämä tulokset sopivat yhteen taulukon 4.4 uskottavuuksien kanssa.

Liitteen 1 sarakkeessa 9 on Mildenhallin käyttämän mallin mukaiset arvot, sarakkeessa 8 McCullagh ja Nelder mallin arvot ja sarakkeissa 6 – 7 Baxter:in, Coutts:in ja Ross:in mallien arvot.

Taulukko 4.7. Laajennettu devianssi ja varianssifunktio $V(\mu; \zeta) = \mu^{\zeta}$ (Mildenhall (1999)).

ζ	Laajennettu devianssi
1.20	372.74
1.45	372.02
1.70	371.42
1.95	370.95
2.20	370.60
2.45	370.37
2.70	370.28
2.95	370.32
3.20	370.49
3.45	370.80

5. HENKILÖAUTOJEN LIIKENNEVAKUUTUKSEN HINNOITTELUA

5.1 Omaisuusvahingon vahinkotiheys

Tarkastellaan yksityishenkilöiden Pohjolassa vakuuttamien yksityiskäyttöisten henkilöautojen liikennevakuutuksen vuonna 2002 sattuneita vahinkoja. Vuoteen 2002 kohdistuneiden vakuutusvuosien lukumäärä on 342 427 vakuutusvuotta. Vakuutuksille sattui 9680 omaisuusvahinkoa , joista maksettiin korvauksia 15 728 tuhatta euroa. Vahinkojen korvausten kertymää on otettu mukaan vuoden 2004 syyskuun loppuun asti, joten aineistoa voidaan pitää lopullisena.

Tariffitekijät , tekijöiden luokkien lukumäärät ja luokittelut ovat

Alue (AL) : 4 : 1 , 2 , 3 , 4

Sylinteritilavuus (SY): 9 : - 1050 , 1051 – 1250 , 1251 – 1350 , 1351 – 1550 , 1551 – 1650 , 1651 – 1850 , 1851 – 2150 , 2151 – 2650 , 2651 -

Auton merkki ja malli (AM): 9 : -4 , -3 , -2 , -1 , 0 , 1 , 2 , 3 , 4

Auton ikä (AI) : 8 : 0 – 2 , 3 – 5 , 6 – 8 , 9 – 11 , 12 – 14 , 15 – 17 , 18 – 20 , 21 –

Sukupuoli (SP) : 2 : mies, nainen

Haltijan ikä (HI) : 12 : - 21 , 22 – 24 , 25 – 29 , 30 – 34 , 35 – 39 , 40 – 44 , 45 – 49 , 50 – 54 , 55 – 59 , 60 – 64 , 65 – 69 , 70 –

Bonus (BO) : 6 : 0 , 5 – 15 , 20 – 30 , 35 – 45 , 50 – 65 , 70 .

Alue määräytyy ajoneuvon haltijan kotikunnan mukaan. Alueessa 1 ovat pääkaupunkiseudun kaupungit ja alueessa 4 ovat maaseudun kunnat. Automalliluokat on määrätty Liikennevakuutuskeskuksen merkki- ja mallitutkimuksessa. Tutkimuksen luokittelun 4 alinta ja 4 ylintä luokkaa on yhdistetty. Auton ikä on laskettu vakuutuskauden alkamisvuoden ja auton käyttöönottovuoden erotuksena. Haltijan ikä on laskettu vakuutuskauden alkamisvuoden ja auton haltijan syntymävuoden erotuksena. Bonusluokittelussa asteikon bonusluokkia on yhdistetty yllä olevan mukaisesti. Tällä tariffitekijöiden luokittelulla solujen lukumäärä on 373 248 .

Sovitetaan aineiston vahinkotiheyksiin Pretium-ohjelmistolla Poisson-malli, jossa linkki-funktiona on logaritmfunktio. Mallissa on tariffitekijöiden päävaikutukset (AL + AM + AI + BO + SY + HI + SP) ja haltijan iän ja sukupuolen (HI · SP), auton iän ja auton merkin ja mallin (AI · AM), sylinteritilavuuden ja auton merkin ja mallin (SY · AM) sekä haltijan iän ja bonuksen (HI · BO) yhteisvaikutukset. Taulukossa 5.1 on esitetty mallien yhteensopivuustilastot.

Taulukko 5.1. Omaisuusvahingon vahinkotiheysmallien devianssianalyysi.

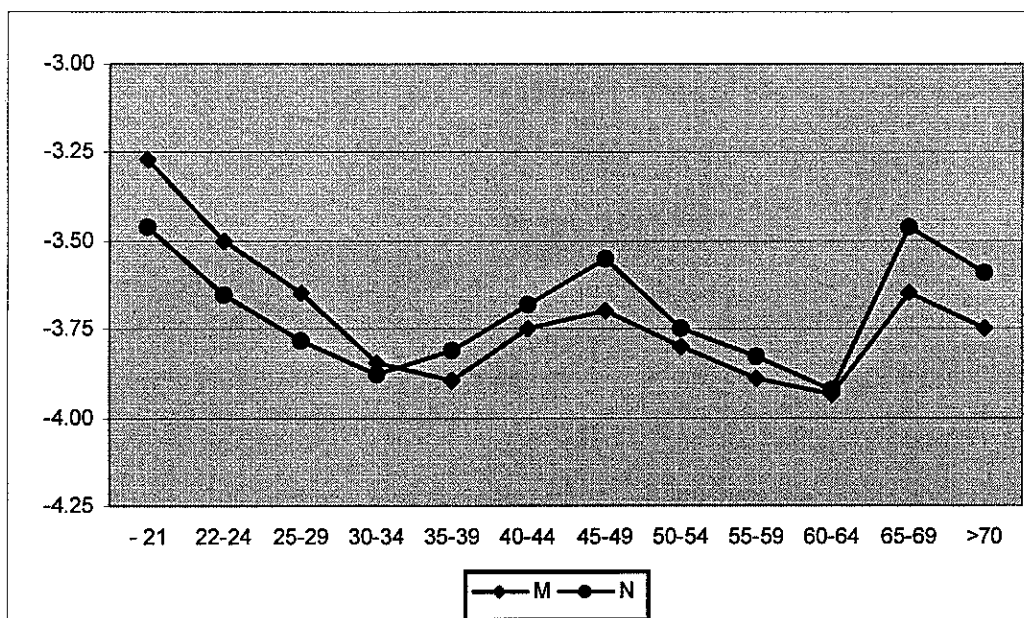
Malli	Devianssi	Vapausasteet	χ^2	tn > χ^2
1	37 835.8			
+ AL	37 621.0	3	214.8	0.0001
+ AM	37 538.3	8	82.7	0.0001
+ AI	37 433.8	7	104.5	0.0001
+ BO	36 609.0	5	824.8	0.0001
+ SY	36 492.9	8	116.1	0.0001
+ HI	36 162.3	11	330.6	0.0001
+ SP	36 159.6	1	2.7	0.1015
+HI · SP	36 137.6	11	22.0	0.0247
+AI · AM	36 050.9	56	86.27	0.0052
+SY · AM	35 964.6	64	86.27	0.0333
+HI · BO	35 872.6	55	92.02	0.0013

Ensimmäisenä on sovitettu aineistoon ainoastaan vakion sisältämä malli, ja sen jälkeisissä malleissa on lisätty tekijöitä yksi kerrallaan taulukon mukaisessa järjestyksessä. Devianssien

erotus on asympotoottisesti χ^2 jakautunut, kun oletuksena on, että malliin lisätty parametri on 0. χ^2 -jakauman vapausasteiden lukumäärä on perättäisten mallien parametrien lukumäärien erotus. Tariffitekijöiden päävaikutukset sukupuolta lukuunottamatta ovat merkitseviä 1 %:n tasolla. Sukupuolen mukaan tehdyssä tilastossa sekä naisten että miesten vahinkotiheys on 0.028. Kun mallissa otetaan muiden tekijöiden vaikutukset huomioon, sukupuolten välillä on eroa, mutta eron merkitsevyys χ^2 -jakauman mukaan on 10 %:n tasolla. Suomalaiset yhtiöt eivät käytä auton ikää liikennevakuutuksen hinnoittelussa, ja haltijan ikä on käytössä vain muutamalla yhtiöllä. Morlock:n ja Zimpelmann:in (1999) mukaan Saksassa 83 % 87:stä tutkitusta moottoriajoneuvovakuutusyhtiöstä käytti auton ikää moottoriajoneuvonvastuun hinnoittelussa. Pitää kuitenkin huomata, että tässä tarkastellaan ainoastaan omaisuusvahingon vahinkotiheyttä. Auton iän vaikutuksen selvittämiseksi riskimaksutasolla pitää ottaa huomioon myös omaisuusvahingon suuruus ja henkilövahingon vahinkotiheys ja suuruus. Bonuksen vaikutus devianssiin on ehkä yllättävän suuri, vaikka mallissa on 6 ensisijaista tariffitekijää. Bonus on määräytynyt aikaisempien vuosien vahinkojen lukumäärien perusteella. Näyttää siltä, että autoa ja vakuutuksenottajaa mittaavien tekijöiden jälkeen hinnoittelusoluihin jää erilaisuutta vakuutuksien välille, ja bonus selittää näistä eroista huomattavan osan.

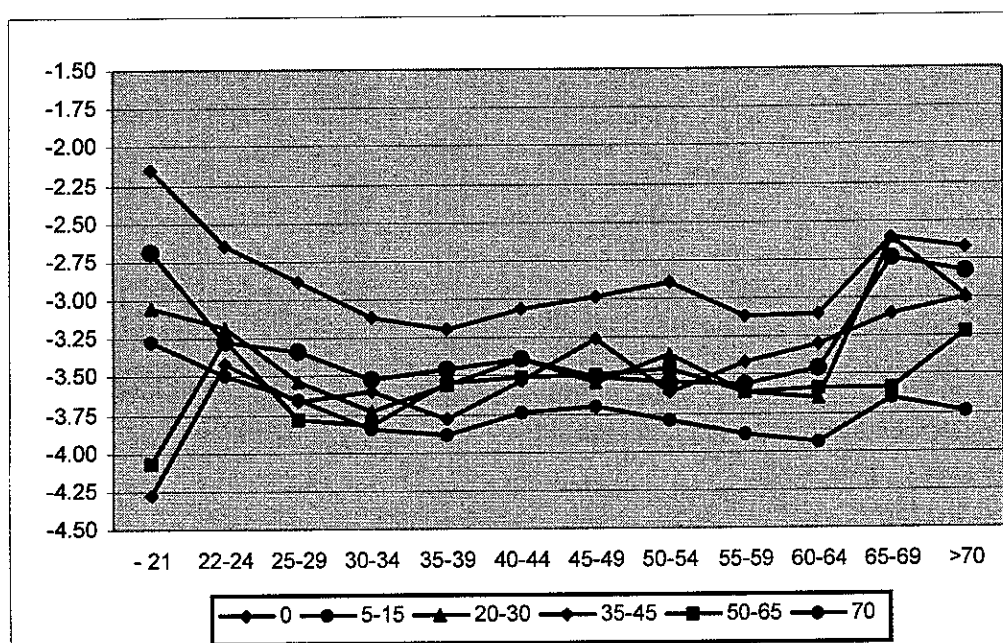
Interaktiomuuttujat ovat merkitseviä 5 %:n tasolla. Kuvissa 5.1 ja 5.2 on esitetty haltijan iän ja sukupuolen sekä haltijan iän ja bonuksen vaikutuksien kuvaajat tekijöiden eri arvoilla. Jos interaktiotermi ei olisi mukana, kuvaajat olisivat samanmuotoisia (siirtymä y-akselin suunnassa), koska tekijöiden vaikutukset ovat additiivisia logaritmiasteikolla.

Kuva 5.1. Omaisuusvahingon vahinkotiheys. Haltijan ikä ja sukupuoli.



Kuvan 5.1 mukaan nuorten naisten vahinkotiheys on pienempi kuin nuorten miesten, mutta 35-39 vuotiaiden ja vanhempien naisten vahinkotiheys on suurempi kuin miesten vahinkotiheys. Sukupuolen vaikutus ei ole yhtä suuri eri ikäisillä henkilöillä.

Kuva 5.2. Omaisuusvahingon vahinkotiheys. Haltijan ikä ja bonus.



Kuvan 5.2 mukaan esimerkiksi 0 bonuksellisista vakuutuksenottajista suurin vahinkotiheys on alle 21 vuotiailla. Nuoria vakuutuksenottajia ei ole käytännössä ylemmissä bonusluokissa, koska bonusta ei ole vielä ehtinyt kertyä. Auton iän ja auton merkin ja mallin interaktiota on ehkä lisännyt se, että merkkejä ja malleja luokiteltaessa auton ikää ei huomioida erillisenä tekijänä, vaan se on vaikuttanut luokitukseen. Esimerkiksi auton vuosimallin vanhetessa, autot ovat siirtyneet nuorisokäyttöön, ja mallin vahingot ovat lisääntyneet, jolloin vahinkotilaston perusteella tehdyssä luokittelussa malli on siirretty ylempään automalliluokkaan. Isolla moottorilla varustettujen autojen riskin vaikutus välittyy ehkä sekä sylinteritilavuuden että automalliluokan välityksellä, mikä on lisännyt näiden tekijöiden interaktiotermin merkitsevyyttä.

Käytännössä interaktiomuuttujista ainoastaan haltijan iän ja sukupuolen voi ottaa mukaan mukaan tarffiin, jos halutaan, että vakuutuksen hinnan muodostus pystytään perustelemaan selkeästi vakuutuksenottajille. Liitteessä 2 on tulokset, kun omaisuusvahinkojen vahinkotiheysaineistoon on sovitettu Poisson-malli, jossa linkkifunktiona on logaritmfunktio ja päävaikutuksista on mukana alue, sylinteritilavuus, auton merkki ja malli, auton ikä, bonus ja haltijan sukupuolesta ja iästä muodostettu yhdistetty tekijä. Tämän mallin logaritminen uskottavuus on $-43\,350$, kun aikaisemmin sovitetun laajemman mallin logaritminen uskottavuus on $-43\,217$. Mallin sovituksen residuaalit ovat liitteen 3 kuvassa 1. Kuvassa on mukana ainoastaan solut, joissa on vähintään 30 vakuutusvuotta. Residuaalit eivät ole normaalisti jakautuneita. Välillä $(-1.9, -1.1)$ on useita soluja, joille ei ole sattunut vahinkoja. Mallissa on 6 tarffimuuttujaa, ja muuttujilla on tiheä luokittelu. Tästä johtuen melko usean solun ha-

vaittu vahinkojen lukumäärä on 0, jolloin havaittu arvo jää mallin arvoa pienemmäksi. Kun solulle, jossa on vähän vakuutuksia, on sattunut 1 tai useampia vahinkoja, solun havaittu vahinkotiheys on suuri. Nämä solut ovat selvästi 0:n oikealla puolella. Toisaalta residuaalit ovat välillä (-2.3, +3.3). Tarkastellaan liitettä 2. Tariffitekijöiden luokkien kertoimien arvojen vaihtelu on melko suuri, sillä alueen, sylinteritilavuuden, auton merkin ja mallin, auton iän, bonuksen ja haltijan sukupuolen ja iän suurimman ja pienimmän kertoimen suhde on 1.52, 1.71, 1.48, 1.56, 2.35, 2.54. Uusilla ja muutaman vuoden ikäisillä autoilla ja vanhoilla yli 18 vuoden ikäisillä autoilla vahinkotiheys on pienempi kuin keski-ikäisillä autoilla. Bonusasteikon ylimmän ja alimman luokan suhde on 3.33, kun se mallin mukaan olisi 2.35. Alle 21 vuotiailla miehillä on suurin vahinkotiheys, ja 30-34 vuotiaiden naisten vahinkotiheys on pienin.

5.2 Omaisuusvahingon keskivahinko

Sovitetaan luvussa 5.1 käytetyn omaisuusvahinkoaineiston keskivahinkoihin gammajakamamalli, jonka linkkifunktio on logaritmfunktio. Mallin tekijät ovat alue, sylinteritilavuus, auton merkki ja malli, auton ikä ja haltijan sukupuolesta ja iästä muodostettu yhdistetty tekijä. Residuaalit ovat liitteen 3 kuvassa 2, devianssianalyysi on taulukossa 5.2 ja mallin sovituksen tulokset ovat liitteessä 4.

Taulukko 5.2. Omaisuusvahingon keskivahinkomallien devianssianalyysi.

Malli	Devianssi	Vapausasteet	χ^2	$tn > \chi^2$
AL	9 733.8			
+ SY	9 720.1	8	13.8	0.0876
+ AM	9 711.8	8	8.3	0.4086
+ AI	9 674.6	7	37.3	< 0.0001
+ HI · SP	9 619.3	23	55.3	0.0002

Residuaalijakauman huippu on 0:n vasemmalla puolella, joten useille soluille mallilla saatu estimaatti on suurempi kuin solun havaittu keskivahinko. Jakauma on selvästi oikealle vino. Kuvasta puuttuu lisäksi 20 solua, joissa residuaali on suurempi kuin 3.9. Residuaalijakauman muotoa selittää se, että aineistossa on muutama iso vahinko, ja useita pieniä vahinkoja. Dispersioparametri ϕ on 0.9989, joten vahinkojen suuruuden variaatiokerroin on 0.9994. Tarkastellaan liitettä 4. Alueen vaikutus on yllättävän pieni. Vahingon koko kasvaa auton sylinteritilavuuden kasvaessa, ja suurimpien autojen aiheuttamat vahingot ovat selvästi isompia kuin pienimpien, mutta keskimmaisissa sylinteritilavuusluokissa vahingon koko ei juuri vaihtele. Sylinteritilavuuden χ^2 -arvo on yli 5%. Automalliluokka ei χ^2 -arvon perusteella ole tilastollisesti merkitsevä. Suurin osa autojen malleista on keskimmaisissa luokissa, joissa vahingon koko ei juuri muutu. Auton ikä ja auton haltijan ikä ovat selvästi tilastollisesti merkitseviä. Suurimmat vahingot ajetaan alle 2-vuotiailla ja yli 21-vuotiailla autoilla. Nuorten miesten ja yli 70-vuotiaiden miesten vahingot ovat suurimpia.

5.3 Henkilövahingon vahinkotiheys

Käytetään henkilövahinkojen aineistona yksityishenkilöiden yksityiskäyttöisten henkilöautojen liikennevakuutuksen Pohjolan vuosien 2000-2002 aineistoa. Vuosiin 2000-2002 kohdistuva vakuutuskanta on 1 061 471 vakuutusvuotta. Vuosina 2000-2002 sattui 4000 henkilövahinkoa, joista vuoden 2004 syyskuun loppuun mennessä on maksettu ja varattu eläkepääomina ja muina vahinkokohtaisina varauksina 35 539 tuhatta euroa.

Sovitetaan vahinkotiheyksiin Poisson malli, jonka linkkifunktio on logaritmfunktio. Tariffitekijät ovat alue, sylinteritilavuus, auton merkki ja malli, auton ikä, haltijan sukupuolesta ja iästä muodostettu yhdistetty tekijä sekä bonus. Tariffitekijöiden luokittelu on sama kuin omaisuusvahinkoa mallinnettaessa luvuissa 5.1 ja 5.2. Mallin residuaalit ovat liitteen 6 kuvassa 1, devianssianalyysi on taulukossa 5.3 ja mallin sovituksen tulokset ovat liitteessä 5.

Taulukko 5.3. Henkilövahingon vahinkotiheysmallien devianssianalyysi.

Malli	Devianssi	Vapausasteet	χ^2	tn > χ^2
1	22 623.8			
+ AL	22 612.2	3	11.7	0.0087
+ SY	22 581.1	8	31.1	0.0001
+ AM	22 529.0	8	52.0	< 0.0001
+ AI	22 497.6	7	31.4	< 0.0001
+ HI · SP	21 963.7	23	533.8	< 0.0001
+ BO	21 753.1	5	210.6	< 0.0001

Residuaaleissa näkyy samat ilmiöt kuin omaisuusvahinkotiheysmallin residuaaleissa. Kaikki tekijät ovat tilastollisesti merkitseviä 1 %:n tasolla. Tarkastellaan liitettä 5. Vähiten henkilövahinkoja ajavat maaseudun pienten kuntien ja pääkaupunkiseuden autoilijat. Henkilövahinkotiheys kasvaa, kun siirrytään merkin ja mallin luokissa luokituksessa ylöspäin. Eniten henkilövahinkoja aiheuttavat 6-11 vuotiaat autot, ja alle 2 vuotiailla ja yli 18 vuotiaalla autoilla vahinkoja on vähemmän. Vahingot lisääntyvät sylinteritilavuuden suuretsa. Nuoret miehet ja naiset ajavat eniten vahinkoja. Vähiten henkilövahinkoja aiheuttavat 30-39 vuotiaat. 0 %:n bonusluokan vahinkotiheyden suhde 70 %:n bonusluokan vahinkotiheyteen on 2.34.

5.3 Henkilövahingon keskivahinko

Aineisto on sama kuin vahinkotiheyksiä mallinnettaessa. Henkilövahingot on tehty lopulliseksi Liikennevakuutuskeskuksen vuoden 2004 riskitutkimuksen selvityskertoimilla. Vahingot on indeksikorotettu vuoden 2005 tasolle riskitutkimuksessa käytetyllä korvaustasoindeksillä. Sovitetaan vahinkoihin gammajakaumamalli, jolla on linkkifunktio logaritmfunktio. Malli sovitetaan katkaistuihin henkilövahinkoihin käyttäen katkaisurajoina Liikennevakuutuskeskuksen riskitutkimuksen rajoja. Vahinkojen yliteosat voidaan riskimaksuja laskettaessa tasoittaa henkilöautojen riskimaksuun tai tasoitus voidaan tehdä myös muiden ajoneuvojen kuin henkilöautojen kesken. Vuoden 2004 syyskuun loppuun mennessä vuosina 2000-2002 sattuneita katkaistuja henkilövahinkoja oli maksettu vuoden 2005 tasossa 39 178

tuhatta euroa. Tariffitekijät ovat muuten samat kuin luvun 5.2 vahinkotiheysmallissa, mutta bonus ei ole mukana. Mallin residuaalit ovat liitteen 6 kuvassa 2, devianssianalyysi on taulukossa 5.4 ja mallin sovituksen tulokset ovat liitteessä 7.

Taulukko 5.4. Henkilövahingon keskivahinkomallien devianssianalyysi.

Malli	Devianssi	Vapausasteet	χ^2	$tn > \chi^2$
AL	20 606.6			
+ SY	20 395.2	8	42.3	< 0.0001
+ AM	20 077.3	8	63.6	< 0.0001
+ AI	19 954.9	7	24.5	0.0009
+ HI·SP	19 755.6	23	39.8	0.0160

Residuaalijakauman huippu on 0:n vasemmalla puolella, ja jakauma on vino oikealle, koska aineistossa on suuria eläkepääomia sisältäviä vahinkoja. Dispersioparametri ϕ on 5.0025, ja vahinkojen suuruuden variaatiokerroin on 2.2366, jotka osoittavat vahingon suuruuden suuren vaihtelun. Sylinteritilavuus, merkki ja malli, auton ikä sekä yhdistetty sukupuoli ja ikä ovat merkitseviä 5 %:n tasolla. Tarkastellaan liitettä 7. Vahingot ovat suurempia maaseudulla kuin pääkaupunkiseudun kaupungeissa. Vahingot kasvavat automalliluokan kasvaessa malliluokkaan 2 asti, jonka jälkeen ne pienevät. Yli 9:n vuotiaat henkilöautot aiheuttavat suurempia henkilövahinkoja kuin nuoremmat autot. Kahden ylimmän sylinteritilavuusluokan autojen vahingot ovat suuria. Nuorten ja yli 65 vuotiaiden vahingot ovat suurempia kuin muiden ikäryhmien autoilijoiden.

6. YHTEENVETO

Kun hinnoitellaan moottoriajoneuvovakuutuksia, yleistetyllä lineaarisella mallilla voidaan valita vahinkoriskiin vaikuttavat tekijät ja laskea tekijöiden vaikutus. Luvussa 2 on esitetty yleistetyllä lineaarisella mallin rakenne, mallin sovituksen hyvyyden arviointimenetelmät, mallin parametrien ratkaisumenetelmä sekä tarkasteltu yleistettyjä lineaarisia malleja kvasiiskottavuutta käyttäen. Luvussa 3 on tarkasteltu tariffin parametrien estimointia myös muilla yleisesti käytetyillä menetelmillä ja näiden muiden menetelmien ja yleistettyjen lineaaristen mallien välistä yhteyttä. Osalle mutta ei kaikille näistä perinteisistä menetelmistä löytyy menetelmää vastaava yleistetty lineaarisen malli. Luvussa 4 on esitetty esimerkkinä autovakuutuksen hinnoittelusta Baxter'in, Coutts'in ja Ross'in (1980) alun perin esittämä vahingon suuruuden mallinnus. Esimerkin avulla on tarkasteltu jakauman ja linkkifunktion valinnan merkitystä mallin sopivuuteen. Luvussa 6 on sovitettu yleistettyjä lineaarisia malleja Pohjo-lan yksityishenkilöiden liikennevakuutuksen henkilöautojen omaisuus- ja henkilövahinkoihin. Osoittautuu, että yleisesti käytössä olevien auton haltijan kotipaikan, auton sylinteritilavuuden ja merkin ja mallin lisäksi auton haltijan iällä ja auton iällä on vaikutusta sekä omaisuusvahinko- että henkilövahinkoriskiin.

LÄHDEVIITTEET

- [1] Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E., Thandi, N., (2004) *Practitioner's Guide to Generalized Linear Models*, Watson Wyatt.
- [2] Bailey, R.A. (1963) Insurance rates with minimum bias, *PCAS* , Vol. L, pp 4 –11.
- [3] Baxter, L.A. , Coutts, S.M. and Ross, G.A.F. (1980) Applications of linear models in motor insurance, in *Proceedings on 21st International Congress of Actuaries*. Zurich. pp 11-29.
- [4] Brockman, M. J., and Wright, T. S. (1992) Statistical model rating: making effective use of your data. *Journal of the Institute of Actuaries* 119, 457 .
- [5] Dobson, A. J. (2001) *An Introduction to Generalized Linear Models, Second Edition*. Chapman Hall/CRC, London.
- [6] van Eeghen, J., Greup, E.K. and Nijssen, J.A. (1983) *Rate Making, Surveys of Actuarial Studies 2*, Nationale-Nederlanden, N.V.
- [7] Firth, D., (1991) *Generalized Linear Models, Statistical Theory and Modelling*, Chapman and Hall.
- [8] Greene, William H., (2003) *Econometric analysis*, Prentice Hall, New Jersey.
- [9] Haberman, Steven and A. R. Renshaw, (1996) Generalized linear models and actuarial science, *The Statistician* 45(4) (1996), 407–436.
- [10] Himmelblau, David M., (1972) *Applied Nonlinear Programming*, McGraw-Hill, Inc.
- [11] Jung, J., (1968) On automobile insurance rate making, *AB*, Vol. 5, No.1.
- [12] Jørgensen, Bent, (1987) Exponential Dispersion Models, *J. R. Statist. Soc. B* 49(2), 127-162.
- [13] Jørgensen, Bent and Marta C. Paes de Souza, (1994) Fitting Tweedie's Compound Poisson Mode to Insurance Claims Data, *Scand. Actuarial J.* 1, 69–93.
- [14] Liikennevakuutuskeskus, (2004) *Liikennevakuutuksen riskitutkimus 2004*.
- [15] Lokki Olli, (1979) *Tilastojakaumien ja tilastollisen testauksen peruskäsitteitä, Tilastoaineiston analysointi*, Suomen Aktuaariyhdistyksen luentosarja, Insinööritieto Oy.
- [16] McCullagh, P. and J. A. Nelder, (1989) *Generalized Linear Models, Second Edition*, Chapman Hall, London.

- [17] Meyer Ulrich, (2002) Third Party Motor Insurance in Europe, Comparative Study of the Economic-Statistical Situation, University of Bamberg.
- [18] Mildenhall, S. J. (1999). A systematic relationship between minimum bias and generalized linear models, Proceedings of the Casualty Actuarial Society 86, 393-487.
- [19] M. Morlock, M. Zimpelmann, (1999) Influence of primary premium differentiation on bonus-malus systems, K Forum, Munich Reinsurance Company.
- [20] Murphy, K. P., Brockman, M. J., and Lee, P. K. W. (2000). Using generalized linear models to build dynamic pricing systems. Casualty Actuarial Forum, Winter 2000.
- [21] Nelder, J.A. and Wedderburn, R.W.M. (1972) Generalized linear Models, J. R. Statist. Soc. A 135 370-84.
- [22] Renshaw A. E., (1994) Modeling the Claims Process in the Presence of Covariates, ASTIN Bulletin 24(2), 265-286.
- [23] SAS/STAT Software, (1996) Changes and Enhancements through Release 6.11, SAS Institute Inc.
- [24] Wedderburn, R.W.M (1974) Quasilikelihood functions, generalized linear models and the Gauss-Newton method, Biometrika 61, 439-47.

Yksityiskäyttöisten henkilöautojen oman ajoneuvon vaunuvahingon keskivahinko											LIITE 1
Baxter, Coutts, Ross 1980											1 (3)
Tariffi- tekijät			Havaitut vahingot	Malli B-C-R OWLS	Malli B-C-R GLIM	Malli McC-N GLIM	Malli M GLIM	(5)-(6)	(5)-(7)	(5)-(8)	
Vak.ot.ikä (1)	Autoryhmä (2)	Auton ikä (3)	Lkm (4)	£ (5)	£ (6)	£ (7)	£ (8)	£ (9)	(10)	(11)	(12)
17-20	A	0-3	8	289	299	315	293	258	-10	-26	-4
17-20	A	4-7	8	282	274	291	267	258	8	-9	15
17-20	A	8-9	4	133	222	237	198	258	-89	-104	-65
17-20	A	10-	1	160	172	184	132	258	-12	-24	28
17-20	B	0-3	10	372	296	310	290	266	76	62	82
17-20	B	4-7	28	249	272	286	264	266	-23	-37	-15
17-20	B	8-9	1	288	219	231	196	266	69	57	92
17-20	B	10-	1	11	170	179	132	266	-159	-168	-121
17-20	C	0-3	9	189	333	346	358	312	-144	-157	-169
17-20	C	4-7	13	288	309	322	319	312	-21	-34	-31
17-20	C	8-9	1	179	256	267	225	312	-77	-88	-46
17-20	C	10-	0		206	215	144	312			
17-20	D	0-3	3	763	407	418	503	389	356	345	260
17-20	D	4-7	2	850	383	394	430	389	467	456	420
17-20	D	8-9	0		330	339	275	389			
17-20	D	10-	0		281	287	163	389			
21-24	A	0-3	18	302	293	308	285	261	9	-6	17
21-24	A	4-7	31	194	269	283	260	261	-75	-89	-66
21-24	A	8-9	10	135	216	229	194	261	-81	-94	-59
21-24	A	10-	4	166	166	176	130	261	0	-10	36
21-24	B	0-3	59	420	291	302	282	270	130	118	138
21-24	B	4-7	96	243	266	278	257	270	-23	-35	-14
21-24	B	8-9	13	196	214	223	192	270	-18	-27	4
21-24	B	10-	3	135	164	171	130	270	-29	-36	5
21-24	C	0-3	44	268	327	338	345	315	-59	-70	-77
21-24	C	4-7	39	343	303	314	309	315	40	29	34
21-24	C	8-9	7	293	250	260	220	315	43	34	73
21-24	C	10-	2	104	201	207	142	315	-97	-103	-38
21-24	D	0-3	24	407	402	410	478	393	5	-3	-71
21-24	D	4-7	18	320	377	386	412	393	-57	-66	-92
21-24	D	8-9	2	205	325	332	267	393	-120	-127	-62
21-24	D	10-	0		275	279	160	393			
25-29	A	0-3	56	268	274	280	266	241	-6	-12	2
25-29	A	4-7	55	285	250	255	244	241	35	30	41
25-29	A	8-9	17	181	197	201	185	241	-16	-20	-4
25-29	A	10-	12	110	147	148	126	241	-37	-38	-16
25-29	B	0-3	125	275	272	274	263	250	4	1	12
25-29	B	4-7	172	243	247	250	242	250	-4	-7	1
25-29	B	8-9	36	179	195	195	184	250	-16	-16	-5
25-29	B	10-	10	264	145	143	126	250	119	121	138
25-29	C	0-3	163	334	308	311	318	295	26	24	16
25-29	C	4-7	129	274	284	286	287	295	-10	-12	-13
25-29	C	8-9	18	208	231	232	208	295	-23	-24	0
25-29	C	10-	8	150	182	179	137	295	-32	-29	13
25-29	D	0-3	72	383	383	383	428	372	0	1	-45
25-29	D	4-7	50	305	358	358	374	372	-53	-53	-69

Tariffi- tekijät	Havaitut vahingot			Malli B-C-R		Malli B-C-R		Malli McC-N		Malli M		2 (3)		
	Vak.ot.ikä (1)	Autoryhmä (2)	Auton ikä (3)	Lkm (4)	£ (5)	OWLS £ (6)	GLIM £ (7)	GLIM £ (8)	GLIM £ (9)	(5)-(6) (10)	(5)-(7) (11)	(5)-(8) (12)		
25-29	D	8-9	6	116	306	304	251	372	-190	-188	-135			
25-29	D	10-	1	636	256	251	154	372	380	385	482			
30-34	A	0-3	43	236	265	264	258	228	-29	-28	-22			
30-34	A	4-7	53	270	241	240	238	228	29	30	32			
30-34	A	8-9	15	160	189	185	181	228	-29	-25	-21			
30-34	A	10-	12	110	139	133	125	228	-29	-23	-15			
30-34	B	0-3	179	259	263	258	256	237	-4	1	3			
30-34	B	4-7	211	226	239	234	236	237	-13	-8	-10			
30-34	B	8-9	39	161	186	180	180	237	-25	-19	-19			
30-34	B	10-	19	107	136	127	124	237	-29	-20	-17			
30-34	C	0-3	197	340	300	295	307	282	41	45	33			
30-34	C	4-7	125	260	275	271	278	282	-15	-11	-18			
30-34	C	8-9	30	189	223	216	204	282	-34	-27	-15			
30-34	C	10-	9	104	173	164	135	282	-69	-60	-31			
30-34	D	0-3	104	400	374	367	408	360	26	33	-8			
30-34	D	4-7	55	349	350	343	359	360	-1	7	-10			
30-34	D	8-9	8	147	297	288	244	360	-150	-141	-97			
30-34	D	10-	2	65	247	236	151	360	-182	-171	-86			
35-39	A	0-3	43	207	211	216	209	180	-4	-9	-2			
35-39	A	4-7	73	129	186	192	195	180	-57	-63	-66			
35-39	A	8-9	21	157	134	137	155	180	23	20	2			
35-39	A	10-	14	113	84	85	112	180	29	28	1			
35-39	B	0-3	191	208	208	210	208	188	0	-2	0			
35-39	B	4-7	219	214	184	186	194	188	30	28	20			
35-39	B	8-9	46	149	131	132	155	188	18	17	-6			
35-39	B	10-	23	137	82	79	111	188	55	58	26			
35-39	C	0-3	210	251	245	247	240	233	6	4	11			
35-39	C	4-7	131	232	221	223	222	233	11	10	10			
35-39	C	8-9	32	203	168	168	172	233	35	35	31			
35-39	C	10-	8	141	118	116	120	233	23	26	21			
35-39	D	0-3	119	233	319	319	298	311	-86	-86	-65			
35-39	D	4-7	43	325	295	295	271	311	30	31	54			
35-39	D	8-9	4	207	243	240	200	311	-36	-33	7			
35-39	D	10-	0		193	188	133	311						
40-49	A	0-3	90	254	232	235	228	195	22	19	26			
40-49	A	4-7	98	213	207	211	212	195	6	2	1			
40-49	A	8-9	35	149	155	157	166	195	-6	-8	-17			
40-49	A	10-	22	98	105	104	117	195	-7	-6	-19			
40-49	B	0-3	380	218	229	230	226	204	-11	-12	-8			
40-49	B	4-7	434	209	205	206	210	204	4	4	-1			
40-49	B	8-9	97	172	152	151	165	204	20	21	7			
40-49	B	10-	59	110	103	99	117	204	7	12	-7			
40-49	C	0-3	401	239	266	266	266	249	-27	-27	-27			
40-49	C	4-7	253	250	242	242	244	249	8	8	6			
40-49	C	8-9	50	174	189	187	185	249	-15	-13	-11			
40-49	C	10-	15	129	139	135	126	249	-10	-6	3			
40-49	D	0-3	199	387	340	338	338	326	47	49	49			
40-49	D	4-7	88	299	316	314	303	326	-17	-15	-4			
40-49	D	8-9	8	325	264	259	217	326	62	66	108			
40-49	D	10-	9	137	214	207	141	326	-77	-70	-4			

Tariffi-tekijät			Havaitut vahingot	Malli B-C-R OWLS	Malli B-C-R GLIM	Malli McC-N GLIM	Malli M GLIM	2 (3)				
Vak.ot.ikä (1)	Autoryhmä (2)	Auton ikä (3)	Lkm (4)	£ (5)	£ (6)	£ (7)	£ (8)	£ (9)	(5)-(6) (10)	(5)-(7) (11)	(5)-(8) (12)	
50-59	A	0-3	69	251	235	223	231	198	16	28	20	
50-59	A	4-7	120	227	211	199	215	198	16	28	12	
50-59	A	8-9	42	172	158	144	167	198	14	28	5	
50-59	A	10-	35	98	109	92	118	198	-11	6	-20	
50-59	B	0-3	366	196	233	218	229	207	-37	-22	-33	
50-59	B	4-7	353	229	209	193	213	207	20	36	16	
50-59	B	8-9	95	164	156	139	166	207	8	25	-2	
50-59	B	10-	45	132	106	86	117	207	26	46	15	
50-59	C	0-3	310	268	269	254	269	252	-1	14	-1	
50-59	C	4-7	148	250	245	230	247	252	5	20	3	
50-59	C	8-9	33	175	193	175	186	252	-18	0	-11	
50-59	C	10-	13	152	143	123	127	252	9	29	25	
50-59	D	0-3	105	391	344	326	344	330	47	65	47	
50-59	D	4-7	46	228	320	302	309	330	-92	-74	-81	
50-59	D	8-9	10	346	267	247	219	330	79	99	127	
50-59	D	10-	1	167	217	195	142	330	-50	-28	25	
60 -	A	0-3	64	264	235	239	231	193	29	25	33	
60 -	A	4-7	100	198	211	215	214	193	-13	-17	-16	
60 -	A	8-9	43	167	159	160	167	193	8	7	0	
60 -	A	10-	53	144	109	108	118	193	35	36	26	
60 -	B	0-3	228	224	233	234	229	202	-9	-10	-5	
60 -	B	4-7	233	193	209	209	213	202	-16	-16	-20	
60 -	B	8-9	73	178	156	155	166	202	22	23	12	
60 -	B	10-	44	101	106	102	117	202	-5	-1	-16	
60 -	C	0-3	183	269	270	270	269	247	-1	-1	0	
60 -	C	4-7	103	258	245	246	247	247	13	12	11	
60 -	C	8-9	20	227	193	191	186	247	34	36	41	
60 -	C	10-	6	119	143	139	127	247	-24	-20	-8	
60 -	D	0-3	62	385	344	342	344	324	41	43	41	
60 -	D	4-7	22	324	320	318	308	324	4	6	16	
60 -	D	8-9	6	192	267	263	219	324	-75	-71	-27	
60 -	D	10-	6	123	218	211	142	324	-95	-88	-19	
									8942	-163	-113	753

Sarakkeet

- 1 - 3 Tariffitekijöiden arvot
4 Havaittu vahinkojen lukumäärä
5 Havaittu vahingon suuruus (£)
6 Baxter, Coutts, Ross : painotettu pienin neliösumma
7 Baxter, Coutts, Ross : normaali, identiteetti
8 McCullagh, Nelder : gamma, käänteis
9 Mildenhall : gamma, identiteetti
10 - 12 Havaittu - mallin arvo , malleille 6 - 8

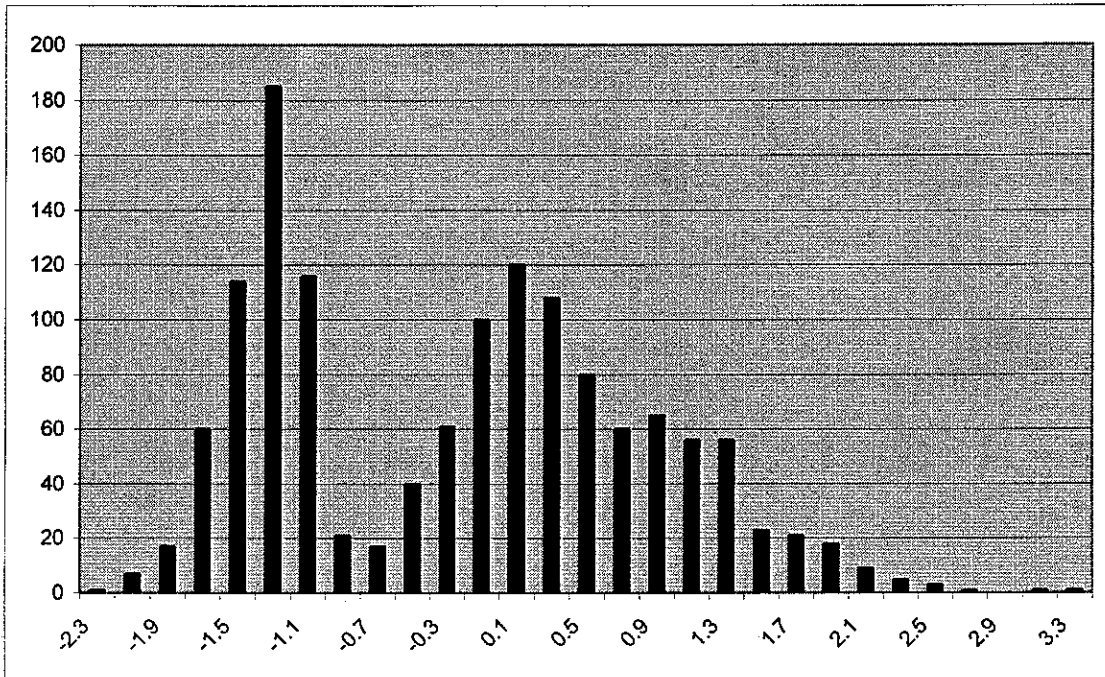
LIITE 2. Liikennevakuutuksen omaisuusvahingon vahinkotiheys. Poisson-log malli.

Parametri	Estimaatti	Keskiahajonta	95 %:n luottamusvälit	
Vakio	-3.7153	0.0502	-3.8138	-3.6168
AL 1	0.1777	0.0274	0.1241	0.2313
AL 2	0.0449	0.0284	-0.0107	0.1005
AL 3	0.0000	0.0000	0.0000	0.0000
AL 4	-0.2400	0.0302	-0.2991	-0.1808
SY -1050	-0.2842	0.0729	-0.4271	-0.1414
SY 1051-1250	-0.1982	0.0501	-0.2964	-0.0999
SY 1251-1350	-0.1097	0.0360	-0.1802	-0.0392
SY 1351-1550	-0.0896	0.0363	-0.1608	-0.0185
SY 1551-1650	0.0000	0.0000	0.0000	0.0000
SY 1651-1850	0.1286	0.0361	0.0579	0.1993
SY 1851-2150	0.1517	0.0309	0.0912	0.2122
SY 2151-2650	0.1560	0.0486	0.0607	0.2513
SY 2651-	0.2505	0.0691	0.1152	0.3859
AM -4	-0.2812	0.0988	-0.4749	-0.0875
AM -3	-0.3107	0.0957	-0.4983	-0.1232
AM -2	-0.1184	0.0786	-0.2725	0.0357
AM -1	-0.1536	0.0438	-0.2394	-0.0678
AM 0	-0.0922	0.0306	-0.1522	-0.0322
AM 1	-0.0579	0.0284	-0.1137	-0.0022
AM 2	0.0000	0.0000	0.0000	0.0000
AM 3	0.0786	0.0344	0.0112	0.1459
AM 4	0.0415	0.0558	-0.0680	0.1509
AI 0-2	-0.1145	0.0367	-0.1865	-0.0425
AI 3-5	-0.0673	0.0344	-0.1346	0.0001
AI 6-8	-0.0302	0.0383	-0.1054	0.0449
AI 9-11	0.0504	0.0355	-0.0192	0.1200
AI 12-14	0.0000	0.0000	0.0000	0.0000
AI 15-17	0.0118	0.0337	-0.0543	0.0779
AI 18-20	-0.0629	0.0475	-0.1559	0.0302
AI 21 -	-0.3946	0.0714	-0.5345	-0.2547
B0 0	0.8523	0.0366	0.7805	0.9241
B0 5-15	0.3767	0.0389	0.3006	0.4529
B0 20-30	0.3179	0.0393	0.2409	0.3949
B0 35-45	0.2908	0.0428	0.2069	0.3748
B0 50-65	0.2627	0.0328	0.1984	0.3271
B0 70	0.0000	0.0000	0.0000	0.0000
HI·SP M -21	0.6833	0.0599	0.5658	0.8007
HI·SP M 22-24	0.1920	0.0660	0.0627	0.3214
HI·SP M 25-29	-0.0769	0.0578	-0.1902	0.0364
HI·SP M 30-34	-0.2068	0.0575	-0.3195	-0.0941
HI·SP M 35-39	-0.1651	0.0539	-0.2708	-0.0594
HI·SP M 40-44	-0.0349	0.0516	-0.1360	0.0662
HI·SP M 45-49	0.0092	0.0509	-0.0905	0.1089
HI·SP M 50-54	0.0000	0.0000	0.0000	0.0000
HI·SP M 55-59	-0.0640	0.0550	-0.1718	0.0438
HI·SP M 60-64	-0.1252	0.0630	-0.2487	-0.0017

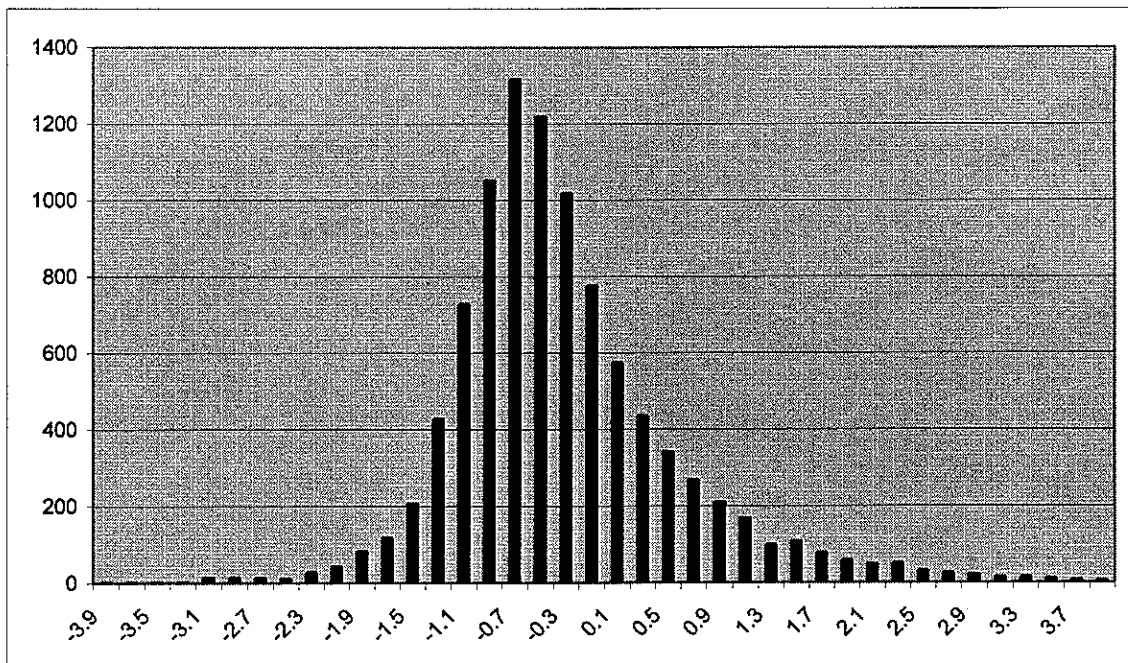
Parametri	Estimaatti	Keskihajonta	95 %:n luottamusvälit	
HI · SP M 65-69	-0.1702	0.0711	-0.3095	-0.0308
HI · SP M 70-	0.1556	0.0552	0.0473	0.2638
HI · SP N -21	0.4876	0.0938	0.3037	0.6715
HI · SP N 22-24	0.0375	0.1061	-0.1704	0.2454
HI · SP N 25-29	-0.1944	0.0824	-0.3559	-0.0329
HI · SP N 30-34	-0.2479	0.0760	-0.3969	-0.0989
HI · SP N 35-39	-0.0879	0.0656	-0.2164	0.0406
HI · SP N 40-44	0.0321	0.0618	-0.0890	0.1532
HI · SP N 45-49	0.1430	0.0602	0.0251	0.2609
HI · SP N 50-54	0.0556	0.0620	-0.0658	0.1771
HI · SP N 55-59	-0.0025	0.0691	-0.1380	0.1330
HI · SP N 60-64	-0.1047	0.0895	-0.2801	0.0707
HI · SP N 65-69	0.0290	0.1030	-0.1730	0.2309
HI · SP N 70-	0.3314	0.0823	0.1702	0.4926

LIITE 3.

Kuva 1. Liikennevakuutuksen omaisuusvahingon vahinkotiheys.
Standardoidut devianssiresiduaalit soluista, joissa vähintään 30 vakuutusvuotta.



Kuva 2. Liikennevakuutuksen omaisuusvahingon keskivahingon standardoidut devianssiresiduaalit.



LIITE 4. Liikennevakuutuksen omaisuusvahingon keskivahinko.Gamma-log malli.

Parametri	Estimaatti	Keskihajonta	95 %:n luottamusvälit	
AL 1	7.3195	0.0504	7.2208	7.4182
AL 2	7.2364	0.0500	7.1383	7.3345
AL 3	7.2777	0.0501	7.1794	7.3759
AL 4	7.3041	0.0515	7.2032	7.4051
SY -1050	-0.0661	0.0730	-0.2091	0.0769
SY 1051-1250	-0.0536	0.0504	-0.1524	0.0452
SY 1251-1350	0.0118	0.0362	-0.0592	0.0828
SY 1351-1550	-0.0401	0.0365	-0.1117	0.0315
SY 1551-1650	0.0000	0.0000	-0.0000	0.0000
SY 1651-1850	-0.0358	0.0363	-0.1070	0.0353
SY 1851-2150	0.0216	0.0309	-0.0389	0.0822
SY 2151-2650	0.0926	0.0498	-0.0050	0.1903
SY 2651-	-0.0164	0.0712	-0.1559	0.1231
AM -4	-0.1125	0.0989	-0.3063	0.0814
AM -3	-0.0156	0.0955	-0.2028	0.1715
AM -2	-0.0237	0.0788	-0.1782	0.1308
AM -1	0.0119	0.0438	-0.0740	0.0978
AM 0	0.0304	0.0309	-0.0302	0.0910
AM 1	0.0030	0.0287	-0.0534	0.0593
AM 2	0.0000	0.0000	0.0000	0.0000
AM 3	0.0904	0.0345	0.0227	0.1581
AM 4	0.0415	0.0560	-0.0684	0.1513
AI 0-2	0.1991	0.0371	0.1264	0.2718
AI 3-5	0.0283	0.0346	-0.0395	0.0961
AI 6-8	0.0596	0.0386	-0.0160	0.1352
AI 9-11	0.0768	0.0356	0.0070	0.1465
AI 12-14	0.0000	0.0000	0.0000	0.0000
AI 15-17	0.0313	0.0339	-0.0351	0.0977
AI 18-20	-0.0199	0.0475	-0.1130	0.0731
AI 21 -	0.1500	0.0699	0.0130	0.2870
HI·SP M -21	0.1303	0.0560	0.0205	0.2401
HI·SP M 22-24	0.0440	0.0640	-0.0813	0.1694
HI·SP M 25-29	0.1165	0.0563	0.0062	0.2269
HI·SP M 30-34	-0.0137	0.0572	-0.1257	0.0984
HI·SP M 35-39	0.0522	0.0539	-0.0534	0.1578
HI·SP M 40-44	0.0169	0.0517	-0.0844	0.1182
HI·SP M 45-49	0.0893	0.0515	-0.0116	0.1901
HI·SP M 50-54	0.0000	0.0000	0.0000	0.0000
HI·SP M 55-59	-0.0301	0.0552	-0.1383	0.0781
HI·SP M 60-64	0.0220	0.0631	-0.1017	0.1457
HI·SP M 65-69	0.0129	0.0711	-0.1263	0.1522
HI·SP M 70-	0.1433	0.0552	0.0351	0.2516
HI·SP N -21	-0.0075	0.0914	-0.1866	0.1717

Parametri	Estimaatti	Keskihajonta	95 %:n luottamusvälit	
HI · SP N 22-24	-0.0313	0.1047	-0.2366	0.1739
HI · SP N 25-29	-0.1004	0.0813	-0.2598	0.0590
HI · SP N 30-34	-0.0974	0.0755	-0.2453	0.0506
HI · SP N 35-39	0.1647	0.0657	0.0359	0.2934
HI · SP N 40-44	0.0567	0.0618	-0.0643	0.1777
HI · SP N 45-49	-0.0387	0.0602	-0.1567	0.0792
HI · SP N 50-54	0.0145	0.0620	-0.1071	0.1361
HI · SP N 55-59	-0.1905	0.0692	-0.3262	-0.0549
HI · SP N 60-64	0.0948	0.0896	-0.0809	0.2704
HI · SP N 65-69	-0.0810	0.1032	-0.2832	0.1212
HI · SP N 70-	0.0242	0.0821	-0.1367	0.1851

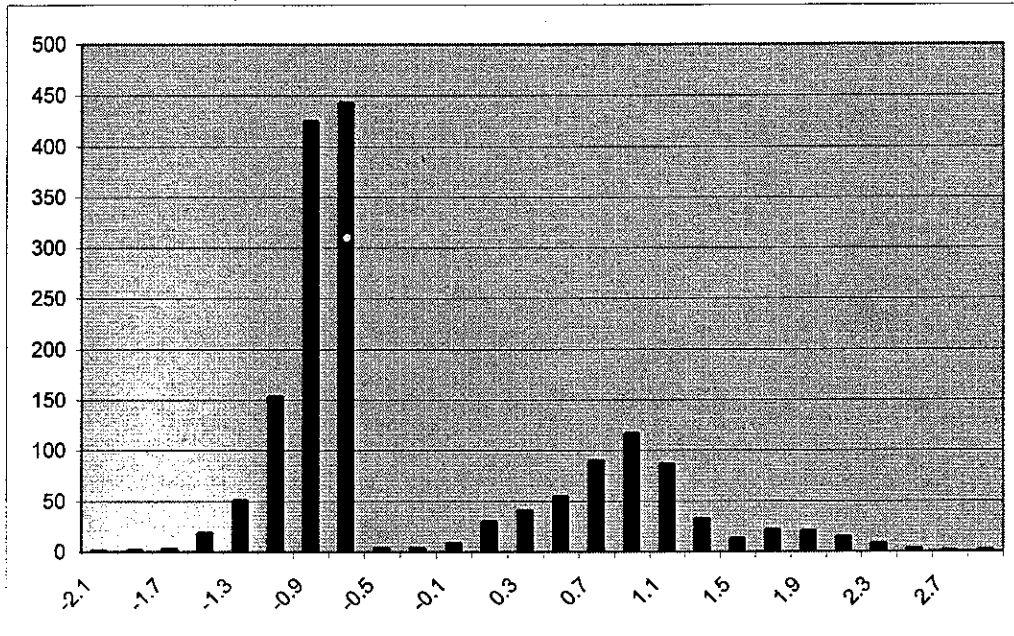
LIITE 5. Liikennevakuutuksen henkilövahingon vahinkotiheys. Poisson-log malli.

Parametri	Estimaatti	Keskihajonta	95 %:n luottamusvälit	
Vakio	-5.8014	0.0783	-5.9548	-5.6480
AL 1	-0.0431	0.0444	-0.1301	0.0439
AL 2	0.0334	0.0417	-0.0484	0.1152
AL 3	0.0000	0.0000	0.0000	0.0000
AL 4	-0.1105	0.0446	-0.1979	-0.0231
SY -1050	-0.4731	0.1201	-0.7085	-0.2377
SY 1051-1250	0.0157	0.0712	-0.1239	0.1554
SY 1251-1350	-0.0147	0.0541	-0.1207	0.0912
SY 1351-1550	-0.0293	0.0562	-0.1394	0.0809
SY 1551-1650	0.0000	0.0000	0.0000	0.0000
SY 1651-1850	0.2003	0.0571	0.0884	0.3122
SY 1851-2150	0.1604	0.0495	0.0634	0.2575
SY 2151-2650	0.2073	0.0783	0.0538	0.3608
SY 2651-	0.1620	0.1185	-0.0702	0.3943
AM -4	-0.7223	0.1866	-1.0881	-0.3565
AM -3	-0.0792	0.1362	-0.3461	0.1878
AM -2	-0.1522	0.1149	-0.3775	0.0731
AM -1	-0.1073	0.0707	-0.2460	0.0313
AM 0	-0.0940	0.0483	-0.1886	0.0006
AM 1	-0.0046	0.0440	-0.0909	0.0816
AM 2	0.0000	0.0000	0.0000	0.0000
AM 3	0.0915	0.0540	-0.0143	0.1973
AM 4	0.1168	0.0834	-0.0465	0.2802
AI 0 - 2	-0.0156	0.0568	-0.1270	0.0958
AI 3 - 5	0.0044	0.0568	-0.1069	0.1158
AI 6 - 8	0.1160	0.0603	-0.0023	0.2342
AI 9 - 11	0.1124	0.0508	0.0129	0.2119
AI 12 - 14	0.0000	0.0000	0.0000	0.0000
AI 15 - 17	0.0090	0.0549	-0.0987	0.1166
AI 18 - 20	-0.0797	0.0785	-0.2335	0.0741
AI 21 -	-0.2888	0.1109	-0.5061	-0.0714
HI · SP M -21	0.8758	0.0876	0.7040	1.0475
HI · SP M 22-24	0.3830	0.0953	0.1963	0.5697
HI · SP M 25-29	-0.1619	0.0922	-0.3427	0.0188
HI · SP M 30-34	-0.2404	0.0889	-0.4147	-0.0662
HI · SP M 35-39	-0.2216	0.0845	-0.3873	-0.0560
HI · SP M 40-44	-0.1400	0.0819	-0.3006	0.0206
HI · SP M 45-49	-0.1330	0.0817	-0.2932	0.0271
HI · SP M 50-54	0.0000	0.0000	0.0000	0.0000
HI · SP M 55-59	-0.1391	0.0896	-0.3148	0.0366
HI · SP M 60-64	-0.1076	0.0988	-0.3013	0.0861
HI · SP M 65-69	-0.0527	0.1065	-0.2614	0.1560
HI · SP M 70-	0.3068	0.0834	0.1434	0.4703
HI · SP N -21	0.8020	0.1276	0.5520	1.0520

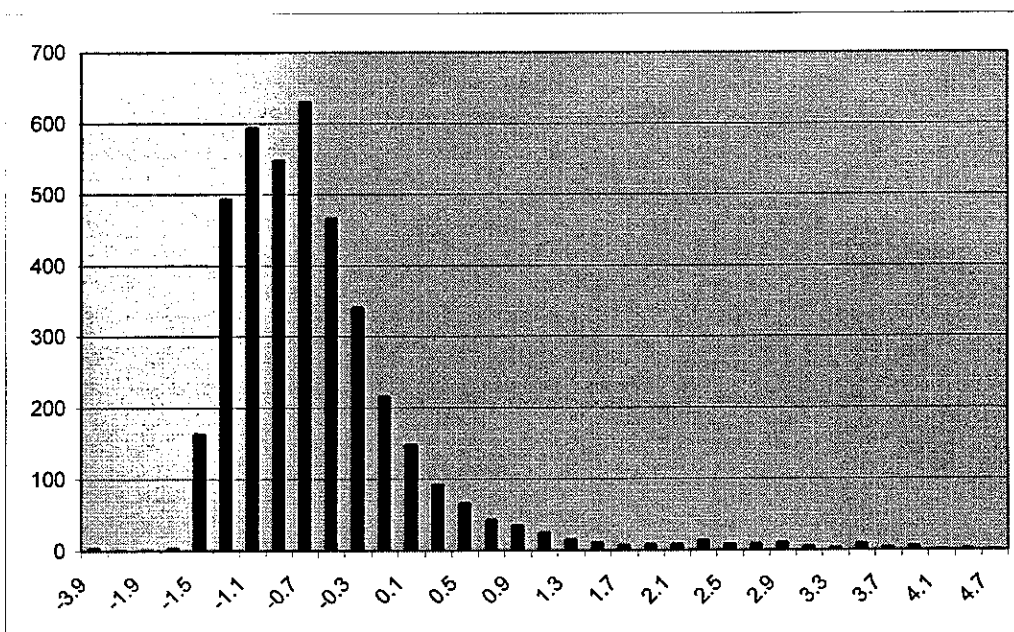
Parametri	Estimaatti	Keskihajonta	95 %:n luottamusvälit	
HI · SP N 22-24	0.0915	0.1573	-0.2169	0.3998
HI · SP N 25-29	-0.0776	0.1220	-0.3168	0.1616
HI · SP N 30-34	-0.2750	0.1166	-0.5034	-0.0465
HI · SP N 35-39	-0.2387	0.1070	-0.4485	-0.0289
HI · SP N 40-44	-0.1032	0.1005	-0.3001	0.0938
HI · SP N 45-49	0.0572	0.0962	-0.1313	0.2457
HI · SP N 50-54	0.1465	0.0933	-0.0364	0.3294
HI · SP N 55-59	0.1134	0.1077	-0.0976	0.3244
HI · SP N 60-64	-0.0762	0.1420	-0.3545	0.2021
HI · SP N 65-69	0.0667	0.1612	-0.2492	0.3826
HI · SP N 70-	0.2993	0.1351	0.0345	0.5641
B0 0	0.8482	0.0561	0.7382	0.9582
B0 5-15	0.3817	0.0598	0.2645	0.4990
B0 20-30	0.2837	0.0627	0.1610	0.4065
B0 35-45	0.1613	0.0708	0.0225	0.3000
B0 50-65	0.2254	0.0518	0.1239	0.3270
B0 70	0.0000	0.0000	0.0000	0.0000

LIITE 6.

Kuva 1. Liikennevakuutuksen henkilövahingon vahinkotiheys.
Standardoidut devianssiresiduaalit soluista, joissa vähintään 80 vakuutusvuotta.



Kuva 2. Liikennevakuutuksen henkilövahingon keskivahingon standardoidut devianssiresiduaalit.



LIITE 7. Liikennevakuutuksen henkilövahingon keskivahinko. Gamma-log malli.

Parametri	Estimaatti	Keskihajonta	95%:n luottamusvälit	
AL 1	8.9977	0.1915	8.6224	9.3730
AL 2	9.1647	0.1788	8.8143	9.5152
AL 3	9.2012	0.1917	8.9013	9.6527
AL 4	9.2770	0.1837	8.8412	9.5613
SY -1050	-0.1857	0.3144	-0.8020	0.4306
SY 1051-1250	0.0935	0.1796	-0.2586	0.4456
SY 1251-1350	0.1090	0.1347	-0.1549	0.3730
SY 1351-1550	-0.1737	0.1362	-0.4407	0.0932
SY 1551-1650	0.0000	0.0000	0.0000	0.0000
SY 1651-1850	-0.2408	0.1416	-0.5183	0.0366
SY 1851-2150	0.0114	0.1162	-0.2165	0.2392
SY 2151-2650	0.5180	0.1874	0.1507	0.8854
SY 2651-	0.7797	0.2810	0.2290	1.3304
AM -4	-0.9740	0.4240	-1.8050	-0.1430
AM -3	-0.3746	0.3240	-1.0097	0.2604
AM -2	-0.9158	0.2718	-1.4485	-0.3832
AM -1	-0.2629	0.1694	-0.5950	0.0692
AM 0	-0.1958	0.1160	-0.4233	0.0316
AM 1	-0.5904	0.1063	-0.7987	-0.3821
AM 2	0.0000	0.0000	0.0000	0.0000
AM 3	-0.6165	0.1349	-0.8809	-0.3521
AM 4	-0.8667	0.1913	-1.2416	-0.4917
AI 0-2	-0.0536	0.1409	-0.3297	0.2224
AI 3-5	0.0280	0.1428	-0.2519	0.3080
AI 6-8	-0.2914	0.1486	-0.5826	-0.0002
AI 9-11	0.2245	0.1269	-0.0243	0.4733
AI 12-14	0.0000	0.0000	0.0000	0.0000
AI 15-17	0.1059	0.1382	-0.1650	0.3768
AI 18-20	-0.3577	0.1908	-0.7317	0.0163
AI 21 -	0.2641	0.2510	-0.2278	0.7561
HI-SP M -21	0.3782	0.1961	-0.0061	0.7624
HI-SP M 22-24	0.0680	0.2246	-0.3721	0.5082
HI-SP M 25-29	0.5666	0.2166	0.1421	0.9911
HI-SP M 30-34	0.0681	0.2156	-0.3544	0.4906
HI-SP M 35-39	0.5526	0.2013	0.1580	0.9471
HI-SP M 40-44	0.2399	0.1916	-0.1356	0.6153
HI-SP M 45-49	0.1118	0.1965	-0.2733	0.4968
HI-SP M 50-54	0.0000	0.0000	0.0000	0.0000
HI-SP M 55-59	-0.1145	0.2101	-0.5262	0.2972
HI-SP M 60-64	-0.1758	0.2345	-0.6354	0.2839
HI-SP M 65-69	0.6410	0.2474	0.1561	1.1259
HI-SP M 70-	0.2505	0.1983	-0.1382	0.6391
HI-SP N -21	0.3894	0.2937	-0.1862	0.9650

HI · SP N 22-24	0.0341	0.3529	-0.6576	0.7258
HI · SP N 25-29	0.6427	0.2948	0.0648	1.2206
HI · SP N 30-34	0.7907	0.2700	0.2616	1.3198
HI · SP N 35-39	0.0966	0.2510	-0.3953	0.5886
HI · SP N 40-44	0.1736	0.2315	-0.2801	0.6273
HI · SP N 45-49	0.2242	0.2258	-0.2184	0.6669
HI · SP N 50-54	0.1790	0.2209	-0.2541	0.6120
HI · SP N 55-59	0.4100	0.2554	-0.0905	0.9106
HI · SP N 60-64	0.0287	0.3733	-0.7030	0.7604
HI · SP N 65-69	0.1506	0.3806	-0.5955	0.8966
HI · SP N 70-	0.1015	0.3158	-0.5175	0.7205

