Tampereen yliopisto

Miikka Virtanen

Vakuutusyhtiön riskiprosessin mallintaminen RiskDemo-ohjelmalla

Informaatioteknologian ja viestinnän tiedekunta
Tilastollisen data-analyysin kandidaattitutkielma
Huhtikuu 2021

Tiivistelmä

Miikka Virtanen: Vakuutusyhtiön riskiprosessin mallintaminen
RiskDemo-ohjelmalla
Kandidaattitutkielma
Tampereen yliopisto
Matematiikan ja tilastollisen data-analyysin tutkinto-ohjelma
Huhtikuu 2021

Tässä tutkielmassa tarkastellaan vakuutusyhtiön riskiprosessin mallintamista RiskDemo-ohjelmalla. Riskiprosessin mallinnuksessa seurataan vakuutusyhtiön pääoman kehitystä. Tämän tutkielman tapauksessa mallinnus tapahtuu klassisen vararikkoteorian ja simuloinnin avulla. Klassisessa vararikkoteoriassa pääoman kehitykseen vaikuttavat vakuutusmaksutulot ja maksettavat vahingonkorvaukset. Siinä vahingonkorvaukset sattuvat Poisson-prosessin mukaisesti ja yksittäisen vahingonkorvauksen suuruus noudattaa jotakin tilastollista jakaumaa. Simuloinnin avulla tuotetaan havainnot vahingonkorvauksista satunnaislukugeneraattoreiden avulla.

RiskDemo vaatii mallinnusta varten tiedot vakuutusyhtiön alkupääomasta, vakuutusmaksutuloon liittyvästä varmuuslisästä, tarkasteltavasta aikavälistä, vahingonkorvausten määrään liittyvästä Poisson-prosessista ja yksittäisen vahingonkorvauksen suuruuden jakaumasta. Ohjelma käyttää yksittäisen vahingonkorvauksen suuruuden jakaumana gamma-jakaumaa. Ohjelman mallinnus tuottaa simulaatiokuvion ja laskee vararikkotodennäköisyyden ja vararikkotodennäköisyyden ylärajaan liittyvän Lundbergin eksponentin.

Tässä tutkielmassa tarkastellaan myös esimerkkiä, jossa RiskDemon avulla luotiin aineisto vararikkotodennäköisyyksistä. Aineistosta luotiin logistiset regressiomallit kymmenen ja kolmen vuoden aikaväleillä tarkasteltuna. Malleilla tarkasteltiin alkupääoman vaikutusta vararikkotodennäköisyyteen. Mallien perusteella alkupääoman kasvaessa vararikkotodennäköisyys pienenee. Lisäksi näyttää siltä, että tarkasteltavan aikavälin pidentyessä alkupääoman vaikutus pienenee. Malli on kuitenkin melko yksinkertainen ja mallien selityskertoimet ovatkin vain kohtalaiset. Varsinkin pienillä alkupääomilla malli ei onnistu selittämään vararikkoa kovinkaan hyvin.

Avainsanat: Klassinen vararikkoteoria, simulointi, R

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

Sisällys

1 Johdanto

Tässä tutkielmassa pyritään selvittämään, miten vakuutusyhtiön riskiprosessia pystytään mallintamaan RiskDemo-ohjelmalla. Riskiprosessin mallintaminen perustuu klassiseen vararikkoteoriaan ja simulointiin. Lisäksi tutkielman lopuksi tarkastellaan esimerkkinä RiskDemolla luotua aineistoa, johon sovitetaan yksinkertainen logistinen regressiomalli. Esimerkin avulla tarkastellaan vakuutusyhtiön alkupääoman vaikutusta vararikon todennäköisyyteen.

Riskiprosessin mallintamisen avulla vakuutusyhtiö pystyy arvioimaan pääomansa kehitystä ja sen avulla tekemään päätöksiä toiminnastaan, esimerkiksi kerättävän vakuutusmaksutulon suuruuteen vaikuttavan varmuuslisän määrästä tai riskien jälleenvakuuttamisesta. Mallintamisen ohessa lasketaan myös vakuutusyhtiön vararikkotodennäköisyys. Mallinnuksella ei kuitenkaan pystytä vastaamaan täysin tosielämän tilanteita, sillä on lähes mahdotonta huomioida kaikki pääoman kehitykseen liittyvät seikat mallissa. Vararikkotodennäköisyydessäkin on kyse vain matemaattisesta vararikosta eli vararikoksi katsotaan tilanne, jolloin pääoma on pienempi kuin nolla. Tämä ei tietenkään vastaa todellista vararikkoa. Vaikkei mallinnuksella saada aikaan tosielämää vastaavaa tilannetta, antaa se kuitenkin riittävän tarkan kuvan vakuutusyhtiön toiminnan kannattavuudesta, jolloin riittävän kattavan mallin perusteella voidaan tehdä johtopäätöksiä.

RiskDemon tuottama mallinnus on kuitenkin varsin yksinkertainen, sillä se ottaa huomioon ainoastaan vakuutusyhtiön alkupääoman, vakuutusmaksutulot sekä vahingonkorvaukset. Ohjelman tarkoituksena onkin lähinnä vain havainnollistaa klassista vararikkoteoriaa ja riskejä todennäköisyyslaskennan ja tilastotieteen kannalta.

Tutkielman lopussa olevan logistisen regressiomallin avulla huomataan alkupääoman ja vararikontodennäköisyyden välinen vaikutus. Alkupääoman kasvaessa vararikko tulee epätodennäköisemmäksi. Lisäksi nähdään, että myös tarkasteltavalla aikavälillä on merkitystä vaikutuksen suuruuteen.

2 Riskiprosessin mallintaminen

Riskiprosessin käsittely voidaan aloittaa tarkastelemalla riskin määritelmää. Riski on käsitteenä hankala, mutta se voidaan määritellä ei-toivotun tapahtuman mahdollisuudeksi. Riski voidaan jakaa puhtaaseen ja spekulatiiviseen riskiin, joista jälkimmäinen ottaa huomioon myös toivotun tapahtuman mahdollisuuden. Vakuutustoiminnassa tämä jako näkyy siten, että vakuutusyhtiö ottaa asiakkaiden puhtaita riskejä itselleen vakuutusmaksua vastaan, jolloin vakuutusyhtiön toiminnan kannalta riski on spekulatiivinen. (Koskinen 2018.) Vakuutusyhtiön kannalta toivottu tapahtuma on, että asiakkaiden riskit eivät toteudu.

Riskin tarkastelussa on tärkeää huomioida myös riskin suuruus. Sen määrittämiseen tarvitaan tietoa mahdollisen tappion suuruudesta ja tapahtuman todennäköisyydestä. (Koskinen 2018.) Tässä tutkielmassa tarkasteltavassa klassisessa vararikkoteoriassa riskin suuruus saadaan Poisson-prosessin ja yksittäisen vahingon suuruuden jakauman avulla.

Tässä luvussa esitetään riskiprosessin mallintamiseen liittyvän klassisen vararikkoteorian lauseita ja määritelmiä. Tämän lisäksi käsitellään simulointia. Ensimmäinen alaluku noudattaa Kaas ym. (2008, s. 89 – 93) esitystä klassisesta vararikkoteoriasta.

2.1 Klassinen vararikkoteoria

Vakuutusyhtiön riskiprosessia voidaan mallintaa klassisen vararikkoteorian avulla. Siinä kuvataan vakuutusyhtiön pääoman kehitystä alkupääoman, vakuutusmaksutulojen ja negatiivisten vahingonkorvausten summana. Riskiprosessi voidaan määritellä seuraavasti:

Määritelmä 2.1 (Riskiprosessi)**.**

$$U(t) = u + ct - S(t), \qquad (t \geq 0),$$

missä

$$\begin{aligned} U(t) &= \text{Vakuutusyhtiön pääoma ajanhetkellä } t, \\ u = U(0) &= \text{alkupääoma}, \\ c &= \text{vakuutusmaksutulo ajanhetkeä kohden ja} \\ S(t) &= X_1 + X_2 + \cdots + X_{N(t)}, \end{aligned}$$

jossa

$$\begin{aligned} N(t) &= \text{Vahingonkorvauksien lukumäärä ajanhetkeen } t \text{ mennessä ja} \\ X_i &= i\text{:nnen korvauksen suuruus, joka oletetaan ei-negatiiviseksi.} \end{aligned}$$

Vakuutusyhtiön pääoman kehitystä kuvataan määritelmän (2.1) mukaisesti. Satunnaismuuttujat $T_1, T_2, \ldots$ määrittävät ajanhetket, joilla vahingonkorvaukset toteutuvat. Prosessin kuvaaja nousee vakuutusmaksutulon c mukaisesti, kun vahingonkorvauksia ei ole. Vahingonkorvauksen toteutuessa ajanhetkellä $t = T_j$, kuvaaja laskee j:nnen vahingonkorvauksen suuruuden X_j verran. Vararikoksi kutsutaan hetkeä, jolloin toteutuneiden vahingonkorvauksien, $X_1, X_2, \ldots$, summa on suurempi kuin alkupääoman ja kerättyjen vakuutusmaksutulojen summa, eli sen hetkinen pääoma on pienempi kuin nolla. Ensimmäinen hetki, jolloin vararikko tapahtuu määritellään seuraavasti:

$$T = \begin{cases} min\{t | t \geq 0 \text{ ja } U(t) \leq 0\}, \\ \infty, \text{jos } U(t) \geq 0 \text{ kaikilla } t. \end{cases} \tag{2.1}$$

Vararikkotodennäköisyys eli todennäköisyys sille, tapahtuuko vararikkoa koskaan on sama kuin todennäköisyys sille, että T on äärellinen. Se voidaan kirjoittaa muotoon:

$$\psi(u) = P(T < \infty). \tag{2.2}$$

Seuraavaksi tarkastellaan prosessia $N(t)$, joka on vahingonkorvausten lukumäärä ajanhetkeen t mennessä. Oletetaan, että se on Poisson-prosessi:

Määritelmä 2.2 (Poisson-prosessi)**.** $N(t)$ on Poisson-prosessi, jos jollekin prosessin intensiteetille $\lambda > 0$ pitää paikkansa, että:

$$N(t+h) - N(t) \sim Poisson(\lambda h)$$

kaikilla $t > 0$, $h > 0$ ja jokaisella $N(s)$, $s \leq t$.

Määritelmästä (2.2) seuraa, että Poisson-prosessin lisäyksillä on seuraavat ominaisuudet:

Riippumattomuus: Lisäykset $N(t_i + h_i) - N(t_i)$ ovat riippumattomia ei-jatkuvilla väleillä $(t_i, t_i + h_i), i = 1, 2, \ldots$

Stationaarisuus: $N(t+h) - N(t)$ on Poisson(λh)-jakautunut jokaisella t:n arvolla.

$N(t)$:n ollessa Poisson-prosessi, $S(t)$ on yhdistetty Poisson-prosessi. Muuttujalla $t = t_0$ yhdistetyt vahingonkorvaukset $S(t_0)$ noudattaa yhdistettyä Poisson-jakaumaa parametrilla λt_0.

Varmuuslisä θ voidaan määritellä vakuutusmaksutulon,

$$c = (1+\theta)\lambda\mu_1, \tag{2.3}$$

avulla seuraavasti:

$$\theta = \frac{c}{\lambda\mu_1} - 1,$$

missä μ_1 on yksittäisen vahingonkorvauksen X_i ensimmäinen momentti $E[X_i^1]$ eli yksittäisen vahingonkorvauksen odotusarvo, kun vahingonkorvauksen X kertymäfunktio ja momentit määritellään seuraavasti:

$$F(x) = P(X_i \leq x); \ \mu_j = E[X_i^j], i, j = 1, 2, \ldots \tag{2.4}$$

Riskiprosessissa, jossa vahingonkorvaukset $X \geq 0$ ja $E[X] = \mu \geq 0$, Lundbergin eksponentti R voidaan määritellä seuraavasti:

Määritelmä 2.3 (Lundbergin eksponentti)**.**

$$1 + (1 + \theta)\mu_1 r = m_x(r),$$

jossa $m_x(r)$ on vahingonkorvauksen x momentit generoiva funktio. Lundbergin eksponentti R saadaan yhtälön positiivisena ratkaisuna muuttujalle r. Käytännössä yhtälöllä on yksi positiivinen ratkaisu, koska momentit generoiva funktio $m_x(r)$ on aidosti konveksi, sillä $m_x''(t) = E[X^2 e^{tX}] > 0$, $m_x'(0) < (1 + \theta)\mu_1$, ja, muutamaa poikkeusta lukuunottamatta, $m_x(t) \to \infty$ jatkuvana.

Vararikkotodennäköisyyden tarkkaa arvoa ei voida aina laskea, joten sille on määritetty seuraavanlainen ehdoton yläraja:

$$\psi(u) \leq e^{-Ru}, \tag{2.5}$$

jossa u on alkupääoma ja R on Lundbergin eksponentti. Epäyhtälö pätee kaikilla $u > 0$. Lisäksi voidaan tarkastella raja-arvoa

$$\lim_{u \to \infty} u^{-1} \log P(T < \infty) = -R, \tag{2.6}$$

jonka avulla voidaan osoittaa, että R on paras mahdollinen arvo, koska mielivaltaiselle $R' > R$ pätee

$$P(T < \infty) > e^{-R'u}, \tag{2.7}$$

jostakin alkupääoman u arvosta lähtien. Rajatarkastelu (yhtälö 2.6) voidaan tehdä, koska suuri alkupääoma u vastaa pientä vararikkotodennäköisyyttä ja usein tätä tarkastelua myös vaaditaan pitkän aikavälin tarkasteluissa. (Nyrhinen 2013, s. 88.)

Todistetaan vielä Lundbergin yläraja (yhtälö 2.5). Todistukseen tarvitaan tieto siitä, että määritelmän (2.3) yhtälö voidaan esittää muodossa $\lambda + cR = \lambda m_x(R)$.

Todistus. Olkoon $\psi_k(u)$ todennäköisyys, että vararikko tapahtuu viimeistään ajanhetkellä k, kun $-\infty < u < \infty$ ja $k = 1, 2, 3, \ldots$. Kun $k \to \infty$, $\psi_k(u)$ kasvaa kohti rajaansa $\psi(u)$ kaikilla alkupääoman u arvoilla. Täten riittää todistaa, että $\psi_k(u) \leq e^{-Ru}$ pätee jokaisella k:n arvolla. Arvolla $k = 0$ epäyhtälö pitää paikkansa, sillä $\psi_0(u) = 1$, kun $u < 0$, ja $\psi_0(u) = 0$, kun $u \geq 0$.

Jaetaan tapahtuma ”vararikko viimeistään ajanhetkellä k” ajan ja ensimmäisen vahingonkorvauksen suuruuden suhteen. Oletetaan, että se tapahtuu aikavälillä t ja $t+dt$. Tällä tapahtumalla voidaan ajatella olevan ’todennäköisyys’ $\lambda e^{-\lambda t} dt$. Oletetaan myös, että ensimmäisen vahingonkorvauksen suuruus on välillä x ja $x + dx$. Tämän todennäköisyys on $dP(x)$. Pääoma heti ajanhetken t jälkeen on siis $u + ct - x$. Integroidaan muuttujien x ja t suhteen:

$$\psi_k(u) = \int_0^\infty \int_0^\infty \psi_{k-1}(u + ct - x)\, dP(x) \lambda e^{-\lambda t} dt.$$

Oletetaan nyt, että induktio-oletus pitää paikkansa arvolla $k-1$, eli $\psi_{k-1}(u) \leq e^{-Ru}$ pätee kaikilla reaalisilla u:n arvoilla. Tästä saadaan

$$\begin{aligned}\psi_k(u) &\leq \int_0^\infty \int_0^\infty e^{-R(u+ct-x)} dP(x)\lambda e^{-\lambda t} dt \\ &= e^{-Ru} \int_0^\infty \lambda e^{-t(\lambda+cR)} dt \int_0^\infty e^{Rx} dP(x) \\ &= e^{-Ru} \frac{\lambda}{\lambda + cR} m_x(R) = e^{-Ru},\end{aligned}$$

jossa viimeinen yhtäsuuruus seuraa Lundbergin eksponentin määritelmän yhtälöstä (2.3). (Kaas et al. 2008, s.92 – 93.) □

2.2 Simuloinnin käyttö riskiprosessin mallinnuksessa

Tässä alaluvussa käsitellään simulaatiota tilastollisena menetelmänä ja sen käyttöä vakuutusyhtiön riskiprosessin mallinnuksessa. Alaluvun päälähteenä on Daykinin, Pentikäisen ja Pesosen (1994) teos *Practical Risk Theory for Actuaries*.

Simuloinnin ideana on jakaa tarkasteltava asia helposti käsiteltäviin osiin ja tuottaa näistä muuttujista havaintoja esimerkiksi satunnaislukugeneraattorien avulla. Satunnaislukugeneraattori voi olla esimerkiksi tietokonealgoritmi, joka tuottaa niin sanottuja pseudo-satunnaislukuja, jotka noudattavat määrättyä jakaumaa riittävällä tarkkuudella. Tarkoituksena on helpottaa monimutkaisten mallien käsittelyä tarjoamalla vaihtoehto analyyttisten ja tavanomaisten numeeristen ratkaisujen tilalle.

Vakuutusyhtiön riskiprosessin mallinnuksessa simulointia tarvitaan varsinkin vahingonkorvausten suuruuksien arvioimiseen. Numeeriset menetelmät saattavat olla monimutkaisia, ja lisää ongelmia tuottavat myös mahdolliset muuttujien väliset korrelaatiot ja yhdysvaikutukset. Monimutkaisten mallien kanssa toimimiseen simulointi tarjoaa joustavan ja tehokkaan ratkaisun.

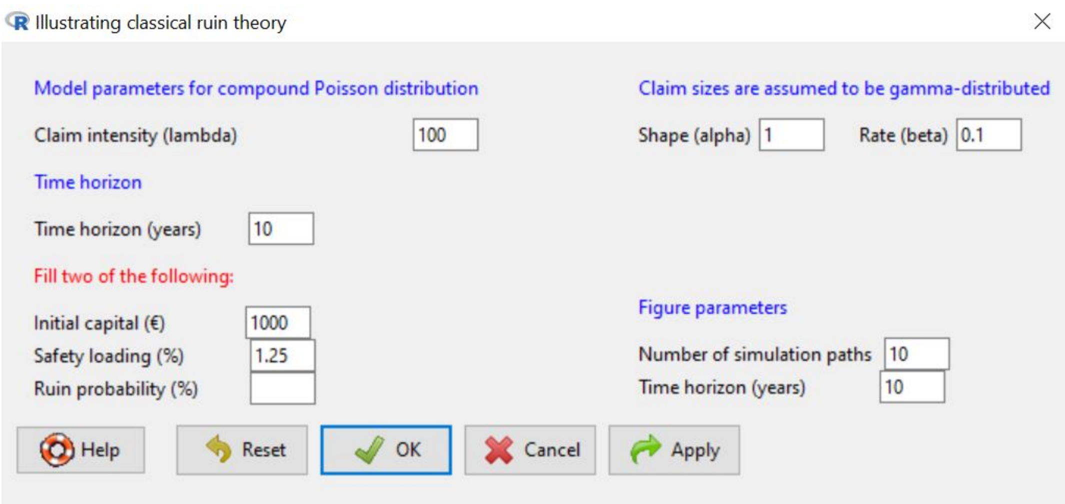
Esimerkiksi vahingonkorvausten kokonaismäärän yhdistetyn painotetun Poisson-muuttujan jakauma saadaan selville tuottamalla ensin havainto struktuurimuuttujasta q, ja sen jälkeen tuotetaan arvo maksettavien korvausten määrälle. Struktuurimuuttuja on painotettuun Poisson-muuttujaan liittyvä arvo, jonka vaihtelun voidaan ajatella selittävän jonkin yksittäisen tekijän vaikutusta vahingonkorvausten määrään (Nyrhinen 2013). Tämän jälkeen saadaan maksettavien korvausten summa simuloimalla korvausten määrän mukaisesti lukuja yksittäisen korvauksen suuruuden jakaumasta ja laskemalla ne yhteen. Tätä toistetaan monta kertaa, jolloin saadaan otos simuloiduista vahingonkorvausten kokonaismääristä. Tästä otoksesta pystytään sitten tilastollisin menetelmin arvioimaan jakaumaa.

3 RiskDemo-ohjelma

Tässä luvussa kerrotaan RiskDemo-ohjelmasta, sen käytöstä ja toiminnasta. Luku pohjautuu Luoman (2018) tekstiin ohjelmasta. Luvussa tarkastellaan myös ohjelman tuottamaa simulointikuviota ja ohjelman valintaikkunaa.

RiskDemo-ohjelma on ohjelmistotyökalu, jolla voidaan havainnollistaa riskejä todennäköisyyslaskennan ja tilastotieteen kannalta. Ohjelma on R-ohjelmiston lisäpaketti, joka käyttää R Commander-käyttöliittymää. RiskDemon avulla riskiä voidaan tutkia erilaisten taulukoiden, riskilukujen ja graafisten esitysten avulla. Tässä tutkielmassa keskitytään klassisen vararikkoteorian sovelluskohteen havainnollistamiseen. Vararikkoteoriaa havainnollistetaan simulointikuviolla ja muutamalla parametrilla, jotka ovat alkupääoma, varmuuslisä, vararikkotodennäköisyys ja Lundbergin eksponentti.

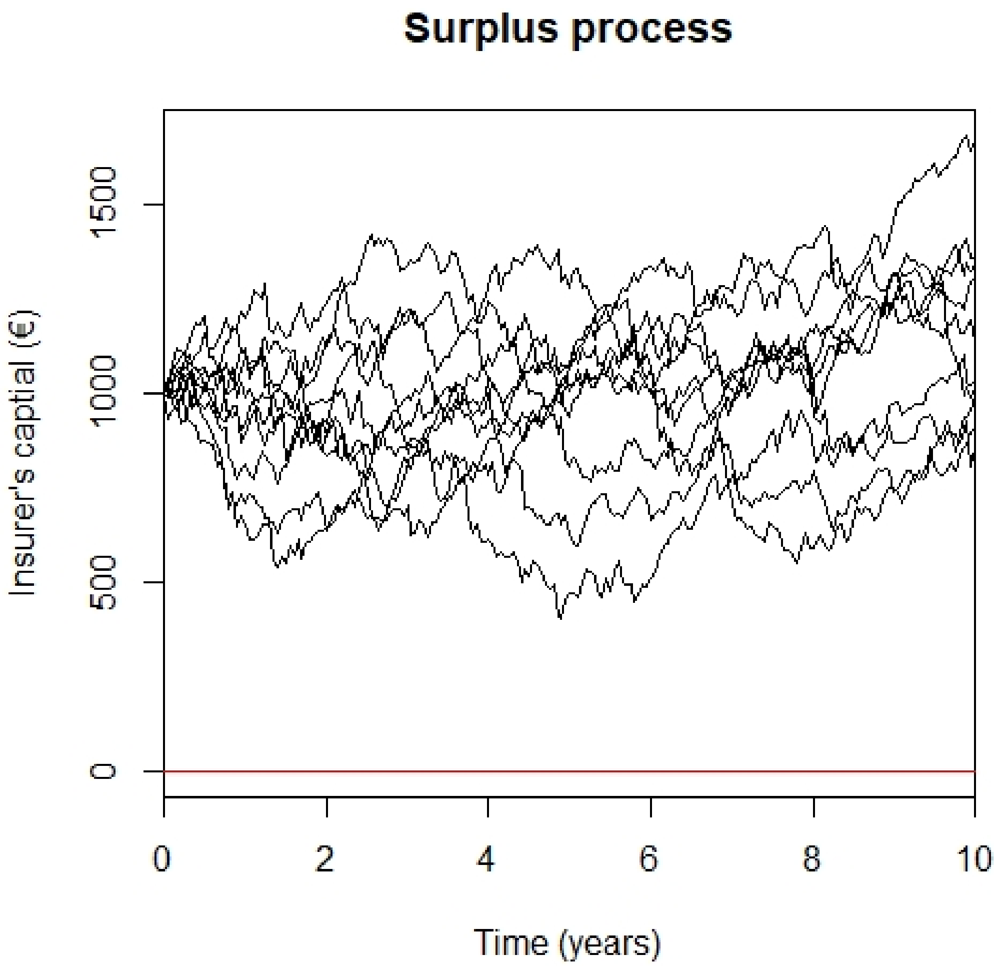
3.1 Ohjelman käyttö



Kuva 3.1. RiskDemon valintaikkuna.

Ohjelma olettaa, että vahinkotapahtumat tapahtuvat Poisson-prosessin mukaisesti. Valintaikkunassa (kuva 3.1) annetaan tätä varten intensiteettiparametri, joka kertoo vahinkojen keskimääräisen lukumäärän vuodessa. RiskDemo olettaa yksittäisten vahinkojen suuruudet gamma-jakautuneiksi. Gamma-jakaumaa käytetään usein yksittäisen korvauksen suuruuden arvioimiseen sen kevythäntäisyyden takia. Valintaikkunassa annetaan myös jakauman parametrit α ja β (shape ja rate). Lisäksi valintaikkunassa voidaan määrittää aikaväli, jolla vararikkoa tarkastellaan. Tämä kohta voidaan kuitenkin jättää tyhjäksi, jolloin tarkastellaan ääretöntä aikaväliä. Tässä tutkielmassa keskitytään kuitenkin tarkastelemaan äärellistä aikaväliä. Valintaikkunaan laitetaan myös kaksi seuraavista: alkupääoma, varmuuslisä ja vararikkotodennäköisyys. Kun ohjelmalle annetaan näistä kaksi, se pystyy laskemaan kolmannen arvon. Varmuuslisä kertoo, kuinka paljon enemmän vakuutusmaksua on kerättävä suhteessa vuotuisen vahinkomenon odotusarvoon, ja se ilmoitetaan prosentteina. Näiden arvojen avulla voidaan tutkia esimerkiksi alkupääoman vaikutusta vararikkotodennäköisyyteen. Valintaikkunassa määritetään myös arvot simulointikokeelle. Määritettävät arvot ovat simulointipolkujen lukumäärä ja tarkasteltava aikajänne.

3.2 Simulointikuvion tuottaminen



Kuva 3.2. Ohjelman oletusarvoilla muodostettu simulointikuvio.

Simulointikuvio (kuva 3.2) näyttää vakuutusyhtiön pääoman kehityksen simulaatiossa. Kuviosta nähdään helposti simuloinnissa käytetty alkupääoma, joka on siis jokaisen simulaation lähtöarvo, tässä kuviossa se on 1000 euroa. Tämän perusteella voidaan myös katsoa, koska vakuutusyhtiö on simulaatiokäyrien perusteella "voitolla", toisin sanoen, milloin käyrän arvo on yli tuhat euroa. Simulaatiokäyrien arvo nousee vakuutusmaksutulojen mukaisesti ja laskee maksettavien korvausten mukaisesti.

RiskDemo luo simulointikuvion määrittämällä ensin vakuutusmaksutulon ja sen jälkeen simuloimalla, satunnaislukugeneraattoreiden avulla, maksettavat vahingonkorvaukset, eli niiden vuosittaiset määrät ja yksittäisten korvausten suuruudet. Vakuutusmaksutulo lasketaan kaavalla (2.3). Tässä tapauksessa se on muotoa:

$$c = (1 + \theta)\lambda\frac{\alpha}{\beta},$$

jossa α/β on yksittäisen vahingonkorvauksen suuruuden odotusarvo.

Vahingonkorvausten simulointia varten RiskDemossa on määritetty, että tarkastelussa on 200 tarkastelupistettä, eli kohtaa, joissa vahingonkorvauksia ikään kuin maksetaan. Ensin tuotetaan matriisi, jossa on 200 kertaa simulaatioiden määrä Poisson-jakautuneita satunnaislukuja. Sen jälkeen lasketaan, kuinka monta niistä on suurempia kuin nolla, ja tuotetaan yhtä monta gamma-jakautunutta satunnaislukua, jotka ovat siis yksittäisten vahingonkorvausten suuruuksia. Lopuksi vakuutusmaksutuloista vähennetään vahingonkorvausten määrät jokaisella tarkastelupisteellä, jolloin pystytään seuraamaan pääoman kehitystä.

4 Alkupääoman vaikutus vararikkotodennäköisyyteen

Lopuksi tarkastellaan Riskdemon käyttöä esimerkin avulla. Tässä esimerkissä tuotetaan Riskdemon *computeRuinFinite* - funktion avulla aineisto, josta luodaan logistinen regressiomalli. Regressiomallin tulosten avulla tarkastellaan alkupääoman vaikutusta vararikkotodennäköisyyteen. Tässä tarkastelussa käytetty koodi on esitetty liitteessä (Liite: R-koodi).

4.1 Aineiston luominen RiskDemo-ohjelmalla

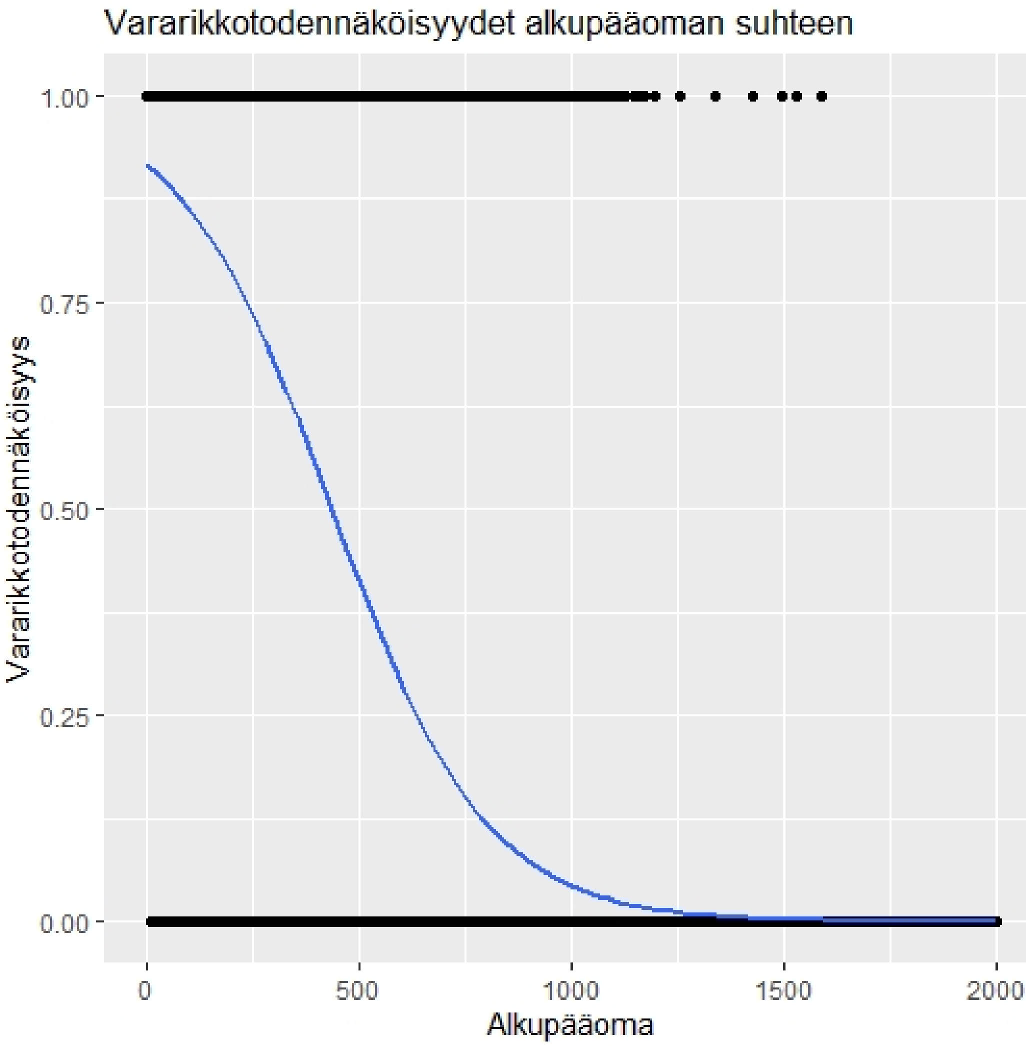
Logistisen regressiomallin selitettävä muuttuja on vararikko ja selittävä muuttuja on alkupääoma. Vararikko-muuttuja on luotu Riskdemon laskemista vararikkotodennäköisyyksistä Bernoullin jakaumaa apuna käyttäen. Vararikko -muuttujan arvo 1 tarkoittaa vararikkoa ja arvo 0 ei-vararikkoa. Tarkasteltavat alkupääomat ovat väliltä 0.2 – 2000 tasavälein. Välin pituus on 0.2. Näin saadaan 10000 eri havaintoa. Väli on valittu niin, että välin pienimmällä arvolla, 0.2, vararikkotodennäköisyys on lähes 1 ja välin suurimmalla arvolla, 2000, se on lähes 0. Vararikkotodennäköisyydet on laskettu 10 ja 3 vuoden aikaväleillä, joissa molemmissa Poisson-prosessin intensiteettiparametri, varmuuslisän arvo ja yksittäisen vahingonkorvauksen suuruuden jakauman parametrit ovat samat. Näin saadaan tarkasteltua myös sitä, onko alkupääoman vaikutuksella eroa eri aikaväleillä tarkasteltuna. Kuvaajissa (4.1, 4.2) mustat pisteet kuvaavat vararikon toteutumista tietyllä alkupääomalla. Pisteen arvoista yksi tarkoittaa vararikkoa ja nolla ei-vararikkoa. Kuvaajien käyrä on luodun mallin mukainen logistinen regressiokäyrä. Kuvaajista (4.1, 4.2) huomataan, että 10 vuoden aikavälillä tarkasteltuna vararikkoja tapahtuu suuremmillakin alkupääomilla, kun taas 3 vuoden tarkastelussa vararikkoja näyttää tapahtuvan vain alkupääomilla, jotka ovat pienempiä kuin 1000. Molemmista kuvaajista voidaan huomata, myös että pieni alkupääoma ei tarkoita välttämättä vararikkoa, sillä molemmissa kuvaajissa vararikko-muuttuja saa arvon 0 pienilläkin alkupääoman arvoilla. Tämä tietysti heikentää mallin toimivuutta.

4.2 Logistiset regressiomallit

Luodusta aineistosta tehtiin logistiset regressiomallit sekä kymmenen että kolmen vuoden aikavälille. Mallit tehtiin logistisen regressioanalyysin kaavan mukaisesti seuraavasti:

$$P(Y_i = 1) = \frac{\exp \beta_0 + \beta_1 u_i}{1 + \exp \beta_0 + \beta_1 u_i},$$

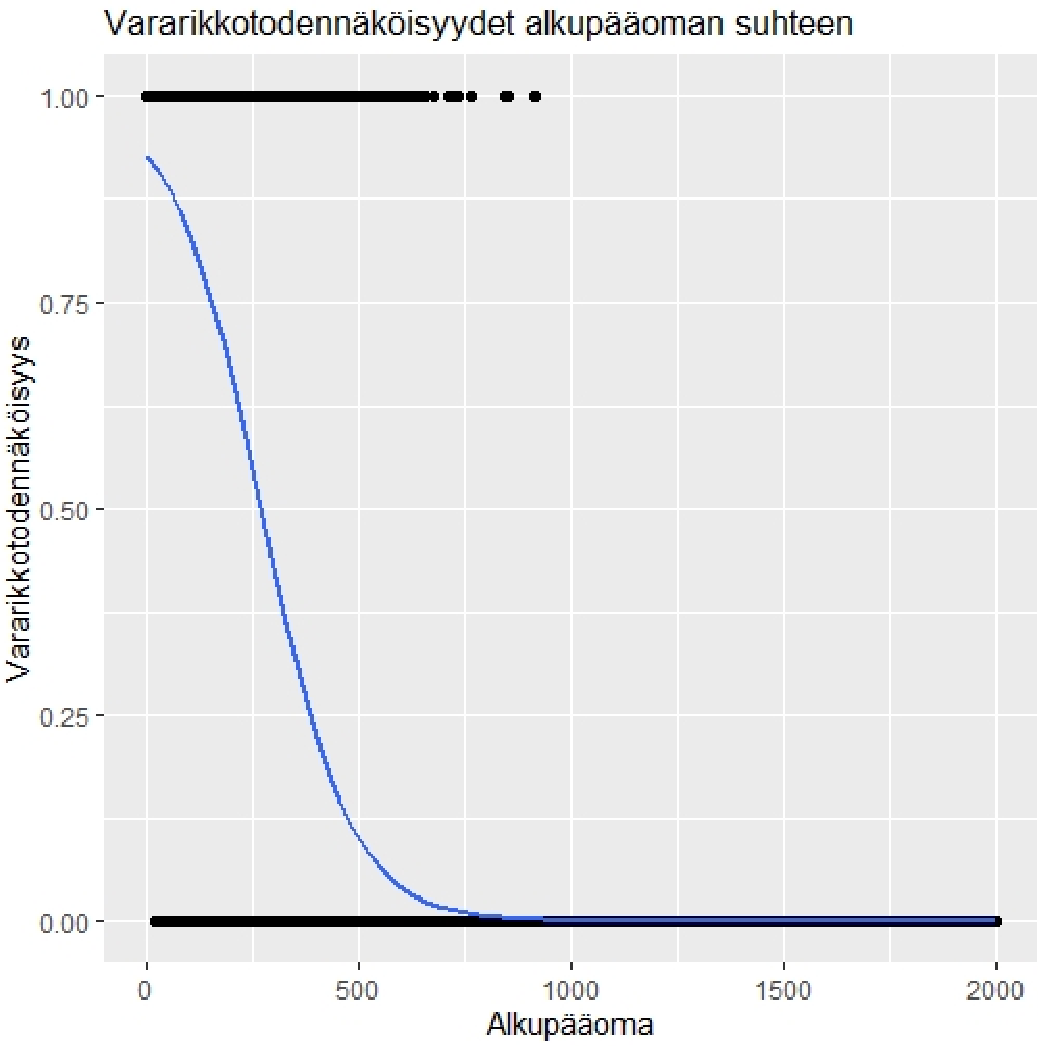
jossa Y_i on vararikko-muuttuja ja u_i on alkupääoma simuloinnilla i. Mallien tuloksista huomataan, että alkupääoma on tilastollisesti merkittävä selittävä muuttuja, sillä sen p-arvo on pieni molemmissa malleissa. Yhden yksikön muutos alkupääomassa on



Kuva 4.1. Vararikkotodennäköisyydet 10 vuoden mallilla. Pisteet yhden ja nollan kohdalla tarkoittavat simuloinnin lopputuloksia, joissa nolla vastaa ei-vararikkoa ja yksi vararikkoa.

kuitenkin vaikutukseltaan pieni. Tämä tulos on myös käytännön kannalta järkevä, sillä yhden euron kasvu alkupääomassa ei vaikuta merkittävästi vararikon todennäköisyyteen. Täytyy kuitenkin huomioida, että vararikkotodennäköisyyden muutokseen vaikuttaa myös alkupääoman suuruus. Yhden yksikön muutoksen pienen vaikutuksen takia saattaisi olla järkevämpää tarkastella isompien muutosten vaikutusta, joten luotiin uudet mallit, joissa selittävänä muuttujana on alkupääoma jaettuna sadalla. Tässä mallissa yhden yksikön muutos selittävässä muuttujassa vastaa siis sadan euron muutosta alkupääomassa.

Näissä malleissa kymmenen vuoden tarkastelussa kertoimeksi saatiin -0.548 ja kolmen vuoden tarkastelussa vastaava kerroin on-0.949. Tulokset on esitetty myös taulukossa (4.1). Logistisen regressiomallin kertoimia on kuitenkin hankala tulkita sellaisenaan, joten niistä laskettiin myös vetosuhteet. Kymmenen vuoden mallissa vetosuhde on 0.578 ja kolmen vuoden mallissa 0.387. Tämä tarkoittaa sitä, että yhden yksikön kasvua kohden, eli tässä mallissa 100 euron kasvua kohden, varari-



Kuva 4.2. Vararikkotodennäköisyydet 3 vuoden mallilla. Pisteet yhden ja nollan kohdalla tarkoittavat simuloinnin lopputuloksia, joissa nolla vastaa ei-vararikkoa ja yksi vararikkoa.

kon veto pienenee 10 vuoden mallissa 0.578-kertaiseksi ja kolmen vuoden mallissa 0.387-kertaiseksi. Prosentteina tämä tarkoittaa sitä, että 10 vuoden mallissa 100 euron kasvu alkupääomassa pienentää vararikon vetoa 42.2 prosenttia ja 3 vuoden mallissa vararikon veto pienenee 61.3 prosenttia. Vararikon vedolla tarkoitetaan siis vararikkotodennäköisyyden suhdetta ei-vararikon todennäköisyyteen.

Kolmen vuoden mallissa alkupääomalla näyttää olevan suurempi vaikutus vararikkotodennäköisyyteen kuin 10 vuoden mallissa. Tämän perusteella näyttäisi siltä, että alkupääoman merkitys vähenee, kun tarkasteltava aikaväli kasvaa. Asiaa täytyisi kuitenkin tutkia tarkemmin, jotta selkeitä johtopäätöksiä pystyttäisiin tekemään.

Malleista laskettiin myös McFaddenin pseudo R^2 - selityskerroin, joka kertoo mallin sopivuudesta aineistoon. Selityskertoimen arvot näkyvät taulukossa (4.1). Mallien p-arvot olivat pienet, mikä kertoo siitä, ettei alkupääoman vaikutus vararikkotodennäköisyyten ole sattumaa. Selityskertoimien arvot ovat kohtalaisia, joten mallien sopivuudessa voi olla parannettavaa. Mallien kuvaajista (4.1, 4.2) nähdään,

Taulukko 4.1. Logististen regressiomallien tulokset.

	Kymmenen vuoden malli	Kolmen vuoden malli
β_0	2.382	2.551
β_1	−0.548	−0.949
McFadden R^2	0.505	0.614

että mallit eivät onnistu selittämään vararikkoa kovinkaan tarkasti pienillä alkupääoman arvoilla, sillä vararikko-muuttujan arvojen vaihtelu on suurinta pienillä alkupääoman arvoilla.

5 Yhteenveto

Tässä tutkielmassa esiteltiin vakuutusyhtiön riskiprosessin mallintamiseen liittyvää teoriaa ja tarkasteltiin kuinka mallintaminen on toteutettu RiskDemo-ohjelmalla. Lisäksi tarkasteltiin esimerkinomaisesti RiskDemolla luodun aineiston avulla alkupääoman vaikutusta vararikon todennäköisyyteen.

RiskDemo mallintaa vakuutusyhtiön riskiprosessia klassisen vararikkoteorian ja simuloinnin avulla. Klassisen vararikkoteorian mukaan vakuutusyhtiön pääomaan vaikuttaa alkupääoma, kerättävät vakuutusmaksutulot ja maksettavat vahingonkorvaukset. Mallissa oletetaan vahingonkorvausten tulevan maksettavaksi Poisson-prosessin mukaisesti. Lisäksi yksittäisen vahingonkorvauksen oletetaan noudattavan jotakin tiettyä jakaumaa, RiskDemossa tähän käytetään gamma-jakaumaa. Näiden oletusten pohjalta RiskDemo simuloi vahingonkorvausten määrän ja suuruudet. Simulointi tarjoaa tehokkaan keinon mallintaa jotakin ilmiötä, kun aineistoa ei ole saatavilla tai kun mallit ovat monimutkaisia.

RiskDemo tuottaa simulaatiokuvion ja laskee vararikkotodennäköisyyden ja Lundbergin eksponentin sille syötettyjen arvojen perusteella. RiskDemon käyttämä malli on kuitenkin varsin yksinkertainen, sillä se jättää paljon vakuutusyhtiön pääomaan vaikuttavia asioita huomiotta. Tästä syystä sen avulla ei voida tehdä luotettavia johtopäätöksiä tosielämän tilanteista, mutta ohjelma kuitenkin havainnollistaa selkeästi riskiprosessin tarkastelua tilastotieteen ja todennäköisyyslaskennan näkökulmasta.

Logististen regressiomallien avulla saatiin tietoa alkupääoman vaikutuksesta vararikkotodennäköisyyteen. Mallien perusteella näyttäisi siltä, että alkupääoman kasvaessa vararikkotodennäköisyys pienenee. Lisäksi pienemmällä aikavälillä tarkasteltaessa alkupääoman vaikutuksen suuruus näyttäisi olevan isompi, eli alkupääoman merkitys pienenee aikavälin kasvaessa. Luotujen mallien selityskertoimet olivat kuitenkin vain kohtalaisia, joten monimutkaisemmilla malleilla voitaisiin saada tarkempia tuloksia.

Aiheen käsittelyä voisi laajentaa koskemaan myös RiskDemo-ohjelman kehittämistä, esimerkiksi tarkastelemalla mallintamisen toteuttamista monimutkaisemmilla malleilla tai käyttäen eri jakaumia yksittäisen vahingonkorvauksen suuruuden kuvaamiseen.

Lähteet

Daykin, C. D., Pentikäinen, T., Pesonen, M. (1994), *Practical Risk Theory for Actuaries*, London: Chapman & Hall

Kaas, R., Goovaerts, M., Dhaene, J. ja Denuit, M. (2008), *Modern Actuarial Risk Theory Using R*, 2nd ed. Berlin, Heidelberg: Springer Berlin Heidelberg.

Koskinen, L. (2018), "Riskienhallinta ja tietämyksen tasot", teoksessa *Riskienhallinnan ajankohtaisia teemoja*, s. 11 – 28

Luoma, A. (2018), ”Riskien havainnollistamisohjelmisto RiskDemo”, teoksessa *Riskienhallinnan ajankohtaisia teemoja*, s. 220 – 245

Nyrhinen, H. (2013), Riskiteoria, Luentomoniste. Helsingin yliopisto. https://wiki.helsinki.fi/display/mathstatKurssit/Riskiteoria,+syksy+2013?preview=/112438755/119538627/rt13.pdf. Luettu 18.2.2021.

Liite: R-koodi

R-koodissa luodaan ensin alkupääomat tasavälille 0.2 – 2000. Sen jälkeen alustetaan muuttujat alkupääomille ja vararikkotodennäköisyyksille. Niihin talletetaan arvot for-silmukassa, jossa lasketaan vararikkotodennäköisyydet jokaiselle alkupääomalle. Seuraavaksi muutetaan vararikkotodennäköisyydet vararikoiksi. Sen jälkeen luodaan ensimmäiset logistiset regressiomallit ja lasketaan selityskertoimet malleille. Lopuksi luodaan vielä uudet mallit, ja lasketaan vetosuhteet malleille.

```
> library(RcmdrPlugin.RiskDemo)

> set.seed(1234)
> paaoma_kaikki <- seq(0.20,2000,by=0.20)
> head(paaoma_kaikki)

[1] 0.2 0.4 0.6 0.8 1.0 1.2

> paaomat10 <- numeric(10000)
> vrtn10 <- numeric(10000)
> paaomat3 <- numeric(10000)
> vrtn3 <- numeric(10000)

> for(i in 0:10000) {
+ ans10 <-
computeRuinFinite(T0=10,U0=paaoma_kaikki[i],theta=1.25/100,lambda=100,alpha=1,
+ beta=0.1) ## 10 vuoden tarkastelu
+ ans3 <-
computeRuinFinite(T0=3,U0=paaoma_kaikki[i],theta=1.25/100,lambda=100,alpha=1,
+ beta=0.1) ## 3 vuoden tarkastelu
+ paaomat10[i] <- ans10$initialCapital
+ paaomat3[i] <- ans3$initialCapital
+ vrtn10[i] <- ans10$ruinProb
+ vrtn3[i] <- ans3$ruinProb
+ }
> vararikko10 <- rbinom(length(vrtn10), size = 1, prob = vrtn10)
> summary(vararikko10)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.0000  0.0000  0.2252  0.0000  1.0000

> vararikko3 <- rbinom(length(vrtn3), size = 1, prob = vrtn3)
> summary(vararikko3)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.0000  0.0000  0.1383  0.0000  1.0000

> data10 <- cbind(vararikko10, paaomat10)
> data10 <- data.frame(data10)
> data10$vararikko10 <- factor(vararikko10)
> data3 <- cbind(vararikko3, paaomat3)
> data3 <- data.frame(data3)
> data3$vararikko3 <- factor(vararikko3)

> ## 10 vuoden malli
> malli10 <- glm(vararikko10 ~ paaomat10, family=binomial(link = "logit"),
data=data10, maxit=100)
> summary(malli10)


Call:
glm(formula = vararikko10 ~ paaomat10, family = binomial(link = "logit"),
```

```
    data = data10, maxit = 100)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2052  -0.3485  -0.0873  -0.0221   3.5608

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.3821608  0.0706816   33.70   <2e-16 ***
paaomat10   -0.0054836  0.0001253  -43.76   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10668.2  on 9999  degrees of freedom
Residual deviance:  5282.5  on 9998  degrees of freedom
AIC: 5286.5

Number of Fisher Scoring iterations: 7

> ll.null10 <- malli10$null.deviance/-2
> ll.proposed10 <- malli10$deviance/-2
>
> # Malli10 selityskerroin ja p-arvo
> Rsq10 <- (ll.null10 - ll.proposed10) / ll.null10
> Rsq10

[1] 0.5048367

> p10 <- 1-pchisq(2*(ll.proposed10 - ll.null10), df =
(length(malli10$coefficients)-1))
> p10

[1] 0

> ## 3 vuoden malli
> malli3 <- glm(vararikko3 ~ paaomat3, family=binomial(link = "logit"),
data=data3, maxit=100)
> summary(malli3)


Call:
glm(formula = vararikko3 ~ paaomat3, family = binomial(link = "logit"),
    data = data3, maxit = 100)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2323  -0.1271  -0.0118  -0.0011   3.5097

Coefficients:
```

```
               Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.5512995  0.0945170   26.99   <2e-16 ***
paaomat3    -0.0094921  0.0002782  -34.12   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8037.3  on 9999  degrees of freedom
Residual deviance: 3102.6  on 9998  degrees of freedom
AIC: 3106.6

Number of Fisher Scoring iterations: 8

> ll.null3 <- malli3$null.deviance/-2
> ll.proposed3 <- malli3$deviance/-2
>
> # Malli3 selityskerroin ja p-arvo
> Rsq3 <- (ll.null3 - ll.proposed3) / ll.null3
> Rsq3

[1] 0.6139736

> p3 <- 1-pchisq(2*(ll.proposed3 - ll.null3), df = (length(malli3$coefficients)-
1))
> p3

[1] 0

> ## Uudet mallit
>
> ## Uusi 10 vuoden malli
> paaomat10_2 <- paaomat10/100
> head(paaomat10_2)

[1] 0.002 0.004 0.006 0.008 0.010 0.012

> data10 <- cbind(data10, paaomat10_2)
> data10 <- data.frame(data10)
> malli10_2 <- glm(vararikko10 ~ paaomat10_2, family=binomial(link = "logit"),
data=data10, maxit=100)
> summary(malli10_2)


Call:
glm(formula = vararikko10 ~ paaomat10_2, family = binomial(link = "logit"),
    data = data10, maxit = 100)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2052  -0.3485  -0.0873  -0.0221   3.5608
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.38216    0.07068   33.70   <2e-16 ***
paaomat10_2 -0.54836    0.01253  -43.76   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10668.2  on 9999  degrees of freedom
Residual deviance:  5282.5  on 9998  degrees of freedom
AIC: 5286.5

Number of Fisher Scoring iterations: 7

> ## Uusi 3 vuoden malli
> paaomat3_2 <- paaomat3/100
> head(paaomat3_2)

[1] 0.002 0.004 0.006 0.008 0.010 0.012

> data3 <- cbind(data3, paaomat3_2)
> malli3_2 <- glm(vararikko3 ~ paaomat3_2, family=binomial(link = "logit"),
data=data3, maxit=100)
> summary(malli3_2)


Call:
glm(formula = vararikko3 ~ paaomat3_2, family = binomial(link = "logit"),
    data = data3, maxit = 100)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2323  -0.1271  -0.0118  -0.0011   3.5097

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.55130    0.09452   26.99   <2e-16 ***
paaomat3_2  -0.94921    0.02782  -34.12   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8037.3  on 9999  degrees of freedom
Residual deviance: 3102.6  on 9998  degrees of freedom
AIC: 3106.6

Number of Fisher Scoring iterations: 8
```

```
> ## Vetosuhteet uusille malleille
>
> ## 10 vuoden malli
> vetosuhde10 <- exp(malli10_2$coefficients[-1])
> vetosuhde10

paaomat10_2
   0.577898

> ## vetosuhde prosentteina
> (vetosuhde10 - 1) * 100

paaomat10_2
   -42.2102

> ## 3 vuoden malli
> vetosuhde3 <- exp(malli3_2$coefficients[-1])
> vetosuhde3

paaomat3_2
 0.3870476

> ## vetosuhde prosentteina
> (vetosuhde3 - 1) * 100

paaomat3_2
 -61.29524
```