Stockholms universitet

Utvärdering av den kollektiva reservsättnings-modellen baserad på simulerad data

Heja Lindgren

Masteruppsats i försäkringsmatematik
Master Thesis in Actuarial Mathematics

Masteruppsats 2019:1
Försäkringsmatematik
Februari 2019

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematisk statistik
Stockholms universitet
Masteruppsats **2019:1**
http://www.math.su.se/matstat

Utvärdering av den kollektiva reservsättnings-modellen baserad på simulerad data

Heja Lindgren*

Februari 2019

Sammanfattning

En försäkringstagare tecknar ett försäkringsavtal hos ett försäkringbolag för att få försäkringsskydd. Från första dag hen tecknar sin försäkring, så har försäkringsbolaget ett åtagande till försäkringstagaren. Detta åtagandet kan gälla även lång tid efter det att avtalet slutat gälla. Tiden från premieinbetalning tills en skada inträffar och är helt slutreglerad kan ta flera år. Därför behöver försäkringsbolaget en reservsättningsmetod som är så korrekt som möjligt för att garantera försäkringstagarens framtida ersättning. Den här uppsatsen har till syfte att avgöra lämpligheten hos en ny reservsättningsmetod baserat på simulerade data. Den nya metoden kallas för den kollektiva reservsättningsmodellen (CRM, eng. the collective reserving model). Utfallet utvärderas mot den kanske mest kända reservsättningsmetoden i den fördelningsfria Chain Ladder- modellen. CRM kan uttryckas som en generaliserad linjär modell (GLM), vilket innebär att modellparametrarna kan skattas genom att använda standardteori för GLM. Analysen visar att om vi har en försäkring där utbetalningstiden är kort, så får vi problem att skatta så många parametrar vi önskar med kvasimaximumlikelihoodanpassning (QML, eng. quasi maximum likelihood) då vi simulerar data 10 000 gånger på grund av att somliga utbetalningsperioder innehåller nollor. Detta är ett numeriskt problem och går i praktiken att undvika,så metoden går att automatisera. På grund av tidsbrist väljer vi dock i denna uppsats att medvetet skatta färre parametrar, vilket leder till en viss underskattning av reserven jämfört med de sanna simulerade utbetalningar och Chain Ladder-modellen men båda metoderna ger bra prediktioner. När vi har en lång utbetalningstid, visar det sig att reservskattningen med CRM är närmare de sanna simulerade utbetalningar än Chain Ladder-modellen.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: perspei@hotmail.com. Handledare: Mathias Lindholm.

1 Abstract

A policyholder signs an insurance contract with an insurance company to obtain insurance cover. From the first day they sign their insurance, the insurance company has a commitment to the policyholder. This commitment may also be valid for a long time after the last premium payment. The time from premium payment until an injury occurs and is completely finalized can take several years. Therefore, the insurance company needs a reserve method that is as accurate as possible to guarantee the policyholder's future compensation. The purpose of this thesis is to determine the suitability of a new reserve method based on a simulated data. It is called the collective reserving model (CRM). The outcome is evaluated against the perhaps most well-known distribution-free Chain Ladder model. CRM can be expressed as a generalized linear model (GLM), which means that the model parameters can be estimated using standard theory for GLM. The analysis shows that if we have an insurance where the payout time is short, then we have problems estimating as many parameters as we want using quasi-maximum likelihood. Since simulated payment data may contain zeros. This is a numerical problem and in practice it can be avoided, so this method can be automated. However due to lack of time, we choose thesis to deliberately estimate fewer parameters, which leads to a slight underestimation of the reserve compared to the true simulated payments and the Chain Ladder model, but both methods give good predictions. When we have a long payout time, it turns out that the reserve estimate with CRM is closer to the true simulated payouts than the Chain Ladder model.

2 Förord

Denna uppsatsen utgör mitt masterarbete i försäkringsmatematik vid Stockhilms Universitet och omfattar 30 hp.

Ett stort tack riktas till min handledare Mathias Lindholm. Tack för hans värdefulla synpunkter och uppmuntran under arbetetsgång.

Innehåll

1 Abstract **ii**

2 Förord **iii**

3 Introduktion **1**
3.1 Syfte och metod 1

4 Teori **2**
4.1 Reservsättning 2
4.2 Chain Ladder-modellen 3
4.3 CRM: Den kollektiva reservsättningsmodellen 5
4.3.1 Den homogena CRM 10

5 Val av fördelningar till simuleringen **16**

6 Dataanalys **18**
6.1 Modellantagande för Chain Ladder-modellen 22
6.2 Modellantagande för CRM 24
6.3 Problem med simulerade trianglar 28

7 Utvärdering av CRM och Chain Ladder-modellen **30**
7.1 Dataset typ 1 30
7.2 Dataset typ 2 40

8 Slutsatser och Diskussion **44**

9 Referenser **46**

10 Appendix **47**

3 Introduktion

3.1 Syfte och metod

Ett sakförsäkringsbolag har åtaganden mot sina försäkringatagare långt efter det att ett försäkringsavtal slutat gälla. Tiden från en premieinbetalning tills det att en skada inträffar och är helt slutreglerad kan ta flera år. Därför behöver försäkringsbolag sätta av kapital för att täcka framtida kostnader för inträffade men ännu ej slutreglerade skador. Detta kapital kallas för en reserv och det krävs en så korrekt reservsättningsmetod som möjligt för att garantera försäkringstagarnas framtida ersättning.

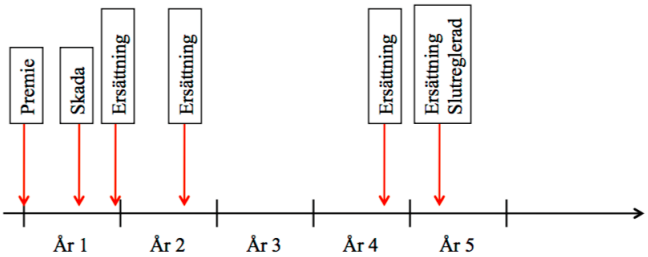
Detta arbete går ut på att försöka hitta en reservsättningsmetod som passar våra simulerade data. Vi testar två reservsättningsmodeller: Den kollektiva reservsättningsmodellen (CRM, eng. the collective reserving model) och den kanske mest kända reservsättningsmodellen Macks Chain Ladder-modell. Vi använder två aggregerade trianglar för att skatta reserven med CRM: En triangel med skadeutbetalningar och en triangel med antal rapporterade skador. Genom att betinga på de rapporterade skadorna kan vi beräkna reserven för kända och okända skador separat. Fördelen med detta är att vi lättare kan se och jämföra om det är de kända skadorna eller de okända skadorna som predikteras bättre. CRM kan uttryckas som en generaliserad linjär modell (GLM) och standardteorier för GLM används för att skatta parametrar.

Vi simulerar data 10 000 gånger, sedan beräknar vi reserverna 10 000 gånger med båda metoderna. Vi jämför sedan de skattade reserverna med de simulerade utbetalningarna. I CRM jämför vi även den skattade kända reserven med de simulerade betalningar som kommer från kända skador, och den skattade okända reserven med den simulerad betalningar som kommer från okända skador. Analysen visar att om vi har en försäkring där utbetalningstiden är kort kan vi med CRM få problem att skatta så många parametrar vi vill på grund av att det kan finnas nollor i somliga betalningsperioder. Detta är ett numeriskt problem och går i praktiken att undvika men på grund av tidsbrist väljer vi dock i denna uppsats att medvetet skatta färre parametrar, viket leder till en viss underskattning av reserven jämfört med de sanna simulerade utbetalningarna och Chain Ladder-modellen men båda metoderna ger bra prediktioner. När vi har en lång utbetalningstid, visar det sig att reservskattingen med CRM är närmare de sanna simulerade utbetalningar än Chain Ladder-modellen.

4 Teori

4.1 Reservsättning

Ett företag erbjuder som bekant individer eller företag att mot en avgift (premien) täcka kostnaderna för eventuella skador av något slag som händer slumpmässigt. Det hela bygger på att kostnaderna kan fördelas över en grupp av försäkringstagare. Principen är att premieintäkterna ska täcka kostnaderna för inträffade skador, men eftersom premierna betalas i förskott kan man inte vara helt säker på detta. Figur 1 visar ett exempel på hur ett skadeförsäkringskontrakt kan se ut över tiden (Figur 1 är hämtad från Skadeförsäkringsredovisning föreläsning 1).



Figur 1: Skadeförsäkring över tiden

Vi ser från Figur 1 att man betalar premie, sedan händer en skada under första året. Den här skadan har fyra utbetalningar och sista utbetalnigen sker år fem. Vi vet att i verkligheten finns det vissa skador som inte leder till någon ersättning men det finns också vissa skador som har mycket längre utbetalningstid med flera utbetalningar. Därför behöver bolaget uppskatta den totala skadekostnaden ifall portföljen skulle drabbas av ovanligt höga kostnader. Vårt problem är att uppskatta den totala reserven.

Reserven kan delas upp i två huvudsakliga poster, nämligen IBNR och RBNS.

IBNR := [Incurred but not reported]

RBNS := [Reported but not settled]

IBNR betyder en reserv för de skador som har inträffat men som vid beräkningstidpunkten ännu inte är rapporterade. RBNS innebär en reserv för de skador som har rapporterats men där alla betalningar ännu inte är gjorda.

Det finns en mängd olika reservsättningsmetoder, men den kanske mest använda modellen är Macks Chain Ladder-modell. Denna modell delar dock inte upp reserverna i IBNR och RBNS. Den bryr sig heller inte om antalet rapporterade skador. Mer om Chain Ladder-modellen kan man läsa i (Mack, 1993). I denna rapport kommer vi även att titta på en ny reservsättningsmodell kallad CRM. Denna modell tillåter flera utbetalningar per skada. Den skattar IBNR och RBNS var för sig. Teori för denna modellen hämtas från (Wahl, 2018).

4.2 Chain Ladder-modellen

I detta avsnitt ska vi kortfattat introducera Macks Chain Ladder-modell. Vi låter C_{ij} beteckna de ackumulerade utbetalda beloppen under utvecklingsår j för skador som inträffade under år i, där $i = 1, 2, 3, ..., m$, och $j = 0, 1, 2, ..., J-1$, och där J är det år då samtliga utbetalningar är slutreglerade. De kumulativa utbetalningarna för känd och framtida skadekostnader kan skrivas enligt Tabell 1.

	Utvecklingsår								
Skadeår	0	1	2	...	$m-1$	m	...	$J-2$	$J-1$
1	C_{10}	C_{11}	C_{12}	...	$C_{1,m-1}$	$C_{1,m}$	...	$C_{1,J-2}$	$C_{1,J-1}$
2	C_{20}	C_{21}	C_{22}	...	$C_{2,m-1}$	$C_{2,m}$	...	$C_{2,J-2}$	$C_{2,J-1}$
3	C_{30}	C_{31}	C_{32}	...	$C_{3,m-1}$	$C_{3,m}$	...	$C_{3,J-2}$	$C_{3,J-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$m-1$	$C_{m-1,0}$	$C_{m-1,1}$	$C_{m-1,2}$	...	$C_{m-1,m-1}$	$C_{m-1,m}$	...	$C_{m-1,J-2}$	$C_{m-1,J-1}$
m	$C_{m,0}$	$C_{m,1}$	$C_{m,2}$	...	$C_{m,m-1}$	$C_{m,m}$	...	$C_{m,J-2}$	$C_{m,J-1}$

Tabell 1: Känd och framtida utveckling av skadekostnaderna, kumulativt skadeutbetalningar

För enkelhetens skull brukar vi titta på fallet då $m = J$. Den kumulativa triangeln över den kända skadekostnaden skrivs då enligt tabell 2.

	Utvecklingsår					
Skadeår	0	1	2	...	$m-2$	$m-1$
1	C_{10}	C_{11}	C_{12}	...	$C_{1,m-2}$	$C_{1,m-1}$
2	C_{20}	C_{21}	C_{22}	...	$C_{2,m-2}$	
3	C_{30}	C_{31}	C_{32}	...		
$\vdots$	$\vdots$	$\vdots$	$\vdots$			
$m-1$	$C_{m-1,0}$	$C_{m-1,1}$				
m	$C_{m,0}$					

Tabell 2: Kumulativ triangel över känd skadeutbetalningar, kumulativa triangel

På diagonalen ligger kalenderåren. Idén med den Chain Ladder-modellen, se t.ex (Wütrich, 2015), är att alla olycksår uppför sig på samma sätt, vilket ger approximationen

$$C_{i,j} \approx C_{i,j-1} f_j.$$

Denna mycket gamla idé omvandlades till den berömda fördelningsfria Chain Ladder-modellen av (Mack, 1993), som i vår notation kan skrivas som:

Antagande 1:

För någon parameter $f_j > 0$, och alla $i = 1, 2, ..., m$ och $j = 1, 2, ..., m-1$, har vi

$$\mathbb{E}[C_{ij} | C_{i0}, ..., C_{i,j-1}] = C_{i,j-1} f_j.$$

(Mack, 1993) har också följande antagande:

Antagande 2:

De slumpmässiga variablerna för olycksåren är oberoende, det vill säga slumpmässiga vektorer $\{C_{i0}, ..., C_{i,m-1}\}$ och $\{C_{j0}, ..., C_{j,m-1}\}$, där $i \neq j$ är oberoende av varandra.

Utvecklingsfaktorerna, $f_1, ..., f_{J-1} > 0$ skattas enligt

$$\hat{f}_j = \frac{\sum_{i=1}^{m-j} C_{i,j}}{\sum_{i=1}^{m-j} C_{i,j-1}}, \qquad j = 1, 2, ..., m-1. \tag{1}$$

(Mack, 1993) visar att, under Antagande 1 och Antagande 2 är estimatorn $\hat{f}_j$ väntevärdsriktig där $\hat{f}_j$ och $\hat{f}_k$ är okorrelerade för olika $j \neq k$.

Värdet i rad i på den senast observerade diagonalen, det vill säga $j = m - i$, är $C_{i,m-i}$. Genom att multiplicera den med de successiva utvecklingsfaktorerna får vi den totala reservskattningen för skadeår i, $C_{i,m-1}$

$$\hat{C}_{i,m-1} = C_{i,m-i}\hat{f}_{m-i+1} \cdots \hat{f}_{m-1} \qquad i = 2, ..., m-1.$$

Som det har noterats av (Mack, 1993) följer att $\hat{C}_{i,m-1}$ är en väntevärderiktig estimator (prediktor) av den ultimokostnaden $C_{i,m-1}$, det vill säga den totala skadeutbetalningar för skadeår i.

Den totala reserven för skadeår i skattas genom att dra av det kända skadebeloppen $C_{i,m-i}$ från det skattade ultimokostnaden $\hat{C}_{i,m-1}$. Den totala reserven är summan av reserverna från samtliga skadeår, vilket beskrivs enligt nedan:

$$\hat{R}_i^{CL} = \hat{C}_{i,m-1} - C_{i,m-i}, \qquad i = 2, 3, ..., m \tag{2}$$

$$\hat{R}^{CL} = \sum_{i=2}^{m} \hat{R}_i^{CL} \tag{3}$$

Notera att beräkningen av reserven enligt ekvationer (2) och (3) gäller för alla modeller.

4.3 CRM: Den kollektiva reservsättningsmodellen

I CRM behöver vi två inkrementella trianglar: en inkrementell triangel med antal rapporterade skador och en inkrementell triangel med skadeutbetalningar.

Låt N_{ij} beteckna det inkrementella antalet rapporterade skador vid skadeår $i = 1, 2, ..., m$, under utvecklingsår $j = 0, 1, ..., m + d - 1$, där d är den maximala fördröjningen med betalningen efter att skadan är raporterad. Då kan vi beskriva känd och framtida utveckling av antalet rapporterade skadorna enligt Tabell 3.

	Utvecklingsår					
Skadeår	0	1	2	...	$m-2$	$m-1$
1	N_{10}	N_{11}	N_{12}	...	$N_{1,m-2}$	$N_{1,m-1}$
2	N_{20}	N_{21}	N_{22}	...	$N_{2,m-2}$	$N_{2,m-1}$
3	N_{30}	N_{31}	N_{32}	...	$N_{3,m-2}$	$N_{3,m-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
m-1	$N_{m-1,0}$	$N_{m-1,1}$	$N_{m-1,2}$	...	$N_{m-1,m-2}$	$N_{m-1,m-1}$
m	$N_{m,0}$	$N_{m,1}$	$N_{m,2}$	...	$N_{m,m-2}$	$N_{m,m-1}$

Tabell 3: Känd och framtida utveckling av antalet rapporterade skador, inkrementellt rapporterade skador

Låt vidare X_{ij} beteckna de inkrementella aggregerade utbetalningarna under skadeår i och utvecklingsår j, Det inkrementella aggregerade utbetalningarna över känd och framtida utbetalningarna kan skrivas enligt Tabell 4.

	Utvecklingsår							
Skadeår	0	1	2	...	$m-1$	m	...	$m+d-1$
1	X_{10}	X_{11}	X_{12}	...	$X_{1,m-1}$	$X_{1,m}$	...	$X_{1,m+d-1}$
2	X_{20}	X_{21}	X_{22}	...	$X_{2,m-1}$	$X_{2,m}$	...	$X_{2,m+d-1}$
3	X_{30}	X_{31}	X_{32}	...	$X_{3,m-1}$	$X_{3,m}$	...	$X_{3,m+d-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$m-1$	$X_{m-1,0}$	$X_{m-1,1}$	$X_{m-1,2}$	...	$X_{m-1,m-1}$	$X_{m-1,m}$	...	$X_{m-1,m+d-1}$
m	$X_{m,0}$	$X_{m,1}$	$X_{m,2}$	...	$X_{m,m-1}$	$X_{m,m}$	...	$X_{m,m+d-1}$

Tabell 4: Känd och framtida utveckling av utbetalningar, inkrementell utbetalningar

Vi använder datan som vi vet idag för att prediktera framtida utbetalningar. Vi definierar delmängden av trianglarna $\mathcal{A}_0$ som

$$\mathcal{A}_0 := \{(i, j) \in \mathbb{N} \times \mathbb{N}_0 : i + j \leq m\}.$$

Enligt denna definitionen kan den inkrementella aggregerade triangeln över de

kända rapporterade skadorna beskrivas som Tabell 5.

	Utvecklingsår					
Skadeår	0	1	2	...	$m-2$	$m-1$
1	N_{10}	N_{11}	N_{12}	...	$N_{1,m-2}$	$N_{1,m-1}$
2	N_{20}	N_{21}	N_{22}	...	$N_{2,m-2}$	
3	N_{30}	N_{31}	N_{32}	...		
$\vdots$	$\vdots$	$\vdots$	$\vdots$			
$m-1$	$N_{m-1,0}$	$N_{m-1,1}$				
m	$N_{m,0}$					

Tabell 5: Känd rapporterade skador, inkrementell triangel

Den inkrementella aggregerade triangeln över de kända utbetalningarna beskrivas som Tabell 6.

	Utvecklingsår					
Skadeår	0	1	2	...	$m-2$	$m-1$
1	X_{10}	X_{11}	X_{12}	...	$X_{1,m-2}$	$X_{1,m-1}$
2	X_{20}	X_{21}	X_{22}	...	$X_{2,m-2}$	
3	X_{30}	X_{31}	X_{32}	...		
$\vdots$	$\vdots$	$\vdots$	$\vdots$			
$m-1$	$X_{m-1,0}$	$X_{m-1,1}$				
m	$X_{m,0}$					

Tabell 6: Känd utbetalningar, inkrementell triangel

Låt N_{ij}^{paid} vara det totala antalet utbetalningar gjorda i $(i,j) \in \mathcal{A}_0$

Vi definierar den nedre högra trapetsoiden av X_{ij} som ska predikteras enligt

$$\mathcal{A}_0^* := \{(i,j) \in \mathbb{N} \times \mathbb{N}_0 : i \leq m, j \leq m-1+d, i+j \geq m+1\}.$$

Sedan definerar vi triangeln för antalet observerade skador som vi hittills känner enligt

$$\aleph_0 := \sigma\{N_{ij} : (i,j) \in \mathcal{A}_0\}.$$

Det finns en reservsättningsmodell som beskrivs i (Verrall, 2010) som benäms för VNJ-modellen. I denna modell antas att alla skador har högst en utbetalning. Med detta antagande visar författarna hur modellen kan anpassas med Quasi-Poisson GLM och hur man predikterar den totala utbetalningen, läs mer om detta i (Verrall, 2010). Man kan ana att detta antagande inte är realistiskt, därför introducerar vi en ny modell som vi kallar för CRM. Skillnaden mellan CRM och VNJ-modellen är att CRM tillåter multipla utbetalningar per skada, i övrigt fungerar den som VNJ-modellen.

Definition av CRM enligt följande:

Låt det totala antalet skador som uppkommer i skadeår i som raporterats i period j, N_{ij}, vara en överspridd Poissonfördelning enligt $N_{ij} \sim ODP(\nu, \phi)$,

$$\mathbb{E}[N_{ij}] = \nu_{ij},$$

$$\mathrm{Var}(N_{ij}) = \phi \mathbb{E}[N_{ij}],$$

där alla N_{ij} antas vara oberoende. Vidare, för varje skada kommer det bli en serie av utbetalningar enligt en diskret-tids Poissonprocess, definerad enligt

$$N_{i,j,k}^{paid} | N_{ij} \sim Po(\lambda_k N_{ij}), \qquad \lambda_k \geq 0, k = 0, 1, 2, ..., d,$$

där N_{ijk}^{paid} betecknar antalet utbetalningar från skador som har rapporterats i (i, j), och som har betalning k perioder efter att de har rapporterats. Låt sedan $\tilde{Y}_{ijkl}$ beteckna den l:te utbetalningen från skador som har rapporterats i (i, j) och som har betalningar k perioden efter att de rapporterats, där alla $\tilde{Y}_{ijkl}$ antas vara oberoende av alla l för fixt (i, j, k) med

$$\mathbb{E}[\tilde{Y}_{ijkl}] = \mu_{ijk},$$

$$\mathrm{Var}(\tilde{Y}_{ijkl}) = \sigma_{ijk}^2.$$

Notera att definitionen för N_{ij}^{paid} är

$$\begin{aligned} N_{ij}^{paid} &= N_{ij0}^{paid} + N_{i,j-1,1}^{paid} + ... + N_{i,j-\min\{j,d\},\min\{j,d\}}^{paid} \\ &= \sum_{k=0}^{\min\{j,d\}} N_{i,j-k,k}^{paid}, \end{aligned}$$

det vill säga antalet utbetalningar i (i,j) är

$$N_{ij}^{paid} | \{N_{il}\}_{l \leq j} \sim Po\left(\sum_{k=0}^{\min\{j,d\}} \lambda_k N_{i,j-k} \right).$$

Vidare kan den totala aggregerade inkrementella utbetalningen i (i,j) skrivs enligt följande

$$\begin{aligned} X_{ij} &:= \sum_{k=0}^{\min\{j,d\}} X_{i,j-k,k} \\ &:= \sum_{k=0}^{\min\{j,d\}} \sum_{l=0}^{N_{i,j-k,k}^{paid}} \tilde{Y}_{i,j-i,k,l}, \end{aligned}$$

viket implicerar följande betalningsdynamik för CRM:

Sats 1. För CRM är

$$\begin{aligned} \mathbb{E}[X_{ij}|\aleph_0] &= \sum_{k=0}^{\min\{j,d\}} \mathbb{E}[X_{i,j-k,k}|\aleph_0] \\ &= \sum_{k=0}^{\min\{j,d\}} \mu_{i,j-k,k} \mathbb{E}[N_{i,j-k,k}^{paid}|\aleph_0] \\ &= \sum_{k=0}^{\min\{j,d\}} \lambda_k \mu_{i,j-k,k} N_{i,j-k}, \end{aligned}$$

och

$$
\begin{aligned}
\mathrm{Var}(X_{ij}|\aleph_0) &= \sum_{k=0}^{\min\{j,d\}} \mathrm{Var}(X_{i,j-k,k}|\aleph_0) \\
&= \sum_{k=0}^{\min\{j,d\}} \frac{\sigma^2_{i,j-k,k} + \mu^2_{i,j-k,k}}{\mu_{i,j-k,k}} \mathbb{E}[X_{i,j-k,k}|\aleph_0],
\end{aligned}
$$

där alla X_{ij} och X_{ijk} antas vara oberoende.

Det finns ett problem med den allmänna formuleringen av Sats 1. Det är att parametrarna $\mu_{i,j,k}$ och $\sigma^2_{i,j,k}$ inte kan beräknas för den framtida trapetsoiden $\mathcal{A}_0^*$ baserat på data i $\mathcal{A}_0$. För att undvika detta kan man välja till exempel sätta $\mu_{ijk} = \mu$ och $\sigma^2_{ijk} = \sigma^2$.

4.3.1 Den homogena CRM

Vi förenklar antagande för fördelning av utbetalningar. Vi börjar med att formerera

Antagande 3.

Alla $\tilde{Y}_{ijkl}$ är oberoende med

$$
\mathbb{E}[\tilde{Y}_{ijkl}] = \mu,
$$

$$
\mathrm{Var}(\tilde{Y}_{ijkl}) = \sigma^2.
$$

Detta förenklade antagande innebär mer eller mindre att utbetalningsbeloppen är oberoende och likafördelade. Givet Antagande 3 erhåller vi:

Följd 1.

$$
\mathbb{E}[X_{ij}|\aleph_0] = \mu \sum_{k=0}^{min\{j,d\}} \lambda_k N_{i,j-k},
$$

och

$$\mathrm{Var}(X_{ij}|\aleph_0) = \frac{\sigma^2 + \mu^2}{\mu}\mathbb{E}[X_{ij}|\aleph_0]. \tag{4}$$

X_{ij} kan alltså uttryckas som en överspridd Poissonfördelning med väntevärde $\mathbb{E}[X_{ij}|\aleph_0]$ och överspriddnings parameter

$$\varphi := \frac{\sigma^2 + \mu^2}{\mu},$$

där alla X_{ij} är oberoende.

Vidare noterar vi att om

$$\Lambda := \sum_{k=0}^{d} \lambda_k$$

$\pi_k = \lambda_k/\Lambda$ och $\tilde{\mu} = \mu\Lambda$, så följer det att

$$\mathbb{E}[X_{ij}|\aleph_0] = \mu \sum_{k=0}^{\min\{j,d\}} \lambda_k N_{i,j-k} = \tilde{\mu} \sum_{k=0}^{\min\{j,d\}} \pi_k N_{i,j-k}.$$

Här kan $\tilde{\mu}$ tolkas som väntevärdet av det totala utbetalningsbeloppet för en enda skada och π_k motsvarar fördelningen av utbetalningar över tiden från den är rapporterad, eftersom $\pi_0 + \pi_1 + ... + \pi_d = 1$. Vidare noterar vi att i CRM kan vi utnyttja standard GLM teori som i (Verrall, 2010) genom att införa

$$\psi_k := \tilde{\mu}\pi_k,$$

där

$$\sum_{k=0}^{d} \pi_k = 1.$$

I vår situation har vi

$$\mathbb{E}[X_{ij}|\aleph_0] = \sum_{k=0}^{\min\{j,d\}} \psi_k N_{i,j-k}. \tag{5}$$

Med formeln (5) ser vi att vi kan ställa upp den på matrisform enligt nedan

$$\begin{bmatrix} \mathbb{E}[X_{10}|\aleph_0] \\ \mathbb{E}[X_{20}|\aleph_0] \\ \mathbb{E}[X_{30}|\aleph_0] \\ \vdots \\ \mathbb{E}[X_{m0}|\aleph_0] \\ \mathbb{E}[X_{11}|\aleph_0] \\ \vdots \\ \mathbb{E}[X_{m-1,1}|\aleph_0] \\ \vdots \\ \mathbb{E}[X_{1,m-1}|\aleph_0] \end{bmatrix} = \begin{bmatrix} N_{10} & 0 & 0 & \dots & 0 \\ N_{20} & 0 & 0 & \dots & 0 \\ N_{30} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ N_{m0} & 0 & 0 & \dots & 0 \\ N_{11} & N_{10} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ N_{m-1,1} & N_{m-1,0} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ N_{1,m-1} & N_{1,m-2} & N_{1,m-3} & \dots & N_{1,0} \end{bmatrix} \cdot \begin{bmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{m-1} \end{bmatrix}, \tag{6}$$

och på samma sätt för ekvation (4).

Nu har vi en överspridd Poisson GLM och parametrarna ψ_k kan estimeras genom att maximera kvasilikelihooden

$$l^p(\boldsymbol{\psi}; \mathcal{F}_0) = \sum_{i,j \in \mathcal{A}_0} \Big[X_{ij} \log \Big(\sum_{k=0}^{\min\{j,d\}} \psi_k N_{i,j-k} \Big) - \sum_{k=0}^{\min\{j,d\}} \psi_k N_{i,j-k} \Big], \tag{7}$$

vilket måste göras numeriskt. Vi antar även $\mathbb{E}[N_{ij}] = \nu_{ij} = \alpha_i \beta_j$, Enligt (Miranda, 2012) kan α_i och β_j skattas som

$$\hat{\alpha}_i = \sum_{j=0}^{m-i} N_{ij} \prod_{j=m-i+1}^{m-1} \hat{f}_j,$$

$$\hat{\beta}_0 = \frac{1}{\prod_{l=1}^{m-1} \hat{f}_l},$$

$$\hat{\beta}_j = \frac{\hat{f}_j - 1}{\prod_{l=j}^{m-1} \hat{f}_l},$$

där $\hat{f}_j$ är utvecklingsfaktor enligt ekvation (1).

Genom att göra ovanstående parametrisering har vi

$$\sum_{k=0}^{d} \psi_k = \tilde{\mu} \sum_{k=0}^{d} \pi_k = \tilde{\mu} = \mu\Lambda.$$

Nu har vi explicit kunskap om $\{N_{ij}^{paid}\}$ men vi kan fortfarande inte identifiera μ och Λ, eftersom vi har uttryckt alla prediktorer av utestående betalningarna i termer av ψ_k. Vi kan däremot estimera μ genom

$$\hat{\mu} = \frac{1}{|\mathcal{A}_0|} \sum_{i,j \in \mathcal{A}_0} \frac{X_{ij}}{N_{ij}^{paid}},$$

där $|\mathcal{A}_0| = \frac{m(m+1)}{2}$ är storleken på den observerade triangeln. Vi uttnyttjar detta för att estimera parametrarna $\hat{\pi}_k, \hat{\Lambda}, \hat{\lambda}_k$ och $\hat{\sigma}^2$ enligt

$$\begin{cases} \hat{\pi}_k = \frac{\hat{\psi}_k}{\sum_{j=0}^{d} \hat{\psi}_j}, \\ \hat{\Lambda} = \frac{\sum_{j=0}^{d} \hat{\psi}_j}{\hat{\mu}}, \\ \hat{\lambda}_k = \frac{\hat{\psi}_k}{\hat{\mu}}, \\ \hat{\sigma}^2 = \varphi\hat{\mu} - (\hat{\mu})^2, \end{cases}$$

där överspridnings parametern φ är skattad genom att använda (11) i (Miranda, 2011)

$$\hat{\varphi} = \frac{1}{|\mathcal{A}_0| - (d+1)} \sum_{i,j \in \mathcal{A}_0} \frac{\left(X_{ij} - \sum_{k=0}^{\min\{j,d\}} \hat{\psi}_k N_{i,j-k}\right)^2}{\sum_{k=0}^{\min\{j,d\}} \hat{\psi}_k N_{i,j-k}.}$$

ϕ kan skattas enligt nedan

$$\hat{\phi} = \frac{1}{|\mathcal{A}_0| - m(m-1)} \sum_{i,j \in \mathcal{A}_0} \frac{(N_{ij} - \hat{\nu}_{ij})^2}{\hat{\nu}_{ij}}.$$

Eftersom vi ska dela upp framtida utbetalningar i RBNS och IBNR, låt

$$N_{ij}^{\mathcal{R}} := \sum_{k=j-(m-i)}^{\min\{j,d\}} N_{i,j-k,k}^{paid}, \qquad (i,j) \in \mathcal{A}_0^*,$$

motsvara det totala antalet av RBNS skadeanmälningar i (i,j), och låt vidare

$$N_{ij}^{\mathcal{I}} := \sum_{k=0}^{j-(m-i)-1} N_{i,j-k,k}^{paid}, \qquad (i,j) \in \mathcal{A}_0^*,$$

vara det totala antalet av IBNR skadeanmälningar i (i,j). Då är de totala RBNS skadebetalningarna i $(i,j) \in \mathcal{A}_0^*$

$$X_{ij}^{\mathcal{R}} := \sum_{l=0}^{N_{i,j}^{\mathcal{R}}} Y_{i,j,l}^{\mathcal{R}}, \qquad (i,j) \in \mathcal{A}_0^*,$$

där $Y_{i,j,l}^{\mathcal{R}}$ är den l:te betalningen i (i,j) som tillhör RBNS-skador. Detta innebär att för ett fixt (i,j) behöver vi beräkna alla $\tilde{Y}_{i,j-k,k,p}$ med $k \geq j-(m-i)$. Liknande princip gäller för IBNR, de totala IBNR skadebetalningarna i $(i,j) \in \mathcal{A}_0^*$ är

$$X_{ij}^{\mathcal{I}} := \sum_{l=0}^{N_{i,j}^{\mathcal{I}}} Y_{i,j,l}^{\mathcal{I}}, \qquad (i,j) \in \mathcal{A}_0^*,$$

där $Y_{ijl}^{\mathcal{I}}$ är den l:te betalningar i (i,j) från IBNR skador. Så för fixt (i,j) beräknar man alla $\tilde{Y}_{i,j-k,k,p}$ med $k < j-(m-i)$. Från detta kan vi slutligen få den totala framtida utbetalningen som tillhör RBNS-skador för skadeår i

$$R_i^{\mathcal{R}} := \sum_{j=m-i+1}^{m-1+d} X_{ij}^{\mathcal{R}}, \tag{8}$$

och motsvarande utryck för IBNR-skador för skadeår i blir

$$R_i^{\mathcal{I}} := \sum_{j=m-i+1}^{m-1+d} X_{ij}^{\mathcal{I}}. \tag{9}$$

Baserat på relationen (8) och (9) får vi väntevärdet av framtida skadeutbetalningar för RBNS och IBNR enligt

$$\begin{aligned} h_i^{\mathcal{R}}(\boldsymbol{\psi}; \mathcal{N}_0) &:= \mathbb{E}\big[R_i^{\mathcal{R}} | \mathcal{N}_0\big] \\ &= \sum_{j=m-i+1}^{m-1+d} \sum_{k=j-m+i}^{\min\{j,d\}} \psi_k N_{i,j-k}, \end{aligned} \tag{10}$$

$$\begin{aligned} h_i^{\mathcal{I}}(\boldsymbol{\psi}; \nu) &:= \mathbb{E}\big[R_i^{\mathcal{I}} | \mathcal{N}_0\big] \\ &= \sum_{j=m-i+1}^{m-1+d} \sum_{k=\max\{0,j-m+1\}}^{\min\{j-m+i-1,d\}} \psi_k \nu_{i,j-k}, \end{aligned} \tag{11}$$

Med hjälp av ekvationerna (10) och (11) kan vi göra plug-in skattningar av RBNS och IBNR skadereserverna genom

$$h_i^{\mathcal{R}}(\hat{\boldsymbol{\psi}}; \mathcal{N}_0)$$

och

$$h_i^{\mathcal{I}}(\hat{\boldsymbol{\psi}}; \hat{\nu})$$

Den totala reserven för RBNS och IBNR blir då summan över alla RBNS och IBNR reserven för varje i, vilket är

$$\hat{R}_{tot}^{\mathcal{R}} = \sum_{i=1}^{m} h_i^{\mathcal{R}}(\hat{\boldsymbol{\psi}}; \mathcal{N}_0); \tag{12}$$

$$\hat{R}_{tot}^{\mathcal{I}} = \sum_{i=2}^{m} h_i^{\mathcal{I}}(\hat{\boldsymbol{\psi}}; \hat{\nu}). \tag{13}$$

Av uttryck (12) och (13) får vi den totala reserven för CRM

$$\hat{R}_{tot} = \hat{R}_{tot}^{\mathcal{R}} + \hat{R}_{tot}^{\mathcal{I}}$$

5 Val av fördelningar till simuleringen

I detta avsnitt ska vi gå igenom vilka fördelningar vi har valt till vår simulering av data.

När vi simulerar data behöver vi minst fem steg. Vi simulerar två trianglar, en för rapportetade skadeanmälningar och en för utbetalningsbelopp för skadeår i och utvecklingsår j. I vårt fall har vi valt att simulera data för 10 skadeår, alltså $m = 10$. Teorin om fördelningarna är hämtad från (Wütrich, 2015).

1. Först simulerar vi antal skador. Låt N beteckna antalet skador som inträffar under skadeår i, vilket vi valt låta vara Poisssonfördelad med parametrar λw, där w motsvarar antalet aktiva kontrakt och λ är kopplat till sannolikheten, att ett enskilt kontrakt får en skadehändelse. Sannolikhetsfunktionen för N är

$$p_k = P(N = k) = e^{-(\lambda w)} \cdot \frac{(\lambda w)^k}{k!}, \qquad k = 0, 1, 2, ...$$

   vidare är Ns väntevärde och varians samma

$$\mathbb{E}[N] = \mathrm{Var}(N) = \lambda w,$$

2. Låt T beteckna tiden då en skada inträffar under skadeår i. Vi har valt att låta T vara Beta-fördelad. Täthetsfunktionen för Beta-fördelningen är

$$f(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1-t)^{\beta-1}, \qquad 0 < t < 1,$$

där α och β är parametrarna i fördelningen. Väntevärdet och variansen av T är

$$\mathbb{E}[T] = \frac{\alpha}{\alpha+\beta},$$

$$\mathrm{Var}(T) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)},$$

3. Varje skada har $0, 1, 2, ...$ utbetalningar. Låt M beteckna antal utbetalningar en skada har. Vi har valt att låta $M \sim NegBin(r, p)$. Sannolikhetsfunktionen för Negativ Binomialfördelningen är

$$P(M = m) = \binom{n-1}{r-1} p^r (1-p)^{m-r}, \qquad m = r, r+1, ...$$

där $0 < p < 1$, är sannolikhet att lyckas, tills man har lyckats r gånger.

Väntevärdet och variansen är

$$\mathbb{E}[M] = \frac{r}{p},$$

$$\mathrm{Var}(M) = \frac{r(1-p)}{p^2}.$$

4. Vidare låter vi U beteckna tiden mellan utbetalningarna. Vi väljer U som Gammafördelning vars täthetsfunktionen är

$$f(u) = \frac{1}{\Gamma(\alpha)\beta^\alpha} u^{\alpha-1} e^{-u/\beta}, \qquad 0 < u < \infty.$$

α och β är parametrarna i fördelningen och Γ betecknar gammafunktionen. Väntevärdet och variansen är

$$\mathbb{E}[U] = \alpha\beta,$$

$$\mathrm{Var}(U) = \alpha\beta^2.$$

5. Till sist simulerar vi Y som är betalningsstorleken. Fördelningen för Y väljer vi som en Lognormalfördelning.

   Täthetsfunktionen för en Lognormalfördelning ges av

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\{-\frac{(\log y - \mu)^2}{2\sigma^2}\}, \qquad y > 0,$$

   Väntevärdet och variansen är

$$\mathbb{E}(Y) = e^{\mu+\sigma^2/2}$$

$$\mathrm{Var}(Y) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$$

Vi antar att $Y_1, Y_2, ...$ är oberoende, att de alla har samma fördelning samt att de är oberoende av N. Observera att genom att simulera enligt steg 1-5 ovan, så simulerar vi en betalningsprocess i kontinuerlig tid, medan Chain Ladder och CRM är definierad i diskret tid.

6 Dataanalys

Enlig denna simulering kan vi generera trianglar med rapporterade skadeanmälningar och utbetalningsbelopp. Fördelen med att simulera data är att vi vet de

faktiska skadeutbetalningarna i den nedre högra trapetsoiden som tillhör $\mathcal{A}_0^*$. Med denna information kan vi jämföra de skattade skadeutbetalningarna med de faktiska simulerade skadeutbetalningarna.

I verkligheten har vissa försäkringar korta utvecklingsår, där tiden mellan det att skadorna inträffar tills det att samtliga utbetalningar är slutreglerade är relativt kort. Ett exempel på detta är husdjursförsäkring. Det finns ocskå andra försäkringar där utvecklingsperioden är lång, som till exempel sjuk- och olycksfallsförsäkring eller barnförsäkring. När vi simulerar data själv så kan vi bestämma fördelningarnas väntevärde och varians och på så sätt få vi olika längd på utvecklingsperioderna.

I detta arbete väljer vi att simulera data för 10 skadeår, $\lambda = 0.2$, väntevärdet av M (antal utbetalningar en skada har) till 1 och standardavvikelsen av M till 1.045, väntevärdet av Y (betalningsstorlek) till 1, standardavvikelsen av Y till 0.556. Vi är intresserade av att jämföra hur bra modellerna fungerar när vi ökar fördröjningstiden U och antalet aktiva kontrakt w. Vi väljer därför att simulera två typer av dataset, där fördröjningstiderna är olika. För varje typ av dataset gör vi en simulering med färre kontrakt och en med fler kontrakt.

Dataset av typ 1 har medel fördröjningstid på 1.76 år, och standardavvikels 0.39929; dataset typ 2 har medel fördröjningstid på 2 år och standardavvikelse 1.408791. Varje typ av dataset simuleras med w=(3 000, 3 500, 4 000, 4 500, 5 000, 6 500, 7 000, 7 500, 8 000, 8 500), där antalet aktiva kontrakt ökar med 500 stycken varje år samt w=(10 000, 15 000, 20 000, 25 000, 30 000, 35 000, 40 000, 45 000, 50 000, 55 000), där antalet aktiva kontrakt ökar med 5 000 stycken per år. Vi simulerar 10 000 datamaterial av typ 1 och typ 2, för att sedan jämföra reserven som skattas med CRM och reserven som skattas med Chain Ladder-modellen med de faktiska simulerade utbetalningarna som referens. För att lättare skilja på de olika dataseten introducerar vi de följande beteckningar:

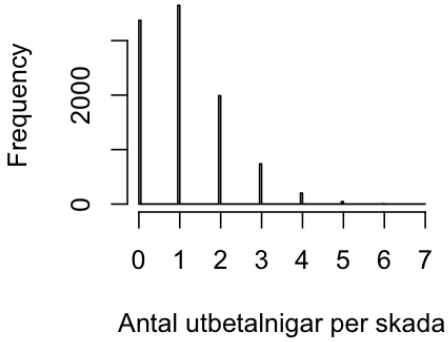
$$D1_{\min} := \text{dataset typ 1 med färre kontrakt}$$

$$D1_{\max} := \text{dataset typ 1 med fler kontrakt}$$

$$D2_{\min} := \text{dataset typ 2 med färre kontrakt}$$

$$D2_{\max} := \text{dataset typ 2 med fler kontrakt}$$

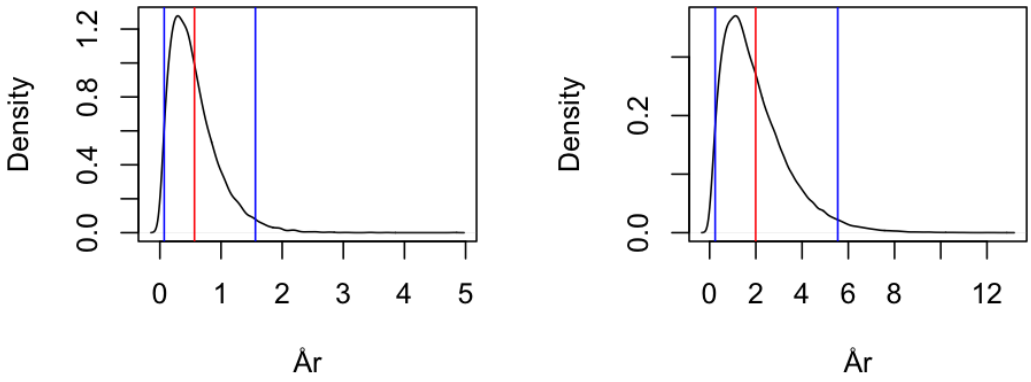
Figur 2 visar ett histogram över antal utbetalningar per skada som vi har använt i denna analys, vilket är samma för de olika typer av datameterial.



Figur 2: Antal utbetlningar per skada

Vi ser från Figur 2 att de flesta skador har mellan 0 till 1 utbetalning. För vissa skador förekommer även 7 utbetalningar.

Vi vill även undersöka hur betalningsfördröjningen ser ut för respektive typ av dataset. Figur 3 visar fördelningen, väntevärdet, 2.75% och 97.5% procentilerna för U.



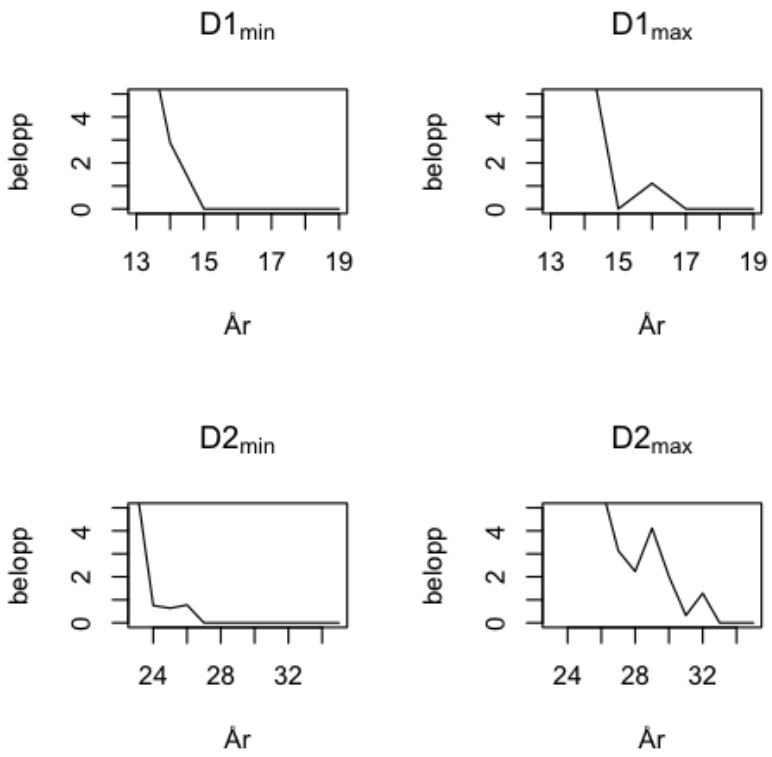
(a) Dataset typ 1

(b) Dataset typ 2

Figur 3: Betalningsfördröjningar, Den röda linjen är väntevärdet. De blåa linjerna är 2.75% och 97.5% percentilerna

Från Figur 3 ser vi att för dataset typ 1 är fördröjningen mycket kortare än dataset typ 2.

Vidare vill vi analysera kassaflödet för vår simulerade data. Vi beräknar kassaflödet för varje år baserat på första simuleringen. Figur 23 (Appendix) visar hela förloppet. Från Figur 23 (Appendix) ser man att kassaflöde ökar i början, sedan minskar det, till sist blir det 0. Vi ser också att dataset typ 1 har kortare betalningstid än dataset typ 2. Eftersom vi i detta fall är intresserade av exakt när utbetalningen är slutreglerad för hela portföljen, så zoomar vi in Figur 23 (Appendix). Figur 4 nedan visar i vilket år kassaflöde blir 0 för respektive dataset.



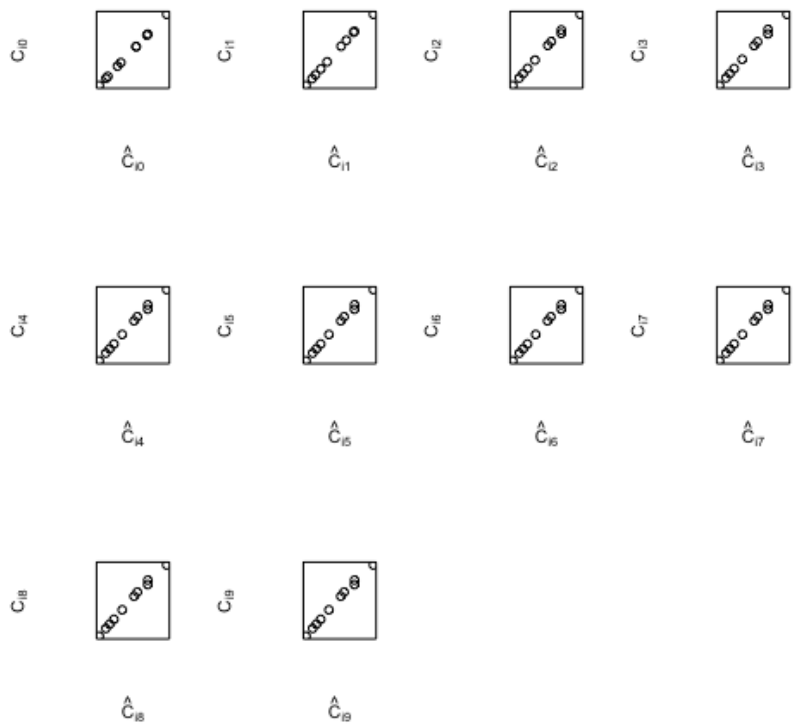
Figur 4: Svansen av kassaflöde för de fyra dataseten baserat på första simulering

Från Figur 4 ser vi att alla utbetalningarna är slutreglerade vid år 15 för $D1_{\min}$. För $D1_{\max}$ slutregleras utbetalningarna vid år 17. Vi noterar vidare att utbetalningarna för $D2_{\min}$ är slutreglerad vid år 27. För $D2_{\max}$ slutregleras utbetalningen vid år 33. Lägg märke till att beloppen är väldigt små för de här senaste åren. Vi noterar att detta gäller för första simulering. Eftersom vi simulerar data 10 000 gånger så kan kassaflöde variera något.

6.1 Modellantagande för Chain Ladder-modellen

Chain Ladder har två antaganden: Antagande 1 är att utbetalningsmönstret för varje utvecklingsår är samma. Antagande 2 är att varje skadeår är oberoende av varandra.

För att undersöka om modellantaganden uppfyller plottar vi de skattade akumulerade utbetalningsbeloppen för utvecklingsår j, alltså $\hat{C}_{i,j}$ mot de faktiska akumulerade utbetalningsbeloppen för utvecklingsår j, alltså $C_{i,j}$ för $j = 0, 1, ..., 9$ baserad på den första simuleringen för varje datatyp. Figur 5 visar resultatet för $D1_{\min}$. Figurerna 24, 25 samt 26 (Appendix) visar resultatet för $D1_{\max}$, $D2_{\min}$ och $D2_{\max}$.



Figur 5: Akumulerade utbetalningsbelopp $\hat{C}_{i,j}$ mot $C_{i,j}$ baserad på den första simuleringen, $D1_{\min}$

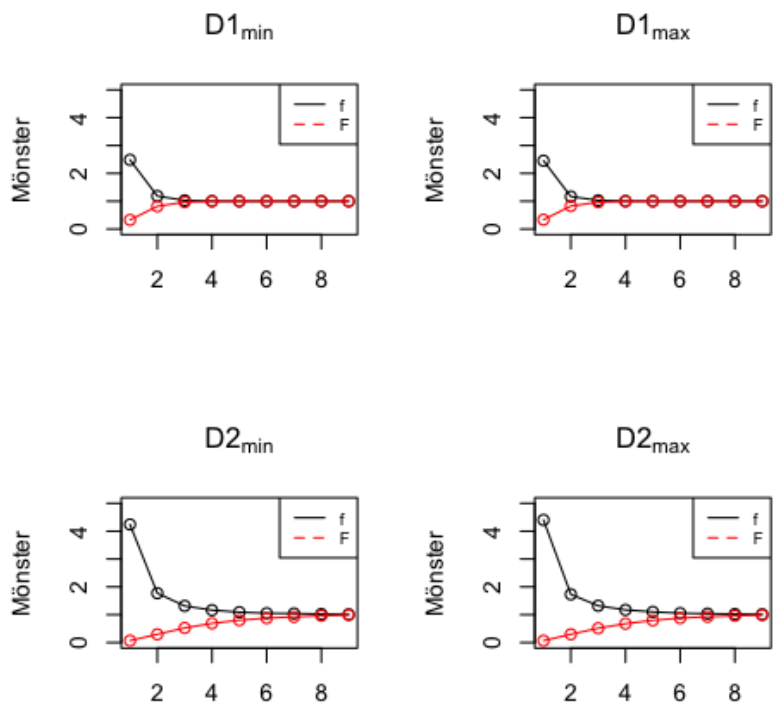
Om modellantaganden är bra ska vi få en rät linje med 45 graders lutning. Från Figur 5 ser vi att linjer för samtliga plottar är ganska raka och har 45 graders lutning. I Figurerna 24, 25 och 26 (appendix) ser vi liknade mönster.

Vidare vet vi att för varje skadeår är antalet skador N Poissonfördelat och utbetalningen $Y_1, Y_2, ...$ är Lognormalfördelad och att de är oberoende av N. Detta ger oss att alla skadeår är oberoende av varandra. Därmed säger vi att antaganden för Chain Ladder-modellen är uppfylld.

Vi beräknar utvecklingsfaktorer f_j, $j = 1, 2, ..., 9$ och andel hittills utbetalt $1/F_j$, $j = 1, 2, ...$, för att kontrollera om utvecklingsmönstret är rimligt. Där Lag-faktor F_j definieras som

$$F_j = f_j \cdot f_{j+1} \cdots f_{J-1}$$

Figur 6 visar utbetalningsmönstret för vår två typer av dataset baserat på första simuleringen. Den övre kurvan är våra utvecklingsfaktorer och den nedre är andel hittills utbetalt.



Figur 6: Utvecklingsfaktorer och andel hittills utbetalt baserad på den första simuleringen

Från Figur 6 ser vi att de största av utbetalningsbeloppen hamnar på de tidigaste utvecklingsåren och avtar sedan successivt. Utifrån denna figur drar vi slutsatsen att utvecklingsmönsteret för vår simulerad data är rimligt. Vi noterar också att betalningen slutregleras betydligt fortare för dataset typ 1 än för dataset typ 2. Tabell 7 visar de siffror som vi har använt till Figur 6 samt det faktiska simulerade andel hittils utbetalt.

	$D1_{\min}$			$D1_{\max}$			$D2_{\min}$			$D2_{\max}$		
År	f	F	Fak.	f	F	Fak.	f	F	Fak.	f	F	Fak.
1	2.486	0.328	0.333	2.453	0.337	0.336	4.241	0.070	0.068	4.403	0.067	0.067
2	1.186	0.816	0.822	1.168	0.826	0.826	1.773	0.296	0.293	1.730	0.297	0.295
3	1.029	0.968	0.968	1.030	0.965	0.965	1.316	0.524	0.516	1.322	0.514	0.514
4	1.004	0.996	0.995	1.005	0.994	0.995	1.164	0.690	0.682	1.173	0.679	0.678
5	1.000	1.000	1.000	1.001	0.999	0.999	1.092	0.803	0.795	1.102	0.797	0.796
6	1.000	1.000	1.000	1.000	1.000	1.000	1.053	0.877	0.873	1.055	0.878	0.876
7	1.000	1.000	1.000	1.000	1.000	1.000	1.044	0.923	0.926	1.038	0.927	0.929
8	1.000	1.000	1.000	1.000	1.000	1.000	1.024	0.964	0.961	1.024	0.962	0.964
9	1.000	1.000	1.000	1.000	1.000	1.000	1.013	0.987	0.985	1.015	0.985	0.986

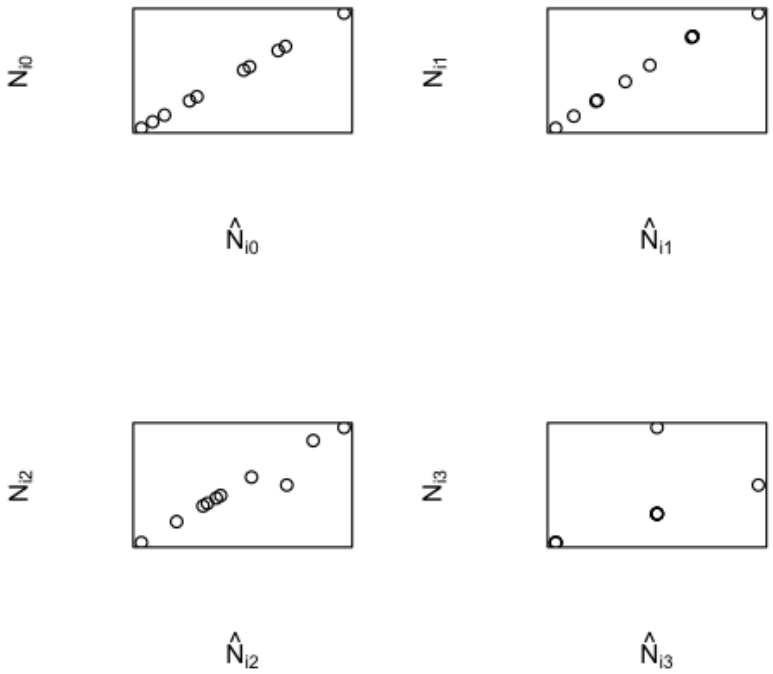
Tabell 7: Utvecklingsfaktorer f, andel hittils utbetalt F och det faktiska simulerade andel hittils utbetalt Fak.

Från Tabell 7 ser vi att andel hittils utbetalt F som är beräknad med Chain Ladder-modellen slutar efter 4 år för $D1_{\min}$. För $D1_{\max}$ slutar det efter 5 år. Alltså utbetalningar för dataset typ 1 slutregleras tidigt. Däremot för dataset typ 2 är situationen annorlunda. Vid år 9 är andel hittils utbetalt F för $D2_{\min} = 98.7\%$ och $D2_{\max} = 98.5\%$. De här siffrorna visar $D2_{\max}$ har mest kvar att betala efter 9 år. Man kan också se att utbetalningamönster för F ligger ganska nära det faktiska simulerade utbetalningsmönstret, vilket betyder att Chain Ladder-modellen fungerar bra på det data vi har simulerat. Eftersom Chain Ladder-modellen bara kan prediktera 10 år framåt, så kan den inte fånga de utbetalningar som inträffar efter utvecklingsår 9.

6.2 Modellantagande för CRM

I CRM skattar vi två trianglar: en är inkrementell triangel för antal rapporterade skador och en är inkrementell triangel för skadeutbetalningar. Vi börjar med att undersöka modellantagendet för antal rapporterade skador. Vi plottar de skattade antal rapporterade skador för varje utvecklingsår j mot de faktiska antal rapporterade skador baserad på den första simuleringen. Om modellantagnadet är bra ska vi få en rät linje med 45 graders lutning. Figur 7 visar

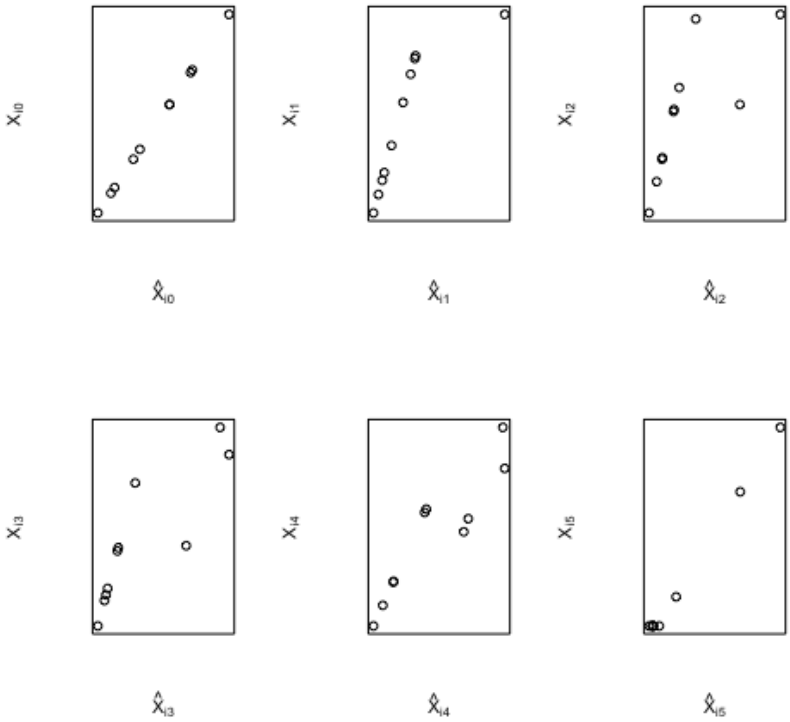
antal rapporterade skador $\hat{N}_{i,j}$ mot $N_{i,j}$ för $D1_{\min}$. Figurerna 27, 28 samt 29 (Appendix) prensenterar för $D1_{\max}$, $D2_{\min}$ och $D2_{\max}$.



Figur 7: Antal rapporterade skador $\hat{N}_{i,j}$ mot $N_{i,j}$ baserad på den första simuleringen, $D1_{\min}$

Vi noterar att både för $D1_{\min}$ och $D1_{\max}$ är utvecklingsperioderna korta och efter utvecklingsår 4 är antal rapporterade skador 0, därför plottar vi bara 4 utvecklingsår. Från Figurerna ser vi att modellantaganena ser ganska OK ut. Ju högre utvecklingsår är desto mindre information får vi för att prediktera framtida antal rapporterade skador. Detta kan vara anledning till att linjerna inte är räta för de högre utvecklingsåren.

Nu vill vi analysera vidare på den inkrementella triangel för skadeutbetalningar. Vi plottar de skattade skadeutbetalningar för varje utvecklingsår j mot de faktiska skadeutbetalningar baserad på den första simuleringen. Om modellantagnadet är bra ska vi få en rät linje med 45 graders lutning. Figur 8 visar antal rapporterade skador $\hat{X}_{i,j}$ mot $X_{i,j}$ för $D1_{\min}$. Figurerna 30-34 (Appendix) prensenterar för $D1_{\max}$, $D2_{\min}$ och $D2_{\max}$.



Figur 8: Skadeutbetalningar $\hat{X}_{i,j}$ mot $X_{i,j}$ baserad på den första simuleringen, $D1_{\min}$

Från Figurerna ser vi att modellantagande ser hyfsat bra ut, men inte lika bra som för N_{ij}. Med ovanstående analys drar vi slutsats om att modellantagande uppfyller någorlunda bra för CRM.

Vidare vill vi undersöka modellantagande för varians med CRM genom att analysera variansantaganderna för X_{ij}. Enligt formeln (4) får vi

$$\frac{\mathrm{Var}(X_{ij}|\aleph_0)}{\mathbb{E}[X_{ij}|\aleph_0]} = \frac{\sigma^2 + \mu^2}{\mu}.$$

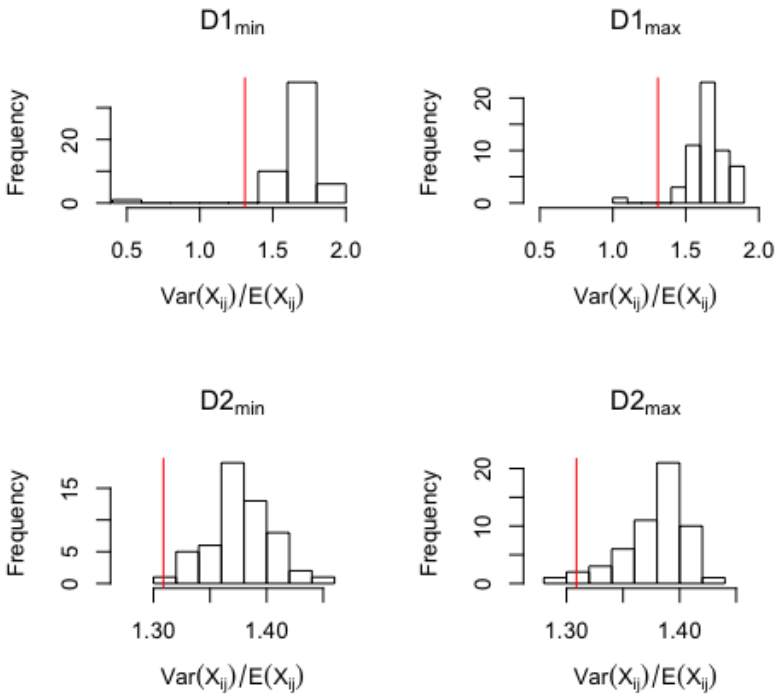
Låt

$$c_{ij} := \frac{\mathrm{Var}(X_{ij}|\aleph_0)}{\mathbb{E}[X_{ij}|\aleph_0]},$$

och

$$l := \frac{\sigma^2 + \mu^2}{\mu}.$$

Vi simulerar datan 10 000 gånger och beräknar c_{ij} som stickprovsvärde. Vi får då totalt 55 stycken. Standardavvikelsen för betalningsstorlekar Y (se Avsnitt 5) är 0.556, vi vet att $E[Y] = 1$, då får vi $l = 1.308792$. I Figur 9 har vi plottat de 55 värden som ett histogram och den röda vertikallinjen är l.



Figur 9: stickprovsvärde mot det teoretiskt värde

Enligt modellen ska de 55 stickprovsvärdena konvergera mot det teoretiska värdet, $l = 1.308792$. Från Figur 9 ser vi att även om vi ökar antal stickprov så verkar de inte konvegera mot det teoretiska värdet, därför säger vi att det finns en signifikant skillnad mellan simulerad data och det teoretiska värdet. Så modellantagande för varians verkar inte vara uppfyllt för modellen. Detta resultat är inte konstigt. Vår data kommer ifrån en kontinuerlig process. Om skadan rapporteras vid årets slut, så är sannolikheten att betalningarna hinner göras under samma år är liten.

6.3 Problem med simulerade trianglar

Vi simulerar data för 10 skadeår men när utbetalningstiden är kort kan utbetalningar bli 0 innan utvecklingsår 9. Eftersom vi simulerar data 10 000 gånger, kan vi får problem med kvasimaximumlikelihoodanpassning (QML, eng. quasi maximum likelihood). Vi kan då inte skatta så många parametrar som vi önskar. Detta är ett numeriskt problem och går i praktiken att undvika genom att för varje sådant dataset hantera dessa nollor korrekt genom att ta bort de utvecklingsperioder som bara består av nollor. På grund av tidsbrist väljer vi dock i denna uppsats att medvetet skatta färre parametrar, vilket leder till en viss underskattning av reserven för CRM. I detta avsnitt tittar vi på ett exempel där problemet uppstår.

I CRM skattas parametrarna ψ_k genom (QML, eng. quasi maximum likelihood). kvasilikelihood kan beskrivas enligt formel

$$l^p(\boldsymbol{\psi}; \mathcal{F}_0) = \sum_{i,j \in \mathcal{A}_0} \big[X_{ij} \log \big(\sum_{k=0}^{\min\{j,d\}} \psi_k N_{i,j-k} \big) - \sum_{k=0}^{\min\{j,d\}} \psi_k N_{i,j-k} \big]. \tag{14}$$

Vi deriverar $l^p(\boldsymbol{\psi}; \mathcal{F}_0)$ med avseendet på ψ, och sätter derivatan till 0, för att lösa ut $\hat{\psi}$. Vi kan se från formen (14) att om X_{ij} är 0, för något j, då när vi deriverar $l^p(\boldsymbol{\psi}; \mathcal{F}_0)$ får vi ett värde som är mindre eller lika med 0. Detta kan alltså bli ett problem, speciellt för stora värden på k.

Det vi kan göra är att vi eliminerar hela ekvationen från dessa utvecklingsåren. Till exempel om X_{18}, X_{28} och X_{19} är 0, raderar vi de ekvationer som motsvarar

$$\mathbb{E}[X_{18} | \aleph_0] = \sum_{k=0}^{\min\{j,d\}} \psi_k N_{1,8-k},$$

$$\mathbb{E}[X_{28} | \aleph_0] = \sum_{k=0}^{\min\{j,d\}} \psi_k N_{2,8-k}$$

och

$$\mathbb{E}[X_{19}|\aleph_0] = \sum_{k=0}^{\min\{j,d\}} \psi_k N_{1,9-k}.$$

I detta exemplet kan vi beräkna $\hat{\psi}_k$, där $k = 0, 1, 2, ..., 7$.

Tabell 8 visar en triangar som vi har problem med.

	Utvecklingsår									
Skadeår	0	1	2	3	4	5	6	7	8	9
1	211.19	320.64	93.23	20.23	0.00	0	0	0	0	0
2	259.30	380.62	122.58	25.27	2.83	0	0	0	0	
3	272.85	426.33	143.54	26.32	2.78	0	0	0		
4	341.91	451.45	144.91	24.36	1.31	0	0			
5	365.94	539.35	188.10	43.42	7.41	0				
6	474.51	678.57	210.42	32.38	7.20					
7	475.39	770.21	190.05	32.93						
8	552.47	821.17	274.69							
9	558.93	830.09								
10	693.83									

Tabell 8: Problemtriangel; Skadeutbetalning

I Tabell 8 är skadebelopp för utvecklingsår 5-9 lika med 0. Vi eliminerar alla ekvationer som tillhör de utvecklingsåren, och skattar sedan $\hat{\psi}_k$, där $k \neq 5, 6, 7, 8, 9$ med QML.

Eftersom vi simulerar data 10 000 gånger, gör att trianglarna kan se väldigt olika ut. I vissa fall går det inte att numerisk skatta så många $\hat{\psi}_k$ som vi vill. Tabell 9 är exempel på en simulering av $D1_{\min}$, där det inte går att optimera $\hat{\psi}_5$.

	Utvecklingsår									
Skadeår	0	1	2	3	4	5	6	7	8	9
1	210.13	357.25	74.11	24.30	5.82	0.00	0	0	0	0
2	286.84	393.13	101.90	24.60	4.33	0.66	0	0	0	
3	308.09	479.05	129.33	25.14	2.26	1.20	0	0		
4	294.27	517.81	153.46	29.97	3.83	0.00	0			
5	399.16	529.91	151.11	38.78	3.99	0.00				
6	452.95	703.44	220.90	54.66	0.77					
7	548.16	753.54	191.03	26.31						
8	553.00	907.48	241.87							
9	4602.01	852.89								
10	651.76									

Tabell 9: Problem triangel; skadeutbetalningar

Vi ser från Tabell 9 att det finns utbetalningar i utvecklingsår 5, då borde QML kunna beräkna $\hat{\psi}_5$. Men GLM-anpassningsmetoden i R varnar för att den numeriska anpassningsmetoden inte har konvergerat, vilket beror på att anpassningen leder till konvergens mot en överspridd Poisson-variabel med väntevärde 0. På grund av tidsbrist väljer vi att endast skatta $\hat{\psi}_k$, där $k = 0, 1, 2, ..., 4$, för $D1_{\min}$, vilket fungerar för samtliga simulerade datamaterial. Liknande problem får vi med $D1_{\max}$. Här väljer vi att endast skatta $\hat{\psi}_k$, där $k = 0, 1, 2, ..., 6$.. I simuleringsstudien kommer vi att se att följden av att vi medvetet anpassar färre parametrar är att reserven för CRM kommer att underskattas något.

7 Utvärdering av CRM och Chain Ladder-modellen

7.1 Dataset typ 1

För att skatta reserven med Chain Ladder-modellen, beräknar vi först den ackumulerade triangeln från den inkrementella triangeln, sedan räknar vi ut utvecklingsfaktorerna f_j. Med hjälp av f_j kan vi sedan beräkna den skattade ultimo för varje skadeår i. Till sist beräknar vi den totala reserven med hjälp av formel (3).

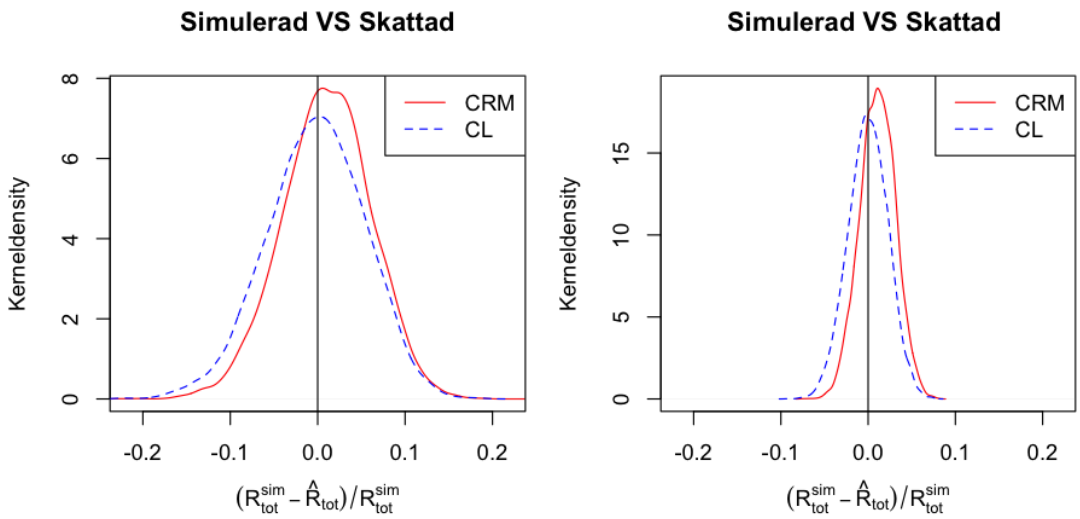
Vi börjar med att analysera och jämföra reserven för båda metoderna. Vi simulerar data 10 000 gånger. För att lättare kunna jämföra resultaten kommer vi att analysera

$$\frac{R^i_{tot} - \hat{R}^{CRM,i}_{tot}}{R^i_{tot}}, \qquad i = 1, 2, ..., 10000 \tag{15}$$

och

$$\frac{R^i_{tot} - \hat{R}^{CL,i}_{tot}}{R^i_{tot}}, \qquad i = 1, 2, ..., 10000 \tag{16}$$

där R^i_{tot} motsvarar totala faktiska framtida utbetalningar från simulering i. Figur 10 visar kerneldensity för (15) och (16). Den röda linjen är CRM (15) och den blåa linjen är Chain Ladder-modellen (16).



(a) $D1_{\min}$ (b) $D1_{\max}$

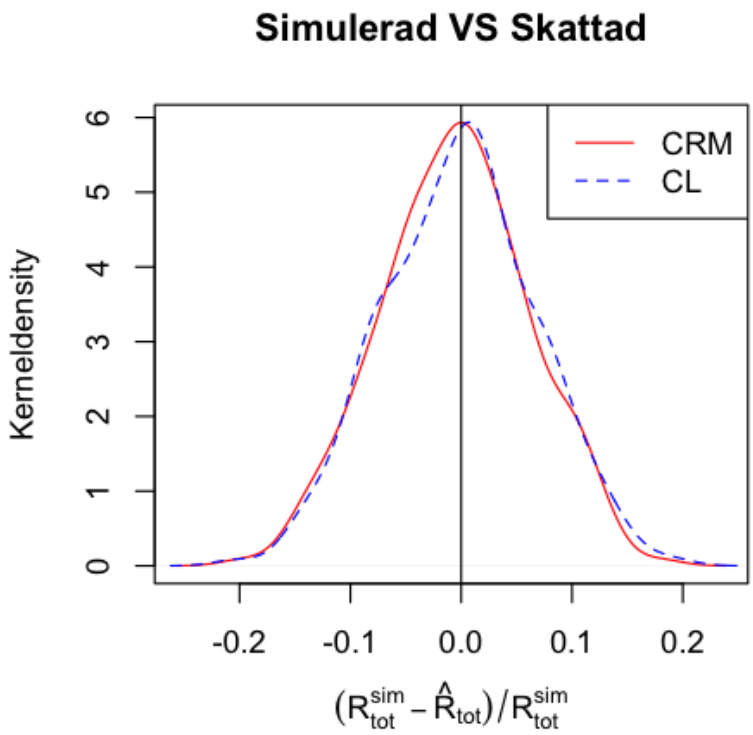
Figur 10: Reserven enligt (15) och (16)

När vi skattar parametrar i CRM nämner vi i Avsnitt 6.3 att vi medvetet anpassar färre parametrar för $D1_{\min}$ och $D1_{\max}$. Detta gör att reserven blir något underskattad. Från Figur 10 ser vi att den skattade reserven med Chain Ladder-modellen är centrerad kring 0 men med CRM lutar fördelningen åt höger något. Detta innebär att flera av reserverna är positiva, vilket leder till

$$\frac{R_{tot}^{sim} - \hat{R}_{tot}^{CRM}}{R_{tot}^{sim}} > 0 \Longrightarrow R_{tot}^{sim} > \hat{R}_{tot}^{CRM}.$$

Alltså reserven med CRM är något underskattad. Detta är orsakat av att vi medvetet anpassar färre parametrar i CRM. Vi noterar att variationen för CRM är något mindre än Chain Ladder-modellen. Vi ser också att variationen för $D1_{\max}$ är mycket mindre än $D1_{\min}$. Detta är rimligt, eftersom $D1_{\max}$ har mycket fler kontrakt än $D1_{\min}$. Ju fler observationer vi har desto närmare kommer den skattade $\hat{\psi}$ till den sanna ψ.

Som vi nämner i Avsnitt 6.3 att problemet med QML går att undvika genom att för varje dataset hanterar dessa nollor korrekt. Detta går att automatisera, men GLM-anpassningsmetoden i R kommer fortfarande att varna för att den numeriska anpassningsmetoden inte har konvergerat. Figur 11 visar ett exempel på reservjämförelse för CRM och CL för ett annat datamaterial som har kort svans och där vi har hanterat problemet med QML. Figur 11 kommer från (Lindholm, 2019).



Figur 11: Reservjämförelse där vi har automatiserat datamaterialet

Från Figur 11 ser vi att CRM-prediktionen är bättre än motsvarande CL-prediktion.

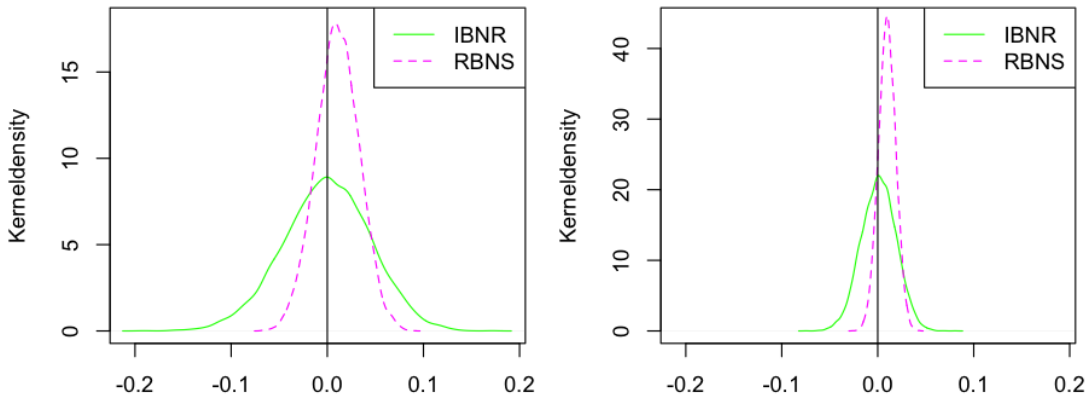
Vi fortsätter analysen med den metoden vi har valt i denna uppsatsen, det vill säga där vi medvetet skatta färre parametrar i CRM. Eftersom CRM skattar IBNR och RBNS reserven var för sig är det intressant att jämföra skattade IBNR och RBNS reserven med simulerade IBNR och RBNS reserven. För jämförelser använder vi följande:

$$\frac{R^{\mathcal{I},i} - R^{\mathcal{I},CRM,i}}{R^{i}_{tot}}, \qquad i = 1, 2, ..., 10000 \tag{17}$$

och

$$\frac{R^{\mathcal{R},i} - R^{\mathcal{R},CRM,i}}{R^{i}_{tot}}, \qquad i = 1, 2, ..., 10000 \tag{18}$$

Vi plottar kernelfördelningarna för (17) och (18) för $D1_{\min}$ och $D1_{\max}$. Figur 12 (a) är kernelfördelningar för $D1_{\min}$, Figur 12 (b) är kernelfördelningar för $D1_{\max}$.



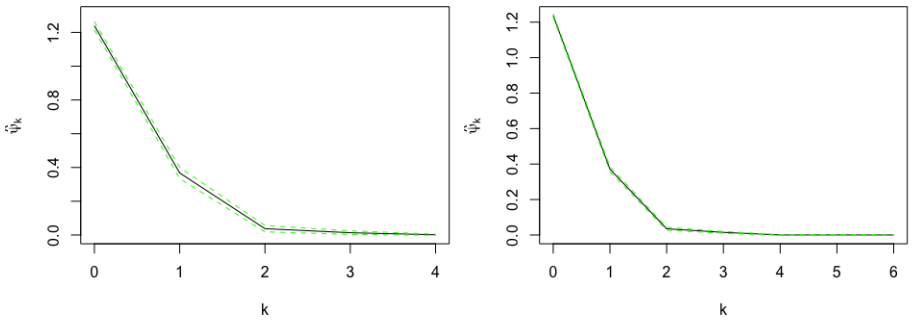
(a) $D1_{\min}$ (b) $D1_{\max}$

Figur 12: Grön linje (17). Lila linje (18).

Från Figur 12 ser vi att IBNR-reserven är centrerad runt 0, men variationen är större än för RBNS-reserven. Fördelningen för RBNS-reserven lutar sig åt höger, vilket innebär att RBNS-reserven är något underskattad. För $D1_{\max}$ är varitationen för både RBNS och IBNR reserven mycket mindre än $D1_{\min}$. Ovanstående analys antyder på att det är RBNS-reserven som leder till att den totala reserven blir något underskattad.

Detta resultat är rimligt, eftersom när vi skattar IBNR-reserven skattar vi först antalet rapporterade skador, sedan utbetalningar. RBNS-reserven däremot använder vi antal rapporterade skador som vi redan vet och skattar bara utbetalningar. Så RBNS-reserven har mer information och färre saker att skatta än IBNR-reserven. Att variationen är mindre för $D1_{\max}$ än $D1_{\min}$ är också förväntat. Samma resonemang som tidigare, ju fler observationer vi har desto närmare kommer den skattade $\hat{\psi}$ till den sanna ψ .

Vidare analysera vi $\hat{\psi}_k$ för CRM. Vi simulerar dataset typ 1, 10 000 gånger, beräknar medelvärdet, 2.75% och 97.5% percentilen av $\hat{\psi}_k$. Figur 13 visar medelvärden, 2.75% och 97.5% percentilen av $\hat{\psi}_k$ för $D1_{\min}$ och $D1_{\max}$.

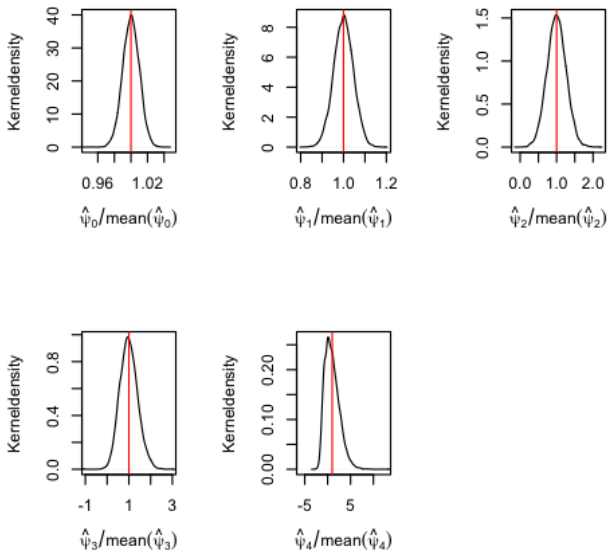


(a) $D1_{\min}$ (b) $D1_{\max}$

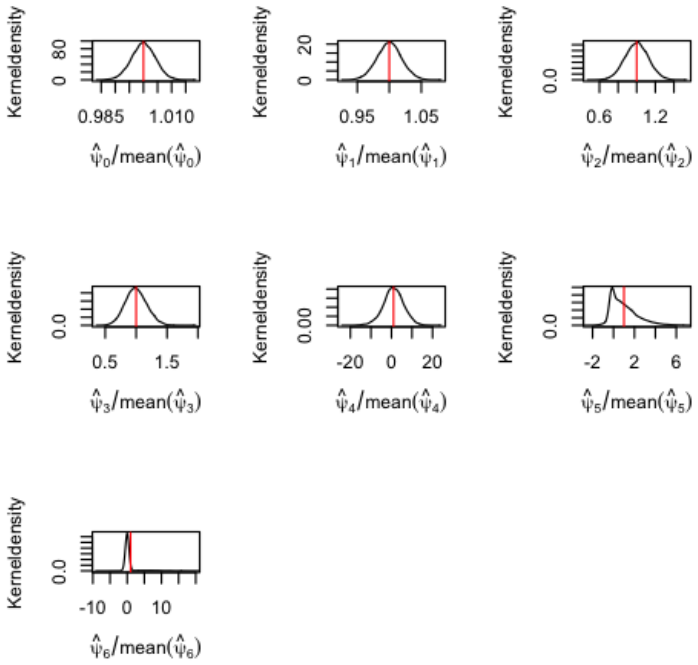
Figur 13: $\bar{\hat{\psi}}_k$ med 2.75% och 97.5% percentilen

Från Figur 13 ser vi att percentilerna ligger ganska nära medelvärdet. $\hat{\psi}_k$ avtar ju större k blir. Man kan ana, för större k blir $\hat{\psi}_k$ negativt för båda $D1_{\min}$ och $D1_{\max}$.

För att titta närmare på vår $\hat{\psi}_k$. plottar vi kernelfördelningen för $\frac{\hat{\psi}_k}{\bar{\hat{\psi}}_k}$, det vill säga $\hat{\psi}_k$ dividerar med medelvärdet av $\hat{\psi}_k$. Figur 14 och Figur 15 visar kernelfördelningarna.



Figur 14: $D1_{\min}; \hat{\psi}_k/\bar{\hat{\psi}}_k$

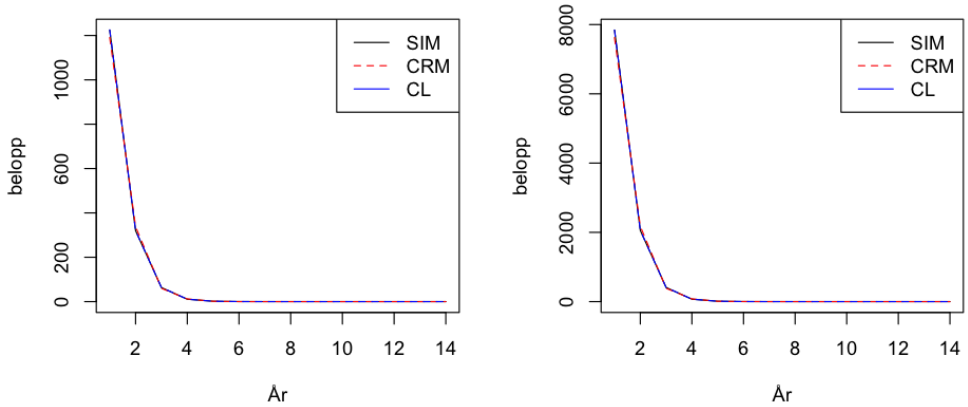


Figur 15: $D1_{\max}; \hat{\psi}_k/\bar{\hat{\psi}}_k$

Vi har flest information för att skatta $\hat{\psi}_0$, och mindre information för att skatta $\hat{\psi}_1$. Informationen avtar ju större k blir. Detta ser man i ekvationssytemet (6) i Avsnitt 4.3.1. Detta medför att skattningen för $\hat{\psi}_0$ kommer vara mer precis, och $\hat{\psi}_{m-1}$ kommer få sämst skattning. I Figur 14 ser vi fördelningen för $\frac{\hat{\psi}_k}{\bar{\hat{\psi}}_k}$, där $k = 0, 1, 2, 3$ är centrerad runt 1. Detta innebär att $\hat{\psi}_k$, där $k = 0, 1, 2, 3$ är nära sitt medelvärde. Fördelningen för $\frac{\hat{\psi}_4}{\bar{\hat{\psi}}_4}$ är inte helt centrerad. Den lutar sig något åt vänster. Vi kan se att variationen blir större ju större k blir, vilket är rimligt, därför ju större k blir desto mindre data vi har. Samtidigt vet vi att värdet för $\hat{\psi}_k$ blir mindre och mindre då k ökar. I vårt fall ser vi från Figur 13 (a) att $\hat{\psi}_3$ och $\hat{\psi}_4$ är väldigt nära 0, Vi noterar också att det förekommer även negativa värden för $\hat{\psi}_3$ och $\hat{\psi}_4$.

I Figur 15 ser vi liknade mönster för $D1_{\max}$, där skattning för $\hat{\psi}_0$ är säkrast och variationen för $\hat{\psi}_4$ är ganska stor. Skattningen för $\hat{\psi}_5$ och $\hat{\psi}_6$ verkar inte vara så bra. Vi ser från Figur 13 (b) att värdet för $\hat{\psi}_4$, $\hat{\psi}_5$ och $\hat{\psi}_6$ är väldigt nära 0. Så att skattningen för de parametrarna är dåliga kommer inte påverka reserven så mycket. Det förekommer negativa värde för $\hat{\psi}_4$, $\hat{\psi}_5$ och $\hat{\psi}_6$. Med negativa $\hat{\psi}_k$ kan framtida skattningar av skadeutbetalningar bli negativa men det är ett troligtvis skattningsfel och är helt acceptabelt.

Vidare vill vi analysera hur kvantilen ser ut för framtida kassaflöde. Vi beräknar medelvärd av kassaflöd, 2.75% och 97.5% percentilerna efter 10 000 simuleringar. Figur 35 och Figur 36 (Appendix) visar avveckling av framtida kassaflöde med 2.75% och 97.5% percentilerna för $D1_{\min}$ och $D1_{\max}$ med respektive metoder. Vi ser från Figurerna hur kassaflöde successivt avtar med tiden. Percentilerna ligger ganska tätt mot medelvärdet. För att lättare jämföra kassaflöde plottar vi Figur 16 som visar kassaflöde utan de 2.75% och 97.5% percentilerna.



(a) $D1_{\min}$ (b) $D1_{\max}$

Figur 16: Framtida kassaflöde (medelvärdet över 10 000 simuleringar); svart linje, simulerad kassaflöde. röd linje, kassaflöde med CRM. blå linje, kassaflöde med Chain Ladder-modellen

Från Figur 16 ser vi att kassaflöde för både CRM och CL ligger relativt nära det simulerade kassaflöde.

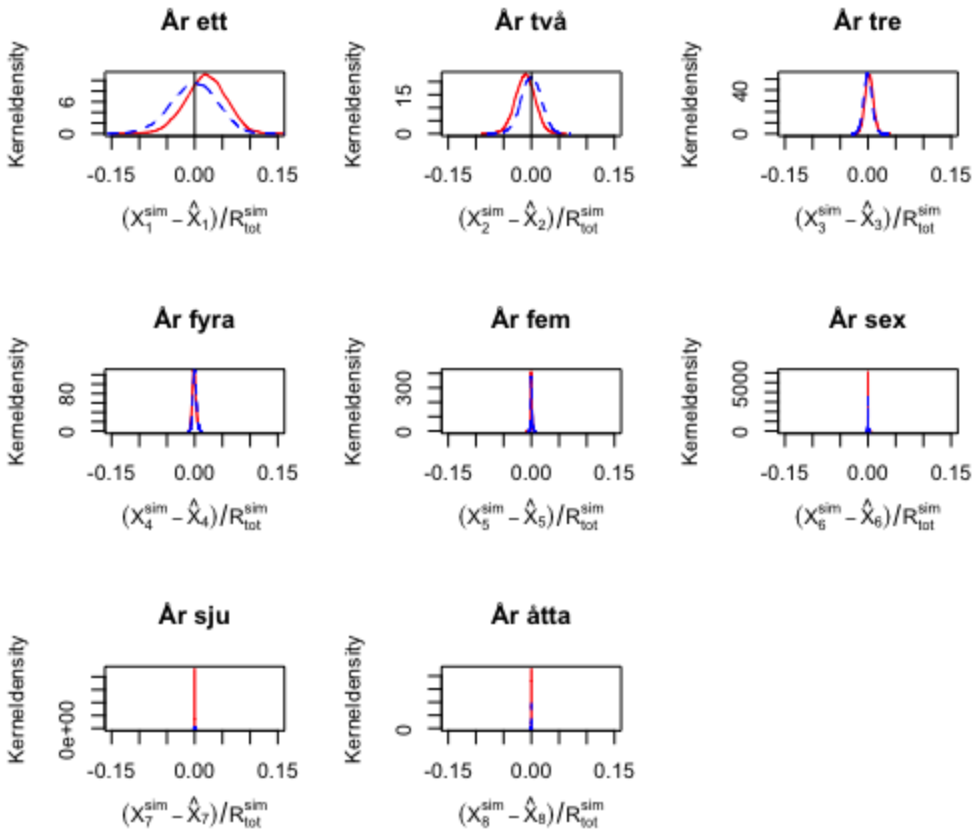
Det är svårt att se hur den skattade kassaflöde följer den simulerade kassaflöde i detalj från Figur 16, därför använder vi följande för jämförelse mellan modellerna

$$\frac{X_k^i - \hat{X}_k^{CRM,i}}{R_{tot}^i}, \qquad k = 1, 2, ..., \quad i = 1, 2, ..., 10000, \tag{19}$$

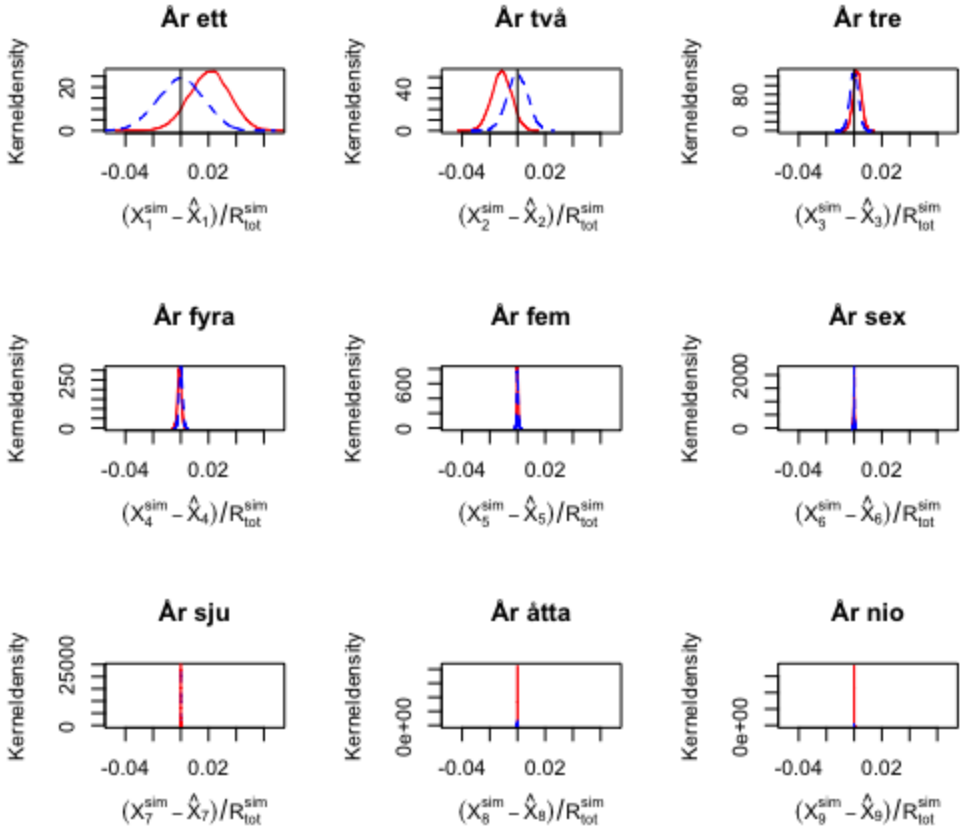
och

$$\frac{X_k^i - \hat{X}_k^{CL,i}}{R_{tot}^i}, \qquad k = 1, 2, ..., \quad i = 1, 2, ..., 10000 \tag{20}$$

där X_k^i motsvarar simlulerade utbetalningar år k i i:te simuleringen. Figur 17 och Figur 18 visar kerneldensity av (19) och (20) för $D1_{\min}$ och $D1_{\max}$. Den röda linjen är CRM, och den blåa linjen är Chain Ladder-modellen.



Figur 17: Kerneldensity av (19) och (20); Röd linje är CRM, Blå linje är CL, $D1_{\min}$



Figur 18: Kerneldensity av (19) och (20); Röd linje är CRM, Blå linje är CL, $D1_{\max}$;

Vi lägger först märke till att i Figur 17 har vi kassaflöde fram till åtta framtida år, men i Figur 18 har vi kassaflöde fram till nio framtida år. Anledningen till detta är att för $D1_{\min}$ finns inga flera utbetalningar efter åtta framtida år.

Vi ser från Figur 17 att skattningen är relativ centrerad med Chain Ladder-modellen för alla framtida år. Med CRM ser vi att kerneldensity för år ett, tre och fyra lutar sig något åt höger. Detta innebär att beloppen är något underskattad med CRM för dessa åren. Däremot för år två lutar fördelningen något år vänster. Detta betyder att med CRM är beloppen överskattade för det året. Kerneldensity är bredast för år ett. Det blir smalare för varje år det går.

Bredden på kerneldensity talar om för oss den relativa förändringsstorleken, det vill säga ju bredare kerneldensity är desto mer påverkar det årets skadeutbetalningar hela reserven. År fem, sex, sju och åtta är kerneldensity som en stolpe runt 0. Detta kan vi tolka som att skadeutbetalningarna är väldigt små för de

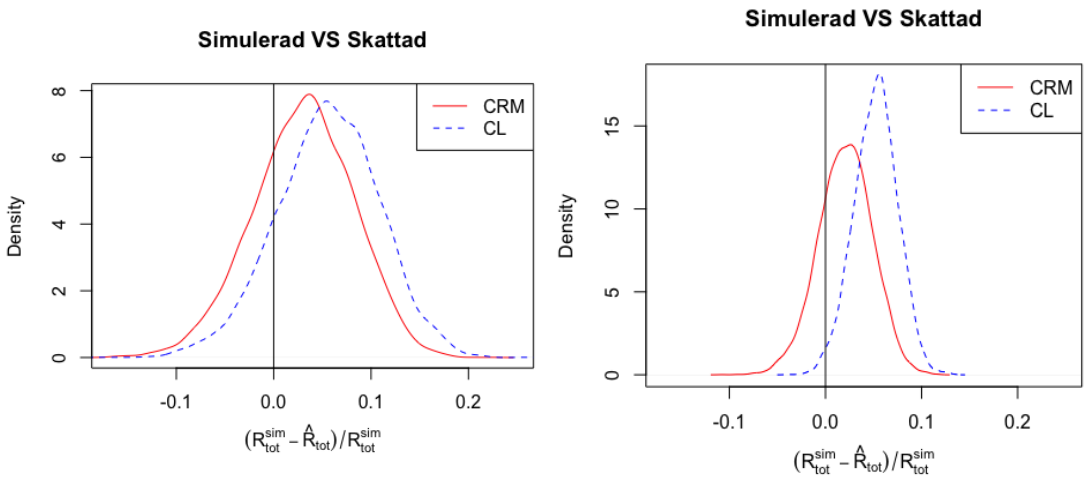
åren. Skattningen för de åren påverkar inte reserven speciellt mycket.

För $D1_{\max}$ ser vi från Figur 18 att med CRM är skadeutbetalningar underskattade för år ett, tre och fyra. År två är det överskattat. År fem, sex, sju, åtta och nio är beloppen väldigt små. så de åren bildrar inte så mycket till hela reserven. Chain Ladder fungerar ganska bra för denna typ av data också.

7.2 Dataset typ 2

Vi ser från Figur 4 i Avsnitt 6 att dataset typ 2 har längre svans än dataset typ 1.

Jämförelsen av reserven, se formler (15) och (16) för dataset typ 2 ser vi i Figur 19.



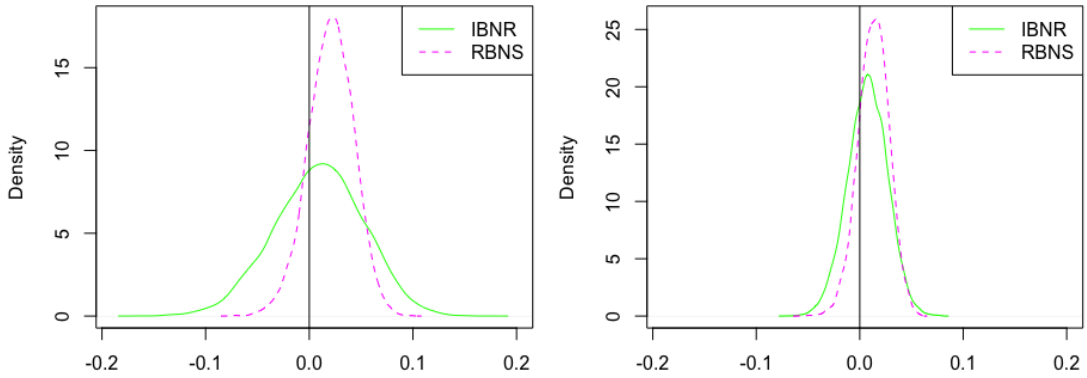
(a) $D2_{\min}$

(b) $D2_{\max}$

Figur 19: Reserven enligt (15) och (16)

Vi ser från Figur 19 (a) att reserven är underskattad för både CRM och med Chain Ladder-modellen men CRM lutar sig något mer åt vänster, vilket betyder att den är en aning mindre underskattad än Chain Ladder. För $D2_{\min}$ verkar CRM variera lite mindre. Om vi tittar på Figur 19 (b) ser vi att Chain Ladder-modellen fungerar ganska dåligt. Nästan hela fördelningen är på höger sida om 0. Reservskattning för CRM är underskattad men är mycket bättre än Chain Ladder-modellen.

Jämförelse för IBNR och RBNS reserven, se formler (17) och (18) för dataset typ 2 visas i Figur 20.

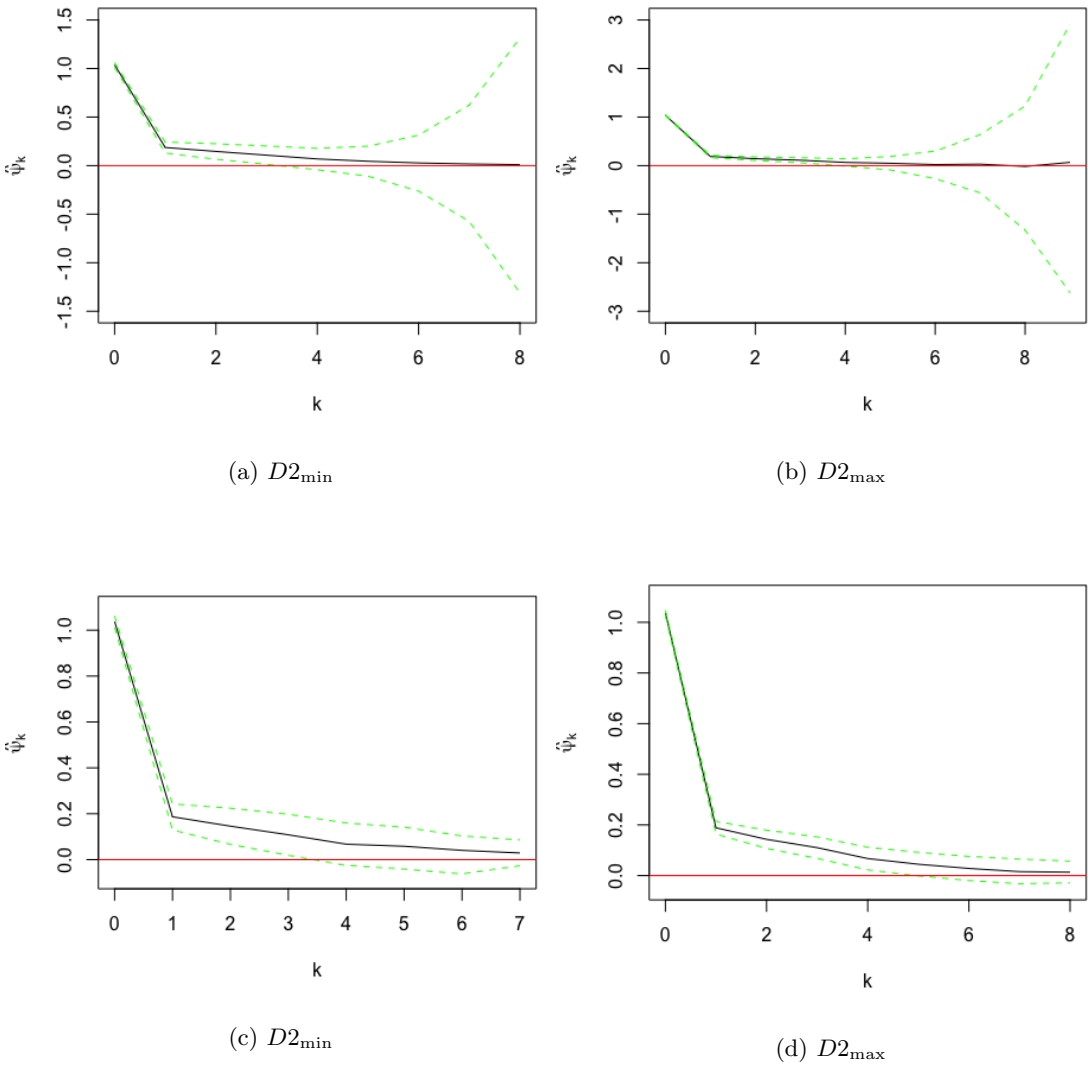


(a) $D2_{\min}$ (b) $D2_{\max}$

Figur 20: Grön linje (17). Lila linje (18).

Från Figur 20 ser vi att fördelningen för RBNS-reserven lutar sig något mer åt höger än IBNR-reserven, och variationen är mindre än IBNR-reserven. Så RBNS-reserven är något underskattad men andel skattningsfel är mindre än IBNS-reserven. Som för datatyp 1, det är RBNS-reserven som orsakar till att den totala reserven blir underskattad.

Vi vill analysera närmare på medelvärdet, 2.75% och 97.5% percentilerna av $\hat{\psi}_k$. Resultatet visas i Figur 21. Figur 21 (a) är $\hat{\psi}_k$ för $D2_{\min}$, där $k = 0, 1, ..., 8$. Figur 21 (c) är $\hat{\psi}_k$ för $D2_{\min}$, där $k = 0, 1, 2, ...7$. Figur 21 (b) är $\hat{\psi}_k$ för $D2_{\max}$, där $k = 0, 1, 2, ..., 9$. Figur 21 (d) är $\hat{\psi}_k$ för $D2_{\max}$, där $k = 0, 1, 2, ..., 8$.



(a) $D2_{\min}$

(b) $D2_{\max}$

(c) $D2_{\min}$

(d) $D2_{\max}$

Figur 21: $\hat{\psi}_k$ med 2.75% och 97.5% percentilerna

Vi ser från Figur 21 (a) och Figur 21 (b) att percentilerna blir väldigt breda när k blir stort. Detta medför att det inte passar att ha $k = 0, 1, 2, ..., 8$ till $D2_{\min}$ och $k = 0, 1, 2, ..., 9$ till $D2_{\max}$. När vi minskar k så ser vi från Figur 21 (c) och Figur 21 (d) att percentilerna ser mycket tätare ut, därför väljer vi $\hat{\psi}_k$, där $k = 0, 1, 2, ..., 7$ för $D2_{\min}$. För $D2_{\max}$ väljer vi $\hat{\psi}_k$, där $k = 0, 1, 2, ..., 8$. Vi ser tydligt här att det uppstår negativa $\hat{\psi}_k$ för båda $D2_{\min}$ och $D2_{\max}$

Nu när vi har bestämt antalet $\hat{\psi}$, kan det vara intressant att veta hur många procent av reserven vi inte fångar med respektive modellerna. Med Chain Ladder-

modellen beräknas reserven bara till och med utvecklingsår 9. Med CRM skattar vi reserven till utvecklingsår 16 för $D2_{\min}$ och till utvecklingsår 17 för $D2_{\max}$.

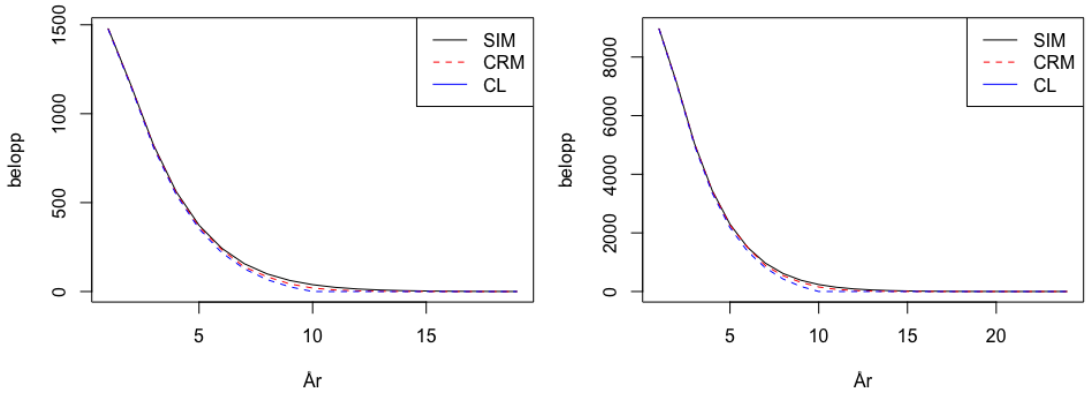
Från simulerade data beräknar vi andel betalningar av total framtida betalningar som båda metoderna missar, Resultaten baserat på 10 000 simuleringar ges i Tabell 10.

Dataset	CRM	CL
$D2_{\min}$	0.256%	5.688%
$D2_{\max}$	0.137%	5.249%

Tabell 10: Andel betalningar av totala framtida betalningar som metoderna inte fånga upp på grund av den långa svansen (Medelvärdet av 10 000 simuleringar)

Vi ser från Tabell 10 att Chain Ladder har betydligt större andel som den inte fångar upp än CRM, vilket inte är konstigt, eftersom CRM predikterar längre än Chain Ladder. Kom ihåg att $D2_{\max}$ har flera kontrakt än $D2_{\min}$. Vi noterar här att andel betalningar av totala framtida betalningar modellerna missar för $D2_{\min}$ är större än $D2_{\max}$. Även så är fallet är reserven modellerna ska prediktera är större för $D2_{\max}$.

Medelvärde av det framtida kassaflödet med 2.75% och 97.5% percentilerna visas i Figur 37 och Figur 38 (Appendix) för dataset typ 2. Man ser att kassaflödet avtar och den 2.75% och 97.5% percentilerna ligger ganska tätt runt medelvärdet. Figur 22 visar kassaflödet utan 2.75% och 97.5% percentilerna för att lättare jämföra det skattade kassaflödet och det simulerade kassaflödet.



(a) $D2_{\min}$ (b) $D2_{\max}$

Figur 22: Framtida kassaflöde (medelvärdet över 10 000 simuleringar); svart linje, simulerat kassaflöde, röd linje, kassaflöde med CRM, blå linje, kassaflöde med Chain Ladder-modellen

Vi tittar på kassaflöde från med år 5 till och med år 14 i Figur 22 (a). Vi ser att kassaflöde som är skattat med CRM (röd linje) ligger närmare det simulerade kassaflödet än Chain Ladder-modellen (blå linje), men reserverna för båda metoderna verkar vara underskattad. Nu tittar vi på kassflödet från och med år 5 till och med år 15 i Figur 22 (b). Kassaflödet för Chain Ladder-modellen är ännu längre ifrån det simulerade kassflödet. Som sagt Chain Ladder skattar inte mer efter år 10, därför är reserven 0 efter det året. För övriga år ser det ut som att det skattade kassaflödet är nära det simulerade kassaflödet.

8 Slutsatser och Diskussion

I detta arbete har vi tittat på två reservsättningsmodeller. Den ena är Macks Chain Ladder-modell, den andra är den kollektiva reservsättningsmodellen (CRM, eng. the collective reserving model). Vi har tillämpat båda modeller på simulerade data. Vi har simulerat data 10 000 gånger.

CRM kräver två tringlar: en är inkrementell triangel för antal rapporterade skador och en är inkrementell triangel för skadeutbetalningar. Parametrarna skattas genom att använda kvasimaximum likelihoodteori (QML, eng. quasi maximum likelihood). Om en försäkring där utbetalningstid är kort, så får vi problem att skatta så många parametrar vi vill med QML då vi simulerar data 10 000 gång-

er på grund av att somliga utbetalningsperioder innehåller nollor. Detta är ett numeriskt problem och går i praktiken att undvika genom att för varje sådant dataset hantera dessa nollor korrekt. Detta går att automatisera, se exempel Figur 11. På grund av tidsbrist väljer vi dock i denna uppsats att medvetet skatta färre parametrar, vilket leder till en viss underskattning av reserven för sådana typer av försäkring jämfört med de sanna simulerade utbetalningar och Chain Ladder-modellen. Chain Ladder-modellen är lätt att implementera. Vi behöver bara veta om den aggregerade triangeln för skadebelopp. Den skattar reserven väldigt bra när utbetalningen slutregleras snabbt.

Om en försäkring där utvecklingsår är längre än skadeår, så kan Chain Ladder modellen inte fånga upp informationen där utveklingsår är längre än skadeår. CRM fungera mycket bättre än Chain Ladder-modeller i detta fall. Eftersom den med hjälp av sina skattade parametrar kan fånga upp en del av informationen efter utvecklingsår som är längre än skadeår. CRM delar upp reserven i två poster, IBNR och RBNS reserven. När vi jämför den simulerad IBNR/RBNS reserven med den skattad IBNR/RBNS reserven, ser vi att RBNS-reserven är något underskattad men andel skattningsfel är mindre än IBNR-reserven. Det antyder på att det är RBNS-reserven som leder till att den totala reserven blir underskattad.

Vi har inte haft tillräckligt mycket tid att undersöka problemet med QML. I framtiden kan man undersöka vidare på detta. Då kommer reserven med CRM inte bli underskattad när utbetalningstiden är kort. I praktiken har vi endast en datamängd och det går att skatta MSEP med metoder som beskrivs i (Wahl, 2018) för att undersöka hur väl modellen passar data. Man kan även tänka sig att beskriva ψ som en funktion av utvecklingsperiod och på så sätt kunna prediktera reserven ännu längre.

9 Referenser

Wahl, F., Lindholm, M., and Verall, R. (2018). Stockholm University Research Report in Mathematical Statistics 2018:19.

Mack, T. (1993). Distribution free calculation of the standard error of chain-ladder reserve setimates. ASTIN Bulletin 23 (2), 213-225.

Lindholm, M. (2019) *Privat korrespondens.*

Miranda, M. D. M., Nielsen, B., Nielsen, J. P., and Verall, R. (2011). Cash flow simulation for a model of outstanding liabilities based on claim amounts and claim numbers. ASTIN Bulletin: The Journal of the IAA: 107-129.

Miranda, M. D. M., Nielsen, J. P., and Verrall, R. (2012). Double Chain Ladder. Astin Bulletin.

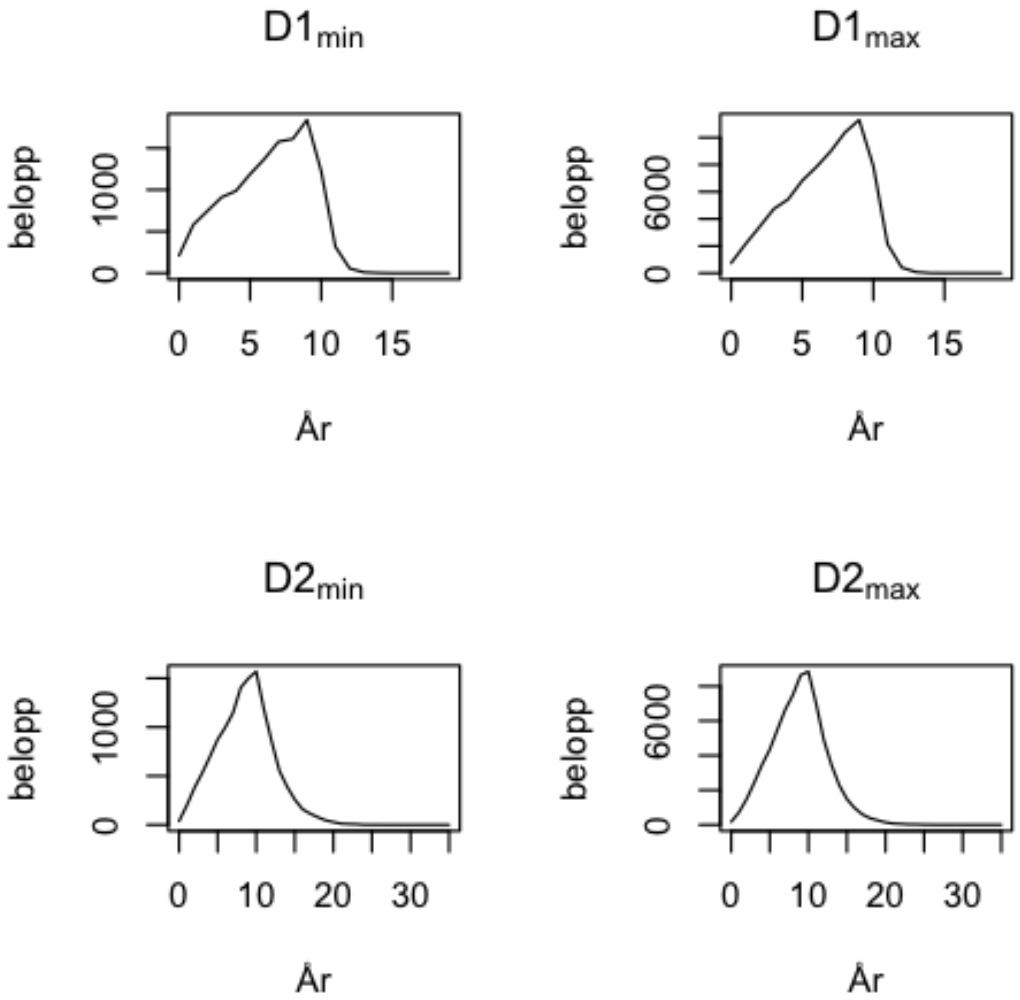
Skadeförsäkringsredovisning föreläsning 1, introduktion, premier.

Verrall, R., Nielsen, J. P., and Jessen, A. H. (2010). Prediktion of RBNS and IBNR slaims using claim amounts and claim counts. Astin Bulletin, 872-887.

Wütrich, M. (2015). *Non-life insurance Mathematics & Statistics.* Lecture Notes.

Wahl, F., Lindholm, M. and Lindskog, F. (2018). *Estimation of condition mean squared error of prediction for claims triangle reserving 2018:6.*

10 Appendix

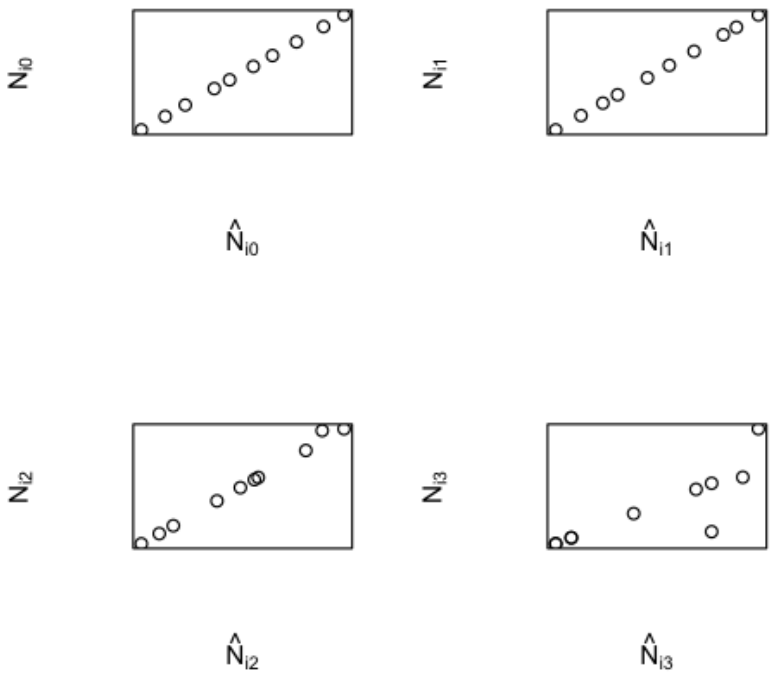


Figur 23: Kassaflöde baserat på första simulering

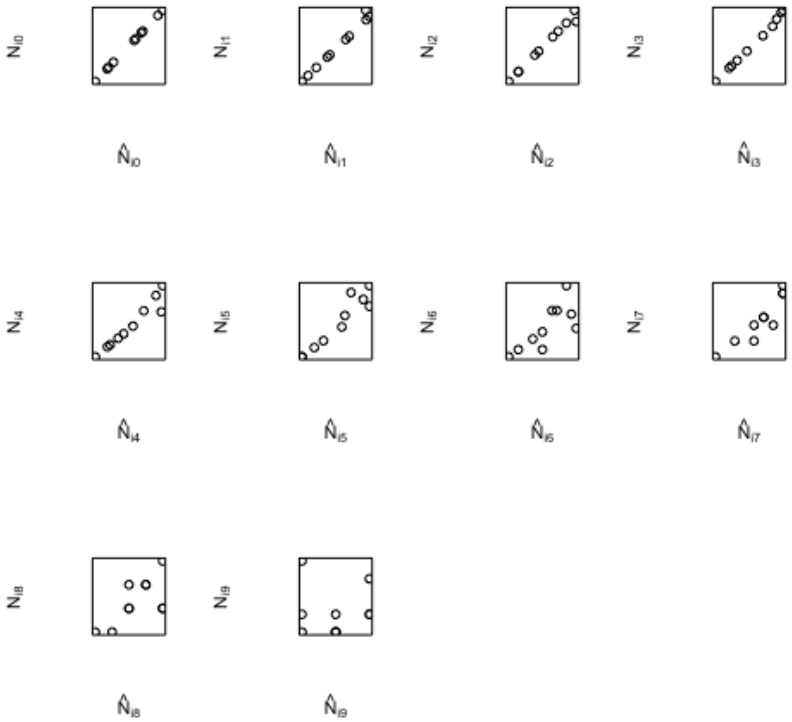
Figur 24: Akumulerade utbetalningsbelopp $\hat{C}_{i,j}$ mot $C_{i,j}$ baserad på den första simuleringen, $D1_{\max}$

Figur 25: Akumulerade utbetalningsbelopp $\hat{C}_{i,j}$ mot $C_{i,j}$ baserad på den första simuleringen, $D2_{\min}$

Figur 26: Akumulerade utbetalningsbelopp $\hat{C}_{i,j}$ mot $C_{i,j}$ baserad på den första simuleringen, $D2_{\max}$

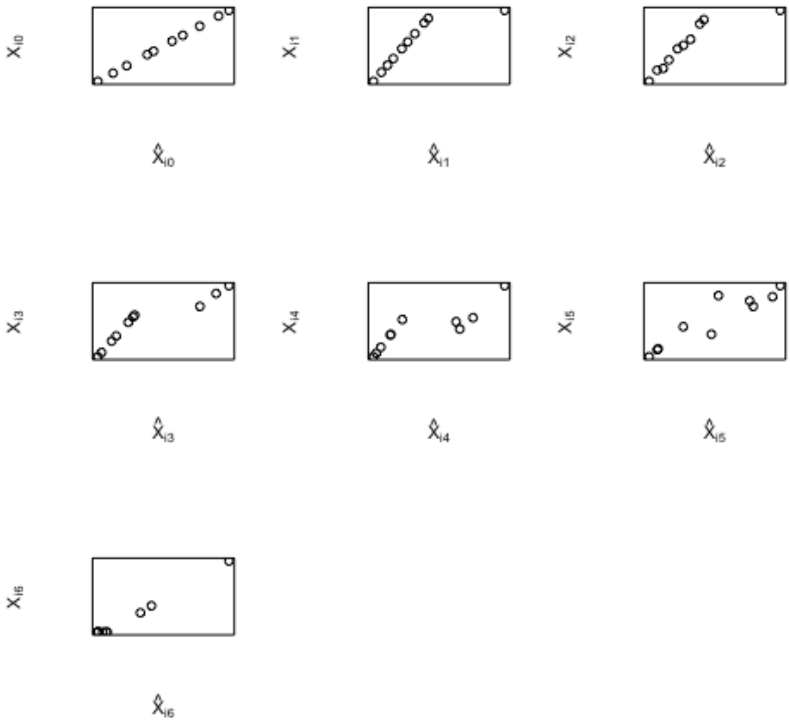


Figur 27: Antal rapporterade skodor $\hat{N}_{i,j}$ mot $N_{i,j}$ baserad på den första simuleringen, $D1_{\max}$

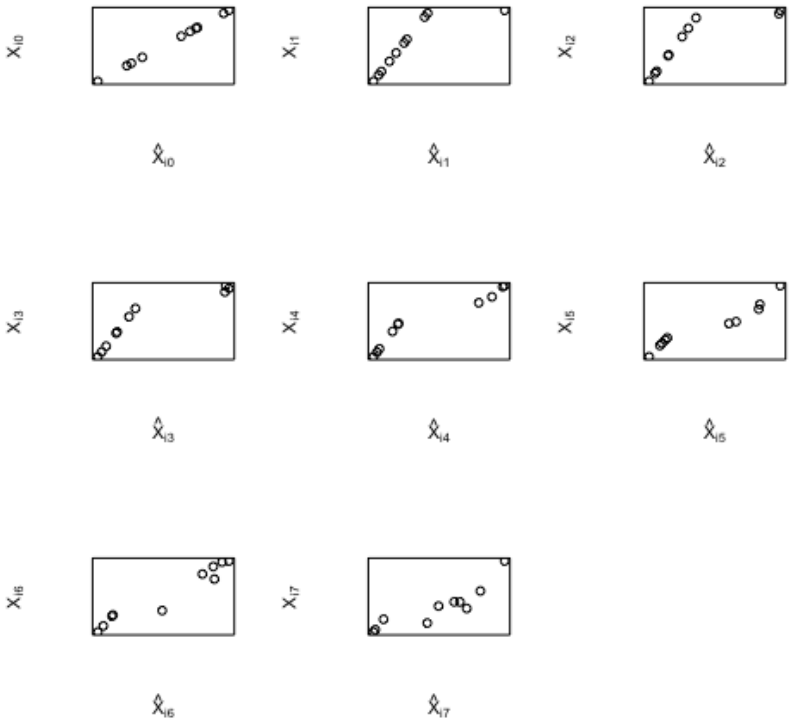


Figur 28: Antal rapporterade skodor $\hat{N}_{i,j}$ mot $N_{i,j}$ baserad på den första simuleringen, $D2_{\min}$

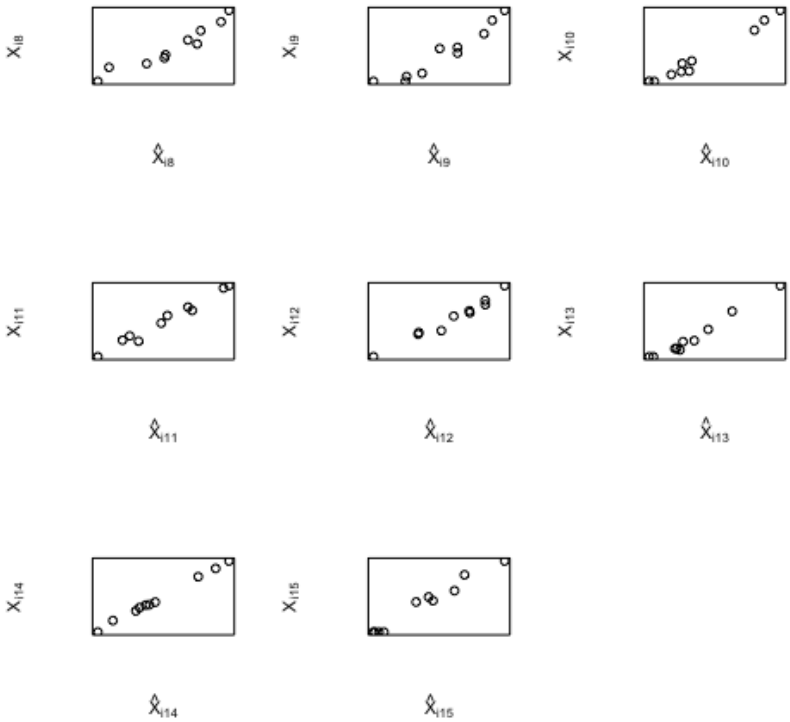
Figur 29: Antal rapporterade skodor $\hat{N}_{i,j}$ mot $N_{i,j}$ baserad på den första simuleringen, $D2_{\max}$



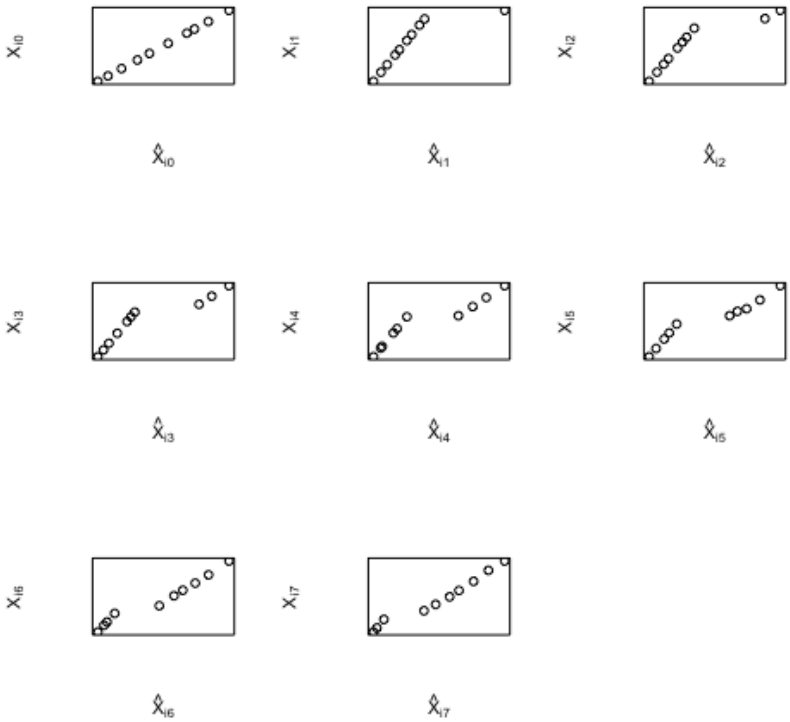
Figur 30: Skadeutbetalningar $\hat{X}_{i,j}$ mot $X_{i,j}$ baserad på den första simuleringen, $D1_{\max}$



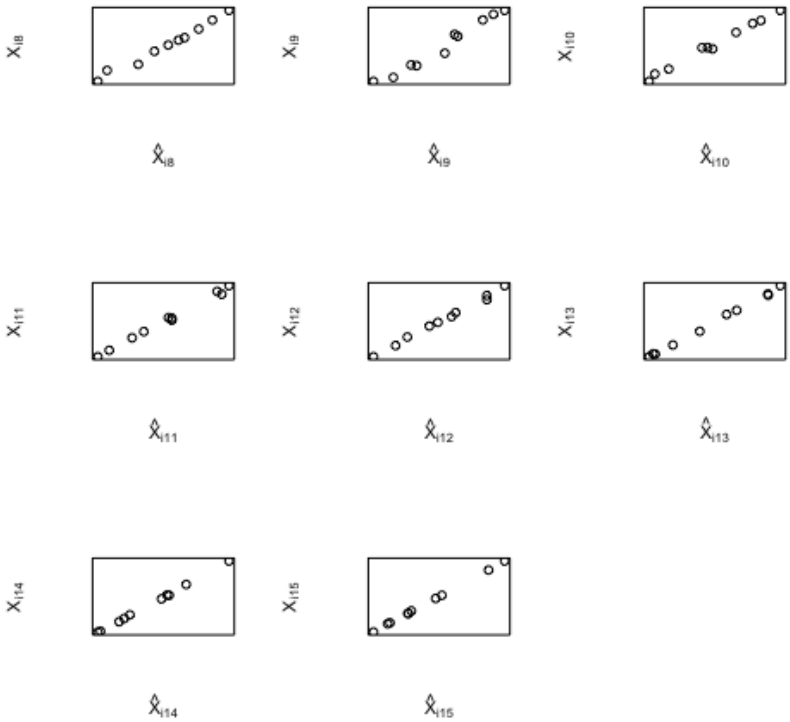
Figur 31: Skadeutbetalningar $\hat{X}_{i,j}$ mot $X_{i,j}, j = 0, 1, ..., 7$ baserad på den första simuleringen, $D2_{\min}$



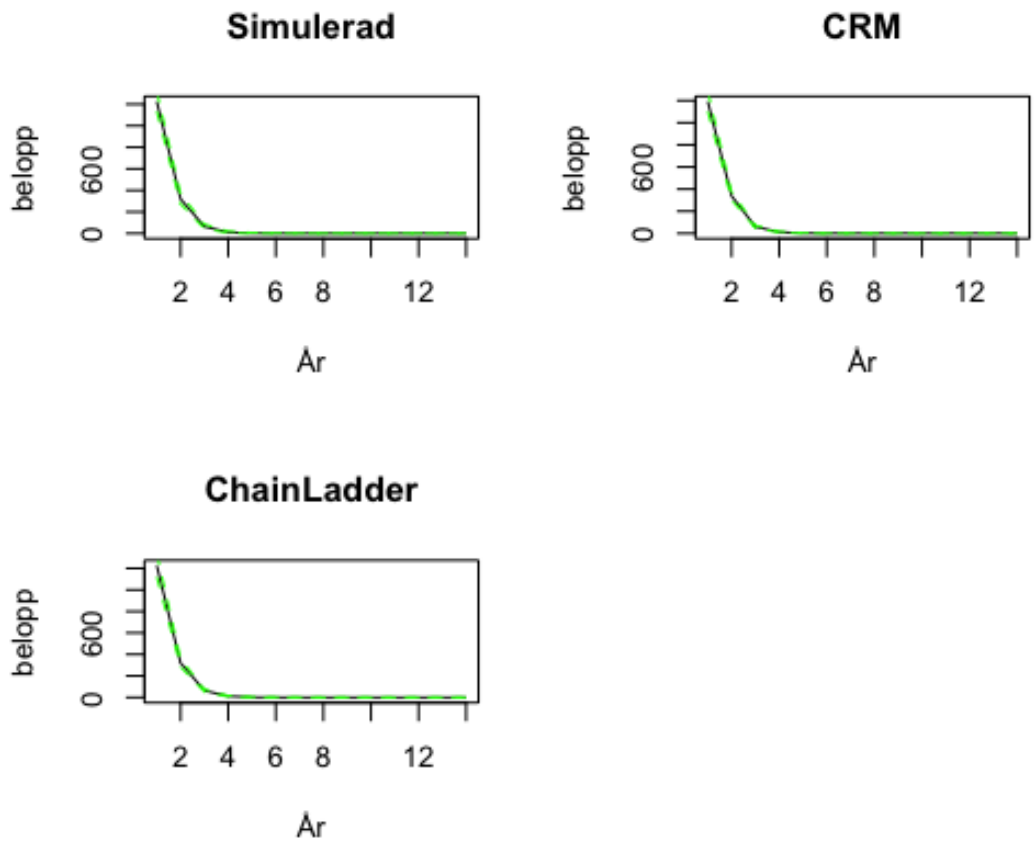
Figur 32: Skadeutbetalningar $\hat{X}_{i,j}$ mot $X_{i,j}, j = 8, 9, ..., 15$ baserad på den första simuleringen, $D2_{\min}$



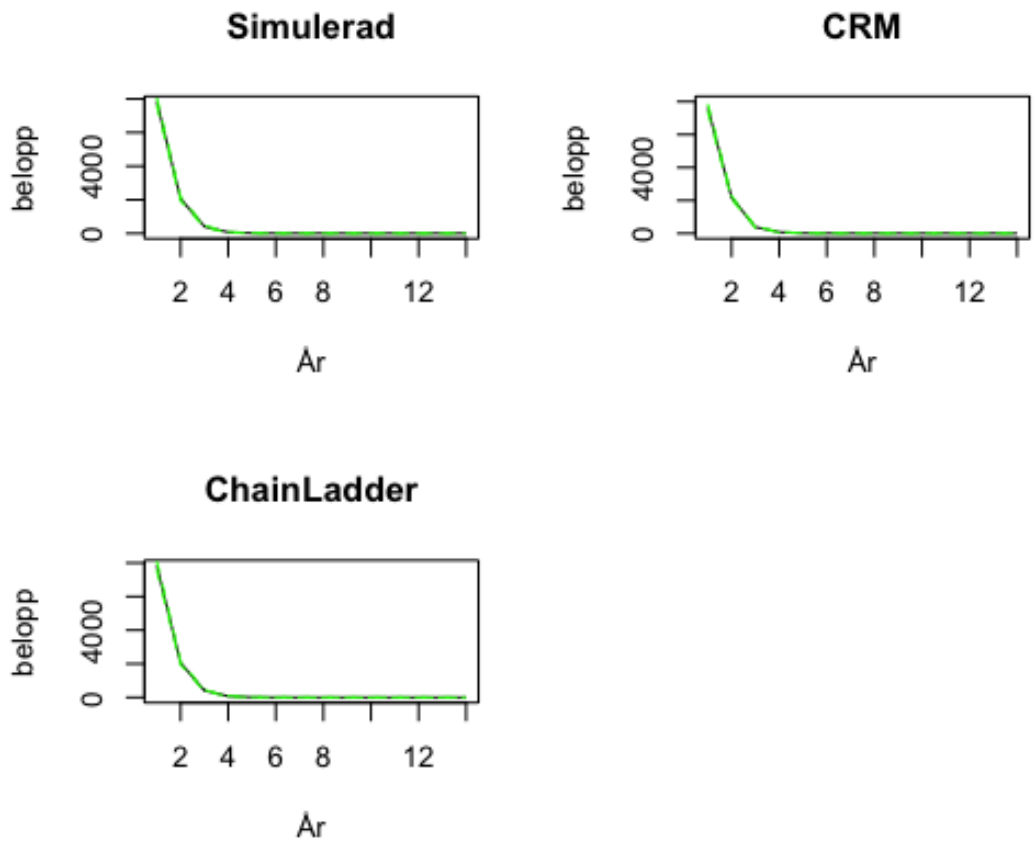
Figur 33: Skadeutbetalningar $\hat{X}_{i,j}$ mot $X_{i,j}, j = 0, 1, ..., 7$ baserad på den första simuleringen, $D2_{\max}$



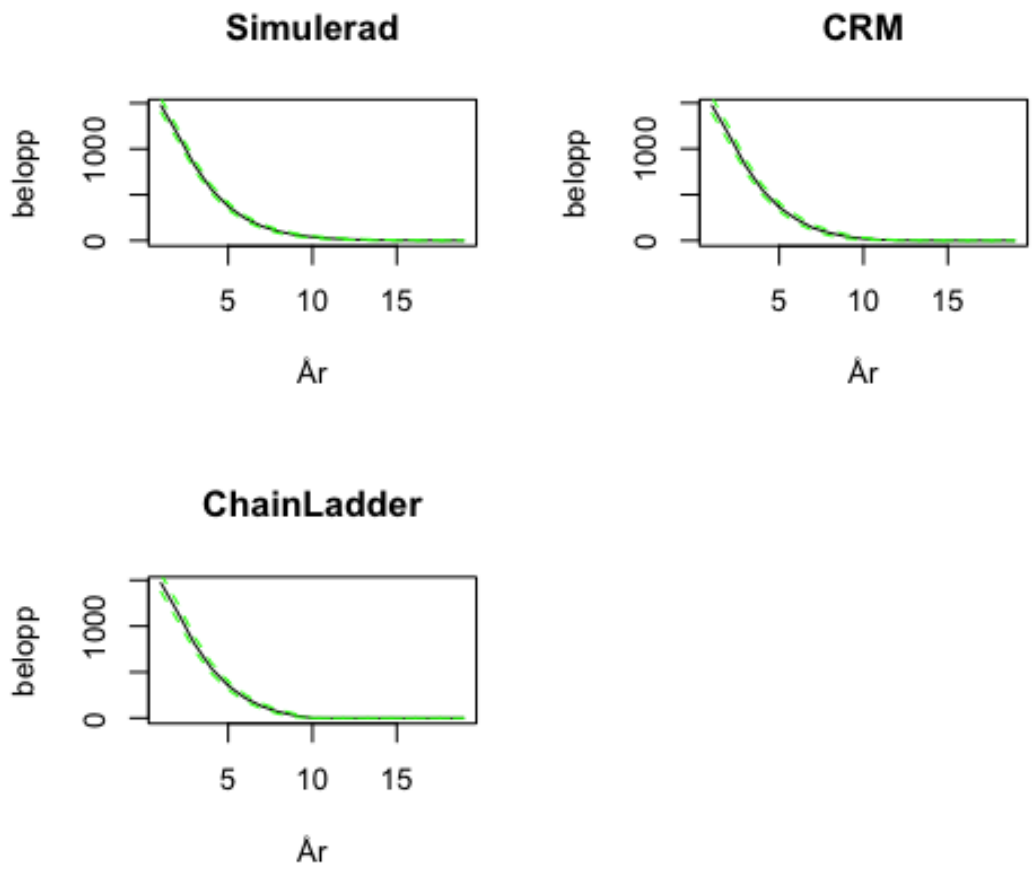
Figur 34: Skadeutbetalningar $\hat{X}_{i,j}$ mot $X_{i,j}, j = 8, 9, ..., 15$ baserad på den första simuleringen, $D2_{\max}$



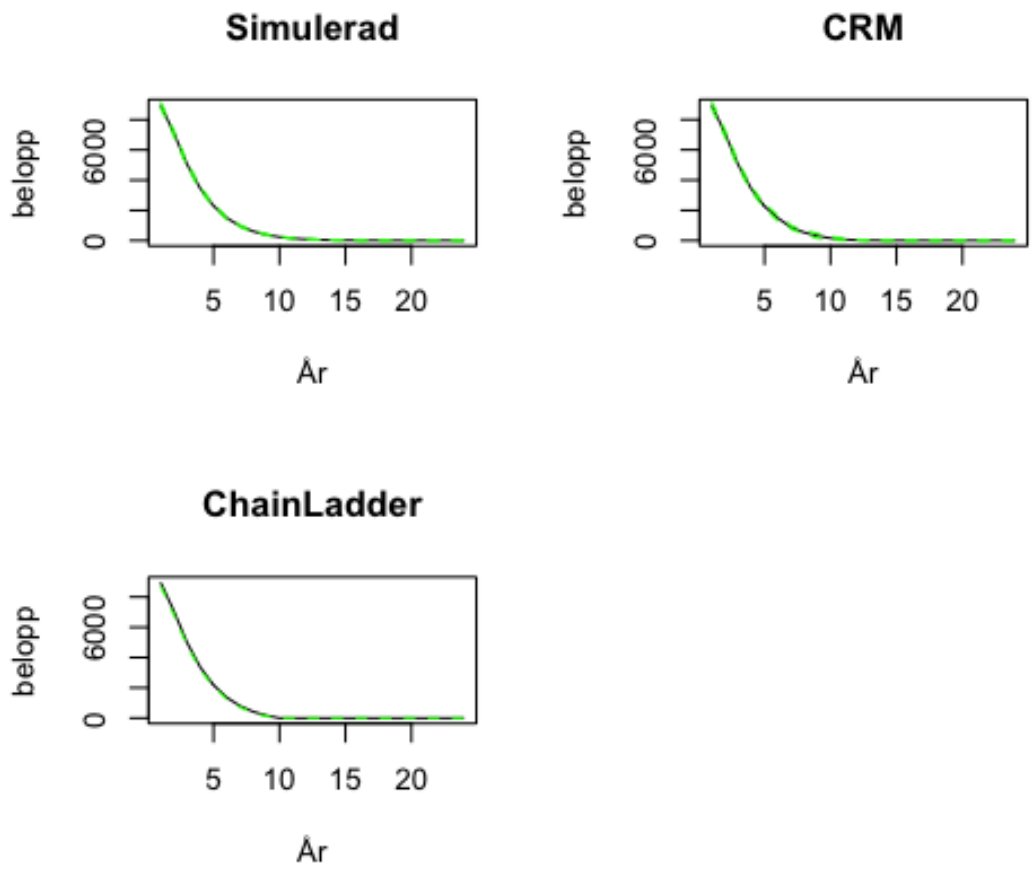
Figur 35: Framtida kassaflöde (medelvärde över 10 000 simuleringar) med 2.75% och 97.5% percentilerna, $D1_{\min}$



Figur 36: Framtida kassaflöde (medelvärde över 10 000 simuleringar) med 2.75% och 97.5% percentilerna, $D1_{\max}$



Figur 37: Framtida kassaflöde (medelvärde över 10 000 simuleringar) med 2.75% och 97.5% percentilerna, $D2_{\min}$



Figur 38: Framtida kassaflöde (medelvärde över 10 000 simuleringar) med 2.75% och 97.5% percentilerna, $D2_{\max}$