Stockholms universitet

En undersökning av hur bra olika reservsättningsmodeller passar simulerad data

Timo Ryhänen

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2015:29
Matematisk statistik
December 2015

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematisk statistik
Stockholms universitet
Kandidatuppsats **2015:29**
http://www.math.su.se/matstat

En undersökning av hur bra olika reservsättningsmodeller passar simulerad data

Timo Ryhänen*

December 2015

Sammanfattning

En av aktuariens huvuduppgifter på ett försäkringsbolag består av att uppskatta skadereserven. Skadereserv är en uppskattning av framtida utbetalningar till försäkringstagarna för redan inträffade skador. I denna uppsats kommer vi att simulera ett försäkringsbestånd med hög variation på utbetalningar mellan skadeår. På detta försäkringsbestånd kommer vi sedan att testa tre olika modeller för att undersöka vilken av dem som bäst predikterar skadekostnaden. De modeller som testas är Chain Ladder, Bornhuetter-Fergusson och Benktander-Hovinen. Slutsatsen av simuleringen blev att Bornhuetter-Fergusson och Benktander-Hovinen gav en bättre estimering av skadereserven och hade en mindre spridning än Chain Ladder och att alla metoder var väntevärdesriktiga.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: timo.ryhanen@live.se. Handledare: Mathias Lindholm.

Abstract

One of the main tasks of the actuary at an insurance company consists of estimating claims reserve. Claims reserve is an estimate of future payments to policyholders that have already occurred. In this paper, we will simulate an insurance portfolio with high variation in payments between the accident years. In this insurance portfolio we will then test three different models to examine which of them best predicts the claim costs. The models tested are Chain Ladder , Bornhuetter-Fergusson and Benktander-Hovinen. The conclusion of the simulation was that Bornhuetter-Fergusson and Benktader-Hovinen gave a better estimation of claims reserves and had a smaller variation than Chain Ladder and that all the methods were unbiased.

Innehåll

1 Förord och tack

Detta arbete utgör en kandidatuppsats i försäkringsmatematik om 15 högskolepoäng vid Stockholms Universitet. Jag vill tacka min handledare, Mathias Lindholm, vars goda ideer och vägledning har varit av stor betydelse. Jag vill även tacka familj och vänner som har uppmuntrat mig i mitt skrivande.

2 Inledning

En av aktuariens huvuduppgifter på ett försäkringsbolag består av att uppskatta skadereserven. Skadereserv är en uppskattning av framtida utbetalningar till försäkringstagarna för redan inträffade skador.

Det är av stor betydelse att göra en korrekt uppskattning av reserven av flera anledningar. Om reserven är för låg så kommer det inte finnas tillräckligt med kapital för att täcka framtida utbetalningar för redan inträffade skador. En annan risk är att eftersom premien inom försäkring beräknas utifrån den totala skadekostnaden så kan det leda till att även premieberäkningen blir felaktig, vilket leder till man underskattar kostnaden för att täcka framtida ej inträffade skador. Inom bolaget är det dessutom viktigt att kapitalförvaltningen får en korrekt uppskattning av skadekostnaden, eftersom de behöver informationen för att veta hur mycket de kan investera på tillgångssidan. Det är även viktigt att kunna ge en korrekt bild av bolagets tillstånd i balans- och resultaträkningen, både internt och externt. Det finns med andra ord en mängd anledningar till att försöka uppskatta skadekostnaden så korrekt som möjligt.

Lite grovt finns det två olika typer av skador inom sakförsäkring, långsvansade och kortsvansade. Det som kännetecknar en långsvansad produkt, till exempel en barnförsäkring, är att tiden mellan det att skadan rapporteras in tills det att den är slutreglerad är lång. En kortsvansad produkt, till exempel sjukvårdsförsäkring, kännetecknas i sin tur av att tiden mellan att skadan inträffas tills det att den är slutreglerad är kort. Det kan även finnas andra olikheter, exempelvis bestånd med hög säsongsvariation, som växer m.m.

För att kunna göra en korrekt uppskattning av den totala skadekostnaden och för att kunna uppskatta en korrekt reserv, så är val av reservmodell av yttersta vikt. Den vanligaste modellen är Chain Ladder, det är en deterministiskt modell, där framtida utbetalningar antas ha samma mönster som de historiska. En brist med Chain Ladder är att den fungerar sämre på bestånd som ändras. Dessa ändringar kan vara exempelvis att beståndet växer, eller att trenden på utbetalningarna ändras, vilket kan bero på villkorsändringar. Även säsongsvariation kan skapa problem, exempelvis beroende på hur högtiderna ligger under ett år. Modellen är dessutom osäker i början av en utvecklingsperiod, när datamängden är liten.

I denna uppsats kommer vi att simulera ett försäkringsbestånd med hög variation på utbetalningar mellan skadeår. På detta försäkringsbestånd så kommer vi sedan att testa tre olika modeller för att undersöka vilket av dem som bäst predikterar skadekostnaden. De modeller som testas är Chain Ladder, Bornhuetter-Fergusson och Benktander-Hovinen. Benktander-Hovinen är en kombination av två modeller, Chain Ladder och Bornhuetter-Fergusson. Både Bornhuetter-Fergusson och Benktander-Hovinen använder sig av en ett antagande om att skadekostnaden vid ultimo är känd, i stället för att använda faktiskt data för att estimera den totala skadekostnaden, och borde ge en jämnare reserv för ett försäkringsbestånd framför allt tidigt i utvecklingen av skadekostnaden när osäkerheten är stor. Däremot är dessa metoder till skillnad från Chain Ladder inte väntevärdesriktiga. Vi ska i denna uppsats undersöka om vi hittar stöd för detta genom simuleringen ovan.

Upplägget för uppsatsen kommer att vara följande. I kapitel 3 presenteras syfte och metod. I kapitel 4 beskrivs triangelnotation, som är det vanligaste sättet att presentera skador i ett sakförsäkringsbolag och de reservsättningsmodeller som används i simuleringen. I kapitel 5 görs antaganden för antal skador och skadeutbetalningar och i kapitel 6 beskrivs simuleringen. Här beskrivs vilka antaganden som ligger till grund för att göra en simulering av skadetrianglarna, samt hur den faktiska skadekostnaden tas fram. I kapitel 7 drar vi slutsatser av simuleringen.

3 Syfte och metod

Syftet med denna uppsats är att undersöka om den finns fördelar med att uppskatta den totala skadekostnaden med Benktander-Hovinen och Bornhuetter-Fergusson jämfört med Chain Ladder för ett bestånd med hög säsongsvariation. Detta gör vi genom att simulera ett bestånd där skadorna antas vara poissonfördelade och storleken på skadeutbetalningen antas var lognormalfördelad. Jämförelsen görs genom att jämföra utfallet av skadekostnaden estimerat med Benktander-Hovinen och Chain Ladder samt Benktander-Hovinen med den faktiska skadekostnaden, som är känd i simuleringen.

4 Reservsättningsmetoder

4.1 Begrepp

I detta kapitel kommer vi att gå igenom metodik för att presentera skador inom sakföräkring samt de reservsättningsmetoder som används i simuleringen, Chain Ladder och Benktander-Hovinen. Vi kommer även att gå igenom Bornhuetter-Fergusson, som används som en del i Benktander-Hovinen.
I uppsatsen används en del försäkringsbegrepp som kan vara bra att känna till:

- *Betalt*: Är det skadebelopp som har betalats hittills av försäkringsbolaget för en inträffad skada. I uppsatsen kommer d att användas för inkrementella utbetalningar och c att användas för kumulativa utbetalningar.

- *Skadekostnad*: Är den totala kostnaden av en skada från det att skadan inträffar tills att den är slutreglerad. Skadekostnaden kommer att uttryckas som U (från Ultimo, den totala skadekostnaden då skadan är slutreglerad).

- *Reserv*: Är det belopp som ett försäkringsbolag sätter av för en inträffad skada, för att kunna täcka sina åtaganden. I ett sakförsäkringsbolag tas reserven fram som differensen mellan den estimerade skadekostnaden och utbetalt belopp. Reserven minskar med utbetalningar tills det att skadan är slutreglerad. I uppsatsen kommer reserven att uttryckas som R. Vi undersöker bara IBNR-reserven (*Incurred but not reported*), skadan har inträffat men har inte börjat utbetalas.

- *Skadeår*: Visar alla utbetalningar för ett visst år då en skada har inträffat. Den uttrycks radvis i en skadetriangel och kommer i denna uppsats att uttryckas som i.

- *Utvecklingsår*: Visar hur utbetalningarna utvecklas från det att skadan har inträffat. Den ligger kolumnvis i en skadetriangel och kommer i denna uppsats att uttryckas som k.

- *Kalenderår*:Visar alla utbetalningar för skador inträffade under aktuellt skadeår och tidigare skadeår. Dessa värden ligger diagonalt i en skadetriangel.

4.2 Triangelnotation

Det vanligaste sättet att redovisa skador inom sakförsäkring är genom så kallade trianglar, där raderna i representerar skadeår och kolumnerna k representerar utvecklingsår.

För att kunna modellera en portfölj med inkrementella skadeutbetalningar, så antar vi att vi har en grupp utbetalningar $d_{i,k}$ tillhör mängden (0,1,2 $\ldots n$) och vi tolkar $d_{i,k}$ som skadeutbetalningen för skadeår i som betalas ut med en fördröjning av k år och alltså på utvecklingsår k och kalenderår i, k.

Vi uttrycker alltså $d_{i,k}$ som den inkrementella utbetalningen för skadeår i och utvecklingsår k. Vi antar att de inkrementella skadeutbetalningarna $d_{i,k}$ är kända för kalenderåret i+$k \leq n$ och att de inte är kända för i+$k \geq n$+1. De kända skadorna får då formen av en inkrimentell triangel. Detta visas i tabell 1 nedan.

Skadeår i/ **Utvecklingsår** k	1	2	3	**...**	$m-1$	m
1	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	**...**	$d_{1,m-1}$	$d_{1,m}$
2	$d_{2,1}$	$d_{2,2}$	$d_{2,3}$	**...**	$d_{2,m-1}$	
3	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$			
.	.	.				
.	.	.				
.	.	.				
$m-1$	$d_{m-1,1}$	$d_{m-1,2}$				
m	$d_{m,1}$					

Tabell 1: Inkrementell triangel

För att modellera en portfölj av kumulativa skadeutbetalningar så antar vi att vi har en grupp kumulativa utbetalningar $C_{i,k}$ som tillhör mängd (0,1,2 $\ldots n$) och vi tolkar $C_{i,k}$ som förlusten under skadeår i som är slutreglerad senast vid tidpunkten k och alltså inte senare en utvecklingsår k.

$C_{i,k}$ är är de kumulativa skadeutbetalningarna för skadeår i och utvecklingsår k, till och med till $C_{i,n-i}$ som är den kumulativa förlusten för aktuell kalenderår n som en aktuell kumulativ förlust och $C_{i,n}$ som den kumulativa ultimoförlusten.

Vi antar att den kumulativa förlusten $C_{i,k}$ är känd för kalenderåret i+$k \leq n$ och att den inte är känd för kalenderåret i+$k \geq n$+1. Som det märks så bygger modelleringen av de inkrementella och kumulativa skadeutbetalningarna på samma princip och det finns ett samband mellan dessa genom att de kumulativa utbetalningarna är den kumulativa summan av de inkrementella skadeutbetalningarna, dvs

$$C_{i,k} = \sum_{l=0}^{j} d_{i,l}$$

De kända kumulativa utbetalningarna för då utseendet av en kumulativ triangel enligt nedan.

Skadeår i/ **Utvecklingsår** k	1	2	3	...	$m-1$	m
1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	...	C_1, m_{-1}	$C_{1,m}$
2	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$	...	C_2, m_{-1}	
3	$C_{3,1}$	$C_{3,2}$	$C_{3,3}$			
.	.	.				
.	.	.				
.	.	.				
$m-1$	$C_{m-1,1}$	$C_{m-1,2}$				
m	$C_{m,1}$					

Tabell 2: Kumulativ triangel

4.3 Chain Ladder

Chain ladder är den vanligaste reservsättningsmetoden. Den är deterministiskt, dvs. antar att mönstret på framtida utbetalningar kommer vara likadan som det historiska. Modellen bygger helt och hållet på erfarenhet av tidigare utbetalningar utifrån den kumulativa triangeln och gör inga antaganden om att det kommer att ske några ändringar i framtiden. Fördelen med modellen är att den är enkel att förstå och bygger bara på ett fåtal antaganden som bör vara uppfyllda:

- Utbetalningarna mellan skadeår antas vara oberoende.

- Mönstret på utbetalningar mellan skadeår ändras inte.

Chain Ladder brukar ge bra skattning med tiden när dataunderlaget är stort efter några års utveckling men i början däremot är dataunderlaget litet, vilket ger stor osäkerhet i skattningen av den totala skadekostnaden. Nedan görs den matematiska härledningen av Chain-Ladder modellen, beskrivningen följder den som ges i [2].

I kapitlet om triangelnotationen användes $C_{i,k}$ för att uttrycka den totala skadekostnaden för skadeår i, där *1*$\leq$ $i \leq I$ och utvecklingsår k, där *1* $\leq j$

$\leq K$. Vi visade att den totala skadekostnaden $C_{i,k}$ till $i+k \leq I+1$ är känd i den kumulativa triangeln.

De ackumulerade utbetalningerna $C_{i,k}$ är oberoende för alla skadeår i. Det som vi vill göra nu är att estimera skadekostnaden för $i+k > I+1$. Vi är intresserade framför allt den totala skadekostnaden vid Ultimo, C_{iI}.

$C_{i,I+1-i}$ är uttrycket för den kumulativa utbetalningarna fram tills i dag. Chain Ladder-metoden bygger på att estimera den totala skadekostnaden till ultimo genom utvecklingsfaktorer, f_k, som skattas enligt det betingade väntavärdet:

$$E(C_{i,k+1} \mid \ C_{iI}, \ldots, C_{i,k}) = C_{i,k} \cdot \hat{f}_k, \ 1 \leq i \leq I, I \leq j \leq I-1$$

I Chain Ladder estimeras $\hat{f}_k$ med

$$\hat{f}_k = \frac{\sum_{j=1}^{I-k} C_{j,k+1}}{\sum_{j=1}^{I-k} C_{j,k}}$$

där $1 \leq$ k $\leq I-1$

Metoden använder samma skattade $\hat{f}_k$ för alla skadeår för att skatta ökningen av skadekostnaden från utvecklingsår j till i$\geq I+1-k$, även om de inviduella faktorerna kan skilja sig mellan olika skadeår. Varje ökning av $C_{i,k}$, $C_{i,k+1}$, antas ha en slumpmässig fördelning av en förväntad ökning av $C_{i,k}$ till $C_{i,k} \cdot f_k$, där C_k, är en okänd sann ökning, som är densamma för alla skadeår och som skattas av tillgänglig data för $\hat{f}_k$.

Eftersom modellen inte tar höjd för beroende mellan skadeår, så kan vi anta att de kumulativa utbetalningarna C_k för de olika skadeåren är oberoende:

$$C_{i1}, ..., C_{iI}, C_{j1}, ..., C_{jI},$$

där $i \neq j$, är oberoende

Vi använder alltså samma utvecklingsfaktorer för alla skadeår och förändringen av skadekostnaden mellan varje utvecklingsår antas vara det betingade väntevärdet av den kumulativa utbetalningen $C_{i,k}$. Om vi använder utvecklingsfaktorer enligt ovan så får vi ut den totala skadekostnaden vid ultimo enligt:

$$U_{i,I} = C_{i,I+1-i} \cdot \hat{f}_{i+1-i} \cdot \ldots \cdot \hat{f}_{i-1}$$

som är ekvivalent med reserven enligt:

$$R_i = C_{i,I+1-i} \cdot (\hat{f}_{I+1-i} \cdot \ldots \cdot \hat{f}_{I-1} - 1)$$

4.4 Bornhuetter-Fergusson

Till skillnad från Chain Ladder så är Bornhuetter-Fergusson en framåtblickande modell. Det är en baysiansk modell, vilket innebär att den gör ett antagande om Ultimo innan vi har några faktiska skadekostnader och nyttjar den information som man får om skadekostnaden när tiden går. Detta gör den genom att använda utvecklingsfaktorerna från Chain Ladder för att kunna avgöra tillväxttakten, som visar hur stor del av skadan som har blivit utbetalt. Modellen gör följande grundantaganden:

- Borhuetter-Fergusson antar att den totala skadekostnaden Ultimo är känd vid beräkningstidpunkten.

- Modellen utnyttjar informationen från utvecklingsfaktorerna f_k från Chain Ladder för information om andel som har blivit utbetald av Ultimo. För att utvecklingsfaktorerna ska vara giltiga måste antaganden om Chain Ladder vara uppfyllda.

Fördelen med modellen är att utvecklingsfaktorerna i Chain Ladder är väldigt osäkra initialt, därför kan det vara bättre att anta en skadekostnad. Bristen med modellen är att eftersom den är framåtblickande och inte utnyttjar informationen om hur utbetalningarna har sett ut så tar din inte hänsyn till ändrade förutsättningar, vilket gör att den inte är väntevärdesriktig och kan bli väldigt missvisande på längre sikt. Modellen bygger även på information om utvecklingsmönstret av utbetalningar, vilket gör att vi är beroende av utvecklingsmönstret i Chain Ladder eller någon liknande utvecklingsmönster.

Nedan görs den matematiska härledningen av Bornhuetter-Fergusson, beskrivningen följer den som ges i [3].

Vi låter U_0 vara vad vi tror om den totala skadekostnaden a priori innan vi har någon erfarenhet av skadekostnaden. Information om U_0 kan hämtas exempelvis från premien, genom att skatta en riskpremie genom att multiplicera premien med en antaganden skadekostnadsprocent.

Vi väljer att titta bara på ett skadeår och tittar hur skadorna utvecklas. Vi inför två nya parametrar $p_k = \frac{1}{f_k}$ och $q_k = 1 - \frac{1}{f_k}$, där f_k är densamma som den definieras i Chain Ladder-modellen. p_k, $0 < p_1 < p_2 < p_n = 1$ är andelen av det totala skadebeloppet som är utbetalt efter k års utveckling. Efter m års utveckling antas skadan vara slutreglerad. Då kan reserven för Bornhuetter-Fergusson, R_{BF}, vid slutet av utvecklingsår k uttryckas som:

$$R_{BF,k} = U_{0,k} \cdot q_k$$

I formeln ovan ser vi att modellen använder samma estimering av den totala skadekostnaden för hela utvecklingsperioden och att den ersätts med faktiska utbetalningar med den andel som har blivit utbetald av skadan. Den totala skadekostnaden till Ultimo, U_{BF},kan uttryckas som:

$$U_{BF,k} = C_k + R_{BF,k}$$

och reserven kan i sin tur uttryckas som:

$$R_{BF,k} = U_0 - C_k$$

Chain Ladder uttrycks i stället som att skadekostnaden till ultimate kan uttryckas som:

$$U_{CL} = \frac{C_j}{f_k}$$

Och reserven kan uttryckas som:

$$R_{CL} = U_{CL} - C_k$$

4.5 Benktander-Hovinen

Vi har nu tittat på två av de vanligaste reservmodellerna inom sakförsäkring, som på sätt och vis är varandras motsatser. Chain Ladder är en bra modell när dataunderlaget börjar bli tillräckligt stort, utbetalningsmönstren mellan olika skadeår tenderar att vara ganska lika. Däremot är modellen osäker i början.

Bornhuetter-Fergusson är i sin tur en bra modell då den inte låter utbetalningarna i början av skadeperioden påverka reserverna. Däremot blir den sämre med tiden eftersom den använder samma estimering av skadekostnaden under hela utvecklingsperioden, vilket gör den väldigt beroende av att man gör en bra initial estimering. Detta är givetvis svårt, framför allt på bestånd som förändras.

Benktander-Hovinen är en viktning av Bornhuetter-Fergusson och Chain Ladder, syftet med modellen är att försöka utnyttja styrkan med båda modellerna, och den gör samma antaganden som dessa två modeller:

- Modellen antar att den totala skadekostnaden Ultimo är känd vid beräkningstidpunkten.

- Modellen antar att betalningsmönstret är densamma som för Chain Ladder och utnyttjar informationen från utvecklingsfaktorerna f_k för information om andel som har blivit utbetald av ultimo för övergången från Bornhuetter-Fergusson till Benktander-Hovinen. För att utvecklingsfaktorerna ska vara giltiga måste antaganden om Chain Ladder vara uppfyllda.

Det modellen gör är att i början av utvecklingsperioden gå mest på den estimerade totala skadekostnaden, U_0, som visades i modellbeskrivningen av Bornhuetter-Fergusson, Chain Ladder får väldigt lite vikt. Med tiden kommer dock att andelen av den estimerade skadekostnaden, U_0, att minska och andelen reserv som beräknas med Chain Ladder blir större.

Fördelen med denna modell jämfört med Bornhuetter-Fergusson är att den utnyttjar den kända informationen om utvecklingsmönstret i trianglarna i stället för att använda sig av samma estimerade skattning av den totala skadekostnaden genom hela utvecklingsperioden, vilket gör att den korrigerar för en möjlig felaktig estimering. Nedan presenteras härledningen för Benktander-Hovinen, beskrivningen följder den som ges i [3].

Vi såg i avsnitten om Chain Ladder och Bornhuetter-Fergusson att Chain Ladder gör inga antaganden om den totala skadekostnaden innan det finns någon data och att Bornhuetter-Fergusson drar i stället inte nytta av erfarenhet av utbetalningar. Benktanders förslag var att slå ihop modellerna genom att utnyttja en kredibilitetsfaktor c:

$$U_c = cU_{CL} + (1 - c)U_0$$

Eftersom c växer när skadeutbetalningen utvecklas, så antar vi att kredibilitetsfaktor c motsvarar p_k i förra avsnittet så $c = \frac{1}{f_k}$ och $1 - c$ motsvar q_k så att $1 - c$=1-$\frac{1}{f_k}$ Detta ger ett uttryck för reserven för Benktander-Hovinen R_{GB}, enligt :

$$R_{GB,k} = R_{BF,k} \cdot \frac{U_{pk}}{U_{0,k}}$$

Detta är metoden som föreslås av Benktander. Observera att vi har:

$$R_{GB,k} = q_k U_{pk}$$

Och

$$U_{0,k} p_k = p_k U_{CL} + q_k U_{0,k} = C_k + R_{BF,k} = U_{BF,k},$$

Dvs

$$R_{GB,k} = q_k U_{BF,k}$$

Vi får alltså att den slutliga Ultimoskattningen blir

$$U_{GB,k} = C_k + R_{GB,k} = (1 - q^2)U_{CL} + q^2 U_0 = U_{(1-q^2)}$$

Esa Hovinen använde sig i stället av kredibilitetsmixen direkt på reserven i stället för på skadekostnaden i ultimate. Han föreslog en reservmodell enligt:

$$R_{EH,k} = cR_{CL} + (1 - c)R_{BF,k},$$

även här med $c = p_k$

Men vid en närmare titt är Hovinens reservmodell identisk med den som Benktander presenterar:

$$R_{EH,k} = p_k q_k U_{CL} + (1 - p_k q_k)U_{0,k} = q_k U_{0,k} p_k = R_{GB,k}$$

5 Modell för antal skador och skadebelopp

I detta kapitel gör vi väljer vi antaganden för antal skador och om skadebeloppen. Antagandena nedan är hämtade från [1, p. 23].

För att kunna modellera antal skador och de individuella skadebeloppen så används en en sammansatt fördelning. Det är en sannolikhetsfördelning av summan av oberoende likafördelade stokastiska variabler, och definieras som att den totala skadekostnaden ges av den sammansatta fördelningen:

$$S = Y_1 + Y_2 + ... + Y_N = \sum_{n=1}^{N} Yi$$

För att kunna använda en sammansatt fördelning så antar vi följande:

1 N är en diskret stokastisk variable som antar bara värden så att $A \subset N_0$. Med detta menas att antal skador är ett heltal om är större än 0. Händelsen $N = 0$ betyder i detta fall att det inte finns några skador som har en skadekostnad där $S = 0$.

2 Y_1, Y_2,... $\overset{i.d.d}{\sim}$ G med $G(0) = 0$

Det andra antagandet innebär att de enskilda skadorna Y_i påverkar inte varandra. Om vi till exempel får en stor skada Y_1, så ger den ingen information om återstående skador Y_i, $i \geq 2$. Utöver det så har vi homogenitet i avseendet att alla skador har samma marginella fördelningsfunktion G där $0 = G(0) = P[Y \leq 1]$, alltså är de enskilda skadorna strikt positiva.

3 N och $(Y_1, Y_2, ...)$ är oberoende

Slutligen så säger det tredje villkoret att de enskilda skadorna inte påverkas av antalet skador och tvärtom. Om vi till exempel har många skador så ger det ingen information om det är små eller stora skador.

Detta är själva grunden i riskmodellering. Utifrån detta kan vi välja modell för antal skador N och utbetalningsmöntret Y_i. Själva grundantagandet är att:

$$E[S] = E[N] \cdot E[Y],$$

och

$$Var[S] = Var[N]E[Y_1]^2 + E[N]Var[Y_1]$$

5.1 Val av fördelning för antal skador

I denna uppsats antas antal skador, N, vara poissonfördelade.

Antagandet om poissonfördelning är vanligt vid simulering av antal skador. En poissonprocess är en stokastisk process i kontinuerlig tid som används vid beskrivning av slumpmässiga händelser som sker med en viss frekvens. Antalet händelser är oberoende och bara en händelse inträffar i taget. Vi väljer en fix volym v>0 och en frekvens λ >0 . Antal skador, N, antas då vara poissonfördelade:

$$N \sim Poi(\lambda v)$$

där

$$p_k = P[N = k] = e^{-\lambda v} \cdot \frac{\lambda v^k}{k!}$$

Utvecklingen av exponentialfunktionen $e^{-\lambda v}$ ger

$$\sum_{k>=0} p_k = 1$$

och vi har då en diskret fördelningsfunktion av $A \in N_0$

Vidare har vi att $E[N] = \lambda v = Var(N)$, se [1, p. 29]

5.2 Val av fördelning för skadebeloppen

Efter att ha hittat en modell för antal skador så vill hitta en modell som kan användas vid simulering för utbetalning av skador. En vanlig egenskap för skadebelopp på ett sakförsäkringsbolag är att ett fåtal stora skador står för en stor del av de sammanlagda skadebeloppen. Ett villkor som bör vara uppfyllt är att ersättningsbeloppen kan betraktas som utfall av oberoende och likafördelade stokastiska variabler. Utifrån antagandena ovan så antas utbetalningarna i simuleringen vara lognormalfördelade, där $Y \sim LN(\mu, \sigma^2)$.

Givet att vi har en lognormal stokastisk variabel Y och parametrarna μ och σ^2. som är väntevärde och standardavvikelse av den naturliga logaritmen av Y, då är Y lognormalfördelad, med väntevärde:

$$E[Y] = e^{\mu + \frac{\sigma^2}{2}}$$

och

variansen

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

se [1, p. 67]

6 Simulering av skadekosntad

I detta kapitel beskrivs simuleringen för att kunna analysera reservmodellerna. Vi gör det genom att simulera fram en antalstriangel med antal skador och utbetalningstriangel med mönstret på utbetalningar för de inträffade skadorna och utifrån det räkna ut den faktiska skadekostnaden och den skattade skadekostnaden utifrån modellen. Simuleringen kommer att göras i programmet R.

I förra kapitlet gjorde vi några grundantaganden om antal skador och skadebelopp som ger oss den information som behövs för att göra en simulering av skadekostnaden. För att kunna simulera skadekostnaden så kommer antal skador och skadebelopp att simuleras med en slumptalsgenerator. Användande av slumptalsgenerator på utbetalningsbelopp och antal skador ger en sammansatt fördelning, där det tar ca 20 år innan skadorna är slutreglerade, så i denna uppsats använder vi en triangel med 20 skadeår. Vi utnyttjar informationen som vi tog fram för den sammanslagna fördelningar, för att simulera en utbetalningstriangel:

$$E[S] = E[N] \cdot E[Y]$$

Varje rad och kolumn är en kombiton av väntevärdet av antal skador, N och utbetalningar, Y. Antal skador, N antas vara Poissonfördelade med $E[N] = 1000$ för varje av de 20 skadeåren. För att fördela ut skadorna med en större tyngd på tidigare utvecklingsår, så fördelas skadorna ut med en multinomial fördelning. Det första året har i snitt 30 procent av skadorna, det andra året har 25 procent, det tredje året har 10 procent och efter det avtar de fort mot 0.

Utbetalningarna, Y antas vara lognormalfördelade , där väntevärde

$$E[Y] = e^{\mu + \frac{\sigma^2}{2}}$$

och

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

Vi väljer μ och σ så att vi får att $E[Y] = 1$ och $Var[Y]$=v0,2 genom att lösa ekvationssystemet. Alla simulerade Y multipliceras med 10000 för att få ut mer realistiska värden och utbetalningarna antas växa med 2 procent per år. Antagandet för $E[S]$ ger oss en utbetalningstriangel och utifrån detta skattas den underliggande triangeln med utvecklingsfaktorerna. Utöver det så uppfyller triangeln villkor 1-3 i kapilet om antaganden för den sammanslagna fördelningen.

1 Antal skador, N, är ett heltal större än 0. Nollvärden tas inte med i simuleringen.

2 Alla skador har samma marginella fördelning och påverkar inte varandra.

3 Antal skador, N, och storleken på skadebeloppen, Y, är oberoende av varandra.

Simuleringen i uppsatsen bygger på 10 000 simuleringar, där faktisk skadekostnad, som är känd, jämförs med skadekostnaden som har estimerats med Chain Ladder, Bornhuetter-Fergusson och Benktander-Hovinen. Utfallen jämförs för att undersöka om Benktander-Hovinen ger en bättre estimering av skadekostnaden. Det som är av speciellt intresse är om Benktander-Hovinen är bättre på att skatta skadekostnaden i ett initialt skede, dvs för tidiga utvecklingsår, när dataunderlaget är litet och osäkerheten är stor. Alla resultat som redovisas i tabellerna nedan är väntevärdet för varje skadeår av de 10 000 körningarna.

Utifrån våra antaganden ovan gällande antal skador, N, och mönstret på skadeutbetalningar, Y så vet vi att antagandena för Chain Ladder är uppfyllda:

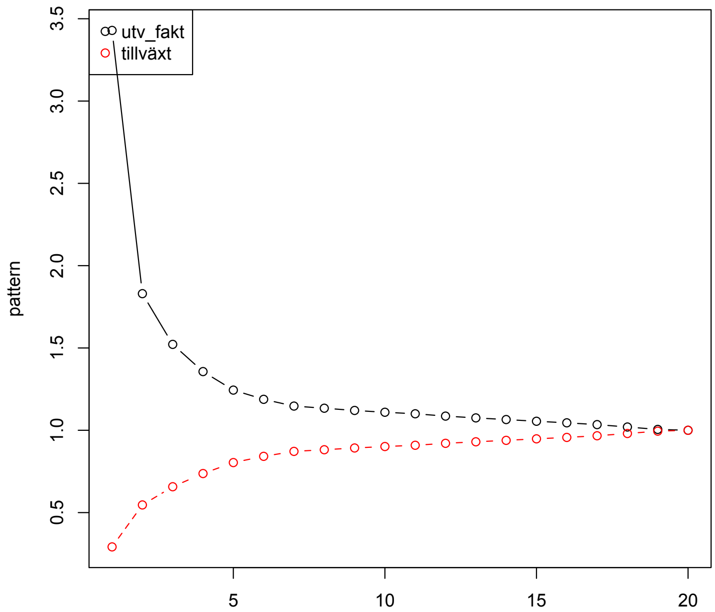
- Varje skadeår är poisson- och lognormalfördelad och oberoende av varandra, vilket är ett av grundantagandena i modellen.
- Att utbetalningsmöntret är lognormalfördelat gör att villkoret är uppfyllt gällande att betalningsmöntret får inte ändras.

Antagandet om den totala skadekostnaden, som används i Benktander-Hovinen, antas vara medelvärdet av den uppskattade skadekostnaden med Chain Ladder för de 5 senaste åren. För de tidigaste åren används medelvärdet av den uppskattade skadekostnaden för antal tillgängliga år. Tanken är att utnyttja den kända informationen för tidigare skadeår, där vi har mer vetskap om utvecklingen av skadekostnaden. Vidare antas att alla avgifter kopplad till premien är 0. Totala skadekostnaden motsvarar alltså medelvärdet av den enligt Chain Ladder estimerade skadekostnaden för de senaste 5 åren. Detta uppfyller grundvillkoren för Benktander-Hovinen och Bornhuetter-Fergusson:

- Den totala skadekostnaden är känd.

- I simuleringen används utvecklingsfaktorerna från Chain Ladder för att bestämma viktningen $\frac{1}{f_k}$, alltså är utvecklingsmönstret känt.

Till att börja med så gjordes en kontroll på om utbetalningsmönstret ser rimligt ut genom att undersöka hur produktsumman av utvecklingsfaktorerna och och tillväxten av utbetalningen ändrades med tiden. Den övre kurvan visar produktsumman av utvecklingsfaktorerna och den nedre visar tillväxten av utbetalningar, som är inversen av produktsumman av utvecklingsfaktorerna. Resultatet visas i figur 1 nedan:



Figur 1: Produktsumman av utvecklingsfaktorer och tillväxt av utbetalningar till Ultimo

Utvecklingsmönstret är realistiskt, där tyngden av utbetalningarna ligger på de tidigaste utvecklingsåren och avtar sedan tills skadorna är slutreglerade efter ca 20 år. Tabell 3 nedan visar de underliggande siffrorna till figur 1. Där ser man att 90 procent av utbetalningarna görs inom 10 år.

Fördelen med denna typ av figur är att den visar tydligare hur stor del av den totala skadekostnaden som har blivit utbetald. Utifrån de två figurerna kan vi dra slutsatsen att de antaganden som används gällande antal skador och skadeutbetalningar är rimliga för att användas för den simuleringen som ligger som för grund i uppsatsen.

	Utvecklingsfaktorer	Andel hittills utbetalt
1	3.43	0.29
2	1.83	0.55
3	1.52	0.66
4	1.36	0.74
5	1.24	0.80
6	1.19	0.84
7	1.15	0.87
8	1.13	0.88
9	1.12	0.89
10	1.11	0.90
11	1.10	0.91
12	1.09	0.92
13	1.08	0.93
14	1.07	0.94
15	1.05	0.95
16	1.05	0.96
17	1.03	0.97
18	1.02	0.98
19	1.01	0.99
20	1.00	1.00

Tabell 3: Produktsumma av utvecklingsfaktorer och andel hittills utbetalt av skadekostnad till Ultimo

6.1 Test av reservsättningsmodeller på simulerad data

I detta kapitel kommer vi att testa våra valda modeller för att undersöka hur bra de är på att estimera skadekostnaden på det på simulerade datamaterialet.Vi börjar med att ta fram den totala faktiska skadekostnaden för skadeår 1995-2014, som är känd i simuleringen.Vi har en skadekostnad som varierar mellan olika skadeår, däremot har vi inget mönster till exempel med en växande affär.

6.1.1 Chain Ladder

Vi börjar med att undersöka hur bra Chain Ladder är på att prediktera skadekostnaden för den valda portföljen. I tabell 4 nedan kan vi se en jämförelse av väntevärden för den faktiska skadekostnaden jämfört med väntevärden skadekostnaden estimerat med Chain Ladder.

Utfallet visar att Chain Ladder är bra på att estimera skadekostnaden ju längre bort vi kommer från aktuellt skadeår, skattningen för äldre år är väntevärdesriktiga. Det syns i tabell 4 att när vi närmar oss aktuellt skadeår så ökar diskrepansen mellan faktiskt skadekostnad och skadekostnad enligt Chain Ladder. Resultatet är väntat och är en av bristerna med Chain Ladder-modellen. Eftersom den totala skadekostnaden beräknas som en faktor utifrån det första årets utbetalning, som är lognormalfördelad och kan alltså variera mellan skadeår, så kan utfallet bli väldigt olika beroende på just denna första utbetalning. Till och med om vi skulle ha två skadeår med samma skadekostnad så skulle skattningen av den totala skadekostnaden enligt Chain Ladder bli väldigt olika beroende på hur de initiala skadebeloppen ser ut.

Detta problem är även känd från verkliga försäkringsportföljer, där faktorer som inflation, trender i utvecklingsfaktorerna och växande och minskande bestånd försvårar estimeringsarbetet, som bygger på antagandet om en oförändrad bestånd.

	Ult_fakt	UltimoCL	UltimoBF	UltimoBH
1995	23750373	23750373	23750373	23750373
1996	27840101	27754742	27744220	27754687
1997	30276452	30112217	30050136	30111005
1998	27103106	27420140	27359002	27418103
1999	30748541	30400370	30256296	30394127
2000	35150296	35435556	35235998	35425187
2001	30946222	30729578	30710685	30728422
2002	32023366	32962934	32769079	32949375
2003	33967064	34039769	33895535	34028336
2004	37443213	37868445	37757630	37858350
2005	32001933	32271873	32195772	32264363
2006	39303842	39876322	39504662	39836362
2007	36926852	36795498	36632656	36776252
2008	37676343	37321347	37356419	37325844
2009	33934009	33556123	33454478	33540033
2010	40764681	39978426	39968410	39976461
2011	40694631	41993482	41310384	41813942
2012	38323548	35636335	35925245	35735407
2013	45714637	47727055	44513567	46269634
2014	44260171	52485194	48055433	49347249

Tabell 4: Väntevärden av faktisk skadekostnad jämfört med väntevärden av estimerad skadekostnad enligt de valda reservsättningsmetoderna för skadeår 1995-2014

För att underlätta analysen så kan vi även jämföra differenser mellan faktisk och estimerad skadekostnad procentuellt, detta visas i tabell 5. Tabellen visar skadekostnad enligt modell dividerad med faktisk skadekostnad.

	UltimoCL	UltimoBF	UltimoBH
1995	1.00	1.00	1.00
1996	1.00	1.00	1.00
1997	0.99	0.99	0.99
1998	1.01	1.01	1.01
1999	0.99	0.98	0.99
2000	1.01	1.00	1.01
2001	0.99	0.99	0.99
2002	1.03	1.02	1.03
2003	1.00	1.00	1.00
2004	1.01	1.01	1.01
2005	1.01	1.01	1.01
2006	1.01	1.01	1.01
2007	1.00	0.99	1.00
2008	0.99	0.99	0.99
2009	0.99	0.99	0.99
2010	0.98	0.98	0.98
2011	1.03	1.02	1.03
2012	0.93	0.94	0.93
2013	1.04	0.97	1.01
2014	1.19	1.09	1.11

Tabell 5: Procentuella diffar mellan väntevärden av faktisk skadekostnad jämfört med väntevärden av estimerad skadekostnad enligt de valda reservsättningsmetoderna för skadeår 1995-2014

Här märks det tydligt att Chain Ladder avviker från faktisk skadekostnad för senast inträffade skadeår, både Bornhuetter-Fergusson och Benktander-Hovinen ger en bättre skattning och dessa skattningar ligger ganska nära varandra. För skadeår 2014 är alla metoder relativt långt i från faktisk utfall.

6.1.2 Bornhuetter-Fergusson

Vi går över till att titta på hur skadekostnaden estimeras med Bornhuetter-Fergusson. Den skattade skadekostnaden enligt Borhhuetter-Fergusson är summan av den faktiskt utvecklade skadekostnaden och den estimerade totala skadekostnaden, och när skadekostnaden närmar sig slutreglering så finns det en liten del kvar att estimera.

I tabell 4 ser vi att modellen ger en bra skattning av skadekostnaden för äldre skadår, precis som Chain Ladder. Det verkar dock även som att skattningen av skadekostnaden för aktuellt skadeår blir bättre än med Chain Ladder. Den estimerade totala skadekostnaden, som bygger på en medelskada av de senaste 5 åren, ser ut att ha en fördel jämfört med Chain Ladder, det verkar finnas en fördel med att utnyttja erfarenhet av tidigare skadeår för att skatta skadekostnaden i ett initialt läge, i stället för att använda empiriskt data.

Skillnaden jämfört med Chain Ladder är att vi inte låter de initiala utbetalningarna bestämma storleken på reserven, utan i början styrs skadekostnaden av ett estimat, som bygger på erfaranhet av hur tidigare års skadekostnader har utvecklats. Om vi lyckas göra en tillräckligt bra skattning av vad vi tror om den totala skadekostnaden, så kan det leda till en bättre skattning av skadekostnaden i det initiala skedet. Det verkar alltså finnas stöd för att använda någon form av utjämnande modell i den typ av portfölj som testas i simuleringen, där vi har variation i skadekostnaden mellan olika skadeår, man vill helst inte ha stora avvikelser i estimeringen av skadekostnaden och att använda sig av en modell av utjämning verkar minska en del av variationen för de senast inträffade skadeåren.

6.1.3 Benktander-Hovinen

Det intressanta blir nu att undersöka om Benktander-Hovinen ger en bättre skattning av skadekostnaden. Som vi nämnde i modellkapitlet så är tanken att utnyttja styrkan i Bornhuetter-Fergusson-modellen för de senast inträffade skadeåren men även styrkan i Chain Ladder-skattningen för tidigare inträffade skadeår, för att utnyttja erfarenheten av den information som vi har om hur skadekostnaden utvecklas.

Nackdelen med Bornhuetter-Fergusson är just att den inte ändrar på den estimerade skadekostnaden som görs innan den första skadeutbetalningen för ett givet skadeår, man skulle kunna tänka sig att Benktander-Hovinen ger en bättre skattning eftersom den även tar hänsyn till hur skadeutbetalningarna utvecklas.

Detta borde ge bättre skattning av skadekostnaden eftersom vi använder oss av både faktiska utbetalningar och en övergång till en estimering av skadekostnaden enligt Chain Ladder. Utfallet av detta visas i tabell 4.

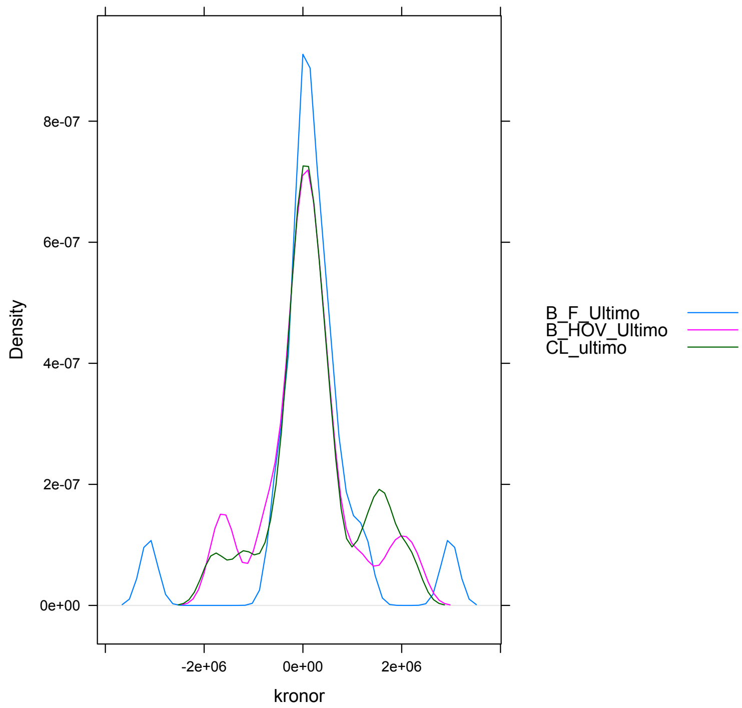
Även Benktander-Hovinen ger en bättre skattning av skadekostnaden än vad Chain Ladder gör. Däremot är skillnaden mellan Benktander-Hovinen och Bornhuetter-Fergusson liten. Vi hitter alltså inte stöd för att Benktander-Fergusson ger en bättre skattning än Bornhuetter-Fergusson. Detta kan dock bero på flera saker, det kan vara så att för det simulerade beståndet så ger den estimerade skadekostnaden som har valts en så bra estimering att man inte har nytta av den erfarenhet som Benktander-Hovinen tillför.

6.2 Reservvolatilitet

Efter att ha jämfört nivåerna på väntevärden av den estimerade skadekostnaden enligt de tre metoderna tittar vi på spridningen. Figur 2 nedan visar avvikelsen mellan väntevärden för estimerat skadekostnad jämfört med väntevärden för faktisk skadekostnad för skadeår 1995-2014 för de 3 metoder som testas i uppsatsen. I figuren används en utjämnad täthetskurva utifrån ett histogram för 10 000 körningar. Valet av en utjämnad täthetskurva i stället för ett histogram görs för att göra resultatet mer lättöverskådligt.

Det som vi vill undersöka är hur svansarna ser ut, det ger en bild av variansen. Bornhuetter-Fergusson lägger sig smalt och högt nära 0 vilket tyder på låg varians. Det som drar ned variansen är att Bornhuetter-Fergusson ger bäst skattning för senast inträffade skadeår, där osäkerheten är stor. Bornhuetter har dock även de sämsta skattningarna, vilket märks på pucklarna längst ut på höger och vänster sida i figur 2 på linjen för bornhuetter-Fergusson. Bornhuetter-Fergusson funkar alltså oftast bra men man kan även hamna väldigt snett.

Även Benktander-Hovinen ger en hög stapel nära 0. Chain Ladder-metoden ger en lägre och bredare spridning, som tyder på högre varians. Detta beror på den höga osäkerheten för senast inträffade skadeår. I figuren ser det ut som att alla 3 metoder är väntevärdesriktiga, vilket känns något överraskande med tanke på att de skadekostnaden för de senast inträffade skadeåren bygger på en antagen estimat för Bornhuetter-Fergusson och Benktander-Hovinen.



Figur 2: Utjämnad täthetskurva utifrån ett histogram för de 3 estimeringsmetoderna

Som en sista kontroll så tittar vi på medelvärdet för alla skadeår för alla de 3 metoderna och jämför med det faktiska utfallet i simuleringen. Detta visas i tabell 6 nedan. Alla metoderna ligger ganska nära varandra. Det intressanta är att i simuleringen verkar Chain Ladder vara den metod som är minst väntevärdesriktig. För att lättare kunna jämföra metoderna så visas i nästa tabell väntevärdet för det faktiska utfallet och väntevärdet av de 3 metoderna.

	fakt	CL	BF	B_HOV
1	34942469.49	35405789.51	34922299.51	35165176.08

Tabell 6: Medelultimo för faktiskt utfall och för alla metoder för alla skadeår

Även en analys av differenserna av medelvärdet mellan metoderna och faktiska utfall visar att Bornhuetter-Fergusson och Benktander-Hovinen är marginellt bättre än Chain Ladder. Detta visas i tabell 7 nedan.

Ett tvåsidig t-test på nivån 99 procent gjordes för att kontrollera om me-

	CL	BF	B_HOV
1	463320	-20169	222706

Tabell 7: Avvikelse i medel från faktiskt utfall för alla metoder för alla skadeår

delvärdet av de valda reservsättningsmetoderna låg inom konfidensintervallet för faktisk utfall.

- Det tvåsidiga t-testet mellan Chain Ladder och faktisk utfall gav ett kondidensintervall mellan -1768385 och 841745. Medlet av differensen var -463320 och låg inom konfidensintervallet.

- Det tvåsidiga t-testet mellan Bornhuetter-Fergusson och faktisk utfall gav ett kondidensintervall mellan -701678 och 742018. Medlet av differensen var -20169 och låg inom konfidensintervallet.

- Det tvåsidiga t-testet mellan Benktander-Hovinen och faktisk utfall gav ett kondidensintervall mellan -1101279 och 655866. Medlet av differensen var -222706 och låg inom konfidensintervallet.

- Alla metoder hade ett p-värde som låg högt över den valda nivån på 1 procent, alltså kan vi inte förkasta nollhypoten som säger att de två valda populationerna är lika. Enligt t-testet ger alltså alla våra reservsättningsmetoder en bra estimering av den faktiska skadekostnaden på den valda konfidensnivån på 99 procent. Testen av konfidensintervallen bekräftar det som visas i figur 2, Bornhuetter-Fergusson har minst spridning och Chain-Ladder har mest.

7 Analys och diskussion

I början av uppsatsen gjordes ett påstående om att Chain Ladder är en bra metod för att skatta reserven för äldre skadeår där man har en större datamängd men ger en hög osäkerhet i början. Detta påstående byggde på att estimeringen av de senaste skadeåren bygger på väldigt lite data, vilket gör att ultimoskattningen, som bygger på just detta data, blir väldigt osäker. Detta blir väldigt tydligt om skadebeloppet för det senaste skadeåret sätts till 0, då blir den estimerade skadekostnaden 0, vilket förstås inte är realistiskt. Detta är ett vanligt problem inom sakförsäkring, till exempel kan säsongsvariation, exempelvis beroende på hur julledigheten ligger, göra att väldigt få skador har registrerats, vilket skulle leda till ett väldigt dåligt estimat av skadekostnaden med Chain Ladder.

I denna uppsats testades om vi får en bättre reservskattning med metoder som inte lägger någon vikt på Chain Ladder initialt. Bornhuetter-Fergusson lägger ingen vikt på Chain Ladder alls utan bygger på antagandet att skadekostnaden är känd vid den initiala beräkningen av den totala skadekostnaden. Benktander-Hovinen Lägger en liten del av vikten på Chain Ladder initialt och viktar över till Chain Ladder med tiden. Båda metoderna gav en lägre varians och bättre skattningar än Chain Ladder för de senast inträffade skadeåren. Alla metoder verkar vara väntevärdesriktiga.

Dock bör inte Chain Ladder diskvalificeras som en reservsättningsmetod, när datamängderna blir större så ger metoden bra skattning av den estimerade skadekostnaden, dessutom ger Benktander-Hovinen och Bornhuetter-Fergusson sällan väntevärdesriktiga skattningar på riktigt data eftersom de bygger på mer eller mindre kvalificerade gissningar. Dessutom används information från Chain Ladder både i Bornhuetter-Fergusson och Benktander-Hovinen. Däremot bör man vara medveten om att osäkerheten är stor i ett initialt läge när datamängden är liten och är ett av de ständigt återkommande problemen inom sakförsäkring, när datamängden börjar bli tillräckligt stor för att göra bra skattningar så börjar utbetalningarna närma sig ultimo.

Referenser

[1] Wuthrich M.V (2015) *Non-Life Insurance: Mathematics and Statistics* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2319328

[2] Mack T. (1994) *Measuring the Variability of Chain Ladder Reserve Estimates* Casualty Actuarial Society Forum, Spring 1994, Vol 1, 101-182.

[3] Mack T. (2000) *Credible Claims Reserves: The Benktander Method* http://www.casact.org/library/astin/vol30no2/333.pdf

[4] Bornhuetter, R.L. and Ferguson, R.E. (1972) *The Actuary and IBNR* Proceedings of the Casualty Actuarial Society, Vol. LIX, pp 181-195.