



Stockholms
universitet

Andel ackumulerade okända belopp - kurvanpassning, osäkerhet och trend

Robert Eriksson

Masteruppsats 2013:4
Matematisk statistik
September 2013

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm



Matematisk statistik
Stockholms universitet
Masteruppsats 2013:4,
<http://www.math.su.se/matstat>

Andel ackumulerade okända belopp - kurvanpassning, osäkerhet och trend

Robert Eriksson*

September 2013

Sammanfattning

Försäkringsbolaget Trygg Hansa intresserar sig varje månad för hittills under året ackumulerat okänt belopp i förhållande till total skadekostnad för samma period. Säkra skattningar kan göras per kvartal men för resterande månader måste en funktion inter- och extrapoleras utifrån kvartalsskattningarna. Idag tenderar funktionen för januari och februari månad ligga för lågt och uppsatsen fokuserar på att hitta en funktion som bättre predikterar avsättningen ackumulerat okänt för dessa månader. I och med att andelen okänt beror direkt på skattade reserver belyser uppsatsen metoder som beräknar spridningsmått. Vi tittar på konfidensintervall för dels reserver per skadeperiod och dels totalt reserven. Parametrisk bootstrap tillämpas och används också i syfte av framförallt spridningsmått. Slutligen utförs en tidsserieanalys av andelen okänt för utvecklingskvartal ett i syfte att se om trender och säsongvariationer går att påvisa.

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

Abstract

The insurance company Trygg Hansa interested in every month so far this year accumulated unknown amount in relation to total claims cost for the same period. Confident estimates can be made per quarter but for the remaining months must be a function inter-and extrapolated from quarterly estimates. At present tend extrapolation to January and February to be too low. The essay focuses on finding a function that better predicts the deposition accumulated unknown for these months.

As the proportion of unknown depends directly on the estimated reserves highlights the essay methodology calculates volatility. We look at the confidence intervals for both reserves per claim period and as a total reserve. Parametric bootstrap application and is also used for the purpose of volatility measure. Finally conducted a time series analysis of the proportion of unknown development first quarter to see if trends and seasonal variations can be demonstrated.

Förord

Denna uppsats utgör ett examensarbete om 30 högskolepoäng och leder till en masterexamen i försäkringsmatematik vid Matematiska institutionen, Stockholms universitet.

Jag vill tacka mina två handledare, Lars Klingberg och Mathias Lindholm, för vägledning, goda råd och värdefull inspiration under arbetets gång. Lars jobbade i skrivandets början som chefaktuarie på Trygg Hansa och jobbar numera som chefaktuarie på Sirius International. Mathias är finansmatematiker på AFA Försäkring och adjungerad universitetslektor på Stockholm universitet. Jag vill också ta tillfället i akt att tacka Lisa Nilsson, aktuarie på Trygg Hansa, för sin hjälpsamhet.

Innehåll

Abstract	1
Förord	2
1 Introduktion	4
1.1 Inledning	4
1.2 Syfte och metod	4
2 Teori och statistiska modeller	7
2.1 Kurvanpassning	7
2.1.1 Splines	7
2.2 Metoder för reserver	8
2.2.1 Chain Ladder	8
2.2.2 Lognormalfördelad modell för utvecklingsfaktorer	11
2.2.3 Lognormalfördelad modell för reserver	12
2.2.4 Parametrisk Bootstrap	13
2.3 Tidsserieanalys	14
2.3.1 Modell	15
2.3.2 Dickey-Fuller-test	15
3 Skadedata	17
4 Resultat	18
4.1 Kurvanpassning	18
4.2 Konfidensintervall för reserver	23
4.3 Parametrisk Bootstrap	27
4.4 Tidsserieanalys	33
5 Slutsatser och diskussion	36
Referenser	37
A Härledning av parameterskattningar	38

1 Introduktion

1.1 Inledning

De flesta har någon gång kommit i kontakt med en försäkring, däremot är det troligtvis få som har kunskap om exakt hur en försäkring fungerar över tid. Vi inleder med en kortare förklaring.

Vid en vanlig enskild skadeförsäkring inleds processen med att försäkringstagaren tecknar en försäkring. Premien betalas för skydd under en period som ofta inom skadeförsäkring är ett år. När premien är inbetald sätter försäkringsgivaren av en s.k. premiereserv dels för ej intjänad premie och dels för kvardröjande risker. Ett försäkringsbolag drabbas givetvis av skador som ibland upptäcks under skadeåret och ibland inte, vilket ger upphov till ytterligare avsättningar. Kostnaderna för inträffade skador ska nämligen periodiseras så att all skadekostnad för ett specifikt år redovisas under just det året i årsredovisningen. För skador som inträffar och slutregleras under räkenskapsåret finns inga större problem. Svårare är att uppskatta skador som skett under året men som inte är slutreglerade vid årets slut. Sådana skador redovisas genom avsättning för oreglerade skador under försäkringsteknisk avsättning, FTA. Avsättning för oreglerade skador delas i sin tur upp i kända och okända skador alternativt IBNR- och RBNS-avsättningar. IBNR är förkortning för *incurred but not reported* och RBNS är förkortning för *reported but not settled*.

RBNS-avsättningar kan bestämmas på olika sätt beroende på situationen. Skadereglerare kan göra en bedömning utifrån sina erfarenheter, alternativt kan matematiska och statistiska metoder användas. En kombination av dessa är också vanligt förekommande. IBNR-avsättningar är svårare att bedöma. Problematiken ligger i att bestämma skador som inte än blivit rapporterade. Här är matematiska och statistiska metoder nödvändiga med historisk utgångspunkt. Uppsatsen kommer till stor del handla om IBNR-avsättningar och hur dessa uppskattas samt hänger ihop med vad Trygg Hansa kallar andel ackumulerat okänt.

1.2 Syfte och metod

Låt oss först definiera skadekostnad för en viss period som det totala skadebeloppet som uppskattningsvis kommer betalas ut under nämnda period. Skadebeloppet definierar vi som det beloppet som betalas ut till försäkringstagare. Skadekostnaden kan vara helt eller delvis känt, då skadekostnaden inte är helt känt beräknas IBNR (och eventuellt RBNS)-avsättningen för perioden som skadekostnad minus det hittills utbetalda skadebeloppet.

Trygg Hansa är varje månad intresserade av andelen ackumulerat skadebelopp i förhållande till skadekostnad hittills inkommet under året. T.ex. står vi i kvartal ett är vi intresserade av ackumulerade skadebelopp för januari, februari och mars i förhållande till slutgiltig skadekostnad för samma period. Andelen ackumulerat okänt ges sedan av ett minus andelen känt.

Då Trygg Hansa använder sig av trianglar per kvartal kan säkra skattningar av andelen ackumulerat okänt bara beräknas per mars, juni, september och december varje år, till resterande månader måste en funktion anpassas till kvartalspunkterna och därefter inter- och extrapoleras till övriga månader. Speciellt svårt är att få en uppskattning av andelen ackumulerat okänt för januari och februari då inga kvartalsdata hunnit komma in. Trygg Hansa använder sig då av en metod som bygger på att de två senaste kvartalen får representera de två kommande, andelen okänt för kvartal fyra föregående år får representera andelen okänt kvartal ett innevarande år och ackumulerat okänt för kvartal tre och fyra föregående år får representera andel ackumulerat okänt kvartal ett och två innevarande år. Utifrån dessa två punkter ska en funktion extrapoleras till månad ett och två, det vill säga januari och februari. I nuläget tenderar andelen ackumulerat okänt för framförallt januari men även februari att underskatta verkliga utfall.

Idag beräknas andelen ackumulerat okänt för januari och februari i princip genom en linjärextrapolation utifrån ackumulerat okänt för kvartal ett och två såsom de togs fram ovan. **Huvudsyftet med rapporten är att finna en funktion som ger en mer säker skattning av andel ackumulerat okänt för januari och februari än vad fallet är idag**, funktionen har som krav att kunna implementeras i Excel. Vi observerar direkt att funktionen för andel ackumulerat okänt är en avtagande funktion från januari till december. Resultatet av kurvanpassningen presenterar i avsnitt 4.1

För resterande månader fungerar dagens metod bra så där behövs i nuläget inget göras. Då däremot andelen ackumulerat okänt för januari och februari tenderar att underskatta vad andelen borde varit ligger bolaget fel i sin uppskattning för nämnda månader vilket nu ska försökas rättas till.

Andelen ackumulerat okänt beror direkt av reservskattning, vi vill därför också titta på reserverna mer i detalj. I en värld full av väntevärden behöver spridningsmått beräknas för att ha koll på sina risker. Framförallt kommer vi i avsnitt 4.2 undersöka hur bra Chain Ladder-metodiken kan sägas ge uppskattningar av reserven genom att beräkna standardavvikelser för reservsättningen samt genom att

förklara hur ett konfidensintervall kan göras för den övre gränsen av totala reserven.

I avsnitt 4.3 visas tekniken med parametrisk bootstrap och exemplifieras på användningsområden i försäkringsbranschen. Vi bygger en lognormalfördelad modell baserat på utvecklingsfaktorer. Vi ser också hur bootstrappingen slår på andel ackumulerat okänt.

I avsnitt 4.4 skapas en tidsseriemodell för andel ackumulerat okänt för rullande förstakvartal. Intresse ligger i att undersöka om trend och säsongskomponenter finns. Om så är fallet bör Trygg Hansa ta dessa i aktning för deras beräkningar. I och med att kvartal representerar varandra är det viktiga att inga stora signifikanta skillnader finns mellan kvartalen.

2 Teori och statistiska modeller

2.1 Kurvanpassning

Kurvanpassning innebär att anpassa en funktion till datapunkter. I avsnitt 4.1 kommer enklare funktioner hanteras och anpassas till två punkter, vilket betyder att vi utan bivillkor endast kan använda funktioner med två fria parametrar. Kravet leder till att funktioner som kan komma i fråga är tämligen enkla. Mer om skattningar av funktionernas parametrar i appendix A.

Splinefunktioner där olika funktioner anpassas till olika datapunkter används allt mer ofta. Då läsaren kanske inte är helt bekant med splines följer en förklaring här under. Vi kommer i resultatdelen tillämpa splinefunktionen på alla kvartalsskattningar av andel ackumulerat okänt till skillnad från övrig kurvanpassning där vi enbart använder mars och juni.

2.1.1 Splines

Idén bakom kubiska splines [2, kap. 5] är att anpassa en funktion på formen

$$S(x) = \begin{cases} s_1(x) & \text{om } x_1 \leq x \leq x_2 \\ s_2(x) & \text{om } x_2 \leq x \leq x_3 \\ \vdots & \\ s_{n-1}(x) & \text{om } x_{n-1} \leq x \leq x_n \end{cases} \quad (2.1)$$

där s_i är ett tredjegradspolynom definierat genom

$$s_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i \quad (2.2)$$

för $i = 1, 2, \dots, n - 1$. Första- och andraderivatans av $s_i(x)$.

$$s'_i(x) = 3a_i(x - x_i)^2 + 2b_i(x - x_i) + c_i \quad (2.3)$$

$$s''_i(x) = 6a_i(x - x_i) + 2b_i \quad (2.4)$$

Det går sedan att visa att funktionen för de olika intervallen ges av ekvationssystemet

$$\begin{bmatrix} 1 & 4 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & 4 & 1 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 4 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 & 4 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 4 & 1 \end{bmatrix} \begin{bmatrix} s_1''(x_1) \\ s_2''(x_1) \\ s_3''(x_1) \\ s_4''(x_1) \\ \vdots \\ s_{n-3}''(x_{n-3}) \\ s_{n-2}''(x_{n-2}) \\ s_{n-1}''(x_{n-1}) \\ s_n''(x_n) \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ y_3 - 2y_4 + y_5 \\ \vdots \\ y_{n-4} - 2y_{n-3} + y_{n-2} \\ y_{n-3} - 2y_{n-2} + y_{n-1} \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix} \quad (2.5)$$

där $y_i = S(X_i)$ samt $h = x_i - x_{i-1}$. Systemet har $n - 2$ rader och n kolumner vilket innebär att det är underdimensionerat. Detta löses vanligen genom att sätta $s_1''(x_1) = s_n''(x_n) = 0$ som kallas naturliga splines eller $s_1''(x_1) = 2s_2''(x_2) - s_3''(x_3)$ samt $s_n''(x_n) = 2s_{n-1}''(x_{n-1}) - s_{n-2}''(x_{n-2})$ som kallas kubiska splines. På grund av uppsatsens omfång kommer vi uteslutande jobba med kubiska splines.

2.2 Metoder för reserver

De försäkringstekniska avsättningarna kan beräknas på många sätt. Vi kommer titta på framförallt fyra olika metoder.

- Chain Ladder.
- Två olika lognormalfördelade modeller.
- Hur parametrisk bootstrap kan tillämpas från den ena lognormalfördelade modellen.

2.2.1 Chain Ladder

Låt $d_{i,j}$ beteckna det skadebelopp som betalats ut under utvecklingsperiod $[j, j+1]$ och som härstammar från skadeperiod i . Vidare låter vi $D_{i,j}$ ha samma tolkning som $d_{i,j}$ men för framtida utbetalningsperioder.

En inkrementell skadekvadrat med båda kända och skattade skadebelopp för senaste skadeperioden m får därmed utseendet i tabell 2.1.

Skadepериод	Utvecklingsperiod					
	1	2	3	...	$m - 1$	m
1	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$...	$d_{1,m-1}$	$d_{1,m}$
2	$d_{2,1}$	$d_{2,2}$	$d_{2,3}$...	$d_{2,m-1}$	$D_{2,m}$
3	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$...	$D_{3,m-1}$	$D_{3,m}$
...
$m - 1$	$d_{m-1,1}$	$d_{m-1,2}$	$D_{m-1,3}$...	$D_{m-1,m-1}$	$D_{m-1,m}$
m	$d_{m,1}$	$D_{m,2}$	$D_{m,3}$...	$D_{m,m-1}$	$D_{m,m}$

Tabell 2.1: Skadekvadrat för historiska och alltså kända skadebelopp, $d_{i,j}$, och framtida okända skadebelopp, $D_{i,j}$.

Definiera nu $C_{i,j} = d_{i,1} + d_{i,2} + \dots + D_{i,j-1} + D_{i,j}$ som den ackumulerade avsättningen för tiden j . $C_{i,j}$ kan vara helt eller delvis känd beroende på j .

Skadepериод	Utvecklingsperiod					
	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 2.2: Exempel på en ackumulerad skadetriangel.

Chain Ladder har använts under en lång tid och är förmodligen den mest populära metoden att skatta reserver. En stor anledning är dess enkelhet samt oberoendet av en specifik fördelning. Beskrivningen härunder baseras på en artikel av Mack [5]. Vi inleder med modellantaganden.

1. De ackumulerade beloppen $C_{i,j}$ för olika skadepерioder är oberoende.
2. Det existerar utvecklingsfaktorer f_j så att för alla i, j gäller

$$E(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}) = E(C_{i,j+1} | C_{i,j}) = C_{i,j} \cdot f_j. \quad (2.6)$$

Observera att f_j inte beror på innevarande skadeår. Vidare så gäller likheten

$$E(C_{i,j+k} | C_{i,1}, \dots, C_{i,j}) = C_{i,j} \cdot f_{j+1} \cdot \dots \cdot f_{j+k-1} \quad (2.7)$$

Bevisas t.ex. genom

$$E(C_{i,j+k}|C_{i,1}, \dots, C_{i,j}) = E[E(C_{i,j+k}|C_{i,1}, \dots, C_{i,j+k-1})|C_{i,1}, \dots, C_{i,j}] \quad (2.8)$$

$$= E(C_{i,j+k-1} \cdot f_{j+k-1}|C_{i,1}, \dots, C_{i,j}) \quad (2.9)$$

$$= E(C_{i,j+k-1}|C_{i,1}, \dots, C_{i,j}) \cdot f_{j+k-1} \quad (2.10)$$

$$= E[E(C_{i,j+k-1}|C_{i,1}, \dots, C_{i,j+k-2})|C_{i,1}, \dots, C_{i,j}] \cdot f_{j+k-1} \quad (2.11)$$

$$= \{\text{Upprepning}\}$$

$$= C_{i,j} \cdot f_{j+1} \cdot \dots \cdot f_{j+k-1}. \quad (2.12)$$

Skattningen för f_j , som betecknas \hat{f}_j behövs för att kunna prediktera framtida skador och därmed årets avsättning. En väntevärdesriktig skattning är av f_j är

$$\hat{f}_j = \frac{\sum_{i=1}^{m-j} C_{i,j+1}}{\sum_{i=1}^{m-j} C_{i,j}}. \quad (2.13)$$

Genom formel (2.6) samt (2.13) så blir reserven för år i

$$\hat{R}_i = \hat{E}(C_{i,m}|C_{i,j}) - C_{i,j} = C_{i,j} \left(\sum_{k=j}^{m-1} \hat{f}_k - 1 \right) \quad (2.14)$$

där $\hat{E}(C_{i,m}|C_{i,j}) = C_{i,j} \cdot \hat{f}_j \cdot \dots \cdot \hat{f}_{m-1}$.

Standardfelet (mått på osäkerhet i skattningen) för ultimoreserven ligger i intresse för alla försäkringsföretag för att dels veta hur bra skattningarna är och dels för hur stor risken är att reserven undervärderas. Standardfelet i kvadrat för ultimoreserven av en skadepperiod ges av

$$(s.e.(C_{i,m}))^2 = (E(C_{i,m}))^2 \cdot \sum_{k=m+1-i}^{m-1} \frac{\sigma_k^2}{f_k^2} \left(\frac{1}{C_{i,k}} + \frac{1}{\sum_{j=i}^{m-k} C_{j,k}} \right) \quad (2.15)$$

där σ_k^2 beräknas genom

$$\sigma_k^2 = \frac{1}{m-k-1} \sum_{j=1}^{m-k} C_{j,k} \left(\frac{C_{j,k+1}}{C_{j,k}} - f_k \right)^2. \quad (2.16)$$

För mer ingående detaljer hänvisas den intresserade läsaren till Macks artikel.

Diskussionen för och emot Chain Ladder är ständigt aktuell. Fördelarna ligger till stor del i att metoden är enkel att förstå och utföra. Det är även relativt enkelt för personer utan försäkringsmatematisk utbildning att förså, vilket är relevant ur redovisningssynpunkt. Historien visar att Chain Ladder predikterar avsättningarna rätt så bra. Nackdelarna går dock inte riktigt att förbise. T.ex. tas ingen hänsyn till kalendertidsperiod, säsongeffekter och trender. Rent matematiskt är principen att algoritmiskt fylla i en fyrkant också lite tveksamt. En annan nackdel är att inga större antaganden om hur uppgifterna tagits fram görs och således används inte allt för mycket riskteori. Vid extremfall analyseras dock givetvis detaljerna alternativt stryks som outliers, här kan argumenteras att man då styr reserven i önskad riktning men det är en annan diskussion utanför ämnet för uppsatsen.

2.2.2 Lognormalfördelad modell för utvecklingsfaktorer

Istället för en fördelningsfri modell som Macks Chain Ladder ovan kan vi tänka oss ett fördelningsantagande [6, kap. 5]. Den här modellen antar ackumulerade belopp där kvoten $F_{i,j} = C_{i,j}/C_{i,j-1}$ är lognormalfördelad. Först några modellantaganden.

1. Kvoten $F_{i,j} = C_{i,j}/C_{i,j-1}$ är lognormalfördelad med parametrar μ_j och θ_j^2 så att

$$\eta_{i,j} = \ln(F_{i,j}) \in N(\mu_j, \theta_j^2) \quad (2.17)$$

för alla i, j .

2. $\eta_{i,j}$ är oberoende för alla i, j .

Under modellantaganden ovan följer

$$E(F_{i,j}) = \exp\left(\mu_j + \frac{1}{2}\theta_j^2\right) \quad (2.18)$$

$$\text{Var}(F_{i,j}) = \exp(2\mu_j + \theta_j^2) (\exp(\theta_j^2) - 1) \quad (2.19)$$

Parametrarna μ_j och θ_j skattas genom

$$\hat{\mu}_j = \frac{1}{I-j+1} \sum_{i=0}^{I-j} \ln\left(\frac{C_{i,j}}{C_{i,j-1}}\right) \quad (2.20)$$

$$\hat{\theta}_j^2 = \frac{1}{I-j} \sum_{i=0}^{I-j} \left(\ln\left(\frac{C_{i,j}}{C_{i,j-1}}\right) - \hat{\mu}_j\right)^2 \quad (2.21)$$

som följer fördelningarna

$$\hat{\mu}_j \in N\left(\mu_j, \frac{\theta_j^2}{I-j+1}\right) \quad (2.22)$$

$$\frac{I-j}{\theta_j^2} \hat{\theta}_j^2 \in \chi_{I-j}^2 \quad (2.23)$$

där χ_{I-j}^2 är χ^2 -fördelningen med $I-j$ frihetsgrader. Vidare är $\hat{\mu}_j$ och $\hat{\theta}_j^2$ oberoende. Bevis finns i [6, kap. 5].

Vi noterar till sist att då utvecklingsfaktorerna följer en lognormal fördelning kommer också reserverna för skadeperioderna (i denna uppsats skadekvartalen) att följa en lognormal fördelning enligt känd teori. Totala reserven som är summan av reserven för skadeperioderna följer däremot inte en lognormal fördelning.

2.2.3 Lognormalfördelad modell för reserver

En annan variant som inte ska blandas ihop med avsnitt 2.2.2 men som ändå fördelningensmässigt hänger ihop med nämnda modell är den som Mack presenterar i en smått berömd text [5]. Mack resonerar att det går att bygga ett konfidensintervall för avsatta reserver med hjälp av lognormalfördelningen. Modellantaganden följer.

1. Varje reserv R_i skattas med Chain ladder (2.14) och dess standardfel i kvadrat med (2.15).
2. Varje reserv, R_i , för skadeår i är lognormalfördelad.

$$\ln(R_i) \in N(\mu_i, \theta_i^2) \quad (2.24)$$

3. Reserverna R_i är sinsemellan oberoende för alla i , men dess skattningar är det inte då de beror på utvecklingsfaktorer som skattas från flera skadeperioder.

Vi ska nu beräkna μ_i och θ_i^2 så att väntevärdet i fördelningen motsvarar Chain Ladder-skattningen av reserverna och att variansen motsvarar standardfelet i Chain Ladder-metoden i kvadrat, det vill säga

$$\hat{R}_i = \exp(\mu_i + \theta_i^2/2) \quad (2.25)$$

$$(s.e.(R_i))^2 = \exp(2\mu_i + \theta_i^2)(\exp(\theta_i^2) - 1) \quad (2.26)$$

vilket leder till parametrarna

$$\hat{\theta}_i^2 = \ln(1 + (s.e.(R_i))^2 / \hat{R}_i^2) \quad (2.27)$$

$$\hat{\mu}_i = \ln(\hat{R}_i) - \hat{\theta}_i^2/2 \quad (2.28)$$

Vi vill hitta en nivå som med en viss sannolikhet reserven för skadeperiod i inte kommer gå över. Konfidensintervallet sägs då vara ensidigt. En rimligt nivå är att vi med 90% sannolikhet kan vara säkra på att den reserven för skadeperiod i inte går över ett visst belopp.

90%-percentilen för \hat{R}_i under modellantaganden är

$$\exp(\hat{\mu}_i + 1.28\hat{\theta}_i) = \quad (2.29)$$

$$\exp(\ln(\hat{R}_i) - \hat{\theta}_i^2/2 + 1.28\hat{\theta}_i) = \quad (2.30)$$

$$\hat{R}_i \exp(1.28\hat{\theta}_i - \hat{\theta}_i^2/2). \quad (2.31)$$

Metoden att beräkna variansmått används också med fördel till reservsättningsosäkerhet. För att ha kontroll på sina reserver finns vi det intressant att beräkna standardfelet för den totala reserven, $\hat{R} = \sum_i \hat{R}_i$. Tyvärr vad det gäller standardfelet för hela reserven är det inte lika enkelt som att summera kvadraten av standardfelet för alla skadeår i och dra roten ur. Då reserverna R_i själva är oberoende är dess skattningar, \hat{R}_i , det inte då skattningarna beror på utvecklingsfaktorerna \hat{f}_j . Vi får istället att ett variansmått för R är lika med det mer komplicerade uttrycket

$$(s.e.(R))^2 = \sum_{i=2}^I \left((s.e.(R_i))^2 + C_{i,I} \left(\sum_{j=i+1}^I C_{j,I} \right) \sum_{k=I+1-i}^{I-1} \frac{2\sigma_k^2/\hat{f}_k^2}{\sum_{n=1}^{I-1} C_{n,k}} \right). \quad (2.32)$$

2.2.4 Parametrisk Bootstrap

Bootstrap (återsampling på svenska) används för att skaffa information om t.ex. en parameter eller liknande genom att dra nya stickprov från observerade data.

Metodiken som används här benämns vanligen parametrisk bootstrap [6, kap. 7] och innebär att oberoende och likafördelade observationer Z_i följer

$$Z_i, \dots, Z_n \in G_{\mu_j, \theta_j} \quad (2.33)$$

där G_{μ_j, θ_j} betecknar en känd fördelning med okända parameter μ_j och θ_j . Valet av två parametrar är för vad som kommer användas i uppsatsen. Självfallet kan parameterantalet vara olika beroende på situation.

För att återsampla skattas parametrarna μ_j med $\hat{\mu}_j$ och θ med $\hat{\theta}_j$ genom t.ex. minsta kvadrat-metoden eller maximum likelihood-metoden. Själva parametriska bootstrappingen använder sedan vad som brukar kallas bootstrapfördelningen,

$$Z_i^*, \dots, Z_n^* \in \hat{G}_{\hat{\mu}_j, \hat{\theta}_j} = G_{\hat{\mu}_j, \hat{\theta}_j} \quad (2.34)$$

I denna text tillämpas en lognormal model för ackumulerade betalningsströmmar. Utgångspunkten är att utvecklingfaktorerna $F_{i,j} = C_{i,j}/C_{i,j-1}$ är lognormalfördelad med parametrar μ_j och θ_j^2 så att

$$\eta_{i,j} = \ln(F_{i,j}) = \ln\left(\frac{C_{i,j}}{C_{i,j-1}}\right) \in N(\mu_j, \theta_j^2) \quad (2.35)$$

där $\eta_{i,j}$ är oberoende för $i = 1, 2, 3, \dots$ och $j = 1, 2, 3, \dots$

Parametrarna μ_j och θ_j skattas med följande.

$$\hat{\mu}_j = \frac{1}{I-j+1} \sum_{i=0}^{I-j} \ln\left(\frac{C_{i,j}}{C_{i,j-1}}\right) = \frac{1}{I-j+1} \sum_{i=0}^{I-j} \eta_{i,j} \quad (2.36)$$

$$\hat{\theta}_j^2 = \frac{1}{I-j} \sum_{i=0}^{I-j} \left(\ln\left(\frac{C_{i,j}}{C_{i,j-1}}\right) - \hat{\mu}_j \right)^2 = \frac{1}{I-j} \sum_{i=0}^{I-j} (\eta_{i,j} - \hat{\mu}_j)^2 \quad (2.37)$$

Genom att successivt återsampla från ursprungstriangeln får vi l estimat för parametrarna genom

$$\hat{\mu}_j^{*l} = \frac{1}{I-j+1} \sum_{i=0}^{I-j} \eta_{i,j}^* \quad (2.38)$$

$$(\hat{\theta}_j^{*l})^2 = \frac{1}{I-j} \sum_{i=0}^{I-j} (\eta_{i,j}^* - \hat{\mu}_j^{*l})^2 \quad (2.39)$$

$$(2.40)$$

På så sätt fås vektorer med återsamplade $\hat{\mu}_j^*$ och $\hat{\theta}_j^*$ som i sin tur ger reservvektorn \hat{R} . Hela reserven för iteration l beräknas av formeln

$$\hat{R}^l = \sum_{i=1}^I C_{i,j-1} \left(\exp \left(\sum_{j=I-i+1}^J \hat{\mu}_j^{*l} + \frac{1}{2} \sum_{j=I-i+1}^J (\hat{\theta}_j^{*l})^2 \left(1 - \frac{1}{I-j+1} \right) \right) - 1 \right) \quad (2.41)$$

enligt samma metod som vanlig triangeluppräknig.

2.3 Tidsserieanalys

Såsom Trygg Hansa jobbar med andel ackumulerat okänt finner vi det intressant att se hur dessa ändras över tid. Då datamängden är något begränsad i och med att fokus ligger på januari och februari månad bör avsnittet ses mer som en diskussion än en strikt teori.

2.3.1 Modell

Vi är intresserade av en modell som innefattar trend och säsong för att se om vi kan påvisa sådana faktorer. En enkel additiv tidsseriemodell med trend och säsong [1, kap. 4] kan skrivas på formen

$$X_t = T_t + S_t + \epsilon_t, \quad t = 1, \dots, n \quad (2.42)$$

där X_t är tidsserievärdet vid tiden t , T_t är den specifika trenden för samma tidpunkt, S_t är en säsongskoefficient och ϵ_t ett slumpfel. Vi har valt följande metoder för att stegvis analysera tidsserien.

1. Skatta trendkomponenten T_t med \hat{T}_t genom glidande medelvärde, d.v.s.

$$\hat{T}_t = \frac{x_{t-q} + x_{t-k+1} + \dots + x_{t+q-1} + x_{t+q}}{d}, \quad q < t \leq n - q \quad (2.43)$$

där $d = 2q$ och d sätter vi till fyra då vi har kvartal som säsongsvariation.

2. Säsongskomponenten S_k skattas med \hat{S}_k genom att för varje $k = 1, \dots, d$ ta medelvärdet, W_k av

$$X_{k+jd} - \hat{T}_{k+jd} \quad (2.44)$$

för heltal j som uppfyller $q < k + jd \leq n - q$. Då säsongskomponenterna bör summeras till noll normeras W_k och komponenterna skattas genom

$$\hat{S}_k = W_k - d^{-1} \sum_{i=1}^d W_i. \quad (2.45)$$

Observerar att $\hat{S}_t = \hat{S}_{t+d}$ gäller.

3. Till sist testas den rensade serien

$$Y_t = X_t - \hat{T}_t - \hat{S}_t \quad (2.46)$$

för stationäritet med Dickey-Fuller-test, se separat teoriavsnitt nedan, det vill säga om det finns indikatorer på trend och säsong.

2.3.2 Dickey-Fuller-test

Låt $\{Y_t\}$ vara en $AR(1)$ -process,

$$Y_t = \phi Y_{t-1} + \epsilon_t \quad (2.47)$$

där ϵ_t är ett slumpfel och ϕ en koefficient. Dickey-Fuller [3, kap. 7] testar följande nollhypotes.

$$H_0 : |\phi| = 1, \text{ d.v.s. tidsserien är ickestationär (har en enhetsrot).}$$

$$H_1 : |\phi| < 1, \text{ d.v.s. tidsserien är stationär (har inte en enhetsrot)}$$

Om vi inte kan förkasta H_0 kan vi inte förkasta att processen $\{Y_t\}$ är en ren slumpvandring och därmed inte stationär. $AR(1)$ -processen kan skrivas om till

$$Y_t - Y_{t-1} = \phi Y_{t-1} - Y_{t-1} + \epsilon_t \quad (2.48)$$

$$\Delta Y_t = \gamma Y_{t-1} + \epsilon_t \quad (2.49)$$

där $\gamma = \phi - 1$. Test för $\phi = 1$ är därmed ekvivalent med att testa om $\gamma = 0$. Formeln kan vid intresse utökas med trend och säsongsvariation.

$$\Delta Y_t = \eta + \beta t + \gamma Y_{t-1} + \epsilon_t \quad (2.50)$$

Under H_0 tittar man sedan på teststatistikan

$$DF = \frac{\hat{\gamma}}{Std(\hat{\gamma})}. \quad (2.51)$$

Om DF är mindre än det kritiska värdet förkastas H_0 . Fördelningen för DF följer inte någon vanlig fördelning utan kvartiler måste beräknas med hjälp av numeriska approximationer eller simulering vilket innebär att testet med fördel görs numeriskt, t.ex. i Matlab.

3 Skadedata

Skadedata kommer från försäkringsbolaget Trygg Hansas SAS-databas för Hemförsäkring som innehåller information om skadehändelser för hem-, rese- och specialförsäkringar. Överlägset största taxerade belopp är hemförsäkringen. Denna försäkringsklass är snabbt känd och reglerad. I datamaterialet ses faktiskt att man ibland övervärderar reserven och blir tvungen att minska skulden efter ett tag. Databasen är beskriven i tabell 3.

Kolumnnamn	Beskrivning
Bolagskod	Vilket bolag skadan gäller
Fromdate	När skadedelen/momentet senast ändrades
Tilldate	Ytterligare ändring skedde detta datum
Jskadenr	Idnummer för skadan
Isk_jmom	Typ av skademoment
Produkttypkod	Typ av händelsekod mest för olycksfall och företag
Skaddat	Datum då skadan inträffade
Skadar	Året då skadan inträffade
Isk_jmnetto	Totalt utbetalt minus totalt inbetalt
Isk_jmtaxbel	Total reserv plus totalt utbetalt minus totalt inbetalt
Isk_restot	Total reserv
Bransch	Vilken bransch för skadan
Regdate_skada	Registreringsdatum
Skdid	Idnummer för skademomentet
Handelsekod	Kod

Tabell 3.1: Beskrivning av kolumnerna i databasen.

Intressant för denna rapport är kolumnerna Fromdate, Tilldate, Jskadenr, Skaddat, Isk_jmtaxbel, Regdate_skada samt Skdid och i viss mån Bolagskod. Regdate_data, Fromdate och Tilldate används till att få in- och utbetalningar ackumulerade. Skade-id används för att separera händelser. Vid t.ex. en brand, krock, olycka eller liknande kan flera försäkringar komma i fråga - dessa måste skiljas. Isk_jmtaxbel är det som i försäkringsjargong benämns incurred, alltså inträffade skadebelopp.

4 Resultat

4.1 Kurvanpassning

Trygg Hansa intresserar sig varje månad för procentsatsen andel ackumulerat okänt som förklarades tidigare i uppsatsen, backa gärna till avsnitt 1.2. Vi börjar med att införa en notation.

Låt $Q_{y,j}^{x,i}$ innebära hittills inkomna ackumulerade skadebelopp för skadeår x , skadekvartal i vid tidpunkt y, j . T.ex. innebär då $Q_{2011,4}^{2011,1}$ de ackumulerade skadebeloppen för skadeår 2011, skadekvartal ett vid tidpunkten 2011, kvartal fyra. I praktiken då vi hela tiden försöker få fram andel ackumulerat okänt för januari och februari kommer notationen vara $Q_{x,4}^{x,3}$ och $Q_{x,4}^{x,4}$.

Utförandet kommer ske stegvis enligt följande tre punkter. Då den stora problematiken är framförallt januari och februari månad kommer vi utgå från dem och *inte* använda rullande kvartal.

1. SAS-programmen som används idag på Trygg Hansa ger trianglar på kvartalsform. För att få en uppfattning om månader måste därför nya program skrivas som istället ger trianglar på månadsbasis.
2. Ur de nya trianglarna beräknas andel ackumulerat okänt för januari och februari direkt ur data och dem får bli facit för kurvanpassningen.
3. Då vi står i början av skadeår $x+1$ används $Q_{x,4}^{x,3}$ och $Q_{x,4}^{x,4}$ för att ta fram andel ackumulerat okänt för mars och juni. Till dessa ska en funktion anpassas och sedan extrapoleras till januari och februari (månad ett och två).

Funktionen bör vara relativt enkel då den ska tillämpas i Excel. Ett dilemma är därmed hur bra ska funktionen passa versus hur enkel bör den vara. Ytterligare problem med själva metoden är hur bra skattningarna av kvartal fyra och kvartal tre *föregående* år kan sägas representera kvartal ett och kvartal två *innevarande* år. Mer analys av trender och säsongsvariationer i avsnitt 4.4.

Utgångspunkt för kurvanpassningen är den linjära funktionen, exponentialfunktionen, potensfunktionen och den logaritmiska funktionen. Det vill säga funktioner

på formen

$$g(z) = A + Bz \quad (4.1)$$

$$g(z) = Ae^{zB} \quad (4.2)$$

$$g(z) = Az^B \quad (4.3)$$

$$g(z) = A + B \ln(z) \quad (4.4)$$

där parametrarna A och B beräknas genom ekvationssystemet

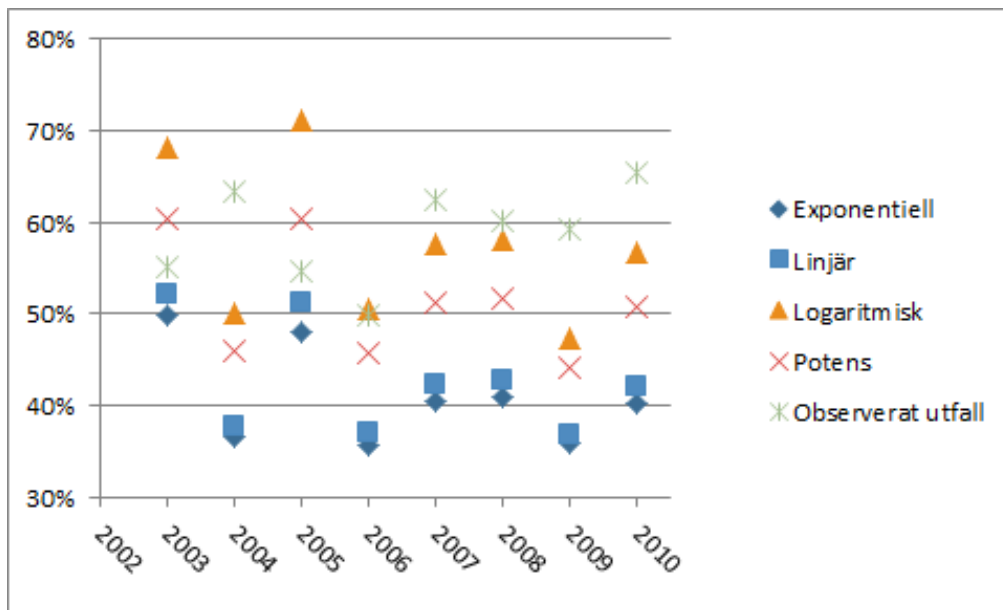
$$\begin{cases} g(3) = Q_{x,4}^{x,4} \\ g(6) = Q_{x,4}^{x,3} \end{cases}$$

och z antar värdena 1, 2 som representerar januari och februari vilka är månaderna som andel ackumulerat okänt för mars och juni ska extrapoleras till.

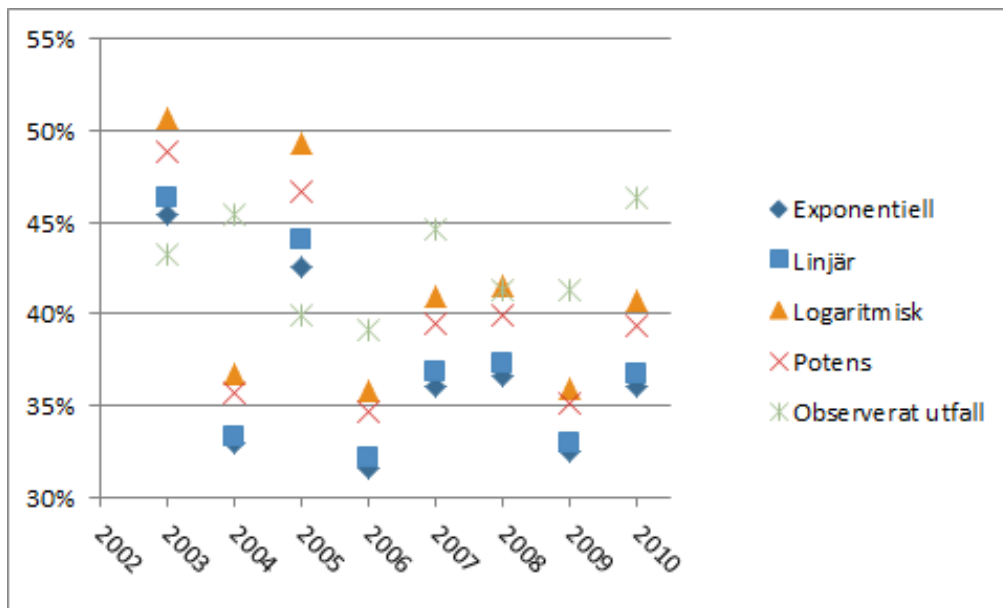
Inom klassen finns data för att prediktera skadekostnaden tillräckligt bra från och med början av 2002 fram till i skrivande stund slutet av 2010. Från och med 2011 tills idag anses som för få punkter för att ge en bra bild av flödena. Totalt finns därmed åtta stycken facit för januari och februari. Något för få för att göra en ordentlig säkerställd analys men är allt som finns att jobba efter.

Funktionerna ovan anpassas, som tidigare nämdes, med hjälp av ett 2×2 -ekvationssystem, se appendix A. Vi utgår från det ackumulerat kända inflödena så andel ackumulerat okänt beräknas genom $1 - g(z)$. Hela tiden anpassar vi en funktion till andel ackumulerat känt och får sedan ackumulerat okänt genom ett minus funktionen.

Resultatet av extrapoleringen av funktionerna $1 - g(z)$ tillsammans med de åtta observerade punkterna för andel ackumulerat okänt beräknade ur månadstränglarna plottas i figur 4.1 för januari månad och figur 4.2 för februari månad.



Figur 4.1: Funktionernas skattning i förhållande till observerade utfall för januari månad.



Figur 4.2: Funktionernas skattning i förhållande till observerade utfall för februari månad.

Ur figurerna är det inte helt lätt att tyda vilka funktioner som ligger närmast andel ackumulerat okänt som beräknades ur månadstrianglarna, vilka benämns ”observerade utfall”. En sammanställning av kvadratroten ur summan av de kvadratiske felen,

$$\sqrt{\sum_i (\text{funktionsutfall}_i - \text{observerade}_i)^2} \quad (4.5)$$

återfinns i tabell 4.1.

	Exponentiell	Linjär	Logaritmisk	Potens
Fel i januari	0.5493	0.5095	0.2943	0.3205
Fel i februari	0.2247	0.2147	0.1749	0.1758
Summa	0.7740	0.7242	0.4692	0.4963

Tabell 4.1: Sammanställning av funktionernas fel som absolutbelopp jämfört med faktiska utfall.

Den logaritmiska funktionen och potensfunktionen är de med minst fel och blir därmed utgångspunkt för fortsatt analys.

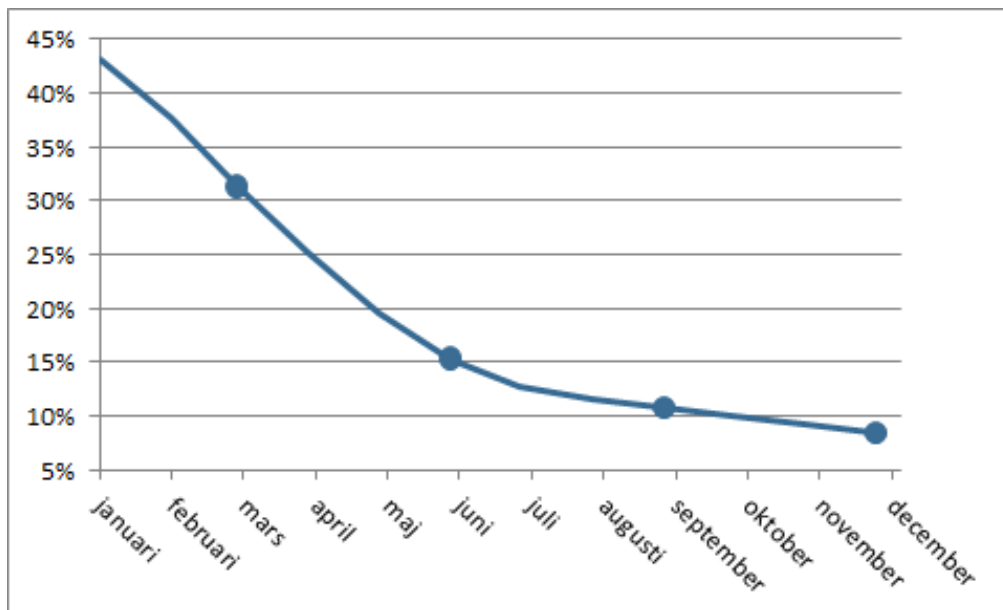
En annan metod är att använda splines som är teoretiskt beskrivet i avsnitt 2.1.1. Figur 4.3 illustrerar utfallet av en splinefunktion som knyter samman de fyra punkterna ackumulerat känt för kvartalen som motsvaras av punkterna 3, 6, 9 och 12 i figuren.

Splinefunktionen tas fram med andel ackumulerat okänt för mars, juni, september och december och extrapoleras sedan till alla övriga månader. Felen definierade enligt ovan för Splinesfunktionen presenteras i tabell 4.2 tillsammans med felen för den logaritmiska fördelningen som är bäst så här långt. Den logaritmiska för-

	Logaritmisk	Kubisk Spline
Fel i januari	0.2943	0.4884
Fel i februari	0.1749	0.2022
Summa	0.4692	0.6906

Tabell 4.2: Jämförelse summa fel mellan logaritmisk funktion och kubisk spline.

delningen ger ett mindre totalfel och är lättare att implementera vilket tyder på att den är att föredra.



Figur 4.3: Exempel på ackumulerat okänt som ges av $1 - S$ där splinesfunktionen S används

Vid en jämförelse med Trygg Hansas nuvarande metod ser vi i tabell 4.3 att de summerade felen blir mindre för vår nya funktion.

	Logaritmisk	Nuvarande metod
Fel i januari	0.2943	0.5493
Fel i februari	0.1749	0.2249
Summa	0.4692	0.7742

Tabell 4.3: Jämförelse summa fel mellan logaritmisk funktion och gammal metod.

Vi ställer oss frågan hur vår bästa funktion tenderar att förhålla sig till andel ackumulerat okänt för januari och februari som tagits fram ur månadstrianglarna. Om funktionen systematiskt underskattar andelen för januari och februari har enligt tabell 4.3 en förbättring skett men fortfarande finns ett stort fel i undervärdering. Resultatet av denna koll är att för den logaritmiska fördelningen i januari månad ligger fyra av åtta utfall under verkliga utfallet. Samma fördelning ger för februari att tre av åtta månader ligger över det verkliga utfallet. I jämförelse med om den gamla funktionen använts så hade alla januariutfall systematiskt underskattats under åttaårsperioden och för februari hade två av åtta överskattats och resten underskattats.

Då kurvanpassningen måste vara relativt simpel stannar vi vid funktion $g(z) = A + B \ln(z)$. Som en vidare diskussion kan man tänka sig att vid ett införande av data på månadsbasis gör att avvikelsen mellan månader sinsemellan blir större än om de summeras till kvartal då skillnaderna i någon mån kan ta ut varandra på sikt, vilket vi också såg i data. Den minsta andelen okänt baserad på månadstrianglarna är 0.4995 för januari och 0.3910 för februari och den största är 0.6530 för januari och 0.4629 för februari. Skillnaderna är som ses väldigt stora vilket gör en funktionsanpassning svår. En enda större skada i januari eller februari kan göra så att andelen okänt fullständigt kommer ur fas vad den borde vara enligt tidigare.

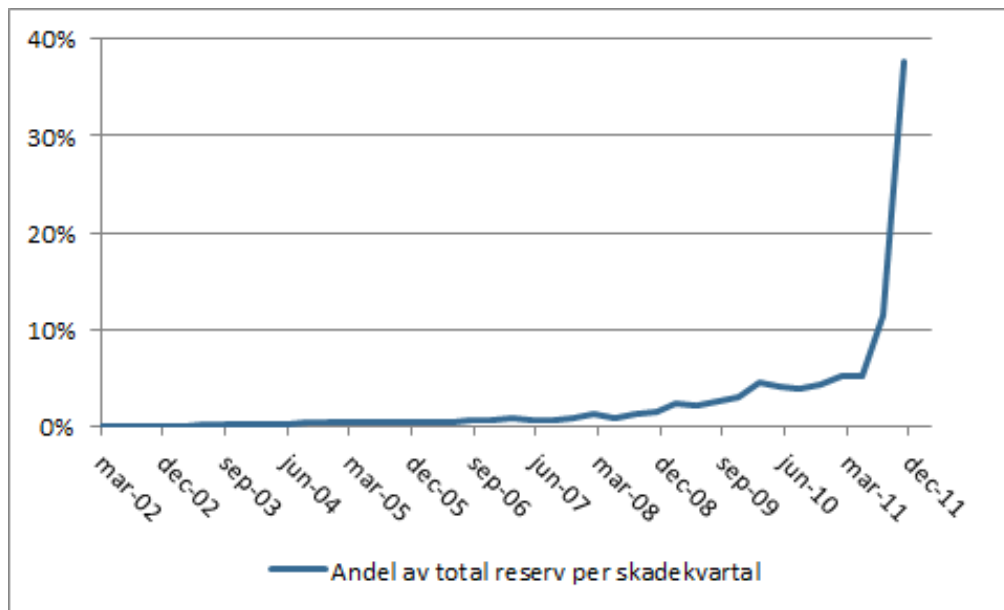
Slutsatsen är därmed att vi kommit fram till en funktion vars fel beräknat genom ett absolutbelopp mellan funktionens skattning och verkligt utfall mindre än vad som blivit med den gamla funktionen vilket måste ses som ett bra resultat. Observera dock att antal verkliga utfall enbart är åtta stycken som går att använda. Denna analys bör därmed upprepas varje år för att få ett säkrare resultat.

4.2 Konfidensintervall för reserver

I avsnitt 2.2.1 diskuterades kort att Chain Ladder-metoden att algoritmiskt fylla i en trekant är matematiskt tveksam. Oberoendeantagandet mellan rader och kolumner är möjligtvis okej mellan år (givet ett konstant flöde där inga större avtalsändringar etc. sker), men när intervallen blir snävare blir också detta antagande mer och mer diskutabelt, jämför år, kvartal och månader. Därför bör man inte blint lita på Chain Ladders väntevärde utan också titta på spridningsmått, här standardavvikelsen för ultimoskattningen som i och med att vi känner vissa betalningsströmmar blir samma sak som standardavvikelsen för reserven.

Avsnittet belyser en metod som belyser hur ett konfidensintervall beräknas för en skadepperiod, här skadekvartal. I och med att andelen ackumulerat okänt beräknas direkt ur reserven ligger det i vårt intresse att se hur procentsatsen förändras om den skattade reserven ligger långt i svansen av konfidensintervallet.

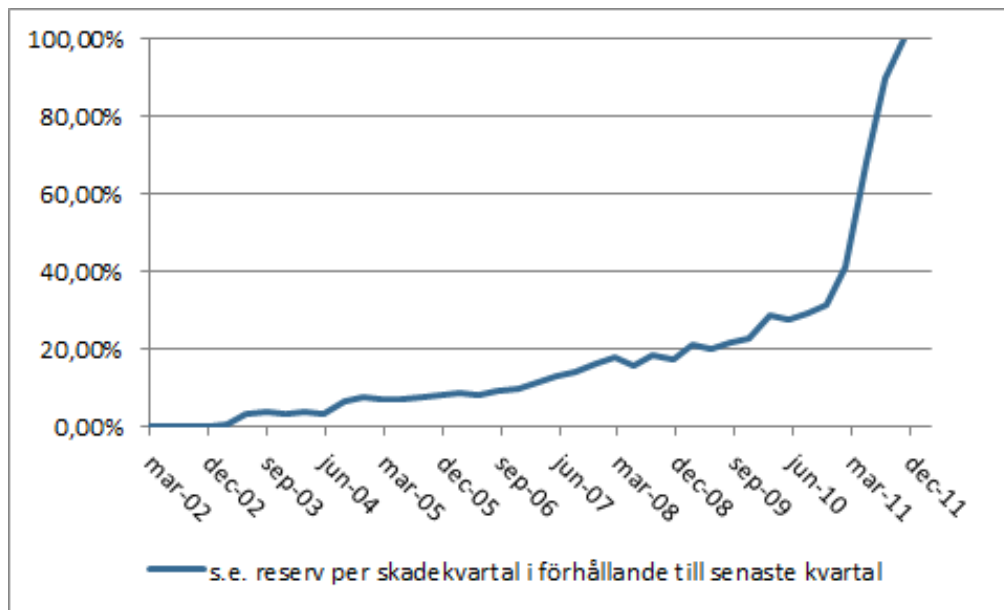
Vi börjar med att beräkna reserven för varje skadekvartal från första kvartalet 2002 till sista kvartalet 2011 med vanlig Chain Ladder per fjärde kvartalet 2011, figur 4.4. Av hänsyn till Trygg Hansa följer vi beloppen och fokuserar på andel av total reserv.



Figur 4.4: Fördelning avsatta reserver beräknade per sista kvartalet 2011. Längst till vänster är reserver för första kvartalet skadeår 2002 i förhållande till total reserv och längst till höger reserver för sista kvartalet skadeår 2011 i förhållande till total reserv. Den absolut största delmängden av totala reserven ligger för sena skadekvartal

För gamla skadekvartal är reserven förhållandevis liten då de stora beloppen redan är utbetalade. Den absolut största delmängden av reserven ligger för nya skadekvartal, vilket naturligtvis är helt logiskt.

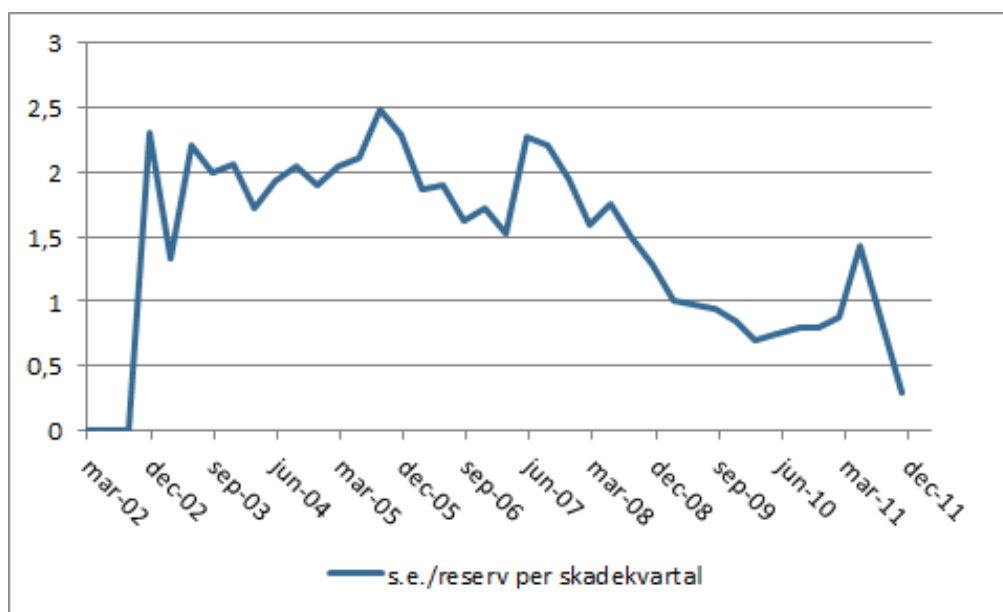
När ett väntevärde har beräknats är vi också intresserade av variansmått. Hur stora avsättningar behövs till exempel göras för att täcka 90% av utfallen som kommer ske? Andelen ackumulerat okänt, samt extrapoleringen till januari och februari, berörs om utbetalningarna ligger nära väntevärdet eller långt ut i svansen. Standardfel för reserver per skadeperiod kan enligt Mack [5] beräknas med formel (2.15) och standardfelen för skadekvartal för klassen Hemförsäkring ges i figur 4.5.



Figur 4.5: Fördelning standardfel för skattningarna av reserverna beräknade per sista kvartalet 2011 i förhållande till standardfelet i pengar för kvartal fyra 2011. Längst till vänster är standardfelet i pengar för första kvartalet skadeår 2002 och längst till höger standardfelet i pengar för sista kvartalet skadeår 2011. Standardfelen tenderar att öka för sena skadekvartal.

Hemförsäkringar är relativt snabbt reglerade och därmed ligger de absolut största standardfelen för skadekvartal nära där vi befinner oss nu. Framförallt de två sista skadekvartalen har större standardfel för reserven än övriga, vilket är föga förvånande då också den största delmängden av totala reserven ligger i de kvartalen.

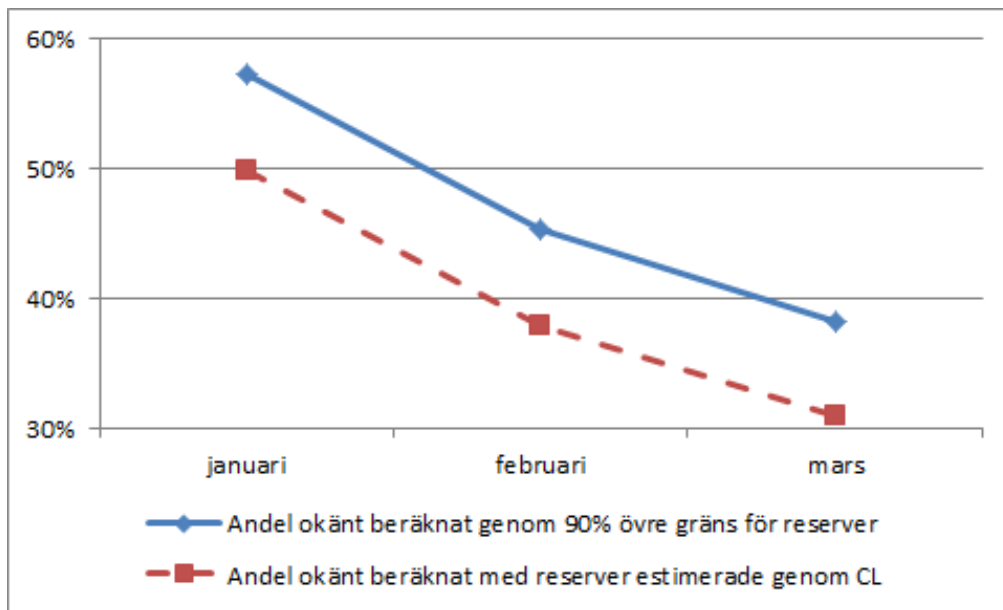
Intressant är också kvoten standardfel dividerat med reserven för de olika perioderna, figur 4.6. Kvoten tenderar att avta ju närmare nuvarande kvartal vi befinner oss. Vi får dock inte glömma att trots att den relativa kvoten är liten för sena skadekvartal ligger den absolut största risken i belopp i just de sena skadekvartalen.



Figur 4.6: Fördelning standardfel/reserv beräknade per sista kvartalet 2011. Längst till vänster är standardfelet/reserven första kvartalet skadeår 2002 och längst till höger standardfelet/reserven för sista kvartalet skadeår 2011. Här tenderar standardfelet/reserven krympa för sena skadekvartal. Vilket beror på de stora reserverna för sena kvartal.

Vi bygger ett konfidensintervall för reserverna för skadekvartalen, med hjälp av ett fördelningsantagande att reserverna är lognormalfördelade för, i syfte att se hur det slår på extrapoleringen av andel ackumulerat okänt. Teorin återfinns i avsnitt 2.2.2. Då vi tyvärr har tillgång till få punkter är det svårt att göra ett vettigt test av modellantaganden. Vi vill här upplysa om en metod som fungerar att tillämpa på vårt datamaterial och är medvetna om att ett hundra procentigt säkert resultat av antaganden inte har givits.

I figur 4.5 beräknades standardfelen för skadekvartal ett år 2002 till skadekvartal fyra år 2011 (med data per senaste kvartal) enligt Macks metod, som vi nu ska använda för att bygga ett uppåt begränsat konfidensintervall på 90%-nivån för reserverna för samma kvartal. Med hjälp av lognormalfördelningsantagandet och extrapoleringen med funktionen $g(z) = A + B \ln(z)$ skapas graferna i figur 4.7 som en uppskattning av andel ackumulerat okänt i januari och februari 2012.



Figur 4.7: Skillnader i andel ackumulerat okänt för januari, februari och mars 2012 om reserven är skattad direkt ur Chain Ladder eller om den når taket på det 90%-iga konfidensintervallet.

Extrapoleringen leder till att en uppskattning av andel okänt i januari och februari 2012 nästan är tio procent högre vid den övre konfidensgränsen än vid väntevärdet för reserven. Ingen jättstor effekt kan tyckas vilket kan ha sin förklaring i att standardavvikelsen för de sena kvartalen är små i förhållande till reserven, $Q_{2011,4}^{2011,3}$ ändras därmed inte så drastiskt vilket gör att extrapolationen inte slår så hårt. För skadekvartal tre är den övre gränsen på konfidensintervallet för reserven ungefär 75% större än den skattade reserven medan samma siffra för kvartal fyra bara är cirka 10%. Vid en större kvot standardavvikelse/reserv för sistakvartalet skulle skillnaden öka mellan graferna i figur 4.7.

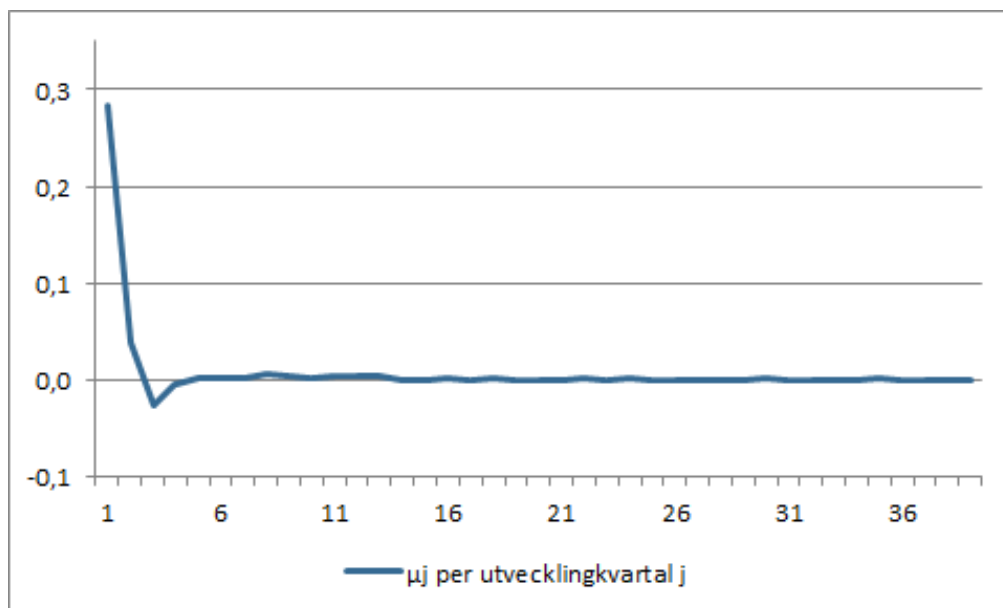
4.3 Parametrisk Bootstrap

Vi ska nu tillämpa en lognormal modell för ackumulerade betalningsströmmar, se modellantaganden i avsnitt 2.2.2. Utgångspunkten är att utvecklingsfaktorerna $F_{i,j} = C_{i,j}/C_{i,j-1}$ är lognormalfördelade.

Modellantaganden testas med hjälp av Kolmogorov-Smirnovs test för normalitet [4, kap. 13], där varje utvecklingsfaktor logaritmeras. För $\eta_{i,j} = \ln(F_{i,j})$ där $j = 1, \dots, 5$ är p -värdet mindre än 0.05 vilket indikerar att på att nollhypotesen

att observationerna kommer från en normalfördelning kan förkastas på 5%-nivån. För övriga j däremot är p -värdet större än 0.05 vilket då innebär att vi inte kan förkasta att dessa kommer från normalfördelningar på 5%-nivån.

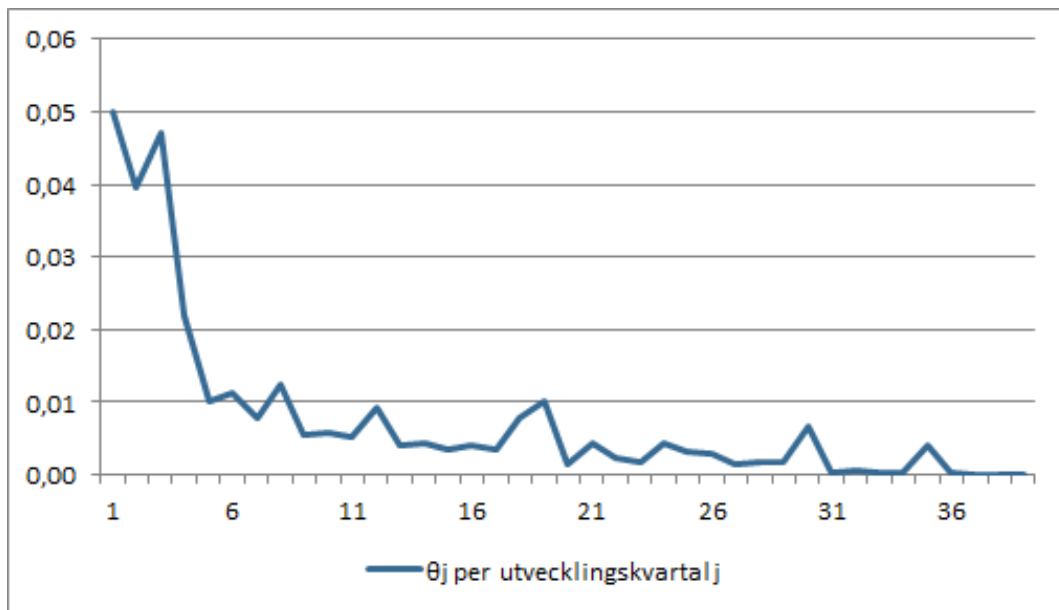
Eftersom modellantaganden verkar approximativt uppfyllda, åtminstone går flertalet av fördelningarna inte att förkasta att så är fallet, börjar vi att skatta μ_j vilka återfinns i figur 4.8. Att utvecklingskvartal tre är något negativt kan bero



Figur 4.8: Skattade $\hat{\mu}_j$ för utvecklingskvartal 1 till 40 beräknat per skadeår 2011, kvartal 4.

på att Trygg Hansa ser över sina försäkringar efter ett halvår och återbetalningar kan då förekomma. En variant är att bortse från det värdet och interpolera mellan kvartal två och fyra. Vi väljer dock att inte inkräkta på modellen med argument att de tidiga utvecklingsfaktorerna inte slår på så många skadeår och att vi då skillnaderna troligtvis blir förhållandevis små oavsett.

Parametern θ_j presenteras i figur 4.9 och visar att standardavvikelsen minskar drastiskt efter första året vilket är väntat då klassen är snabbreglerad.

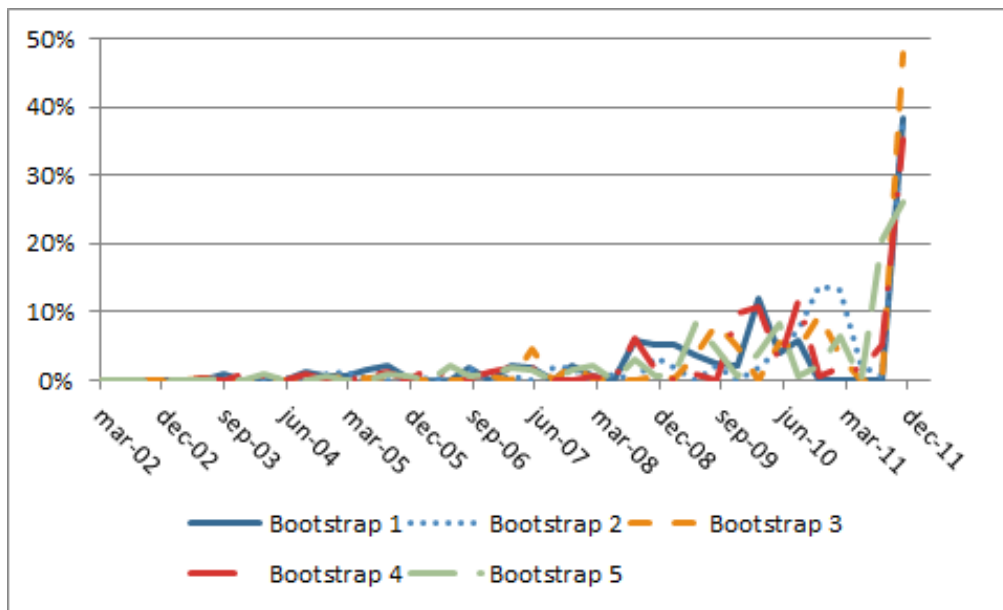


Figur 4.9: Skattade $\hat{\theta}_j$ för utvecklingskvartal 1 till 40 beräknat per skadeår 2011, kvartal 4.

Eftersom intresse finns för hur bootstrappingen beter sig undersökts fem återsamplingar närmre för att se hur reserver, andel ackumulerat okänt och extrapolation beter sig. Bootstrappingen sker genom att slumpstal dras från lognormalfördelningen genom den s.k. inversmetoden som utförs genom tre steg.

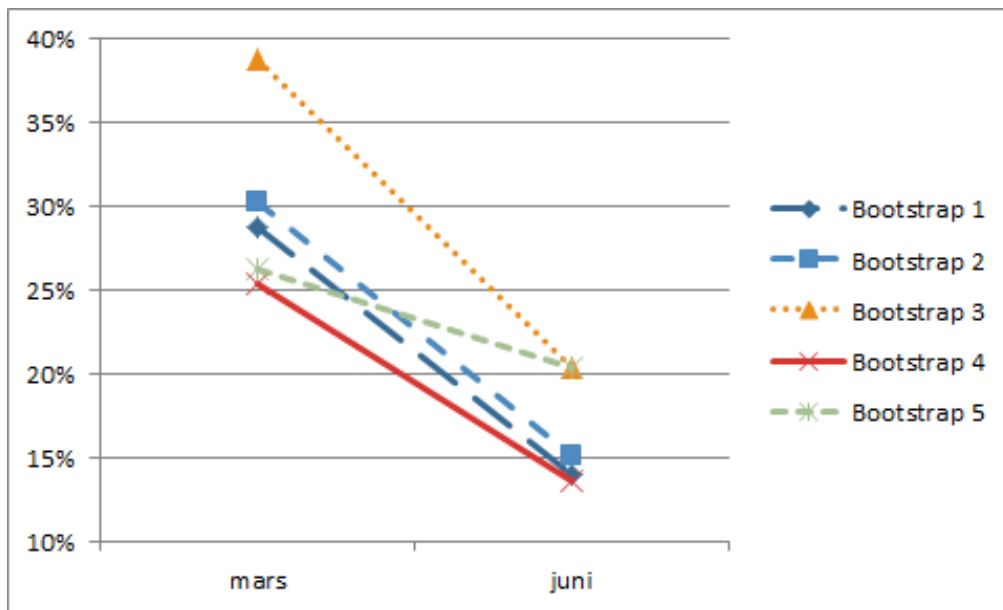
1. Låt $X \in F$ vara vår fördelning.
2. Låt $u \in U(0, 1)$.
3. Beräkna det minsta x så att $P(X \leq x) \leq u$.

Fördelningen för de bootstrappede reserverna plottas i figur 4.10. Framförallt vill vi titta på hur stora skillnader det är för sena skadekvartal för de olika återsamplingarna. Figuren är lite otydlig men vi upptäcker vissa skillnader framförallt för det senare skadekvartalen. Högsta reserven för sista skadekvartalen är den som benämns bootstrap 3 vilken är betydligt högre för sena skadekvartalen än övriga. Skillnaderna är dock ganska naturliga då de stora flödena sker just under de sena skadekvartalen och spridningen blir därmed störst just där. En liten differens i början kan ge stora utslag i slutet för de multipelt uppräknade beloppen.

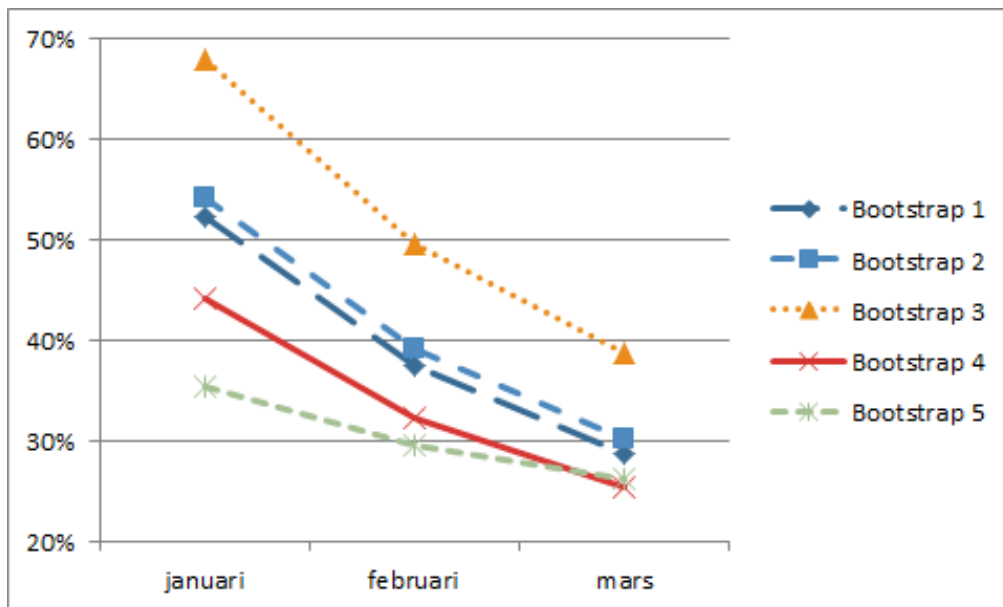


Figur 4.10: Fördelning bootstrapade reserver beräknade per sista kvartalet 2011. Längst till vänster är reserver för första kvartalet skadeår 2002 och längst till höger reserver för sista kvartalet skadeår 2011.

I figur 4.11 ser vi hur andelen ackumulerat okänt drabbas av skillnaden i återsamlingen. Baserat på de olika utfallen fås att andelen okänt för mars i spannet cirka 25% till 38% och för juni mellan cirka 13% till 21%. Skillnaden mellan återsamlingarna ger utslag också på extrapolationen. Figur 4.12 belyser att skillnaderna blir stora då funktionen $g(z) = A + B \ln(z)$ anpassas som estimat för andelen ackumulerat okänt januari och februari 2012. Minsta värdet på andelen ackumulerat okänt för januari månad är cirka 35% och största nästan 70%. Det här belyser en tveksamhet i metoden. Ett extremt utfall slår oerhört hårdt på framförallt extrapoleringen.



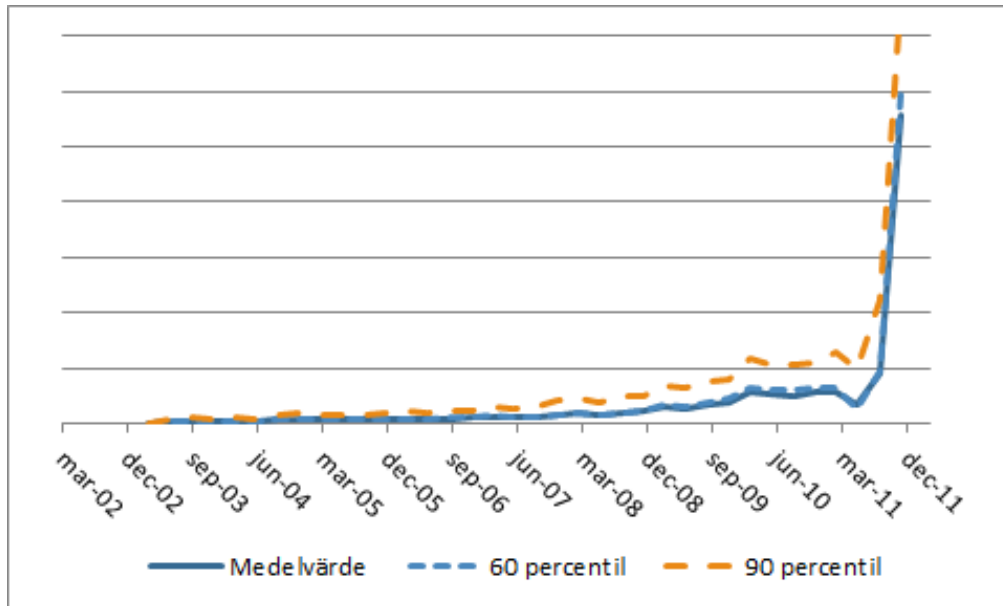
Figur 4.11: Andel ackumulerat okänt för mars och juni, 5 återsamplningar.



Figur 4.12: Andel ackumulerat okänt för januari till mars, 5 återsamplningar.

En annan tillämpning av bootstrap är att bygga vidare på avsnitt 4.2 och ta fram

en övre konfidensgräns för reserven för varje hittills inträffat år. Om reserven åter-samlas ett stort antal gånger, låt oss säga n stycken, så fås en form av övre gräns genom att hitta lämplig percentil. Valet av percentiler är beroende på tillämpning men vi ska här titta på 60-percentilen och 90-percentilen, det vill säga den första säger oss vilken summa som 60% av utfallen är mindre än och för 90-percentilen gäller 90%. Resultatet av 5 000 återsamlingar för varje skadeår plottas i figur 4.13. Vi ser att 60-percentilen nästan ligger exakt med medelvärdet medan 90-



Figur 4.13: Bootstrapade percentilnivåer för reserverna för skadekvartal ett skadeår 2002 längst till vänster och skadekvartal fyra skadeår 2011 längst till höger, beräknade på helår 2011 års data. Vi döljer beloppen med hänsyn till Trygg Hansa och fokuserar på skalor.

percentilen såklart ligger på en högre nivå, figuren tyder att det ibland är fråga om dubbla belopp mot väntevärdet. Dock kommer den stora ökningen precis som innan för de senaste kvartalen. Den gula streckade linjen kan sägas vara en övre konfidensgräns på 90%-nivån.

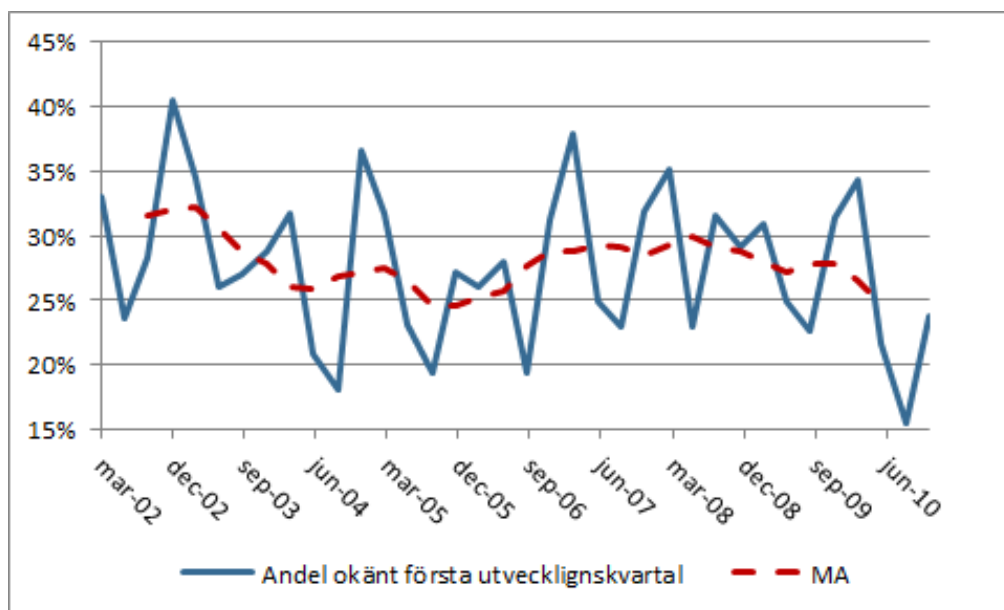
Återsamlingen används också som komplement till ren reservsättning. Vid n stycken återsamlingar kan en täthetsfunktion och en fördelningsfunktion byggas med hjälp av utfallen som delas upp i intervall likt ett vanligt histogram. Om vi dessutom låter n vara stort ges en bra bild om vart reserven ungefär bör ligga. Då klassen Hemförsäkring återsamlas 5 000 gånger får vi en fördelning vars medianvärde stämmer bra med vad som givits om t.ex. Chain Ladder används.

4.4 Tidsserieanalys

Trender i andel ackumulerat okänt ska nu undersökas för första utvecklingskvartalet genom att titta på hur dessa sett ut från första skadekvartalet 2002 till sista skadekvartalet 2010 genom tidsserieanalys. Tidsserien börjar med vad som inte var inkommet i mars 2002 per mars 2002 i förhållande till slutgiltig skadekostnad i mars 2002, den fortsätter sedan med vad som inte var inkommet i juni 2002 per juni 2002 i förhållande till slutgiltig skadekostnad för juni 2002. Tidsserien slutar med vad som inte var inkommet i december 2010 per december 2010 i förhållande till slutgiltig skadekostnad för december 2010. Det är ett successivt rullande av utvecklingskvartal ett.

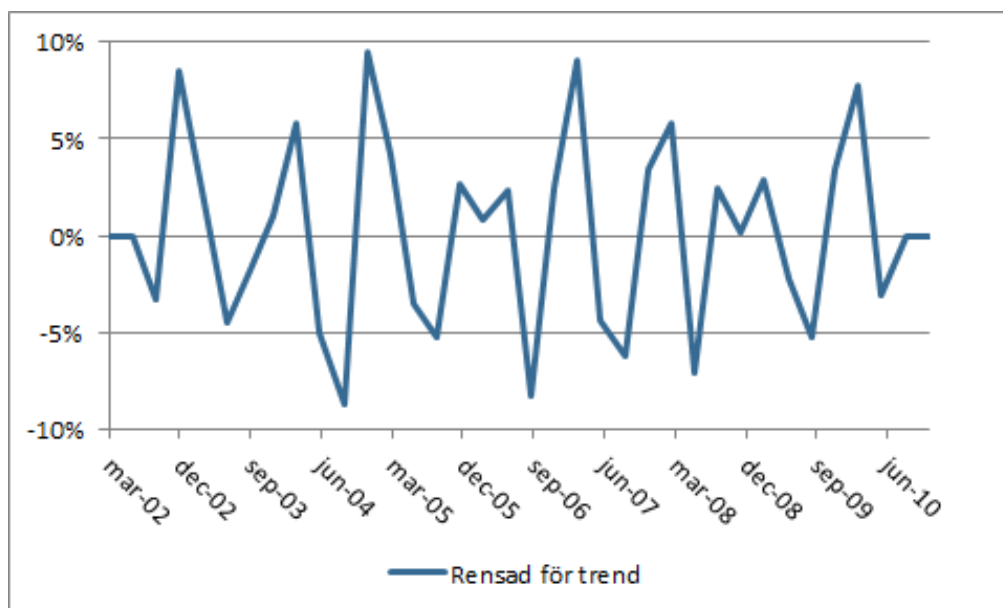
Om det går att visa att säsongskomponenterna för andel okänt är stora för kvartal ett, två, tre och fyra bör en viss försiktighetsåtgärd åtminstone reflekteras över i och med Trygg Hansas metod att kommande kvartal representeras av det senast inkomna.

Vi följer stegen i teoriavsnittet 2.3.1 och börjar med att ta fram trenden som beräknas med hjälp av glidande medelvärde. Figur 4.14 visar det glidande medelvärdet tillsammans med tidsserien i sin helhet.



Figur 4.14: Glidande medelvärde av andel okänt för utvecklingskvartal ett.

Serien rensas nu för trend och plottas i figur 4.15.



Figur 4.15: Tidsserie rensad för trend.

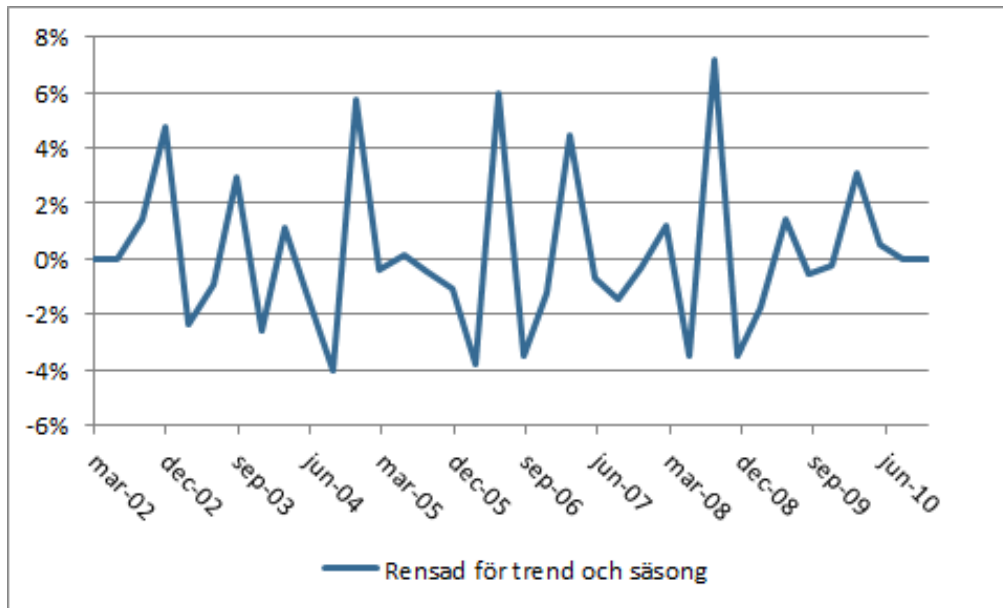
I och med hur Trygg Hansas jobbar med procentsatser okänt ligger det i vårt intresse att undersöka om andel ackumulerat okänt för senast inkommit har säsongseffekter. Säsongskomponenterna för andel okänt presenteras i tabell 4.4.

	Säsongskomponent
Kvartal 1	0.0463
Kvartal 2	-0.0362
Kvartal 3	-0.0472
Kvartal 4	0.0371

Tabell 4.4: Säsongskomponenter av andelen okänt för utvecklingskvartal ett.

Enligt säsongskomponenterna fungerar andel okänt för kvartal fyra som en relativt bra representation för kvartal ett. Naturligtvis kan lokala avvikelser förekomma men på lång sikt (givet de relativt få datapunkterna som finns att arbeta med här) verkar metoden vara rimlig.

Slutligen rensas serien både för trend och säsongskomponenter, figur 4.16.



Figur 4.16: Rullande andel okänt rensad för trend och säsongsvariationer.

Antag att den rensade serien $Y_t = X_t - \hat{T}_t - \hat{S}_t$ kan skrivas på formen,

$$Y_t = c + \phi Y_{t-1} + \epsilon_t, \quad (4.6)$$

alltså som en autoregressiv process av ordning ett, en $AR(1)$ -process. Då serien är rensad för trend och säsongskomponenter testar Dickey-Fuller om processen har en drift eller ej oberoende av tidigare nämnda faktorer.

Dickey-Fuller kräver en avancerad fördelning så vi valde därför att göra testet i Matlab med funktionen `adftest`. Resultatet blev att andel ackumulerat okänt för senast inkommit kvartal förkastas nollhypotesen att $\phi = 1$ på 5%-nivån och serien är därmed stationär och alltså inte en ren slumpvandring vilket indikerar att trend och säsong finns.

Samma metodik kan såklart göras för övriga utvecklingskvartal andel ackumulerat okänt. Vi nöjer oss dock med en tidsserie för utvecklingskvartal ett då metoden framgår tydligt samt att nämna andel är den som slår hårdast på extrapoleringen. Vid en utökad analys bör man fundera på hur de transformationer av data som görs påverkar data i form av eventuellt utökad spridning för den rensade modellen etc.

5 Slutsatser och diskussion

Huvudsyftet med uppsatsen var att hitta lämpliga funktioner för hur stor ackumulerat okänt-procent som ska bokföras vid januari och februari månad. I och med att Trygg Hansa använder kvartalsdata finns inga observationer förrän i mars varje år. Utgångspunkten för skattningarna är därför att använda sig av de två föregående kvartalsskattningarna. För att få en bild av vad som är rimligt skrevs ett SAS-program som gjorde om kvartalsdata till månadsdata vilket var ett ganska omfattande arbete.

Idag använder Trygg Hansa en inter/extrapolationsmetod som har vissa brister. Vi har undersökt olika alternativa funktioner och kommit fram till att en funktion på formen $g(z) = A + B \ln(z)$ gav mindre kvadratiska fel än funktionen som används idag. Felen tenderar inte heller att konsekvent underskatta verkligt utfall såsom var fallet förut. Tyvärr fanns bara tillgång till åtta punktskattningar vilket naturligtvis gör det svårt att göra säkerställd statistik. Utifrån vad som fanns förbättrades dock metoden avsevärt.

Vidare har vi undersökt och beräknat spridning för reserver med mera. Andel ackumulerat okänt baseras på den skattade ultimoreserven för de olika skadekvartalen vilket gör att en uppskattning av denne måste göras. Avsnitt 4.2 visar en metod för att bygga konfidensintervall för reserverna och hur det påverkar andelen okänt. Framförallt extrapolation är känslig för outliers, ett extremt utfall kan slå hårt på funktionen som anpassas. Modellantaganden är svåra att kolla vilket stör en aning, men med vad som fanns att tillgå belyser kapitlet en metod att få bättre kännedom om spridningsmått.

En annan variant att få en uppfattning om andelen okänt är att bygga en tidsserie. I Avsnitt 4.4 tittar vi på den skattade andelen okänt för utvecklingskvartal ett över tid. Trend och säsong rensas och med hjälp av Dickey-Fuller-testet visar på att serien verkar ha en konvergens runt noll vilket indikerar på att en trend och säsong finns. Avsnittet antyder också att kvartal fyra är en okej approximation av kvartal ett vilket Trygg Hansas metod för andel okänt bygger på.

I och med ökad datakraft går bootstrap idag att göra bättre med fler simuleringar än förr. Avsnitt 4.3 visar exempel på återsampling med ett normalfördelningsantagande, vilken styrks av Kolmogorov-Smirnov-test, och hur resultaten blir för andelen okänt. Ännu ett exempel som visar hur relativt små skillnader i kvartalsskattningar gör att extrapolationen för funktion $g(z) = A + B \ln(z)$ ändras. Tveksamheter till metoden kan därmed eventuellt diskuteras vid outliers.

Referenser

- [1] Brockwell, P. & Davis, R. (2003): *Introduction To Time Series and Forecasting*. Springer-Verlag New York Inc.
- [2] Johansson, B. & Ohlsson, E. (2010): *Non-life Insurance Pricing With Generalized Linear Models*. Springer-Verlag Berlin and Heidelberg GmbH & Co. K.
- [3] Johnston, J. & DiNardo, J. (1997): *Econometric Methods*. McGraw Hill Higher Education
- [4] Lindgren, B.W. (1993): *Statistical Theory*. Chapman & Hall/CRC.
- [5] Mack, T. (1993): *Measuring The Variability Of Chain Ladder Reserve Estimates*. CAS Prize paper competition.
- [6] Wuthrich, M.V. & Merz, M. (2008): *Stochastic Claims Reserving Methods In Insurance*. John Wiley & Sons Ltd.

A Härledning av parameterskattningar

För att anpassa kurvorna i sektion 4.1 ska ekvationssystemet

$$\begin{cases} y_1 &= f(z_1) \\ y_2 &= f(z_2) \end{cases} \quad (\text{A.1})$$

lösas för vart av funktionerna.

Vi börjar med den linjära varianten. 2 okända parametrar A och B ska skattas med hjälp av 2 utfall. Ekvationssystemet blir därmed

$$\begin{cases} y_1 &= A + Bz_1 \\ y_2 &= A + Bz_2 \end{cases} \quad (\text{A.2})$$

Den övre ekvationen i systemet A.2 kan skrivas om till $A = y_1 - Bz_1$ som sättes in i den undre ekvationen som då kan formuleras till

$$\begin{aligned} y_2 &= y_1 - Bz_1 + Bz_2 = y_1 - B(z_1 - z_2) \\ \Leftrightarrow B(z_1 - z_2) &= y_1 - y_2 \\ \Leftrightarrow B &= \frac{y_1 - y_2}{z_1 - z_2}. \end{aligned}$$

B kan nu sättas in i valfri ekvation, t.ex. den övre, som gör att vi får A genom följande.

$$\begin{aligned} y_1 &= A + \frac{y_1 - y_2}{z_1 - z_2} \cdot z_1 \\ \Leftrightarrow A &= y_1 - \frac{z_1(y_1 - y_2)}{z_1 - z_2} = \frac{y_1(z_1 - z_2) - z_1(y_1 - y_2)}{z_1 - z_2} \\ \Leftrightarrow A &= \frac{z_1y_2 - z_2y_1}{z_1 - z_2} \end{aligned}$$

För den logaritmiska funktionen är tillvägagångssättet precis likadant så den utelämnas. Parametrarna för exponentialfördelningen fås genom att lösa systemet

$$\begin{cases} y_1 &= Ae^{Bz_1} \\ y_2 &= Ae^{Bz_2} \end{cases} \quad (\text{A.3})$$

Genom att logaritmera A.3 så erhålls det nya ekvationssystemet

$$\begin{cases} \ln y_1 &= \ln A + Bz_1 \\ \ln y_2 &= \ln A + Bz_2 \end{cases} \quad (\text{A.4})$$

där den övre ekvationen kan skrivas $\ln A = \ln y_i - Bz_1$ som insatt i den undre ekvationen ger parametern B .

$$\begin{aligned}\ln y_2 &= \ln y_1 - Bz_1 + Bz_2 = \ln y_1 - B(z_1 - z_2) \\ \Leftrightarrow B &= \frac{\ln y_1 - \ln y_2}{z_1 - z_2}\end{aligned}$$

A löses sedan genom att sätta in B i valfri ekvation.

Till sist potensfunktionen vars parametrar ges genom att lösa systemet

$$\begin{cases} y_1 &= Az_1^B \\ y_2 &= Az_2^B \end{cases} \quad (\text{A.5})$$

och som ovan logaritmerar vi bägge ekvationerna

$$\begin{cases} \ln y_1 &= \ln A + B \ln z_1 \\ \ln y_2 &= \ln A + B \ln z_2 \end{cases} \quad (\text{A.6})$$

vilket ger ett linjärt ekvationssystem som vi sett förslag på lösningsmetod ovan.