



Universiteit  
Leiden  
The Netherlands

## **Kijken in een glazen bol. Voorspellen van IBNR schade**

Meer, M.B.N. van der

### **Citation**

Meer, M. B. N. van der. (2006). *Kijken in een glazen bol.: Voorspellen van IBNR schade.*

Version: Not Applicable (or Unknown)

License: [License to inclusion and publication of a Bachelor or Master thesis in the Leiden University Student Repository](#)

Downloaded from: <https://hdl.handle.net/1887/3597546>

**Note:** To cite this publication please use the final published version (if applicable).

Manja van der Meer

# Kijken in een glazen bol Voorspellen van IBNR schade

Doctoraalscriptie, verdedigd op 4 augustus 2006

Afstudeerdocent: Prof. Dr. L.C.M. Kallenberg

Afstudeerbegeleider: Dr. E.W. van Zwet

Begeleider bij Delta Lloyd: Drs. N.A.A. Marquinie AAG

**delta lloyd**



Mathematisch Instituut, Universiteit Leiden

# Inhoudsopgave

<b>Inleiding</b>	<b>4</b>
<b>Hoofdstuk 1 De zorgverzekeraar</b>	<b>6</b>
1.1 Wat is IBNR	6
1.2 Factoren die de schadelast beïnvloeden	6
1.3 Schadedriehoeken	6
1.4 Onderverdeling in de verstrekkingen	9
1.5 Wat zijn DBC's	10
1.6 Nieuwe basisverzekering	11
1.7 IBNR methoden	11
1.7.1 Voorbeeld	12
<b>Hoofdstuk 2 De Chain Ladder methode</b>	<b>13</b>
2.1 Klassieke Chain Ladder	13
2.2 Chain Ladder en maximum likelihood	14
2.3 Gegeneraliseerde Lineaire Modellen	20
2.3.1 Chain Ladder en GLM	23
2.4 Chain Ladder en overdispersie	23
2.5 Chain Ladder en de Normale verdeling	28
2.6 Expectation Maximization methode	29
<b>Hoofdstuk 3 Foutmarge</b>	<b>32</b>
3.1 Methode van Mack	32
3.2 Bootstrap	33
3.2.1 Parametrische bootstrap	33
3.2.2 Niet parametrische bootstrap	35
<b>Hoofdstuk 4 Loglineaire modellen</b>	<b>36</b>
4.1 Schatters	36
4.2 IBNR Voorziening	37
4.3 Bootstrap	38
<b>Hoofdstuk 5 Alternatieve methoden</b>	<b>40</b>
5.1 Variant I	41
5.2 Accident Year Incurred Loss Development methode	43
5.3 Loss Ratio methode	43
5.4 Bornhüter Ferguson methode	44
5.5 Aritmetische separatie methode	46

<b>Hoofdstuk 6</b>	<b>Het meest geschikte model</b>	<b>50</b>
6.1	Modelkeuze . . . . .	50
6.2	Forward Stagewise Regression . . . . .	50
6.3	Cross Validation . . . . .	54
6.4	F-toets . . . . .	55
<b>Hoofdstuk 7</b>	<b>Resultaten</b>	<b>56</b>
7.1	Alle data . . . . .	56
7.2	Chain Ladder schatter . . . . .	57
7.2.1	Chain Ladder met Variant I . . . . .	59
7.3	Loglineair model en FSR schatter . . . . .	59
7.4	Per Verstrekking . . . . .	61
7.4.1	Specialistische hulp . . . . .	61
7.4.2	Ziekenhuisverpleging . . . . .	63
7.4.3	CHA tapes . . . . .	65
7.4.4	Overige verstrekkingen . . . . .	67
7.5	Samenvattend . . . . .	69
<b>Hoofdstuk 8</b>	<b>Samenvatting</b>	<b>76</b>
<b>Appendix A</b>		<b>78</b>
<b>Appendix B</b>		<b>80</b>
<b>Appendix C</b>		<b>81</b>
<b>Appendix D</b>		<b>82</b>
<b>Literatuur</b>		<b>83</b>

## Inleiding

De scriptie die voor u ligt, is een product van een wiskunde opleiding aan de Universiteit Leiden en een afstudeerstage bij Delta Lloyd Zorg. In de stage heeft onderzoek plaats gevonden naar diverse modellen om schattingen te maken voor kosten in de toekomst. Hiervoor zijn diverse statistische methoden gebruikt. Als eerste zal de achtergrond en de context worden geschetst waarin het onderzoek heeft plaatsgevonden. Vervolgens worden verwachtingen en foutmarges van diverse modellen van verschillende kanten bekeken. Tot slot zal gezien laten worden wat het meest geschikte model is en wat de resultaten zijn wanneer het model wordt toegepast.

In de wereld van verzekeraars worden ieder (kalender)jaar schattingen gemaakt van de verwachte schadelast. De verwachte schadelast bestaat als eerste uit claims van oudere jaren die nog binnen moeten komen, als tweede uit de verwachte schadelast van het huidige jaar en als derde uit de verwachte kosten van het aankomende jaar. Op deze schatters worden de voorzieningen en premiestelling gebaseerd. Een voorziening is het bedrag dat de verzekeraar als actieve middelen beschikbaar houdt om toekomstige schadeclaims uit te betalen. Het is dus van belang om zowel naar het verleden, het heden, als naar de toekomst te kijken. Het wordt steeds belangrijker om de verwachte schadelast goed in te schatten. Hiervoor zijn verschillende redenen te noemen.

Om te beginnen hangt de hoogte van de *premie* die de klanten betalen samen met de verwachte schadeclaims. Omdat de hoogte van de premie aan het begin van ieder kalenderjaar wordt vastgesteld, is het belangrijk om aan het begin van ieder kalenderjaar een goed beeld te hebben van de verwachte schadeclaims. Een verkeerde schatting van de hoogte van de totale schadeclaims levert een verkeerde premie op. Dit kan enerzijds tot gevolg hebben dat de premie te laag is en de verzekeraar verlies maakt en anderzijds dat de premie te hoog is en er verzekeringen worden opgezegd.

Ten tweede baseert het management de *voorzieningen* en de nieuwe *ontwikkelingsplannen* op schatters van de verwachte schadelast. Als de verwachte schadelast en daarmee de voorzieningen bijvoorbeeld te hoog worden ingeschat, dan zit je goed qua betrouwbaarheid om aan de toekomstige verplichtingen te kunnen voldoen. Deze hoge schadelast gaat ten koste van het deel van het vermogen dat gebruikt wordt voor de langdurige beleggingen. Bij een lage inschatting gaat de verzekeraar achteruit op het onderdeel betrouwbaarheid. Daarnaast zou het zelfs kunnen gebeuren dat de verzekeraar niet aan zijn verplichtingen kan voldoen en niet uit kan betalen. Daarentegen zijn er hierbij meer mogelijkheden om vermogen langdurig te investeren en is er meer rendement te behalen. Het is dus erg belangrijk om een goed

evenwicht te bepalen.

Als derde bestaat er *regelgeving* zoals IFRS voor beursgenoteerde bedrijven, waardoor beursgenoteerde verzekeringsmaatschappijen gebonden zijn aan stabiele marges. Marges worden hierdoor beperkt. Omdat Delta Lloyd een dochter is van de beursgenoteerde Engelse verzekeringsmaatschappij Aviva, heeft ze met IFRS te maken.

Het management wil een beslissing van de hoogte van de voorziening kunnen funderen op een degelijke schatting met een percentage van de mate van zekerheid. Dus naast een getal voor de schatter van de voorzieningen, willen ze ook weten dat deze schatter bijvoorbeeld met 90% zekerheid aan de verplichtingen kan voldoen. Wanneer een schatter van bijvoorbeeld 300 miljoen voor 90% adequaat is, betekent dit dat  $P(R \leq 300 \cdot 10^6) = 0,9$ , waarbij  $R$  de benodigde voorziening. Hoe hoger de aangehouden voorzieningen, hoe hoger het percentage zal worden.

De Delta Lloyd Groep is een verzekeringsmaatschappij welke bestaat uit Delta Lloyd (DL), OHRA en ABN AMRO Verzekeringen (AAV). Een onderdeel hiervan is Delta Lloyd Zorg (DLZ), welke verantwoordelijk is voor de zorgverzekeringen van Delta Lloyd. Voor de afdeling Actuarieel binnen DLZ is een van de belangrijkste taken het maken van schadeprognoses. Voor de schadeprognose wordt gebruik gemaakt van een wiskundig model. Binnen het Actuarieel is de vraag ontstaan naar een nieuw model dat een schatter oplevert voor de verwachte schadeclaims. Naast een schatter moet een nieuw model ook een foutmarge of variantie opleveren, zodat er ook een percentage van de mate van zekerheid bepaald kan worden. Door naast het huidige model een nieuw model te gebruiken, wordt de objectiviteit van de schatters groter.

In de afgelopen decennia hebben diverse wiskundigen zich over dit probleem gebogen en zijn er verschillende modellen ontwikkeld. In deze scriptie zullen een aantal modellen met de bijbehorende aannames worden uitgelicht en worden er enkele voors en tegens genoemd. Vervolgens worden enkele modellen wiskundig gezien aan elkaar gekoppeld. Hierdoor zal er iets meer duidelijk worden over de verdeling en variantie. Ten slotte zal er een praktisch model worden opgebouwd, dat binnen DLZ ingezet kan worden om geschikte schadeprognoses te maken. Hiervoor zal een computertool ontwikkeld worden.

# 1 De zorgverzekeraar

## 1.1 Wat is IBNR

Een (schade)verzekeraar is geïnteresseerd in de schadeontwikkeling per risicoperiode. Meestal wordt hierbij naar een schadejaar gekeken (alternatieven zijn kwartalen en maanden). De premie die de verzekerde betaalt, wordt toegerekend aan het jaar waarvoor de polisdekking geldt. Zo wordt ook de schade toegerekend aan dit jaar, ongeacht het tijdstip waarop de schade uiteindelijk bij de verzekeraar geclaimd wordt. Schade die na afloop van een kalenderjaar geclaimd is, wordt IBNR-schade genoemd, IBNR staat voor 'Incurred But Not Reported'. Bij een zorgverzekering worden schadeclaims voor het grootste deel in het eerste en tweede jaar afgehandeld. Een klein deel wordt nog tot drie à vier jaar na dato geclaimd. De verzekeraar krijgt schade dan vertraagd in beeld, maar moet dit wel voorzien. De schadevoorziening IBNR is een belangrijk onderdeel van de balans van een (schade)verzekeraar.

## 1.2 Factoren die de schadelast beïnvloeden

Er zijn veel verschillende factoren die de hoogte van de verwachte schadelast kunnen beïnvloeden. Als eerste is er de verzekerde zelf. De verwachte kosten die een persoon zal maken, hangen onder andere samen met zijn of haar gezondheidstoestand. In het ziekenfonds wordt de gezondheidstoestand gekoppeld aan persoonsgebonden factoren: leeftijd, geslacht, regio, rechtsgrond (aard van inkomen) en ziektebeeld. Het ziekenfonds werkt met een budgetteringssysteem. Dit houdt in dat de verzekeraar voor haar verzekerden budget van de overheid krijgt, waarbij de hoogte van het budget samen valt met de gezondheidstoestand van de verzekerden. Omdat niet alle factoren bekend zijn bij de verzekeraar, is het budget per verzekerde lastig in te schatten.

Ten tweede heeft het aantal verzekerden per jaar invloed op de hoogte van de totale schade. Daarnaast zijn er nog diverse andere zaken die de totale schadelast beïnvloeden, zoals (gewone) inflatie, zorginflatie, jurisprudentie, verandering in wetgeving en portefeuille, werkprocedures, claimgedrag.

## 1.3 Schadedriehoeken

Om een voorspelling te kunnen maken van verwachte kosten in de toekomst, wordt er gebruik gemaakt van data uit eerdere jaren. Historische data is bijvoorbeeld te gebruiken om trends te bepalen die zich, naar verwacht, ook in de toekomst zullen voortzetten. Trends kunnen zich voordoen in een bepaald schadejaar, afwikkeljaar of kalenderjaar.

Een *schadejaar* is het jaar waarin de schade zich heeft voortgedaan. We kunnen bijvoorbeeld schadejaar 2001 bekijken, hiertoe behoren alle kosten die samenhangen met schade die in 2001 gemaakt is.

Een *afwikkeljaar* is het 1<sup>e</sup>, 2<sup>e</sup>, 3<sup>e</sup> enz. jaar waarin schadeclaims kunnen worden ingediend nadat ze hebben plaatsgevonden. Er wordt bijvoorbeeld in totaal 100 euro schade gemaakt in het schadejaar 2001. Er kan dan bijvoorbeeld 75 euro worden geclaimd in het eerste afwikkeljaar, dat is het jaar 2001. Vervolgens zou er 24 euro in het tweede afwikkeljaar geclaimd kunnen worden, dat is dan het jaar 2002. De laatste euro wordt nog in het derde (2003) en vierde (2004) afwikkeljaar geclaimd.

Er kunnen ook claims bekeken worden die in een bepaald *kalenderjaar* worden ingediend. Als we kalenderjaar 2003 bekijken, dan worden er in dit jaar claims ingediend van schade die gemaakt is in het eerste afwikkeljaar van 2003. Daarnaast kunnen er claims ingediend worden over schade gemaakt in 2002, dit is dan het tweede afwikkeljaar van 2002. Verder kunnen er claims ingediend worden over schade gemaakt in 2001, dit is dan het derde afwikkeljaar.

Het totaal aan incrementele claims wordt aangeduid met  $C_{ij}$  waarbij  $i$  staat voor het schadejaar en  $j$  voor het afwikkeljaar,  $i + j - 1$  is het kalenderjaar.

Om historische data overzichtelijk weer te geven, wordt de data in een schadedriehoek gezet, de schadedriehoek wordt ook wel afwikkeldriehoek genoemd, zie figuur 1.1. Hierin zijn horizontaal de schadejaren af te lezen, verticaal de afwikkeljaren en diagonaal de kalenderjaren.

schadejaar	afwikkeljaar						
	1	2	...	j	...	t - 1	t
1	$C_{11}$	$C_{12}$	...	$C_{1j}$	...	$C_{1, t-1}$	$C_{1,t}$
2	$C_{21}$	$C_{22}$	...	$C_{2j}$	...	$C_{2, t-1}$	
⋮	⋮	⋮		⋮			
$i$	$C_{i1}$	$C_{i2}$	...	$C_{ij}$			
⋮	⋮	⋮					
$t - 1$	$C_{t-1,1}$	$C_{t-1,2}$					
$t$	$C_{t,1}$						

Figuur 1.1

De index van de laatst bekende data wordt aangeduid met  $i, t + 1 - i$ . Dit correspondeert met de onderste diagonaal in figuur 1.1.

Normaal gesproken hebben we het over de incrementele claims ( $C_{ij}$ ). Regelmatig wordt er in modellen ook gekeken naar cumulatieve claims, deze wor-



den aangeduid met  $D_{ij}$ . Er geldt

$$D_{ij} = \sum_{k=1}^j C_{ik} \quad \text{met } D_{i,1} = C_{i,1}$$

Binnen de verzekeringen geldt er voor de zorgsector dat na 3 à 4 jaar nage-  
noeg alle claims zijn ingediend door de verzekerden. De ervaring in de  
zorgsector tot en met het jaar 2004 is dat grofweg driekwart van de kosten  
wordt geclaimd in het eerste jaar, bijvoorbeeld het jaar 2000, dit is dus het  
eerste afwikkeljaar. In het tweede afwikkeljaar wordt er iets minder dan een  
kwart geclaimd. Grofweg 1% van de totaal gemaakte schade in 2000 wordt  
in het derde afwikkeljaar, in dit geval 2002, ingediend. Nog een klein deel,  
minder dan 0,2%, wordt in het vierde en latere afwikkeljaren ingediend. Zie  
figuur 1.2 voor een voorbeeld, hierin staan incrementele schades.

schadejaar	afwikkeljaar				totaal
	1	2	3	4	
1 (2000)	1000	350	15	1	1366
2 (2001)	970	360	17	.	.
3 (2002)	1010	344	.	.	.
4 (2003)	986	.	.	.	.

Figuur 1.2

Uit bovenstaande schadedriehoek kun je verschillende informatie halen. Bij-  
voorbeeld dat er in schadejaar 2000 (horizontaal) in totaal 1366 (1000+350+  
15 + 1) is geclaimd en dat er in kalenderjaar 2003 (diagonaal) in totaal 1348  
(986+344+17+1) is geclaimd. Een derde mogelijk te onderscheiden richting  
is die van de afwikkeljaren (verticaal). In figuur 1.2 is bijvoorbeeld te zien  
dat er gemiddeld 991,5 ( $\frac{1000+970+1010+986}{4}$ ) in het eerste afwikkeljaar wordt  
ingediend.

Uiteindelijk is het de bedoeling om bovenstaande driehoek vol te schatten.  
De som van de schatters:  $\hat{C}_{4,2}$ ,  $\hat{C}_{3,3}$ ,  $\hat{C}_{4,3}$ ,  $\hat{C}_{2,4}$ ,  $\hat{C}_{3,4}$  en  $\hat{C}_{4,4}$ , is dan de ver-  
wachting van de IBNR-voorzieningen die aangehouden moeten worden aan  
het begin van 2004.

In veel gevallen is er meer historische data bekend dan er in figuur 1.2  
genoemd is. Wanneer er meer data bekend is, is er waarschijnlijk meer  
te zeggen over een trend in bijvoorbeeld de verschillende schadejaren of  
afwikkeljaren. We kunnen er dan voor kiezen om een model te maken dat  
gebaseerd is op de grotere hoeveelheid data. De data is dan als volgt weer  
te geven.

schadejaar	afwikkeljaar				totaal
	1	2	3	4	
1 (1997)	972	298	10	2	1282
2 (1998)	964	316	13	1	1294
3 (1999)	980	338	16	2	1336
4 (2000)	1000	350	15	1	1366
5 (2001)	970	360	17	.	.
6 (2002)	1010	344	.	.	.
7 (2003)	986	.	.	.	.

Figuur 1.3

Voor het ziekenfonds is het acceptabel om aan te nemen dat claims binnen vier jaar zijn ingediend. Echter in andere branches, zoals Leven en Schade, kan de doorlooptijd vele malen langer zijn. Een voorbeeld hiervan is aansprakelijkheid, waarbij soms veel tijd nodig is om erachter te komen of, en zo ja op welk bedrag een persoon recht heeft. Om toch een duidelijk overzicht te kunnen geven, zonder bijvoorbeeld 30 jaren te hoeven weergeven, worden alleen de jaren weergegeven waarin het grootste deel van de claims zijn ingediend. Dit kan oplopen tot bijvoorbeeld tien jaar. De claims die in de jaren daarna nog worden ingediend vallen onder de categorie 'staart', waarin dus meerdere jaren zijn verwerkt. Het schatten van de staart is een studie op zich, dit is voor de zorg niet van belang omdat de staart erg klein is.

#### 1.4 Onderverdeling in de verstrekkingen

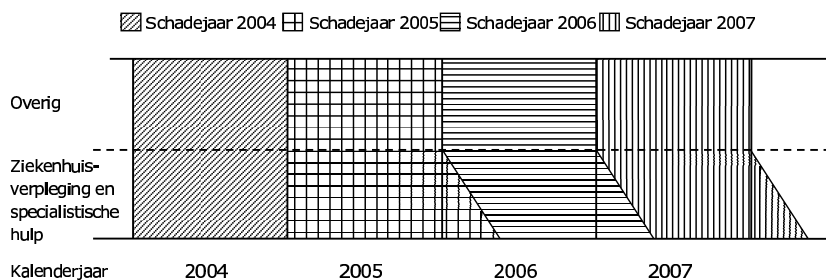
De zorgverzekeraar maakt in de zorg onderscheid tussen verschillende verstrekkingen. Dit maakt de kosten overzichtelijker, omdat er per verstrekking een eigen afwikkelpatroon is. Er kan dan bij het schatten van toekomstige claims gebruik gemaakt worden van trends in de verschillende verstrekkingen. Voordat er geschat gaat worden, wordt de schadedriehoek altijd gecontroleerd op incidentele afwijkingen. Omdat deze afwijkingen meestal zijn terug te leiden tot een bepaalde verstrekking, kan men met correctie van de incidentele afwijkingen op verstrekkingenniveau een duidelijkere trend zichtbaar maken.

Daarnaast zijn er voor het ziekenfonds landelijke regels aangaande nacalculatie en verevening, welke zorgen voor spreiding van het risico over alle zorgverzekeraars aan het einde van het boekjaar. Hierdoor worden enerzijds resultaten herverdeeld over verzekeraars en anderzijds komen risico's een deel voor rekening van de centrale overheid. Ook hierbij is het essentieel dat er een onderverdeling wordt gemaakt in de diverse types verstrekkingen.

Tot en met 2004 bestonden de verstrekkingen als eerste uit de *Ziekenhuisverpleging*, hierbij wordt onderscheid gemaakt tussen vaste en variabele kosten. Daarnaast is er de *Specialistische hulp* en als laatste zijn er de *Overige verstrekkingen*. Claims met deze onderverdeling lopen uiterlijk nog door tot en met 2007, dan zijn vier afwikkeljaren van 2004 achter de rug. De data die onderzocht zal worden heeft bovenstaande structuur. Hierdoor zullen we voor de wiskundige modellen, die gemaakt gaan worden om schatters te genereren, gebruik maken van deze onderverdeling.

### 1.5 Wat zijn DBC's

Per 1 januari 2005 is landelijk een nieuwe methode van declareren ingevoerd voor de rubrieken Ziekenhuisverpleging en Specialistische hulp: de Diagnose Behandel Combinaties (DBC's). Door de invoering van DBC's ontstaat er met ingang van schadejaar 2005 een structurele verandering in de financiering van de Zorg. Een aantal samenhangende, losse verrichtingen worden, binnen de categorie Ziekenhuisverpleging en Specialistische hulp, samengevoegd tot één DBC. De verrichtingen worden niet meer afzonderlijk, maar pas na het beëindigen van het volledige DBC-traject afgerekend. De verzekeraar waar de verzekerde bij aanvang van de DBC actief verzekerd was, krijgt het volledige bedrag te betalen, ongeacht het feit of de verzekerde bijvoorbeeld een dag na aanvang van het DBC-traject vertrokken is naar een andere verzekeraar. Dit laatste speelt vaak bij de overgang naar een nieuw kalenderjaar en zeker bij de start van de basisverzekering. Door de invoering van DBC's is er een verschuiving van betalingen ontstaan vanaf het schadejaar 2005. In figuur 1.4 is te zien dat tot en met 2004 de zorgkosten die gemaakt zijn in 2004 ook in dit jaar worden geboekt. Vanaf 2005 is te zien dat een deel van de in 2006 gemaakte kosten worden toegeschreven aan het jaar 2005. Er vindt dus een verschuiving plaats, hierdoor zijn schadeclaims uit 2005 eenmalig hoger. Vanaf 2006 is het weer stabiel maar verschoven ten opzichte van eerdere jaren.



Figuur 1.4

## 1.6 Nieuwe basisverzekering

Bij de invoering van de nieuwe basisverzekering in januari 2006, is het onderscheid tussen het ziekenfonds en een particuliere verzekering weggefallen. Hierdoor is er een grote verandering in verzekerden ontstaan, er zijn verzekerden weggegaan en er zijn nieuwe verzekerden bijgekomen. Bekende trends over het gedrag van de verzekerden tot 2006 zijn nu niet meer van toepassing. Daarnaast zijn er veranderingen gekomen in de types dekkingen. Door de verschuiving van vergoeding voor medische hulp voor de ex-particuliere verzekerden, hebben we ook te maken met een zogenaamde zorginflatie. Dat wil zeggen de verschuiving van kosten doordat medische hulp niet meer of juist wel wordt vergoed. Daardoor is het medio 2006 nog lastig om goede voorspellingen te doen over de verwachte schadeclaims.

De invoering van DBC's en de nieuwe basisverzekering zijn de redenen dat we data en trends van de jaren 1999 tot en met 2003 van het ziekenfonds bekijken. Hiervan is goede data bekend en is te toetsen of schatters acceptabel zijn. De reden dat we het ziekenfonds bekijken, is omdat de opbouw hiervan het meeste lijkt op de opbouw van de nieuwe basisverzekering.

## 1.7 IBNR methoden

Een IBNR methode is een methode die een schatting maakt van de te verwachte schade, ofwel de 'Incurred But Not Reported'-schade. Het gaat om schaden die al gemaakt zijn, maar waarvan de claims nog worden verwacht. Met een IBNR methode wordt dus de schadedriehoek uit figuur 1.1 volgeschat. In de loop der jaren zijn diverse methoden ontwikkeld om de schadedriehoek vol te schatten. Omdat het geen éénduidig proces is, blijft ieder model subjectief.

$R_i$  is de IBNR voorziening voor jaar  $i$  en is gelijk aan het totale schadebedrag van schadejaar  $i$  min de reeds geboekte claims van dat schadejaar. Uitgedrukt in incrementele schade geldt:

$$\hat{R}_i = \sum_{j=1}^t \hat{C}_{i,j} - \sum_{j=1}^{t+1-i} C_{i,j} = \sum_{j=t+2-i}^t \hat{C}_{i,j}$$

Uitgedrukt in cumulatieve schade geldt:

$$\hat{R}_i = \hat{D}_{i,t} - D_{i,t+1-i}$$

Hieronder zullen enkele veel gebruikte methoden worden beschrijven. Hierbij wordt onderscheid gemaakt tussen de volgende drie methoden:

Als eerste zijn er methoden die gebaseerd zijn op de verwachting van de

incrementele (of cumulatieve) schade, waarbij een *lineair model* gemaakt wordt. Het meest gebruikte model van dit type is de Chain Ladder. In hoofdstuk 2 wordt de Chain Ladder uitgebreid besproken. Aanvullend worden in Hoofdstuk 3 enkele methoden toegelicht om de foutmarge te bepalen.

Ten tweede zijn er methoden die gebaseerd zijn op de verwachting van de incrementele (of cumulatieve) schade, waarbij een *loglineair model* gemaakt wordt. Deze modellen zullen in hoofdstuk 4 worden besproken.

In de loop der jaren zijn er ook veel varianten van lineaire modellen ontwikkeld. Op deze varianten wordt in hoofdstuk 5 dieper ingegaan.

### 1.7.1 Voorbeeld

Bij de verschillende methoden zal iedere keer een getallenvoorbeeld worden gegeven ter illustratie. Vanwege de concurrentiegevoeligheid van de data, maken we gebruik van onderstaand voorbeeld of van fictieve data. Het voorbeeld staat in onderstaande schadedriehoek.

schadejaar	afwikkeljaar			
	1	2	3	4
1	60	25	10	5
2	63	24	11	.
3	58	23	.	.
4	65	.	.	.

Figuur 1.5

In de verschillende voorbeelden wordt er gebruik gemaakt van diverse stochastische grootheden. Hieronder wordt een overzicht geven van de variabelen die bij alle modellen terug komen. Variabelen die specifiek zijn voor een bepaalde methode, zullen ook alleen bij die methode beschreven worden.

$C_{i,j}$  = Incrementele schadeclaims in schadejaar  $i$  en afwikkeljaar  $j$ ,  
deze zijn bekend voor  $i = 1, \dots, t$  en  $j = 1, \dots, t + 1 - j$ .

$D_{i,j}$  = Cumulatieve schadeclaims in schadejaar  $i$  en afwikkeljaar  $j$ ,  
deze zijn bekend voor  $i = 1, \dots, t$  en  $j = 1, \dots, t + 1 - j$ .

$D_{i,t}$  = Totale schade voor schadejaar  $i$ .

$D_{i,t+1-i}$  = Totale bekende schade voor schadejaar  $i$ .

$t$  = Totale tijd waarin alle claims zijn ingediend.

Tenzij anders vermeld, wordt gebruik gemaakt van de aanname dat de  $C_{ij}$  onderling onafhankelijk zijn.

## 2 De Chain Ladder methode

### 2.1 Klassieke Chain Ladder

De Chain Ladder methode is een veel gebruikte methode om voorzieningen te schatten en maakt gebruik van trends in de richting van het afwikkeljaar en het schadejaar. Er wordt gebruik gemaakt van de aanname dat er een vaste verhouding bestaat tussen de afwikkelkolommen van de schadedriehoek. Voor het bepalen van de verhouding wordt gebruik gemaakt van de cumulatieve schade, zodat ook in de laatste afwikkeljaren er volume overblijft om de stabiliteit zoveel mogelijk te behouden. De verhoudingen tussen kolom  $j - 1$  en kolom  $j$  wordt de proportionaliteitsfactor  $f_j$  genoemd.

#### Schatters

$f_j$  = proportionaliteitsfactor tussen afwikkeljaar  $j - 1$  en  $j$

De basis aanname voor de Chain Ladder methode is dat er proportionaliteitsfactoren  $f_j$  voor  $j = 2, \dots, t$  bestaan zodanig dat:

$$E(D_{i,j} | D_{i,1}, \dots, D_{i,j-1}) = D_{i,j-1} f_j, \quad i = 1, \dots, t \quad j = 2, \dots, t$$

Bij de Chain Ladder is het gebruikelijk om de volgende schatter te gebruiken voor  $f_j$ :

$$\hat{f}_j = \frac{\sum_{i=1}^{t+1-j} D_{i,j}}{\sum_{i=1}^{t+1-j} D_{i,j-1}} = \frac{\sum_{i=1}^{t+1-j} \sum_{k=1}^j C_{i,k}}{\sum_{i=1}^{t+1-j} \sum_{k=1}^{j-1} C_{i,k}}$$

Wanneer er bovenop de bekende data uit de schadedriehoek nog  $m$  extra schadejaren bekend zijn, zoals bijvoorbeeld in figuur 1.3 waarbij  $m = 3$ , dan kan de proportionaliteitsfactor ook over deze extra  $m$  jaren worden berekend. Een schatter van  $f_j$  wordt dan:

$$\hat{f}_j = \frac{\sum_{i=1}^{t+1-j+m} D_{i,j}}{\sum_{i=1}^{t+1-j+m} D_{i,j-1}}$$

In Mack (1993) wordt de zuiverheid en de ongecorrleerdheid van de schatter  $\hat{f}_j$  aangetoond. Ook wanneer er  $m$  extra jaren aan de schadedriehoek worden toegevoegd, blijft het bewijs van Mack voor de zuiverheid en ongecorrleerdheid van toepassing. De verwachting van de totale schade uit schadejaar  $i$  is:

$$E(D_{i,t} | D_{i,j}, i + j \leq t + 1) = D_{i,t+1-i} \cdot \prod_{k=t+2-i}^t f_k$$

#### IBNR Voorziening

Hiermee hebben we de voorziening:

$$\begin{aligned} \text{IBNR voorziening : } \hat{R}_i &= \hat{D}_{it} - D_{i,t+1-i} \\ &= \left( \prod_{j=t+2-i}^t \hat{f}_j - 1 \right) D_{i,t+1-i} \end{aligned} \quad (1)$$

De Chain Ladder is een veel gebruikte methode omdat deze goed te begrijpen, toe te passen en uit te leggen is. Een nadeel is dat de Chain Ladder methode gebruik maakt van veel parameters. Doordat er in de zorgsector gebruik wordt gemaakt van maar 3 à 4 ontwikkeljaren, is er relatief weinig data ten opzichte van het aantal parameters, waardoor overparameterisatie kan ontstaan.

### Voorbeeld

Met de formule voor de schatter van  $\hat{f}_j$  en het voorbeeld uit figuur 1.5, krijgen we de volgende waarden:

afwikkeljaar	2	3	4
$\hat{f}_j$	$\frac{85+87+81}{60+63+58} \approx 1,40$	$\frac{95+98}{85+87} \approx 1,12$	$\frac{100}{95} \approx 1,05$

We kunnen nu de driehoek volschatten met de formule

$$\hat{D}_{i,j} = D_{i,t+1-i} \prod_{k=t+2-i}^j \hat{f}_k$$

schadejaar	afwikkeljaar			
	1	2	3	4
1	60	85	95	100
2	63	87	98	<b>103,2</b>
3	58	81	<b>90,9</b>	<b>95,7</b>
4	65	<b>90,9</b>	<b>101,9</b>	<b>107,3</b>

Figuur 2.1

Ofwel de totale voorziening wordt:

$$\hat{R} = (103,2 - 98) + (95,7 - 81) + (107,3 - 65) = 62,2.$$

## 2.2 Chain Ladder en maximum likelihood

De Chain Ladder is een van de meest gebruikte methoden voor het bepalen van voorzieningen. Vooral de voordelen, zoals de eenvoud en het snelle rekenwerk wegen hierbij zwaar. Nadelen zoals de gevoeligheid voor variatie in de data, kunnen tegemoet worden gekomen door de data eerst met de hand te corrigeren op incidentele afwijkingen. Zelfs voor het niet onbelangrijke nadeel dat de methode niets zegt over de variantie, zijn diverse pogingen gedaan om toch een oplossing te vinden. Mack is daar een goed voorbeeld van, zelfs zonder verdeling kan de methode van Mack iets zeggen over de standaard afwijking, dit zal in 3.1 aan de orde komen. Ook andere wiskundigen hebben modellen ontworpen welke dezelfde verwachting opleveren als de Chain Ladder en die ook iets zeggen over de variantie. Hieronder zal ik de methode van Hachemeister en Stanard (1975) beschrijven, welke

een voorbeeld is van de maximum likelihood methode. De methode levert dezelfde verwachting op als de Chain Ladder methode.

### Schatters

Dit model werkt met incrementele betalingen in de betalingsdriehoek en maakt gebruik van de volgende aannames:

- (1)  $E(C_{ij}) = x_i y_j$  met onbekende parameters  $x_i$  en  $y_j$ , zodanig dat  $\sum_{j=1}^t y_j = 1$  en  $y_j \geq 0 \forall j$
- (2) Elke  $C_{ij}$  heeft de Poisson verdeling
- (3) Alle  $C_{ij}$  zijn onafhankelijk

Aan de hand van (1) kunnen we de verwachting van de voorzieningen als volgt uitdrukken:

$$\begin{aligned}
 E(R_i) &= E\left(\sum_{j=t+2-i}^t C_{i,j}\right) \\
 &= \sum_{j=t+2-i}^t E(C_{i,j}) \\
 &= x_i \sum_{j=t+2-i}^t y_j
 \end{aligned} \tag{2}$$

Ofwel de verwachte voorziening voor schadejaar  $i$  is totale verwachte schade in schadejaar  $i$  ( $x_i$ ) vermenigvuldigd met de fractie van de nog te verwachte kosten.

Om na te gaan of deze aannames leiden tot dezelfde verwachting als de Chain Ladder, beginnen we met het nagaan van de mle. Als eerste geldt:  $C_{ij} \sim \text{Poisson}(x_i y_j)$

$$P(C_{i,j}) = e^{-x_i y_j} \frac{(x_i y_j)^{C_{ij}}}{C_{ij}!}$$

Hiermee kunnen we de likelihood functie  $L$  opstellen:

$$\begin{aligned}
 L &= f(C_{ij}, 1 \leq j \leq t, 1 \leq i \leq t+1-j \mid x_i y_j) \\
 &= e^{-\sum_{j=1}^t \sum_{i=1}^{t+1-j} (x_i y_j)} \cdot \frac{\prod_{j=1}^t \prod_{i=1}^{t+1-j} (x_i y_j)^{C_{ij}}}{\prod_{j=1}^t \prod_{i=1}^{t+1-j} C_{ij}!}
 \end{aligned}$$

De log likelihood ( $l$ ) wordt daarmee de volgende:

$$l = -\sum_{j=1}^t \sum_{i=1}^{t+1-j} x_i y_j + \sum_{j=1}^t \sum_{i=1}^{t+1-j} C_{ij} \cdot \log(x_i y_j) -$$



$$\log\left(\prod_{j=1}^t \prod_{i=1}^{t+i-j} C_{ij}\right)$$

We stellen de verschillende afgeleiden gelijk aan 0:

$$\begin{aligned} \frac{\partial l}{\partial y_j} &= - \sum_{i=1}^{t+1-j} x_i + \sum_{i=1}^{t+1-j} \frac{C_{ij} x_i}{x_i y_j} = 0 \\ \text{ofwel} \\ \hat{y}_j &= \frac{\sum_{i=1}^{t+1-j} C_{ij}}{\sum_{i=1}^{t+1-j} \hat{x}_i} \end{aligned} \quad (3)$$

Omdat  $E(C_{ij}) = x_i y_j$  net als de schadedriehoek symmetrisch is, geldt bovengaande mle schatter van  $\hat{y}_j$  ook voor  $\hat{x}_i$ :

$$\hat{x}_i = \frac{\sum_{j=1}^{t+1-i} C_{ij}}{\sum_{j=1}^{t+1-i} \hat{y}_j} \quad (4)$$

Deze vergelijkingen zijn stapsgewijs op te lossen, waarbij gebruik wordt gemaakt van  $\sum_{j=1}^t \hat{y}_j = 1$ .

$$\begin{aligned} \hat{x}_1 &= \frac{\sum_{j=1}^t C_{1,j}}{\sum_{j=1}^t \hat{y}_j} = \sum_{j=1}^t C_{1,j} \\ \hat{y}_t &= \frac{C_{1,t}}{\hat{x}_1} \\ \hat{x}_2 &= \frac{\sum_{j=1}^{t-1} C_{2,j}}{\sum_{j=1}^{t-1} \hat{y}_j} = \frac{\sum_{j=1}^{t-1} C_{2,j}}{1 - \hat{y}_t} \\ \hat{y}_{t-1} &= \frac{\sum_{i=1}^2 C_{i,t-1}}{\sum_{i=1}^2 \hat{x}_i} \\ \hat{x}_3 &= \frac{\sum_{j=1}^{t-2} C_{3,j}}{\sum_{j=1}^{t-2} \hat{y}_j} = \frac{\sum_{j=1}^{t-2} C_{3,j}}{1 - \hat{y}_{t-1} - \hat{y}_t} \\ &\vdots \end{aligned}$$

En zo verder totdat alle  $x_i$  en  $y_j$  bekend zijn.

Als eerste zal worden aangetoond dat de schatter van de proportionaliteitsfactor  $\hat{f}_j$  zoals deze in het Chain Ladder model gedefinieerd is, overeen komt met de factor in de methode van Hachemeister en Stanard. Wanneer dit is aangetoond, komen hiermee automatisch de schatters voor de voorzieningen overeen.

$$\hat{f}_j = \frac{\sum_{i=1}^{t-j+1} \sum_{k=1}^j C_{i,k}}{\sum_{i=1}^{t-j+1} \sum_{k=1}^{j-1} C_{i,k}} \quad (\text{de Chain Ladder schatter})$$

De data  $C_{i,j}$  is te schatten met  $\widehat{x}_i \widehat{y}_j$ . Het is niet vanzelfsprekend toegestaan om de data te schatten met  $\widehat{x}_i \widehat{y}_j$ , dit geldt alleen voor speciale gevallen. Samen met de mle schatter uit (3) en (4) kunnen we zo een speciaal geval afleiden.

$$\sum_{i=1}^{t-j+1} \sum_{k=1}^j C_{i,k} = \sum_{i=1}^{t-j+1} \sum_{k=1}^j \widehat{x}_i \widehat{y}_k$$

Hiermee kunnen we een uitdrukking voor  $\widehat{f}_j$  afleiden:

$$\begin{aligned} \widehat{f}_j &= \frac{\sum_{i=1}^{t-j+1} \sum_{k=1}^j C_{i,k}}{\sum_{i=1}^{t-j+1} \sum_{k=1}^{j-1} C_{i,k}} \\ &= \frac{\sum_{i=1}^{t-j+1} \sum_{k=1}^j C_{i,k}}{\sum_{i=1}^{t-j+2} \sum_{k=1}^{j-1} C_{i,k} - \sum_{k=1}^{j-1} C_{t-j+2,k}} \\ &= \frac{\sum_{i=1}^{t-j+1} \widehat{x}_i \sum_{k=1}^j \widehat{y}_k}{\sum_{i=1}^{t-j+2} \widehat{x}_i \sum_{k=1}^{j-1} \widehat{y}_k - \sum_{k=1}^{j-1} C_{t-j+2,k}} \\ &= \frac{\sum_{i=1}^{t-j+1} \widehat{x}_i \sum_{k=1}^j \widehat{y}_k}{\sum_{i=1}^{t-j+2} \widehat{x}_i \sum_{k=1}^{j-1} \widehat{y}_k - \widehat{x}_{t-j+2} \sum_{k=1}^{j-1} \widehat{y}_k} \\ &= \frac{\sum_{k=1}^j \widehat{y}_k}{\sum_{k=1}^{j-1} \widehat{y}_k} \end{aligned} \quad (5)$$

Om te laten zien dat deze factor dezelfde IBNR voorziening oplevert (1), gaan we na of deze overeen komt met (2) als we de Hachemeister en Stanard schatter voor de proportionaliteitsfactor gebruiken.

$$\begin{aligned} (1): \widehat{R}_i &= \left( \prod_{j=t+2-i}^t \widehat{f}_j - 1 \right) D_{i,t+1-i} \\ &= \left( \prod_{j=t+2-i}^t \frac{\sum_{k=1}^j \widehat{y}_k}{\sum_{k=1}^{j-1} \widehat{y}_k} - 1 \right) \sum_{j=1}^{t+1-i} C_{i,j} \\ &= \left( \frac{\sum_{k=1}^{t+2-i} \widehat{y}_k}{\sum_{k=1}^{t+1-i} \widehat{y}_k} \cdot \frac{\sum_{k=1}^{t+3-i} \widehat{y}_k}{\sum_{k=1}^{t+2-i} \widehat{y}_k} \cdot \dots \cdot \frac{\sum_{k=1}^t \widehat{y}_k}{\sum_{k=1}^{t-1} \widehat{y}_k} - 1 \right) \sum_{j=1}^{t+1-i} C_{i,j} \\ &= \left( \frac{\sum_{k=1}^t \widehat{y}_k}{\sum_{k=1}^{t+1-i} \widehat{y}_k} - 1 \right) \sum_{j=1}^{t+1-i} C_{i,j} \\ &= \left( \frac{1}{\sum_{k=1}^{t+1-i} \widehat{y}_k} - 1 \right) \sum_{j=1}^{t+1-i} C_{i,j} \end{aligned}$$

We maken nu gebruik van (4) waarin  $\sum_{j=1}^{t+1-i} C_{i,j} = \widehat{x}_i \sum_{j=1}^{t+1-i} \widehat{y}_j$ .

$$\widehat{R}_i = \left( \frac{1}{\sum_{k=1}^{t+1-i} \widehat{y}_k} - 1 \right) \widehat{x}_i \sum_{k=1}^{t+1-i} \widehat{y}_k$$

$$= \hat{x}_i \left( 1 - \sum_{k=1}^{t+1-i} \hat{y}_k \right) \quad (6)$$

Wat precies de schatter is van de verwachting van de IBNR voorziening bij de methode van Hachemeister en Stanard uit (2).

Tot slot zullen we nog de zuiverheid van de schatter  $\hat{x}_i$  bekijken door na te gaan of  $E(\hat{x}_i) = x_i$ . Omdat de schatter van de IBNR voorziening is gegeven door  $\hat{R} = \sum_{i+j>t+1} \hat{x}_i \hat{y}_j = \sum_i \hat{x}_i - \sum_{i+j \leq t+1} \hat{x}_i \hat{y}_j = \sum_i x_i - \sum_{i+j \leq t+1} C_{i,j}$  en derhalve alleen van de schatter  $\hat{x}_i$  afhangt, is het voldoende om alleen de zuiverheid van  $\hat{x}_i$  na te gaan. Hierbij is  $\{C_{i,j}, i+j \leq t+1\}$  een collectie onderling onafhankelijke stochastische grootheden, zodat  $E(C_{i,j}) = x_i y_j$ .

Nu geldt voor  $k = 1$ :

$$E(\hat{x}_1) = E\left(\sum_{j=1}^t \hat{x}_1 \hat{y}_j\right) = \sum_{j=1}^t x_1 y_j = x_1$$

Voor  $k = 2, \dots, t$  hebben we:

$$\hat{x}_k = \sum_{j=1}^{t+1-k} C_{k,j} \prod_{j=t+2-k}^t \hat{f}_j$$

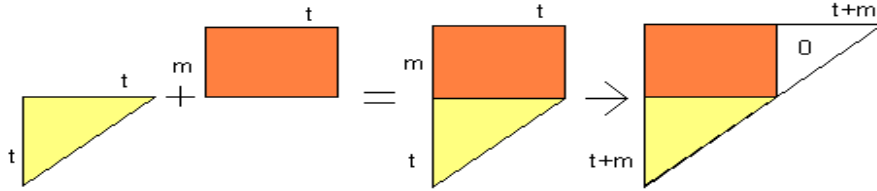
Merk nu op dat  $\sum_{j=1}^{t+1-k} C_{k,j}$  en  $\prod_{j=t+2-k}^t \hat{f}_j$  onafhankelijk zijn, omdat  $\prod_{j=t+2-k}^t \hat{f}_j$  een functie is van  $\{C_{i,j} : i+j \leq t+1, i \leq k-1\}$ . Er geldt dus

$$E(\hat{x}_k) = \left(x_k \sum_{j=1}^{t+1-k} y_j\right) E\left(\prod_{j=t+2-k}^t \hat{f}_j\right) = \frac{x_k}{\prod_{j=t+2-k}^t f_j} E\left(\prod_{j=t+2-k}^t \hat{f}_j\right)$$

Mack (1993) laat zien dat de schatters  $\hat{f}_j$  zuiver en ongecorrleerd zijn, dus geldt er de zuiverheid van de schatter:  $E(\hat{x}_k) = x_k$ .

Er is nog op te merken dat de hierboven beschreven methode voor het verkrijgen van schatters ook op gaat wanneer er  $m$  schadejaren meer bekend zijn dan alleen de data uit de schadedriehoek, zoals bijvoorbeeld in figuur 1.3. Doordat we ervan uitgaan dat  $C_{i,j} \sim Poisson(x_i y_j)$  en de Poisson verdeling de eigenschap heeft dat de som van Poisson verdeelde stochasten ook weer Poisson verdeeld zijn, geldt ook dat  $\sum_{j=1}^t C_{i,j} = x_i \sim Poisson(x_i)$ . We kunnen nu het eerste schadejaar in de schadedriehoek vervangen door de som van alle bekende schadejaren. De som van het eerste schadejaar had eerst de verdeling  $Poisson(x_1)$ , dit wordt nu vervangen door  $Poisson(\sum_{i=1}^{m+1} x_i)$ . Vervolgens kunnen we analoog verder gaan aan de hierboven beschreven methode.

In de praktijk is het gebruikelijk om toch te doen alsof we met een schadedriehoek werken, in plaats van met een schadedriehoek plus een blok extra data. In plaats van de schadedriehoek met  $t$  afwikkeljaren plus  $m$  extra shadejaren te beschouwen, kunnen we ook een schadedriehoek beschouwen met  $t + m$  afwikkeljaren en  $t + m$  shadejaren, zoals in figuur 2.2. Hierbij krijgen we extra  $C_{i,j}$  voor  $i + j \leq t + m + 1$  en  $j > t$ , deze stellen we gelijk aan nul. Wanneer we dan met (3) en (4) de schatters gaan bepalen voor  $\hat{x}_i$  en  $\hat{y}_j$ , dan krijgen we voor  $j = t + 1, \dots, t + m$  dat  $\hat{y}_j = 0$ . Hierdoor geldt ook voor  $i = 1, \dots, t + m$  dat  $\hat{x}_i = \sum_{j=1}^{t+m} C_{i,j} = \sum_{j=1}^t C_{i,j}$ . Doordat we nu nog steeds met een driehoek werken, blijven alle formules van toepassing. De extra  $y_j$ 's worden gelijk aan nul en beïnvloeden de schatters niet. De waarden van  $\hat{C}_{i,j}$  zijn gelijk aan nul voor  $j = t + 1, \dots, t + m$  en  $i = 1, \dots, t + m$ , doordat de bijbehorende  $\hat{y}_j = 0$ . Het is gebruikelijk om deze nullen niet weer te geven in de resultaten schadedriehoek.



Figuur 2.2

### IBNR Voorziening

De voorzieningen kunnen nu als volgt geschat worden:

$$\begin{aligned}
 \text{IBNR voorziening : } \hat{R}_i &= \sum_{j=t+2-i}^t \hat{C}_{it} \\
 &= \hat{x}_i \sum_{j=t+2-i}^t \hat{y}_j \quad (7)
 \end{aligned}$$

### Voorbeeld

Bij dit voorbeeld ga ik uit van de aannames van Hachemeister en Stanard, namelijk  $E(C_{ij}) = x_i y_j$  met de schatters van  $\hat{x}_i$  en  $\hat{y}_j$  uit (3) en (4).

$$\begin{aligned}
 \hat{x}_1 &= \sum_{j=1}^4 C_{1j} = 100 \\
 \hat{y}_4 &= \frac{C_{14}}{\hat{x}_1} = \frac{5}{100} = 0,05 \\
 \hat{x}_2 &= \frac{\sum_{j=1}^3 C_{2j}}{1 - \hat{y}_4} = \frac{98}{95/100} = 103,16 \\
 \hat{y}_3 &= \frac{\sum_{i=1}^2 C_{i,3}}{\sum_{i=1}^2 \hat{x}_i} = \frac{10 + 11}{100 + 103,16} = \frac{399}{3860} = 0,10
 \end{aligned}$$

$$\begin{aligned}\hat{x}_3 &= \frac{\sum_{j=1}^2 C_{3,j}}{1 - \hat{y}_3 - \hat{y}_4} = \frac{58 + 23}{1 - 0,10 - 0,05} = 95,67 \\ \hat{y}_2 &= \frac{\sum_{i=1}^3 C_{i,2}}{\sum_{i=1}^3 \hat{x}_i} = \frac{25 + 24 + 23}{100 + 103,16 + 95,67} = 0,24 \\ \hat{x}_4 &= \frac{C_{4,1}}{1 - \hat{y}_2 - \hat{y}_3 - \hat{y}_4} = \frac{65}{1 - 0,24 - 0,10 - 0,05} = 107,32 \\ \hat{y}_1 &= \frac{\sum_{i=1}^4 C_{i,1}}{\sum_{i=1}^4 \hat{x}_i} = \frac{60 + 63 + 58 + 65}{100 + 103,16 + 95,67 + 107,32} = 0,61\end{aligned}$$

Hiermee krijgen we de volgende incrementele schadedriehoek:

schadejaar	afwikkeljaar			
	1	2	3	4
1	60	25	10	5
2	63	24	11	<b>5,2</b>
3	58	23	<b>9,9</b>	<b>4,8</b>
4	65	<b>25,9</b>	<b>11,1</b>	<b>5,4</b>

Figuur 2.3

Ofwel de totale voorziening wordt:

$$\hat{R} = 5,2 + 9,9 + 4,8 + 25,9 + 11,1 + 5,4 = 62,2.$$

### 2.3 Gegeneraliseerde Lineaire Modellen

Een GLM is een generalisatie van het lineaire model op twee verschillende vlakken. Als eerste is het toegestaan dat de afwijking van het gemiddelde een andere verdeling hebben dan de normale verdeling. Iedere verdeling is toegestaan zolang het een verdeling van de exponentiële familie is, bijvoorbeeld de Poisson-, Binomiale-, Gamma- of de inverse Gaussian verdeling. Ten tweede is het niet meer nodig dat het gemiddelde een lineaire functie is van de data, er kan bijvoorbeeld ook sprake zijn van een logaritmische functie. Hier komen we in hoofdstuk 4 nader op terug. Een algemene introductie is te vinden in McCullagh and Nelder (1989).

#### IBNR Voorziening

Een GLM is opgebouwd uit drie verschillende componenten. Als eerste is er de **(1) stochastische component**, welke gekoppeld is aan de verdeling van de exponentiële familie. Als tweede is er de **(2) systematische component**, welke een uitdrukking is van het lineaire deel van het model. Ten slotte is er de **(3) linkfunctie**, die de grondslag legt voor bijvoorbeeld een additief of multiplicatief model. Hieronder zal ik de drie componenten nader beschrijven.

## (1) Stochastische component

$C_{ij} \sim f(y_{ij}; \theta_{ij}; \phi)$  met  $C_{ij}$  onafhankelijk. Verder is  $f$  de dichtheidsfunctie van een exponentiële familie,  $y_{ij}$  de geobserveerde data,  $\phi$  een schaalparameter en  $\theta_{ij}$  is afhankelijk van de verdeling.

In het algemeen kan de dichtheidsfunctie worden geschreven als:

$$f(y_{ij}; \theta_{ij}; \phi) = e^{[y_{ij}\theta_{ij} - b(\theta_{ij})]/a(\phi) + c(y_{ij}, \phi)} \quad (8)$$

waarbij  $a(\cdot)$ ,  $b(\cdot)$  en  $c(\cdot)$  bekende functies zijn die per dichtheid kunnen verschillen. De functie  $a(\phi)$  heeft meestal de vorm  $a(\phi) = \phi$ ,  $\phi$  wordt ook wel de dispersie parameter genoemd. Als  $\phi$  een bekende constante is, dan kan  $f$  eenvoudiger worden geschreven als:

$$f(y_{ij}; \theta_{ij}) = \tilde{a}(\theta_{ij}) \tilde{b}(y_{ij}) e^{y_{ij} Q(\theta_{ij})} \quad (9)$$

met  $Q(\theta) = \theta/a(\phi)$ ,  $\tilde{a}(\theta) = e^{-b(\theta)/a(\phi)}$  en  $\tilde{b}(y) = e^{c(y, \phi)}$ .

Formule (8) is geschikt voor tweeparameter families zoals de normale of de gamma verdeling. Formule (9) kan worden gebruikt voor bijvoorbeeld de Poisson verdeling.

Met behulp van de log likelihood functie  $l(\phi; \theta; y) = \log f_Y(y; \theta; \phi)$  en de twee bekende relaties  $E\left(\frac{\partial l}{\partial \theta}\right) = 0$  en  $E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left(\frac{\partial l}{\partial \theta}\right)^2 = 0$ , krijgen we, als we  $\mu_{ij}$  schrijven voor de verwachting van  $C_{ij}$ :

$$E(C_{ij}) = \mu_{ij} = b'(\theta_{ij}) \quad (10)$$

$$Var(C_{ij}) = b''(\theta_{ij})a(\phi) \quad (11)$$

De variantie kan worden uitgedrukt als een functie van het gemiddelde, namelijk:  $Var(C_{ij}) = a(\phi)V(\mu_{ij})$ .  $V(\cdot)$  heet hier de variantie functie.

Stel dat de verdeling van de functie niet bekend is, maar dat de eerste twee momenten wel gegeven zijn, dan kan er gebruik worden gemaakt van de quasi (log)likelihood in plaats van de (log) likelihood. Stel  $X_1, X_2, \dots, X_n$  een onafhankelijke steekproef uit een kansverdeling met verwachting  $E(X) = \mu$  en variantie  $Var(X) = V(\mu)$  voor een zekere bekende functie  $V$ . De zogeheten quasi likelihood functie

$$Q(\mu) = \sum_i \int_{t=x_i}^{\mu} \frac{x_i - t}{V(\mu)} dt \quad (12)$$

gedraagt zich in veel opzichten als een likelihood. Als we  $\mu$  schatten door de quasi log likelihood te maximaliseren, mogen we verwachten dat het resultaat de meeste optimaliteits eigenschappen van een maximum log likelihood

schatte typisch heeft.

## (2) Systematische component

De systematische component is gebaseerd op een lineaire voorspeller:  $\eta_{ij} = (X\vec{\beta})_{ij}$  met  $\vec{\beta}$  de regressieparameters van het model van dimensie  $q$  en  $X$  de regressiematrix van dimensie  $t^2q$ . Er zijn veel verschillende mogelijkheden voor de keuze van  $\vec{\beta}$  en  $X$ . Een aantal bekende zijn:

'Chain Ladder-type':  $\eta_{ij} = \alpha_i + \beta_j$ ,  
waarbij  $\alpha_i$  een parameter voor schadejaar  $i$  en  $\beta_j$  een parameter voor afwikkeljaar  $j$ . Om een niet-singuliere regressiematrix te verkrijgen, moet er een extra eis worden gesteld voor  $\beta$ , een veel gebruikte keuze is  $\beta_1 = 0$  of  $\sum_{j=1}^t \beta_j = 0$ .

'Barnett & Zehnwirth (1998)-type':  $\eta_{ij} = \alpha_i + \sum_{k=1}^{j-1} \beta_k + \sum_{h=1}^{i+j-2} \gamma_h$ ,  
waarbij  $\alpha_i$  een parameter voor schadejaar  $i$ ,  $\beta_j$  een parameter voor afwikkeljaar  $j$  en  $\gamma_h$  een parameter voor kalenderjaar  $i + j - 1$ . Meer algemene modellen met parameters in de drie richtingen, schadejaar, afwikkeljaar en kalenderjaar, zijn als toepassing in de GLM voor het eerst gebruikt door De Vylder en Goovaerts (1979).

## (3) Linkfunctie

De linkfunctie koppelt de verwachting  $\mu_{ij}$  van  $C_{ij}$  aan de lineaire voorspeller. Deze derde component koppelt dus de stochastische component aan de systematische component. Dit gebeurt door middel van:  $\eta_{ij} = g(\mu_{ij})$  met  $g(\cdot)$  een monotone, differentieerbare functie. Veel gebruikte linkfuncties zijn de identiteit en de logaritmische linkfunctie, het is dus hiermee mogelijk een additief dan wel multiplicatief model te bouwen.

De parameters van het model kunnen worden geschat met de maximum (log)likelihood functie van de verdeling uit de stochastische component. Wanneer de verdeling niet bekend is, maar wel de verwachting en de variantie, dan kunnen de parameters worden geschat met behulp van de quasi likelihood functie.

Er zijn inmiddels diverse computerprogramma's ontwikkeld die met GLM kunnen werken, waaronder GLIM, SAS en R. De input zijn de stochastische component, de systematische component en de linkfunctie. De output zijn de schatters.

### 2.3.1 Chain Ladder en GLM

Gegeneraliseerde Lineaire Modellen zijn breed te gebruiken. De drie componenten zijn namelijk zodanig te kiezen dat het GLM overeenkomt met een ander model. In Renshaw en Verrall (1994) laten zien dat ook de Chain Ladder te schrijven is als GLM. De keuzes voor de componenten moeten dan als volgt gemaakt worden:

(1) Stochastische component: We specificeren de eerste twee momenten  $E(C_{i,j}) = x_i y_j$  en  $Var(C_{i,j}) = \phi E(C_{i,j}) = \phi x_i y_j$ . Een lineair verband tussen de verwachting en de variantie wordt in de literatuur soms aangeduid als over-dispersed Poisson. De parameters  $\phi$ ,  $x_i$  en  $y_j$  schatten we met behulp van de quasi (log)likelihood.

(2) Systematische component: lineaire voorspeller  $g(\mu_{i,j}) = \eta_{ij} = \alpha_i + \beta_j$

(3) Linkfunctie: logaritmisch, ofwel:  $\log E(C_{i,j}) = \alpha_i + \beta_j = \log x_i + \log y_j$

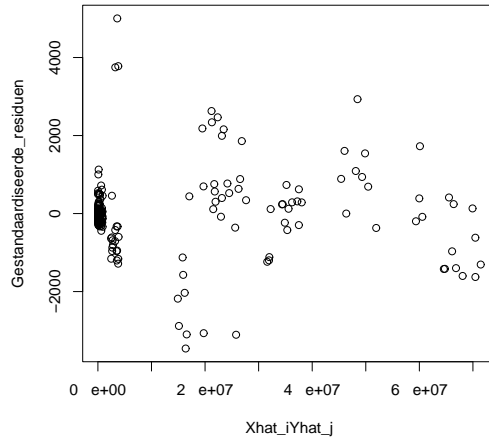
Met behulp van bovenstaande relatie tussen de Chain Ladder en het GL model, kan er iets worden gezegd over de schatters van  $C_{i,j}$  voor  $i+j > t+1$  en de foutmarge van het Chain Ladder model.

### 2.4 Chain Ladder en overdispersie

Een eigenschap van de Poisson verdeling is dat de variantie gelijk is aan de verwachting. In de praktijk geldt voor zorgverzekeraars dat de variantie vaak groter is dan de verwachting. Dit noemt men *overdispersie*, er geldt dan:  $E(C_{i,j}) = \lambda_{ij}$  en  $Var(C_{i,j}) = \phi E(C_{i,j}) = \phi \lambda_{ij}$ . Overdispersie is in principe geen bezwaar als men de Chain Ladder alleen wil gebruiken om de reserve  $R$  te voorspellen. De schatters zijn namelijk zuiver wanneer de  $C_{i,j}$  onafhankelijk zijn. De schatters kunnen dan worden berekend met de quasi likelihood (12), hier is namelijk alleen de verwachting en de variantie voor nodig. Om de verdeling van de voorspellingsfout te bepalen, zal met de overdispersie rekening gehouden moeten worden.

Met behulp van de Bootstrap methode, welke in paragraaf 3.2 uitgelegd wordt, kan er een foutmarge gevonden worden. Maar voordat we de bootstrap toepassen, is het verstandig om na te gaan of ons model wel past bij de data. Hiervoor zetten we  $\hat{x}_i \hat{y}_j$  uit tegen de gestandaardiseerde residuen  $(C_{i,j} - \hat{x}_i \hat{y}_j) / \sqrt{\hat{x}_i \hat{y}_j}$ . Onder het Poisson model zouden deze residuen ongeveer verwachting 0 en variantie 1 moeten hebben. Als er sprake is van overdispersie, dan is de spreiding van de residuen groter. In Figuur 2.4 zijn de residuen te zien, het is duidelijk dat de variantie groter is dan 1 en er dus sprake is van overdispersie.





Figuur 2.4

### Schatters

In het geval van overdispersie, is de negatief binomiale verdeling een mogelijk alternatief voor de Poisson verdeling. Deze verdeling maakt gebruik van twee parameters  $p$  en  $r$  en is gegeven door:

$$P(X = k) = \binom{r+k-1}{r-1} p^r (1-p)^k, \quad k = 0, 1, 2, \dots \quad \text{en er geldt:}$$

$$E(X) = \frac{r(1-p)}{p}$$

$$Var(X) = \frac{r(1-p)}{p^2}$$

De negatief binomiale verdeling is op te vatten als een mengsel van Poisson verdelingen.

- (1) Trek  $\Lambda$  uit de Gamma verdeling met parameters  $r$  en  $(1-p)/p$
- (2) Gegeven  $\Lambda = \lambda$ , trek  $X$  uit de Poisson verdeling met parameter  $\lambda$

Door gebruik te maken van een her-parameterisatie,  $p = \frac{\omega}{\lambda+\omega}$  en  $r = \omega$ , geldt er:

$$P(X = k) = \binom{\omega+k-1}{k-1} \left(\frac{\lambda}{\lambda+\omega}\right)^k \left(\frac{\omega}{\lambda+\omega}\right)^\omega$$

$$= \frac{\lambda^k}{k!} \cdot \frac{(\omega+k-1)!}{(\omega-1)!(\lambda+\omega)^k} \cdot \frac{1}{(1+\lambda/\omega)^\omega}$$

en er geldt:

$$E(X) = \lambda$$

$$Var(X) = \frac{\lambda(\lambda+\omega)}{\omega}$$

Als  $\omega$  groot wordt, convergeert de tweede term van  $P(X = k)$  naar 1 en de laatste term naar  $e^{-\lambda}$  en ontstaat er de Poisson verdeling met parameter  $\lambda$ . In deze context wordt de negatief binomiale verdeling ook wel gebruikt als over-dispersed Poisson verdeling met dispersieparameter  $\phi = \frac{\lambda + \omega}{\omega} = 1 + \frac{\lambda}{\omega}$ . De parameter  $\omega$  controleert de afwijking van de variantie bij een over-dispersed Poisson verdeling.

Stel nu dat  $C_{i,j}$  negatief binomiaal verdeeld met parameters  $\lambda_{i,j} = x_i y_j$  en  $\omega$ , dan worden de maximum likelihood schatters:

$$P(C_{i,j}) = \frac{(x_i y_j)^{C_{i,j}}}{C_{i,j}!} \cdot \frac{(\omega + C_{i,j} - 1)!}{(\omega - 1)!(x_i y_j + \omega)^{C_{i,j}}} \cdot \frac{1}{(1 + x_i y_j / \omega)^\omega}$$

$$\log \text{lik} : l = \sum_{i+j \leq t+1} \left( C_{i,j} (\log x_i + \log y_j) - \log C_{i,j}! + \log \frac{(\omega + C_{i,j} - 1)!}{(\omega - 1)!} \right. \\ \left. - (C_{i,j} + \omega) \log(x_i y_j + \omega) + \omega \log \omega \right) \quad (13)$$

$$\frac{\partial l}{\partial x_i} = \sum_{j=1}^{t+1-i} \left( \frac{C_{i,j}}{x_i} - \frac{C_{i,j} y_j}{x_i y_j + \omega} - \frac{\omega y_j}{\omega + x_i y_j} \right)$$

$$= \sum_{j=1}^{t+1-i} \left( \frac{C_{i,j}}{x_i} - \frac{C_{i,j} y_j + \omega y_j}{x_i y_j + \omega} \right) = 0$$

$$\hat{x}_i = \frac{\sum_{j=1}^{t+1-i} \omega C_{i,j}}{\sum_{j=1}^{t+1-i} \omega y_j}$$

$$= \frac{\sum_{j=1}^{t+1-i} C_{i,j}}{\sum_{j=1}^{t+1-i} \hat{y}_j} \quad (14)$$

wegens de symmetrie geldt ook:

$$\hat{y}_j = \frac{\sum_{i=1}^{t+1-j} C_{i,j}}{\sum_{i=1}^{t+1-j} \hat{x}_i} \quad (15)$$

Deze schatters komen precies overeen met de Chain Ladder schatters uit (3) en (4). De overdispersie blijkt dus geen rol te spelen bij het bepalen van een voorspelling voor de benodigde reserve. Om de parametrische bootstrap te kunnen gebruiken voor de verdeling van de voorspellingsfout, moeten we wel  $\omega$  kunnen schatten. We kunnen hierbij gebruik maken van de maximum likelihood (Fisher 1941, Piegorsch 1990), maximum quasi likelihood (Clark en Perry, 1989) of de momenten methode. We kiezen de quasi likelihood methode omdat deze het beste lijkt te presteren, vooral bij kleine steekproeven (Piegorsch, 1990).

In ons model vinden we als quasi log likelihood

$$Q(x, y, \omega) = \sum_{i+j \leq t+1} \left( C_{i,j} \log \left( \frac{x_i y_j}{\omega + x_i y_j} \right) + \omega \log \left( \frac{\omega}{\omega + x_i y_j} \right) \right)$$

Als we de quasi likelihood vergelijken met de echte likelihood uit (13), zien we dat ze gelijk zijn op de term  $\log\left(\frac{\omega + C_{i,j}}{C_{i,j}}\right)$  na. Het is duidelijk dat dit geen effect heeft op het schatten van de  $x_i$  en  $y_j$  zodat we kunnen concluderen dat voor deze parameters de maximum likelihood en maximum quasi likelihood schatters hetzelfde zijn.

Om de schatter voor  $\omega$  te bepalen nemen we de partiële afgeleide:

$$\frac{\partial}{\partial \omega} Q(x, y, \omega) = \sum_{i+j \leq t+1} \left( \log\left(\frac{\omega}{\omega + x_i y_j}\right) + \frac{C_{i,j} + \omega}{x_i y_j + \omega} + 1 \right)$$

Vervolgens kunnen we de schatters voor  $x_i$  en  $y_j$  invullen en het resultaat gelijk aan nul stellen. Om de vergelijking die we krijgen op te lossen, gebruiken we het iteratieproces van Newton. Hierbij is

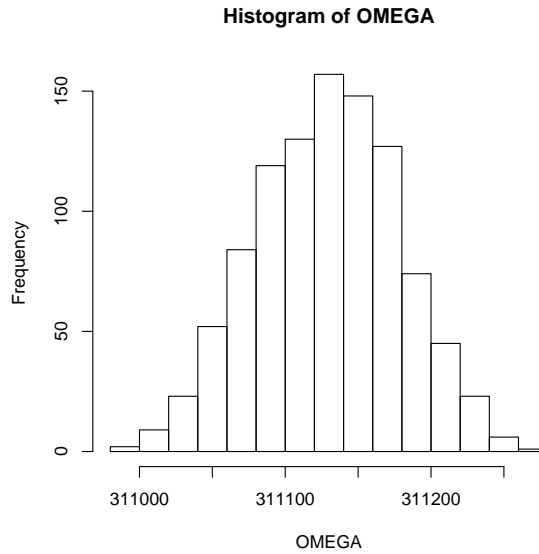
$$\frac{\partial^2}{\partial \omega^2} Q(x, y, \omega) = \sum_{i+j \leq t+1} \frac{x_i^2 y_j^2 + \omega C_{i,j}}{\omega(x_i y_j + \omega)^2}$$

We kiezen een startwaarde  $\omega_0$  en itereren

$$\omega_{n+1} = \omega_n - \frac{\frac{\partial}{\partial \omega} Q(\hat{x}, \hat{y}, \omega_n)}{\frac{\partial^2}{\partial \omega^2} Q(\hat{x}, \hat{y}, \omega_n)} \quad (16)$$

In appendix A staat een programma om in R de schatter voor  $\omega$  te berekenen.

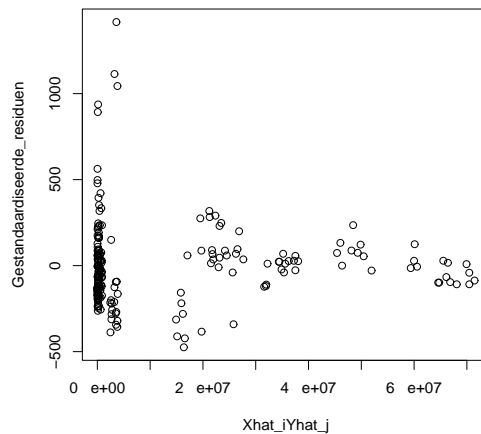
Wanneer we voor de fictieve kwartaaldata een schatter gaan bepalen voor  $\omega$ , krijgen we  $\hat{\omega} = 314.435$ . Om te controleren of  $\hat{\omega}$  acceptabel is en constant blijft voor steekproeven van  $C_{i,j}^*$  uit de negatief binomiale verdeling met schatters  $\hat{x}_i \hat{y}_j$  en  $\hat{\omega}$ , simuleren we 1000 steekproeven en bepalen we iedere keer  $\hat{\omega}$  met behulp van het iteratieproces. Het histogram geeft de 1000 waarden voor  $\hat{\omega}$  weer.



Figuur 2.5

Hierbij is het gemiddelde:  $\bar{\omega} = 311.129$  en  $var(\omega) = 2385$  en  $\sigma = 48,8$ . Omdat  $\bar{\omega}$  toch veel afwijkt van  $\hat{\omega}$ , maar relatief gezien de afwijking maar 1,1% is, hebben we te maken met een onzuivere schatter welke door de kleine procentuele afwijking wel acceptabel is.

Tot slot willen we nog kijken of het model bij de data past door  $\hat{x}_i \hat{y}_j$  uit te zetten tegen de gestandaardiseerde residuen  $(C_{i,j} - \hat{x}_i \hat{y}_j) / \sqrt{\frac{\hat{x}_i \hat{y}_j (\hat{x}_i \hat{y}_j + \hat{\omega})}{\hat{\omega}}}$ . Onder de aanname van de negatief binomiale verdeling, zouden deze residuen ongeveer verwachting 0 en variantie 1 moeten hebben.



Figuur 2.6

Hoewel er voor kleine  $\hat{x}_i \hat{y}_j$  nog best wat uitschieters zijn, wordt de variantie voor grote waarden van  $\hat{x}_i \hat{y}_j$  wel kleiner. In ieder geval wordt de variantie geen 1. Wel is het een verbetering ten opzichte van de gestandaardiseerde Poisson residuen uit figuur 2.4.

Wanneer we  $m$  extra schadejaren aan de schadedriehoek toevoegen, kunnen we net als in 2.2 is beschreven weer een driehoek maken met  $m + t$  afwikkeljaren door  $C_{i,j}$  gelijk aan nul toe te voegen voor  $j > t$  en  $i + j \leq t + m + 1$ . Doordat we dan weer een driehoek hebben, blijven de hierboven afgeleide formules van toepassing. Voor de extra toegevoegde nullen verandert de waarde van de log likelihood functie niet, waardoor de waarden van de schatters  $\hat{x}_i$ ,  $\hat{y}_j$  en  $\hat{\omega}$  niet worden beïnvloed.

## 2.5 Chain Ladder en de Normale verdeling

Een andere benadering van een model dat de eigenschappen  $E(C_{i,j}) = x_i y_j$  en  $Var(C_{i,j}) = \phi x_i y_j$  heeft, is met de normale verdeling waarbij  $C_{i,j} \sim N(x_i y_j, \phi x_i y_j)$ . Hieronder zullen we onderzoeken of dit een bruikbare benadering oplevert.

### Schatters

Stel nu dat we uitgaan van deze aannames. Dan gaan we als volgt te werk:

In het algemeen geldt:  $X \sim N(\mu, \sigma^2)$ :  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ , hier geldt:

$$P(C_{i,j}) = \frac{1}{\sqrt{2\pi\phi x_i y_j}} \cdot e^{-\frac{1}{2}\left(\frac{C_{i,j} - x_i y_j}{\phi x_i y_j}\right)^2}$$

$$\log \text{lik} : l = \sum_{i+j \leq t+1} \left( -\frac{1}{2} \log(2\pi\phi x_i y_j) - \frac{1}{2} \left( \frac{C_{i,j} - x_i y_j}{\phi x_i y_j} \right)^2 \right) \quad (17)$$

$$\begin{aligned} \frac{\partial l}{\partial x_i} &= \sum_{j=1}^{t+1-i} -\frac{1}{2x_i} - \left( \frac{C_{i,j}}{\phi x_i y_j} - \frac{1}{\phi} \right) \left( \frac{-C_{i,j}}{\phi x_i^2 y_j} \right) \\ &= \sum_{j=1}^{t+1-i} -\frac{1}{2x_i} + \frac{C_{i,j}^2}{\phi^2 x_i^3 y_j^2} - \frac{C_{i,j}}{\phi^2 x_i^2 y_j} \\ &= \sum_{j=1}^{t+1-i} \frac{-\phi^2 x_i^2 y_j^2 + 2C_{i,j}^2 - 2C_{i,j} x_i y_j}{2\phi^2 x_i^3 y_j^2} = 0 \end{aligned}$$

ofwel:

$$0 = \sum_{j=1}^{t+1-i} \left( -\phi^2 y_j^2 x_i^2 - 2C_{i,j} y_j x_i + 2C_{i,j}^2 \right)$$

$$\hat{x}_i = \frac{\sum_{j=1}^{t+1-i} C_{i,j} \hat{y}_j}{-\sum_{j=1}^{t+1-i} \hat{\phi}^2 \hat{y}_j^2} \pm$$

$$\frac{\sqrt{(\sum_{j=1}^{t+1-i} C_{i,j} \hat{y}_j)^2 + \sum_{j=1}^{t+1-i} \hat{\phi}^2 \hat{y}_j^2 \cdot \sum_{j=1}^{t+1-i} 2C_{i,j}^2 - \sum_{j=1}^{t+1-i} \hat{\phi}^2 \hat{y}_j^2}}{}$$

Als eerste kan worden opgemerkt dat deze schatters ongelijk zijn aan de schatters voor  $x_i$  en  $y_j$  uit de Chain Ladder methode. Als tweede zijn deze mle schatters veel moeilijker om uit te rekenen dan bij de Chain Ladder. Deze schatters zijn dus niet goed bruikbaar.

Een andere manier om een benadering van de schatters van  $x_i$ ,  $y_j$  en  $\phi$  te vinden, is om gebruik te maken van de quasi likelihood. Hierbij komen de schatters van  $x_i$  en  $y_j$  overeen met de schatters van  $\hat{x}_i$  en  $\hat{y}_j$  uit de Chain Ladder. Vervolgens is  $\phi$  te schatten met

$$\hat{\phi} = \frac{1}{N - k} \sum_{i+j \leq t+1} \frac{(C_{i,j} - x_i y_j)^2}{x_i y_j} \quad (18)$$

Hierbij is  $N$  het aantal waarnemingen  $N = \frac{1}{2}t(t+1)$  en  $k$  is het aantal parameters  $k = 2t - 1$ .

Wanneer we dit toepassen op de fictieve zorgdata, dan krijgen we  $\hat{\phi} = 1.020.683$ . Als standaardafwijking krijgen we  $\hat{\sigma} := \sqrt{\sum_{i+j > t+1} \hat{\phi} \hat{x}_i \hat{y}_j} = 10.678.569$ . Dit komt overeen met ongeveer 10% van de met de Chain Ladder uitgerekende IBNR voorziening.

Net als bij de Poisson verdeling is op te merken dat deze methode ook op gaat wanneer er  $m$  schadejaren meer bekend zijn dan alleen de data uit de schadedriehoek. Net als bij de Poisson verdeling, geldt namelijk bij de normale verdeling dat als  $C_{i,j} \sim N(x_i y_j, \phi x_i y_j)$ , dat  $\sum_{j=1}^t C_{i,j} \sim N(x_i, \phi x_i)$ . We kunnen nu het eerste schadejaar in de schadedriehoek vervangen door de som van alle bekende schadejaren. De som van het eerste schadejaar had eerst de verdeling  $N(x_1, \phi x_1)$ , dit wordt nu vervangen door  $N(\sum_{i=1}^{m+1} x_i, \phi \sum_{i=1}^{m+1} x_i)$ . Vervolgens kunnen we analoog verder gaan aan de hierboven beschreven methode.

## 2.6 Expectation Maximization methode

De Expectation Maximization (EM) methode, Dempster et al. (1977), is een iteratief proces dat in een eindig aantal stappen convergeert naar de mle. Het EM algoritme wisselt af tussen een 'expectation' stap, waarin bij schadedriehoeken de verwachte waarde van de onderdriehoek wordt berekend en tussen een 'maximization' stap, waarin de mle wordt berekend van de volgeschatte driehoek. De methode is met name handig wanneer de schatters voor  $\hat{x}_i$  en  $\hat{y}_j$  niet bepaald kunnen worden aan de hand van de schadedriehoek, maar wel geschat kunnen worden aan de hand van de volgeschatte

driehoek. Ook werkt dit algoritme als er een  $C_{i,j}$  voor  $i + j \leq t + 1$  onbekend is. Het algoritme zal uitgelegd worden aan de Chain Ladder aanname waarbij  $E(C_{i,j}) = x_i y_j$ .

### Algoritme

- (1) Begin met het bepalen van  $x_i^0$  en  $y_j^0$ , dit mag in principe willekeurig. Een mogelijke keuze zou zijn  $x_i^0 = \sum_{j=1}^t C_{1,j}$  voor  $i = 1, \dots, t$  en  $y_j^0 = \frac{C_{1,j}}{\sum_{j=1}^t C_{1,j}}$  voor  $j = 1, \dots, t$ .
- (2) Schat de schadedriehoek vol met  $x_i^0$  en  $y_j^0$ , dit resulteert in  $C_{i,j}^0$ . Het volschatten van de schadedriehoek is de 'expectation' stap.
- (3) Bepaal vervolgens  $x_i^1$  en  $y_j^1$  aan de hand van de bij het model behorende formules. In dit geval:  $x_i^1 = \sum_{j=1}^t C_{i,j}^0$  en  $y_j^1 = \frac{\sum_{i=1}^t C_{i,j}^0}{\sum_{i=1}^t \sum_{j=1}^t C_{i,j}^0}$ . Het bepalen van een nieuwe  $x_i$  en  $y_j$  is de 'maximization' stap.
- (4) Herhaal de stappen 2 en 3 totdat  $|x_i^n - x_i^{n+1}| < \epsilon$  en  $|y_j^n - y_j^{n+1}| < \epsilon$ ,  $\epsilon > 0$ .

Een programma om dit algoritme in R uit te voeren staat in appendix B.

### Voorbeeld

We beginnen weer met de schadedriehoek uit Figuur 1.5. We beginnen met de grove schatters  $x^0 = (100 \ 100 \ 100 \ 100)$  en  $y^0 = (\frac{65}{100} \ \frac{25}{100} \ \frac{10}{100} \ \frac{5}{100})$ . Hiermee schatten we de schadedriehoek vol:

$$C^0 = \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 60 & 25 & 10 & 5 \\ 2 & 63 & 24 & 11 & \mathbf{5} \\ 3 & 58 & 23 & \mathbf{10} & \mathbf{5} \\ 4 & 65 & \mathbf{25} & \mathbf{10} & \mathbf{5} \end{array}$$

Vervolgens kunnen we  $x^1$  en  $y^1$  bepalen:  $x^0 = (100 \ 103 \ 96 \ 105)$  en  $y^1 = (0,61 \ 0,24 \ 0,10 \ 0,05)$ . Hiermee wordt  $C^1$ :

$$C^1 = \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 60 & 25 & 10 & 5 \\ 2 & 63 & 24 & 11 & \mathbf{5,10} \\ 3 & 58 & 23 & \mathbf{9,74} & \mathbf{4,75} \\ 4 & 65 & \mathbf{25,21} & \mathbf{10,66} & \mathbf{5,20} \end{array}$$

En verder  $x^2 = (100,00 \ 103,10 \ 95,50 \ 106,06)$  en  $y^2 = (0,61 \ 0,24 \ 0,10 \ 0,05)$  met:

$$C^2 = \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 60 & 25 & 10 & 5 \\ 2 & 63 & 24 & 11 & \mathbf{5,11} \\ 3 & 58 & 23 & \mathbf{9,77} & \mathbf{4,73} \\ 4 & 65 & \mathbf{25,48} & \mathbf{10,85} & \mathbf{5,26} \end{array}$$

Dit kan zo nog een tijdje verder gaan. Achtereenvolgens hebben we in de verschillende stappen de voorziening:

stap $k$	0	1	2	...	10	...	15	16
$R = \sum_{i=2}^t \sum_{j=t+2-i}^t C_{i,j}^k$	60	60,66	61,19	...	62,10	...	62,15	62,15

Na ongeveer 15 stappen blijft de voorziening constant op 2 decimalen. De uiteindelijke waarden zijn dan geworden:  $x = (100,00 \ 103,16 \ 95,67 \ 107,31)$ ,  $y = (0,606 \ 0,241 \ 0,103 \ 0,050)$  en

	1	2	3	4
$C =$	60	25	10	5
2	63	24	11	<b>5,16</b>
3	58	23	<b>9,89</b>	<b>4,78</b>
4	65	<b>25,86</b>	<b>11,09</b>	<b>5,37</b>

De voorziening is hierbij:  $\hat{R} = 5,2 + 9,9 + 4,8 + 25,9 + 11,1 + 5,47 = 62,2$ .

Doordat de Chain Ladder schatters overeen komen met de mle in het Poisson model en in het Poisson model de mle van de volledige data worden gegeven door  $\hat{x}_i = \sum_{j=1}^t C_{i,j}$  en  $\hat{y}_j = \frac{\sum_{i=1}^t C_{i,j}}{\sum_{i=1}^t \sum_{j=1}^t C_{i,j}}$ , zijn de schatters verkregen met het hierboven beschreven EM algoritme gelijk aan de Chain Ladder schatters.

Voor het EM algoritme maakt het niet uit of er gebruik wordt gemaakt van eerdere schadejaren bij het volschatten van de schadedriehoek. Zolang de formules in de derde stap van het algoritme overeen komen met de mle van de volledige data, zal het proces uiteindelijk convergeren naar de meest waarschijnlijke schatter.



### 3 Foutmarge

Een belangrijk onderdeel bij het genereren van schatters, is het vinden van een verwachte fout. In enkele gevallen is de verdeling bekend en kan er iets over de variantie gezegd worden. Bootstrap is ook een methode om een foutmarge te genereren, dit kan zowel parametrisch als niet-parametrisch. De verwachte kwadratische fout is een andere mogelijkheid om naar de foutmarge te kijken. Hieronder zal ik beginnen met Mack, die een methode geeft voor het bepalen van een fout bij de Chain Ladder methode.

#### 3.1 Methode van Mack

In Mack (1993) is een methode beschreven om een standaard fout te bepalen bij de Chain Ladder methode, zonder gebruik te maken van een verdeling. Het uitgangspunt van Mack is de Chain Ladder methode vanwege diens simpliciteit. Deze methode werkt namelijk al met een minimale hoeveelheid aan aannames omdat er geen gebruik gemaakt hoeft te worden van een kansverdeling. Om zonder kansverdeling toch iets te kunnen zeggen over de standaard fout, maakt Mack gebruik van een aantal aannames die dicht liggen bij de aannames van de Chain Ladder.

#### Model

De methode van Mack (1993) is gebaseerd op de volgende aannames:

$$1 \quad E(C_{i,j+1} | C_{i1}, \dots, C_{ij}) = C_{ij} f_j, \quad i = 1, \dots, t, \quad j = 1, \dots, t-1 \quad (19)$$

$$2 \quad C_{i,j} \text{ zijn onderling onafhankelijk} \quad (20)$$

$$3 \quad \text{Var}(\sum_{k=1}^{j+1} C_{i,k} | C_{i1}, \dots, C_{ij}) = \sum_{k=1}^j C_{i,k} \sigma_j^2 \quad (21)$$

$$i = 1, \dots, t, \quad j = 1, \dots, t-1$$

Waarbij  $f_j$  gedefinieerd is zoals bij de Chain Ladder methode en  $\sigma_j^2$  ( $j = 1, \dots, t-1$ ) zijn onbekende parameters. Verder leveren (19) en (20) de zuiverheid op.

Uitgaande van bovenstaande punten, kan bewezen worden dat de mean square error (mse) van de IBNR voorziening voor schadejaar  $i$  de volgende is:

$$\widehat{mse}(\hat{R}_i) = \hat{D}_{i,t}^2 \sum_{j=t+1-i}^{t-1} \frac{\hat{\sigma}_j^2}{\hat{f}_{j+1}^2} \left( \frac{1}{\hat{D}_{i,j}} + \frac{1}{\sum_{k=1}^{t-j} D_{k,j}} \right) \quad (22)$$

waarbij

$$\hat{\sigma}_j^2 = \frac{1}{t-j-1} \sum_{i=1}^{t-j} D_{i,j} \left( \frac{D_{i,j+1}}{D_{i,j}} - \hat{f}_{j+1} \right)^2 \quad j = 1, \dots, t-2 \quad (23)$$

De parameter  $\hat{\sigma}_{t-1}$  wordt geschat door aan te nemen dat  $\frac{\hat{\sigma}_{t-3}}{\hat{\sigma}_{t-2}} = \frac{\hat{\sigma}_{t-2}}{\hat{\sigma}_{t-1}}$  als  $\hat{\sigma}_{t-3} > \hat{\sigma}_{t-2}$ . Anders wordt  $\hat{\sigma}_{t-1}$  geschat door  $\hat{\sigma}_{t-3}$ .

De methode van Mack heeft als voordeel dat het bij een eenvoudige methode als de Chain Ladder, zonder dat er een aanname wordt gemaakt over de verdeling, er toch een schatter gegenereerd kan worden voor de standaardafwijking. De inzetbaarheid van de methode van Mack wordt beperkt door de derde aanname welke een redelijk sterke aanname is. Een nadeel is dat de motivatie van deze aanname onduidelijk is. Wel kan onder deze aanname de formule van de variantie worden afgeleid.

### Voorbeeld

Deze methode maakt gebruik van de Chain Ladder om de schadedriehoek vol te schatten, deze is terug te vinden in Figuur 2.1. Als eerste kunnen we  $\hat{\sigma}^2$  bepalen. Met behulp van formule (23) krijgen we:

$j$	1	2	3
$\hat{\sigma}_j^2$	0,01966	0,00332	0,00056

Nu kunnen we ook de  $\widehat{mse}(R_i)$  bepalen, deze staan in onderstaande tabel, waarin ook nog  $\widehat{R}_i$  herhaald staat.

$i$	2	3	4
$\widehat{R}_i$	5,2	14,7	42,3
$\widehat{mse}(R_i)$	0,1118	0,5385	3,05159
fout %	2,2%	3,6%	7,2%

## 3.2 Bootstrap

Het bepalen van een schatter alleen is meestal niet toereikend genoeg, daar is minstens een mate van de fout of variantie bij nodig. De Bootstrap, Efron (1979), is een methode om inzicht te krijgen in de verdeling van de fout  $R - \widehat{R} = \sum_{i+j>t+1} (C_{i,j} - \hat{x}_i \hat{y}_j)$ .

Bij de bootstrap bepalen we  $n$ , bijvoorbeeld  $n = 10.000$ , schadedriehoeken. Deze schadedriehoeken zijn gebaseerd op de originele schadedriehoek, dit kan parametrisch en niet-parametrisch. We kunnen nu  $n$  driehoeken volschatten en de verdeling bepalen van de schatters  $\widehat{C}_{i,j}$  voor  $i + j > t + 1$ .

### 3.2.1 Parametrische bootstrap

Het principe dat hierboven staat uitgelegd, is te gebruiken voor een parametrische bootstrap. In dit geval is er een verdeling van de  $C_{i,j}$  gespecificeerd. Stel dat de  $C_{i,j}$  onafhankelijk en

$$(a) \quad C_{i,j} \sim Pois(x_i y_j) \text{ en}$$

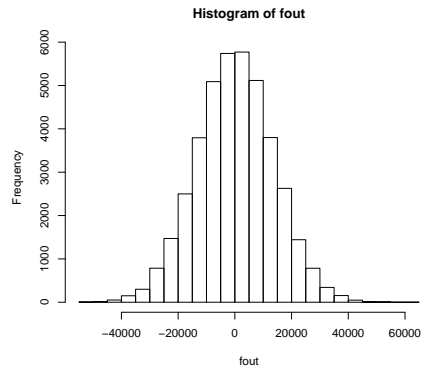
$$(b) \quad C_{i,j} \sim NegBinom(x_i y_j, \omega)$$

dan is  $E(C_{i,j}) = x_i y_j$  en gaan we als volgt te werk:

- (1) Bepaal  $\hat{x}_i$  en  $\hat{y}_j$  op basis van  $\{C_{i,j}, i + j \leq t + 1\}$
- (2) (a)  $\sim$ Poisson: Trek een steekproef  $C_{i,j}^*$  uit de Poisson( $\hat{x}_i \hat{y}_j$ ) verdeling
- (2) (b)  $\sim$ Negatief Binomiaal: Trek een steekproef  $C_{i,j}^*$  uit de Negatief Binomiale ( $\hat{x}_i \hat{y}_j, \hat{\omega}$ ) verdeling
- (3) Bepaal  $\hat{x}_i^*$  en  $\hat{y}_j^*$  op basis van  $\{C_{i,j}^*, i + j \leq t + 1\}$
- (4) Bepaal de fout  $R^* - \hat{R}^* = \sum_{i+j>t+1} (C_{i,j}^* - \hat{x}_i^* \hat{y}_j^*)$
- (5) Herhaal de stappen 2 tot en met 4  $n$  keer, bijvoorbeeld  $n = 10.000$

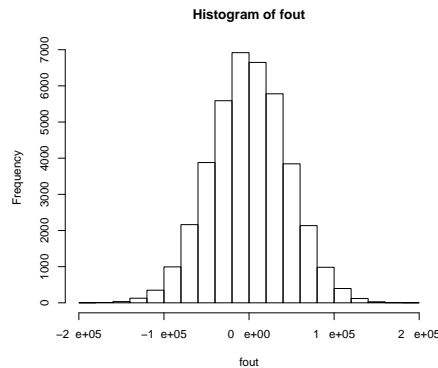
Aan het einde van de procedure hebben we ook hier een steekproef van  $n$  'bootstrap fouten'. De verdeling van de fout komt overeen met de verdeling van de steekproef indien de schatters  $\hat{x}_i$  en  $\hat{y}_j$  dicht bij de ware  $x_i$  en  $y_j$  liggen en als het Poisson model of Negatief Binomiaal model juist is.

In Appendix C staat een programma om in R de hierboven beschreven parametrische bootstrap uit te rekenen met het Poisson model. Als we dit op het voorbeeld toepassen uit Figuur 1.5 met  $n = 40.000$ , krijgen we in figuur 3.1 de fout met Poisson aanname en in figuur 3.2 de fout met Negatief Binomiaal aanname.



Figuur 3.1

De gemiddelde fout is 85,41 met standaardafwijking 13575.



Figuur 3.2

De gemiddelde fout is 35,02 met standaardafwijking 45170.

De standaardafwijking bij de Negatief Binomiaal aanname is ongeveer drie keer zo groot als de standaardafwijking bij de Poisson aanname. Ondanks dat we een kleine fout willen, is het aannemelijker om uit te gaan van de Negatief Binomiale verdeling. De reden hiervan is dat de verdeling van de fout in figuur 3.2 realistischer in de context te plaatsen is. Daarnaast past het Poisson model, waarbij de verwachting gelijk is aan de variantie, niet bij de data en is daardoor geen goed model om in deze context toe te passen.

### 3.2.2 Niet parametrische bootstrap

Stel dat  $C_{i,j} = x_i y_j \epsilon_{i,j}$  met  $x_i$  een factor voor schadejaar  $i$ ,  $y_j$  een factor voor afwikkeljaar  $j$  en  $\epsilon_{i,j}$  een multiplicatieve fout met  $\epsilon_{i,j}$  i.i.d,  $E(\epsilon_{i,j}) = 1$ ,  $Var(\epsilon_{i,j}) = \sigma^2$ . Daarmee is de verwachting  $E(C_{i,j}) = x_i y_j$  en  $Var(C_{i,j}) = x_i^2 y_j^2 \sigma^2$ . Als we aannemen dat de  $C_{i,j}$  onafhankelijk zijn, dan gaan we als volgt te werk:

- (1) Bepaal  $\hat{x}_i$  en  $\hat{y}_j$  op basis van  $\{C_{i,j}, i + j \leq t + 1\}$
- (2) Trek een steekproef  $\epsilon_{i,j}^*$  en bepaal  $C_{i,j}^* = \hat{x}_i \hat{y}_j \epsilon_{i,j}^*$
- (3) Bepaal  $\hat{x}_i^*$  en  $\hat{y}_j^*$  op basis van  $\{C_{i,j}^*, i + j \leq t + 1\}$
- (4) Bepaal de fout  $R^* - \hat{R}^* = \sum_{i+j > t+1} (C_{i,j}^* - \hat{x}_i^* \hat{y}_j^*)$
- (5) Herhaal de stappen 2 tot en met 4  $n$  keer, bijvoorbeeld  $n = 10.000$

Aan het einde van de procedure hebben we een steekproef van  $n$  'bootstrap fouten'. De verdeling van de fout komt overeen met de verdeling van de steekproef indien de schatters  $\hat{x}_i$  en  $\hat{y}_j$  dicht bij de ware  $x_i$  en  $y_j$  liggen en als de aanname van de multiplicatieve fout  $\epsilon$  juist is.

## 4 Loglineaire modellen

Een standaard lineair model gaat ervan uit dat de data additief is opgebouwd. Omdat in de praktijk veel data multiplicatief is opgebouwd, zoals bijvoorbeeld bij zorgverzekeringen, kan gebruik worden gemaakt van een loglineair model. Loglineaire modellen hebben de eigenschap dat de logaritmes van de incrementele schadeclaims, additief van opbouw worden, waardoor het gemakkelijker wordt om de data te 'ontleden'. De GLM is ook een voorbeeld van een loglineair model wanneer de logfunctie als linkfunctie gekozen wordt.

Het loglineaire model is opgebouwd uit een deterministische component en een normaal verdeelde storingsterm, McCullagh and Nelder. Daarnaast moet het loglineaire model voldoen aan de Gauss-Markov condities.

### 4.1 Schatters

$C_{ij}$  = incrementele schadeclaims met de extra eis:  $C_{ij} > 0$   
 $\varepsilon_{ij}$  = normaal verdeelde storingsterm, waarbij  $\varepsilon_{ij} \sim N(0, \sigma^2)$   
 $\eta_{ij}$  = deterministische component met  $\eta_{ij} = (X\vec{\beta})_{ij}$ , waarbij:  
 $\vec{\beta}$  = vector van onbekende parameters van dimensie  $q$   
 $X$  = designmatrix van dimensie  $\frac{1}{2}t(t+1)q$

Dan is het model:

$$Z_{ij} = \log(C_{ij}) = \eta_{ij} + \varepsilon_{ij}$$

De Gauss-Markov condities waaraan nog voldaan moet worden zijn:

1.  $Var(\varepsilon_{ij}) = \sigma^2$
2.  $Cov(\varepsilon_{ij}, \varepsilon_{kl}) = 0, \quad i \neq k, j \neq l$

Om vervolgens de verwachte schadeclaims te vinden, hebben we de formule:  $E(C_{ij}) = e^{(\eta_{ij} + \frac{1}{2}\hat{\sigma}^2)}$  met  $\hat{\sigma}^2$  de steekproef variantie en

$$\hat{\beta} = (X'X)^{-1}X'Z \tag{24}$$

$$\hat{\sigma}^2 = Var(Z - X\hat{\beta}) \tag{25}$$

Een voordeel van loglineaire modellen is dat je mogelijk sneller trends kunt herkennen als de data een enigszins exponentiële opbouw heeft. Daarnaast kun je zelf kiezen van welke parameters en variabelen uit de data je gebruik wil maken. Door deze zo gunstig mogelijk te kiezen, kun je al snel een redelijk goed model verkrijgen en overparameterisatie voorkomen. Verder is het handig dat een niet symmetrische (skewed) verdeling is toegestaan. Een nadeel is dat  $\log(C_{ij})$  ongedefinieerd is voor  $C_{ij} \leq 0$ . Bij grote onzekerheid is er grote variantie en kan er een explosie effect optreden. Hierdoor krijg je enerzijds geen juist beeld van de mogelijke schadeclaims, maar anderzijds

is het een goede waarschuwing dat er bij onzekerheid een grotere reserve aangehouden moet worden.

## 4.2 IBNR Voorziening

In dit model kun je verschillende keuzes maken voor welke parameters je wil gebruiken. Mogelijkheden zijn bijvoorbeeld het schadejaar, afwikkeljaar of kalenderjaar. Ongeacht de keuze hiervan, is de voorziening als volgt uit te drukken:

$$\begin{aligned} \text{IBNR voorziening: } \hat{R}_i &= \sum_{j=t+2-i}^t \hat{C}_{it} \\ &= \sum_{j=t+2-i}^t e^{\eta_{ij} + \frac{1}{2}\hat{\sigma}^2} \end{aligned} \quad (26)$$

In het voorbeeld is te zien hoe  $\eta_{ij}$  kan worden berekend voor  $i, j = 1, \dots, t$ .

### Voorbeeld

In dit voorbeeld zullen we gebruik maken van parameters per schadejaar ( $\xi_i$ ) en per afwikkeljaar ( $\gamma_j$ ). Om te zorgen dat  $X'X$  regulier is, maken we gebruik van de extra eis dat  $\sum_{j=1}^t \gamma_j = 0$ . De vector  $\vec{\beta}$  kan nu worden gedefinieerd als  $\beta = (\vec{\xi}, \vec{\gamma})^T = (\xi_1, \xi_2, \dots, \xi_t, \gamma_1, \dots, \gamma_{t-1})^T$ . De designmatrix  $X$  wordt dan voor het voorbeeld, waarbij  $t = 4$ :

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

$X$  is hierbij gebaseerd op  $i + j \leq t + 1$  ofwel op bekende data. De volgende stappen moeten worden ondernomen om de schatter te bepalen:

(1) Bepaal  $Z$ , bij het voorbeeld wordt het:

$$Z = (\log 60, \log 25, \log 10, \log 5, \log 63, \log 24, \log 11, \log 58, \log 23, \log 65)^T$$

(2) Bepaal  $\hat{\beta}$  met (24)

(3) Bereken de schatter voor  $\eta_{ij}$  met  $\hat{\eta}_{ij} = (X\hat{\beta})_{ij}$ , dit levert  $\hat{\eta}_{ij}$  op voor  $i + j \leq t + 1$ . Om ook een schatter te berekenen voor  $i + j > t + 1$ , moet

gebruik worden gemaakt van designmatrix  $\tilde{X}$  welke gebaseerd is op alle data:  
 $\hat{\eta}_{i,j} = (\tilde{X}\hat{\beta})_{ij}$ .

$$\tilde{X} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & -1 & -1 \end{pmatrix}$$

(4) Bepaal de schatters voor  $C_{i,j}$  met  $\widehat{C}_{i,j} = e^{\hat{\eta}_{i,j} + \frac{1}{2}\hat{\sigma}^2}$  met  $\hat{\sigma}^2$  als in (25).  
 Dit resulteert in de schatters in figuur 4.1.

schadejaar	afwikkeljaar			
	1	2	3	4
1	60	25	10	5
2	63	24	11	<b>5,2</b>
3	58	23	<b>9,9</b>	<b>4,8</b>
4	65	<b>25,9</b>	<b>11,1</b>	<b>5,4</b>

Figuur 4.1

De totale voorziening wordt:  $\hat{R} = 5,2 + 9,9 + 4,8 + 25,9 + 11,1 + 5,4 = 62,2$

Wanneer er behalve de data uit de schadedriehoek meer data uit eerdere schadejaren bekend is, kan dit zonder problemen worden meegenomen. Deze methode gaat namelijk uit van een designmatrix met bijbehorende datavector. Als we de eerdere schadejaren vertalen naar de datavector en hier de designmatrix op aanpassen, krijgen we een schatter waar ook de extra toegevoegde data in wordt meegenomen.

### 4.3 Bootstrap

Om de foutmarge van de IBNR-schade schatter te bepalen, kunnen we gebruik maken van de bootstrap methode. De volgende stappen moeten dan

uitgevoerd worden.

- (1) Schat  $\hat{\eta}_{i,j}$  en de steekproefvariantie  $\hat{\sigma}^2$  zoals beschreven in de punten 1 tot en met 4 hierboven
- (2) Trek een steekproef  $\epsilon^*$  uit de  $N(0, \hat{\sigma}^2)$  verdeling en bepaal  $C_{i,j}^* = e^{\hat{\eta}_{i,j} + \frac{1}{2}\epsilon}$
- (3) Schat  $\hat{\eta}_{i,j}^*$  aan de hand van  $C_{i,j}^*$
- (4) Bepaal de fout  $R^* - \hat{R}^* = \sum_{i+j > t+1} (C_{i,j}^* - e^{\hat{\eta}_{i,j} + \frac{1}{2}\hat{\sigma}^2})$
- (5) Herhaal de stappen 2 tot en met 4  $n$  keer, bijvoorbeeld  $n = 10.000$



## 5 Alternatieve methoden

In de verschillende modellen die in dit hoofdstuk worden besproken, maken we gebruik van diverse aannames. Bijvoorbeeld dat alle  $C_{i,j}$  onafhankelijk of een aanname over  $E(C_{i,j})$ . Voor eenvoudige modellen kunnen we vervolgens schatters bepalen aan de hand van een formule waarin alleen enkele waarden ingevuld moeten worden. Deze methoden gaan niet specifiek in op de variantie of marge van de schatter. Ze zijn redelijk 'recht toe recht aan' opgebouwd en daardoor goed uit te leggen aan de niet-technici. In de praktijk is het van belang dat bijvoorbeeld het management grofweg begrijpt waar schatters van voorzieningen vandaan komen. Van dit soort methoden wordt vaak in eerste instantie gebruik gemaakt om een 'gevoel' te krijgen van de verwachte schadeclaims.

In de loop der jaren zijn er redelijk veel van dit type IBNR methoden ontwikkeld. In 'David Skurnick, A Survey of Loss Reserving Methods' wordt een uitgebreid overzicht gegeven van ruim 30 verschillende methoden (zie hiervoor Appendix D).

Onder de minder eenvoudige methoden verstaan we Maximum-likelihood methoden, dit is een aanvulling op de eenvoudige methoden. Naast gebruik te maken van aannames over onderlinge onafhankelijkheid, wordt er ook gebruik gemaakt van een kansverdeling. Hierdoor kan zowel de meest waarschijnlijke schatter (maximum likelihood estimator) als de variantie worden bepaald.

De laatst genoemde methode heeft de voorkeur boven de recht-toe-recht-aan methode. Een reden hiervan is dat bijvoorbeeld een advies aan het management over een hoogte van de voorziening meer waarde heeft als erbij vermeld kan worden dat dit bedrag voor 80 % adequaat is. Als een bedrag van 300 miljoen 80 % adequaat is, ofwel  $P(R \leq 300 \cdot 10^6) = 0,8$ , en een bedrag van 320 miljoen voor 90 % adequaat, ofwel  $P(R \leq 320 \cdot 10^6) = 0,9$ , dan kan het management beslissen welk risico ze nemen en afwegen welke voorzieningen ze treffen. Een nadeel is dat maximum likelihood methoden ingewikkelder zijn opgebouwd en daardoor moeilijker uit te leggen aan niet-technici.

Ongeacht de methode is het van belang dat de schatter zuiver is. Hieronder zal ik ingaan op de maximum likelihood schatter en de zuiverheid van een aantal veel gebruikte methoden. Daarnaast zal ik met een voorbeeld de methoden illustreren. Voor het voorbeeld zal ik gebruik maken van de cumulatieve schadedriehoek uit figuur 1.5, waarbij  $t = 4$ . Er is op te merken dat veel methoden uiteindelijk op hetzelfde neerkomen.

Doordat we bij de diverse methoden gebruik maken van de aanname dat de

data een Poisson verdeling heeft, mogen we gebruik maken van extra data uit eerdere schadejaren. De reden hiervan is in 2.2 toegelicht.

## 5.1 Variant I

We beginnen met een variant op de Chain Ladder methode. De reden van de variant is om tegemoet te komen aan het nadeel van de grote hoeveelheid parameters. We passen het Chain Ladder model zodanig aan dat deze niet meer afhankelijk is van het schadejaar. In dit geval maken we gebruik van de aanname  $C_{i,j} \sim Pois(xy_j)$  waarbij  $E(C_{i,j}) = xy_j$ . Bij dit model is de fractie gedefinieerd als  $y_j = \frac{\sum_{i=1}^t C_{i,j}}{\sum_{i=1}^t \sum_{j=1}^t C_{i,j}}$  en  $x = \frac{1}{t} \sum_{i=1}^t \sum_{j=1}^t C_{i,j}$ . Wanneer we de data eerst corrigeren op jaareffecten zoals aantallen verzekerden en zorginflatie, dan is de aanname dat de schade onafhankelijk is van het schadejaar een hele redelijke. We kunnen het volgende afleiden:

$$\begin{aligned}
 P(C_{i,j}) &= \frac{(xy_j)^{C_{i,j}} \cdot e^{-xy_j}}{C_{i,j}!} \\
 \log \text{lik} : l &= \sum_{i+j \leq t+1} (C_{i,j} \log xy_j - xy_j - \log C_{i,j}!) \\
 \frac{\partial l}{\partial y_j} &= \sum_{i=1}^{t+1-j} \left( \frac{C_{i,j}}{y_j} - x \right) = 0 \\
 \hat{y}_j &= \frac{\sum_{i=1}^{t+1-j} C_{i,j}}{(t+1-j)\hat{x}} \tag{27}
 \end{aligned}$$

In de praktijk kunnen we als volgt te werk gaan om  $\hat{x}$  en  $\hat{y}_j$  uit te rekenen. We maken gebruik van  $\sum_{j=1}^t \hat{y}_j = 1$  en (27), hiermee:

$$\sum_{j=1}^t \frac{\sum_{i=1}^{t+1-j} C_{i,j}}{(t+1-j)\hat{x}} = 1$$

ofwel

$$\hat{x} = \sum_{j=1}^t \frac{\sum_{i=1}^{t+1-j} C_{i,j}}{(t+1-j)} \tag{28}$$

We merken op dat er geen stapsgewijze procedure meer nodig is om  $\hat{x}$  te bepalen, zoals dat eerder bij met de schadejaar-afhankelijke  $\hat{x}_i$  nog wel het geval was. Voor het schatten van de IBNR voorziening is het ook niet meer nodig om  $\hat{y}_j$  te bepalen.

Dat de schatter  $\hat{x}_i$  zuiver is, volgt uit:

$$E(\hat{x}) = E\left(\sum_{j=1}^t \frac{\sum_{i=1}^{t+1-j} C_{i,j}}{(t+1-j)}\right)$$

$$\begin{aligned}
&= \sum_{j=1}^t \frac{\sum_{i=1}^{t+1-j} E(C_{i,j})}{(t+1-j)} \\
&= \sum_{j=1}^t \frac{\sum_{i=1}^{t+1-j} xy_j}{(t+1-j)} \\
&= \sum_{j=1}^t \frac{(t+1-j)xy_j}{(t+1-j)} \\
&= \sum_{j=1}^t xy_j \\
&= x
\end{aligned}$$

### IBNR Voorziening

De voorziening voor schadejaar  $i$  kan worden geschat door de som te nemen over de nog te verwachte schadeclaims  $\sum_{j=t+2-i}^t \hat{C}_{i,j}$ :

$$\begin{aligned}
\text{IBNR voorziening: } \hat{R}_i &= \sum_{j=t+2-i}^t \hat{C}_{i,j} \\
&= \hat{x} \sum_{j=t+2-i}^t \hat{y}_j \quad (29)
\end{aligned}$$

### Voorbeeld

Voor het voorbeeld uit figuur 1.5 bepalen we eerst  $\hat{x}$  met  $\hat{x} = \sum_{j=1}^4 \frac{\sum_{i=1}^{5-j} C_{i,j}}{(5-j)}$ :

$$\hat{x} = \frac{60 + 63 + 58 + 65}{4} + \frac{25 + 24 + 23}{3} + \frac{10 + 11}{2} + \frac{5}{1} = 101$$

Vervolgens bepalen we  $\hat{y}_j$  met (27):

$j$	1	2	3	4
$\hat{y}_j$	0,609	0,238	0,104	0,049

We kunnen nu de driehoek volschatten door gebruik te maken van  $E(C_{i,j}) = xy_j$ :

schadejaar	afwikkeljaar			
	1	2	3	4
1	60	25	10	5
2	63	24	11	<b>5,0</b>
3	58	23	<b>10,5</b>	<b>5,0</b>
4	65	<b>24,0</b>	<b>10,5</b>	<b>5,0</b>

Figuur 5.1

Ofwel de totale voorziening wordt:  $\hat{R} = 24,0 + 2 \cdot 10,5 + 3 \cdot 5,0 = 60,0$ .

## 5.2 Accident Year Incurred Loss Development methode

Bij de Accident Year Incurred Loss Development methode (AYILD) wordt gebruik gemaakt van de aanname dat er op bepaalde tijdstippen (bijvoorbeeld na ieder afwikkeljaar) ook bepaalde percentages van de totale schade per schadejaar zijn geclaimd. Er wordt dus vanuit gegaan dat de afwikkeling van schaden in tijd redelijk stabiel is. Een aanname kan in deze bijvoorbeeld zijn dat na het eerste afwikkeljaar er 60%, na het tweede afwikkeljaar 85%, na het derde jaar 95% en na het vierde jaar er 100% geclaimd is.

### Schatters

$y_j$  = fractie van de totale schade in afwikkeljaar  $j$ , hierbij geldt  $\sum_{j=1}^t y_j = 1$  en  $y_j \geq 0 \forall j$ .

Stel dat we gebruik maken van de aanname dat  $C_{i,j}$  Poisson verdeeld is en er geldt  $E(C_{i,j}) = y_j \sum_{j=1}^t C_{i,j}$ . Dan hebben we  $C_{i,j} \sim Poisson(y_j \sum_{j=1}^t C_{i,j})$  en is er op te merken dat de AYILD methode exact overeen komt met de Chain Ladder vanwege  $E(\sum_{j=1}^t C_{i,j}) = x_i$ . In de literatuur wordt de AYILD methode als een op zichzelf staande methode gezien, wat dus niet waar blijkt te zijn.

## 5.3 Loss Ratio methode

Bij de Loss Ratio methode wordt gebruik gemaakt van de aanname dat de verhouding tussen de som van de netto schade en de bruto (of netto) ontvangen premie voor dat schadejaar, constant blijft en eenvoudig te bepalen is. De factor tussen deze twee waarden heet de schadefactor en kan worden berekend door de som van de netto schade te delen door de bruto (of netto) premie voor dat schadejaar. Samen met de reeds bekende geboekte schade, is de IBNR voorziening uit te rekenen.

### Schatters

$p_i$  = verdiende premie (bruto of netto) van schadejaar  $i$

$h$  = schadefactor, waarbij  $h = \frac{D_{i,t}}{p_i}$

Een lineair model is als volgt te definiëren:  $D_{i,t} = \sum_{j=1}^t C_{i,j} = hp_i + \epsilon$  met  $\epsilon$  de ruis van het model. Een mogelijke schatter voor  $h$  is  $\hat{h} = \frac{D_{1,t}}{p_1}$  met  $E(\hat{h}) = E(E(\hat{h}|p_1)) = E(h) = h$  een zuivere schatter van  $h$ . Verder geldt  $E\left(\frac{\sum_{j=1}^t C_{i,j}}{p_i}\right) = h$ . Dit model correspondeert met het Chain Ladder model zonder schadejaarafhankelijkheid zoals Variant I in 5.1 met  $E\left(\frac{C_{i,j}}{p_i}\right) = xy_j$ , waarbij  $x = h$ .

### IBNR Voorziening

Met bovenstaande gegevens hebben we voor de Loss Ratio methode het volgende model:

$$\text{IBNR voorziening: } \hat{R}_i = \hat{h} \cdot p_i - D_{i,t+1-i} \quad (30)$$

Een voordeel is dat deze methode gemakkelijk is om uit te voeren en uit te leggen. Daarnaast is de methode onafhankelijk van de verwerkingssnelheid van schadeclaims. Een nadeel is dat de methode afhankelijk is van het goed kunnen schatten van  $h$ , de schadefactor. Wanneer de premie niet goed is aangepast aan de zorginflatie, dan krijg je een andere verhouding en andere schatters voor de IBNR voorziening.

### Voorbeeld

Stel dat onderstaande tabel een overzicht is van de verdiende premies van schadejaar 1 (2000) tot en met 4 (2003):

schadejaar	1	2	3	4
$p_i$	120	125	129	131

Wanneer we de grove schatter  $\hat{h} = \frac{D_{1,t}}{p_1}$  toepassen, dan  $\hat{h} = \frac{100}{120} = \frac{5}{6}$ . Hiermee kunnen we de schadedriehoek invullen en de IBNR Voorziening bepalen.

schadejaar	afwikkeljaar			
	1	2	3	4
1	60	85	95	100
2	63	87	98	<b>104,2</b>
3	58	81		<b>107,5</b>
4	65			<b>109,2</b>

Figuur 5.2

Ofwel de totale voorziening wordt:

$$\hat{R} = (104,2 - 98) + (107,5 - 81) + (109,2 - 65) = 76,9$$

Door de grove manier van het schatten van  $h$ , krijgen we ook hier een grove schatter voor de totale voorziening. In de volgende paragraaf wordt dit model iets verder verfijnd, zodat we een beter model krijgen.

### 5.4 Bornhüter Ferguson methode

De Bornhüter Ferguson methode combineert de Loss Ratio methode met de Accident Year Incurred Loss Development methode. Er wordt dus zowel vanuit gegaan dat de verhouding tussen de som van de netto schade en de bruto (of netto) premie voor dat schadejaar, constant blijft als dat er op bepaalde tijdstippen ook bepaalde percentages van de totale schade per

schadejaar zijn geclaimd.

Als eerste wordt de som van de totale schade in een schadejaar bepaald met behulp van de schadefactor, zoals in de Loss Ratio methode. Vervolgens wordt met behulp van de verwachte fractie van de totaalschade bekeken welk deel nog geclaimd gaat worden. De schatter van de incrementele schadeclaims  $C_{i,j}$  kunnen worden berekend door de fractie van de nog te verwachte claims ( $\sum_{j=t+2-i}^t \hat{y}_j$ ) te vermenigvuldigen met het geschatte totaalbedrag ( $h \cdot p_i$ ).

### Schatters

$y_j$  = fractie van de totale schade in afwikkeljaar  $j$

$h$  = schadefactor, waarbij  $h = \frac{D_{i,t}}{p_i}$

$p_i$  = verdiende premie (bruto of netto) van schadejaar  $i$

De verwachting van de incrementele schade is te formuleren als  $E(C_{i,j}) = hp_i y_j$ , ofwel  $E\left(\frac{C_{i,j}}{p_i}\right) = h y_j$ . Er valt op te merken dat dit overeen komt met variant I zoals beschreven in 5.1 met als enige verschil dat er nu gekeken wordt naar een op premie gecorrigeerde verwachting van de schade en  $x = h$ . Analoog aan Variant I, vinden we hier met de maximum likelihood methode de schatters:

$$\hat{y}_j = \frac{\sum_{i=1}^{t+1-j} \frac{C_{i,j}}{p_i}}{(t+1-j)\hat{h}}$$

en

$$\hat{h} = \sum_{j=1}^t \frac{\sum_{i=1}^{t+1-j} \frac{C_{i,j}}{p_i}}{(t+1-j)}$$

### IBNR Voorziening

De voorziening kan als volgt worden geschat:

$$IBNR \text{ voorziening} : \hat{R}_i = h \cdot p_i \sum_{j=t+2-i}^t y_j \quad (31)$$

Aangezien deze methode een combinatie is van twee redelijk eenvoudige methodes, is dit opnieuw ook een eenvoudige methode die goed uitgelegd kan worden. Een nadeel is dat er voorbij wordt gegaan aan de correlatie tussen de reeds geboekte schade en de voorziening voor toekomstige schadeclaims, terwijl deze zeker aanwezig is.

### Voorbeeld

Bij dit model zal ik ook gebruik maken van de verdiende premies  $p_i$  zoals deze in voorbeeld 5.3 worden gebruikt. We kunnen nu de schatter  $\hat{h}$  uitrekenen met  $\hat{h} = \sum_{j=1}^4 \frac{\sum_{i=1}^{5-j} \frac{C_{i,j}}{p_i}}{(5-j)} = 0,808$ , hiermee kunnen we  $y_j$  uitrekenen:

$y_j = (0,604 \ 0,239 \ 0,106 \ 0,052)$ . Met  $E(C_{i,j}) = hp_i y_j$  kunnen we nu de schadedriehoek volschatten.

schadejaar	afwikkeljaar			
	1	2	3	4
1	60	25	10	5
2	63	24	11	<b>5,21</b>
3	58	23	<b>11,05</b>	<b>5,38</b>
4	65	<b>25,27</b>	<b>11,22</b>	<b>5,46</b>

Figuur 5.3

Ofwel de totale voorziening wordt:

$$\hat{R} = 5,21 + 11,05 + 5,38 + 25,27 + 11,22 + 5,46 = 63,6$$

## 5.5 Aritmetische separatie methode

De Aritmetische separatie methode houdt rekening met trends zowel in de richting van het afwikkeljaar als het kalenderjaar. De aanname is dat  $C_{ij}$  de gemiddelde schade bevat en de realisatie is van een onafhankelijke Poissonverdeling.

### Schatters

$y_j$  = fractie van de totale schade in afwikkeljaar  $j$

$\lambda_{i+j-1}$  = parameter voor het kalenderjaar  $i + j - 1$

De verwachting van de incrementele schade is  $E(C_{ij}) = y_j \lambda_{i+j-1}$ .

Met bovenstaande aanname, ziet de schadedriehoek met  $t = 4$  er als volgt uit:

schadejaar	afwikkeljaar			
	1	2	3	4
1 (2000)	$y_1 \lambda_1$	$y_2 \lambda_2$	$y_3 \lambda_3$	$y_4 \lambda_4$
2 (2001)	$y_1 \lambda_2$	$y_2 \lambda_3$	$y_3 \lambda_4$	.
3 (2002)	$y_1 \lambda_3$	$y_2 \lambda_4$	.	.
4 (2003)	$y_1 \lambda_4$	.	.	.

Figuur 5.4

Om schatters voor  $y_j$  en  $\lambda_{i+j-1}$  te vinden, gebruiken we weer de maximum likelihood methode, met  $C_{i,j} \sim Pois(y_j \lambda_{i+j-1})$ .

$$P(C_{i,j}) = \frac{(y_j \lambda_{i+j-1})^{C_{i,j}} e^{-y_j \lambda_{i+j-1}}}{C_{i,j}!}$$

$$\begin{aligned}
\log lik : l &= \sum_{i+j \leq t+1} C_{i,j} \log y_j + \sum_{i+j \leq t+1} C_{i,j} \log \lambda_{i+j-1} \\
&\quad - \sum_{i+j \leq t+1} y_j \lambda_{i+j-1} - \sum_{i+j \leq t+1} \log C_{i,j}! \\
\frac{\partial l}{\partial \lambda_{i+j-1}} &= \frac{\sum_{k=1}^{i+j-1} C_{k,i+j-k}}{\lambda_{i+j-1}} - \sum_{k=1}^{i+j-1} y_k = 0 \\
\hat{\lambda}_{i+j-1} &= \frac{\sum_{k=1}^{i+j-1} C_{k,i+j-k}}{\sum_{k=1}^{i+j-1} \hat{y}_k} \tag{32} \\
\frac{\partial l}{\partial y_j} &= \frac{\sum_{i=1}^{t+1-j} C_{i,j}}{y_j} - \sum_{i=1}^{t+1-j} \lambda_{i+j-1} = 0 \\
\hat{y}_j &= \frac{\sum_{i=1}^{t+1-j} C_{i,j}}{\sum_{i=1}^{t+1-j} \hat{\lambda}_{i+j-1}} \tag{33}
\end{aligned}$$

In de praktijk zullen  $y_j$  en  $\lambda_{i+j-1}$  als volgt worden geschat:

$$\begin{aligned}
i+j-1 = t : \quad \hat{\lambda}_t &= \frac{\sum_{k=1}^t C_{k,t+1-k}}{\sum_{k=1}^t \hat{y}_k} = \sum_{k=1}^t C_{k,t+1-k} \\
j = t : \quad \hat{y}_t &= \frac{C_{1,t}}{\hat{\lambda}_t} \\
i+j-1 = t-1 : \quad \hat{\lambda}_{t-1} &= \frac{\sum_{k=1}^{t-1} C_{k,t-k}}{\sum_{k=1}^{t-1} \hat{y}_k} = \frac{\sum_{k=1}^{t-1} C_{k,t-k}}{1 - \hat{y}_t} \\
j = t-1 : \quad \hat{y}_{t-1} &= \frac{\sum_{i=1}^2 C_{i,t-1}}{\sum_{i=1}^2 \hat{\lambda}_{t-1}} \\
&\quad \vdots
\end{aligned}$$

En zo verder totdat alle  $\hat{y}_j$  en  $\hat{\lambda}_{i+j-1}$  bekend zijn. Om vervolgens een schatting te maken van de komende jaren, zullen nog waarden van  $\lambda_k$  voor  $k = t+1, \dots, 2t-1$  geschat moeten worden. Dit kan worden gedaan door lineaire of exponentiële extrapolatie.

Opgemerkt kan worden, dat op een parameterisatie op kalenderjaren versus een parameterisatie op shadejaren na, dit model overeen komt met de Chain Ladder. Omdat hierdoor de mle schatters overeenkomen, geldt automatisch dat de schatters  $\hat{y}_j$  en  $\hat{\lambda}_{i+j-1}$  zuiver zijn vanwege de zuiverheid van de schatters  $\hat{y}_j$  en  $\hat{x}_i$  uit de Chain Ladder.

### IBNR Voorziening

Uiteindelijk kunnen we de shadedriehoek verder invullen met  $\hat{C}_{ij} = \hat{y}_j \hat{\lambda}_{i+j-1}$



en kunnen we de voorziening bepalen.

$$\begin{aligned}
 \text{IBNR voorziening : } \widehat{R}_i &= \sum_{j=t+2-i}^t \widehat{C}_{it} \\
 &= \sum_{j=t+2-i}^t \widehat{y}_j \widehat{\lambda}_{i+j-1} \quad (34)
 \end{aligned}$$

Een voordeel van deze methode, is dat het zowel het afwikkeljaar als het kalenderjaar meeneemt. Een nadeel dat hieruit voorkomt, is dat je genoeg data moet hebben om overparameterisatie te voorkomen. Daarnaast blijf de methode onduidelijk over wat een goede methode is om  $\lambda_k$  te extrapoleren.

### Voorbeeld

Met behulp van (32) en (33) leiden we de schatters af voor  $\lambda_{i+j-1}$  en  $y_j$ :

$$\begin{aligned}
 \widehat{\lambda}_4 &= \frac{\sum_{i=1}^t C_{i,t+1-i}}{1} = 5 + 11 + 23 + 65 = 104 \\
 \widehat{y}_4 &= \frac{C_{1,t}}{\lambda_t} = \frac{5}{104} \approx 0,05 \\
 \widehat{\lambda}_3 &= \frac{\sum_{i=1}^{t-1} C_{i,t-i}}{1 - y_4} = \frac{10 + 24 + 58}{\frac{99}{104}} \approx 96,65 \\
 \widehat{y}_3 &= \frac{C_{1,t-1}}{\lambda_t - 1} = \frac{10 + 11}{104 + 96,6} \approx 0,10 \\
 \widehat{\lambda}_2 &= \frac{\sum_{i=1}^{t-2} C_{i,t-1-i}}{y_1 + y_2} = \frac{63 + 25}{1 - y_4 - y_3} = \frac{88}{0,85} \approx 103,87 \\
 \widehat{y}_2 &= \frac{C_{1,t-2} + C_{2,t-2} + C_{3,t-2}}{\lambda_t + \lambda_{t-1} + \lambda_{t-2}} = \frac{25 + 24 + 23}{104 + 96,65 + 103,87} \approx 0,24 \\
 \widehat{\lambda}_1 &= \frac{\sum_{i=1}^{t-3} C_{i,t-2-i}}{y_1} = \frac{60}{1 - y_2 - y_3 - y_4} \approx 98,36 \\
 \widehat{y}_1 &= 1 - y_2 - y_3 - y_4 \approx 0,61
 \end{aligned}$$

Ofwel:

$k$	$y_k$	$\lambda_k$
1	0,61	98,4
2	0,24	103,9
3	0,10	96,6
4	0,05	104

Er is niet direct een duidelijk verband af te leiden tussen de verschillende  $\lambda_k$ 's. Stel dat in dit geval de  $\lambda_k$ 's worden aangeleverd:

$k$	5	6	7
$\lambda_k$	97	105	99

Dan kunnen we als volgt de driehoek volschatten:

schadejaar	afwikkeljaar			
	1	2	3	4
1	60	25	10	5
2	63	24	11	<b>4,9</b>
3	58	23	<b>9,7</b>	<b>5,3</b>
4	65	<b>23,3</b>	<b>10,5</b>	<b>5,0</b>

Figuur 5.5

Ofwel de totale voorziening wordt:

$$\widehat{R} = 4,9 + 9,7 + 4,3 + 23,3 + 10,5 + 5,0 = 57,7.$$

### Toevoeging

Omdat de hierboven beschreven Aritmetische separatie methode onduidelijk blijft over het vinden van goede schatters voor  $\lambda_k$  voor  $k > t$ , is er op dit punt van deze methode verbetering mogelijk. De methode beschrijft dat de benodigde  $\lambda$ 's te vinden zijn door ervan uit te gaan dat (i) er een lineair verband is of (ii) een exponentieel verband. Bij deze twee mogelijkheden is het model  $E(C_{i,j}) = y_j \lambda_{i+j-1}$  scherper te formuleren, namelijk als volgt:

$$\begin{aligned} (i) \quad E(C_{i,j}) &= y_j \cdot (\theta_{i+j-1} + \rho) \quad \text{of} \\ (ii) \quad E(C_{i,j}) &= y_j \theta e^{\rho_{i+j-1}} \end{aligned}$$

Ook hier is dan na te gaan wat een geschikte mle is en of deze zuiver is.

## 6 Het meest geschikte model

### 6.1 Modelkeuze

In de loop der tijd zijn er vele verschillende modellen gemaakt en gebruikt. De verschillende modellen maken gebruik van zeer diverse parameters of regressoren. De meest voor de hand liggende regressoren zijn:

1. schadejaar
2. afwikkeljaar
3. ontvangen premie per jaar
4. aantallen verzekerden
5. (zorg) inflatie
6. kalenderjaar

Om overparameterisatie te voorkomen, kunnen nooit alle mogelijke regressoren gebruikt worden. De meeste modellen maken gebruik van maximaal drie typen regressoren zoals ze hierboven genoemd zijn. Bijna alle modellen zijn zodanig opgebouwd dat als ze bijvoorbeeld gebruik maken van schadejaren, dat alle schadejaren worden meegenomen. Terwijl het soms handiger is om afwijkende schadejaren weg te laten om geen afwijkende schatters te krijgen vanwege de gevoeligheid van dergelijke modellen.

De vraag is wat nu het meest geschikte model is in de verschillende contexten, in het bijzonder in de context 'Zorgverzekeraar'. Hierover zijn diverse artikelen verschenen welke pleiten voor allemaal verschillende modellen. Om een model te maken dat alleen maar gebruik maakt van de meest bijdragende parameters, gaan we gebruik maken van Forward Stagewise Regression.

### 6.2 Forward Stagewise Regression

Forward Stagewise Regression (FSR) is een methode om de parameters te bepalen die het meest bijdragen aan het model. Deze methode begint met een model dat maar één parameter bevat, FRS bepaalt welke ene parameter zorgt voor het minst slechte model bij de geobserveerde data op basis van de kleinste kwadraten. In de tweede stap bepaalt FRS welke tweede parameter het model zo goed mogelijk verbetert. Dit gaat zo verder totdat je achtereenvolgens weet welke parameters je kunt toevoegen om het model zo goed als mogelijk te verbeteren. Tot slot kun je dan kiezen of je gebruik wil maken van de eerste drie, vier of tien parameters. De keuze van het aantal parameters hangt samen met de hoeveelheid geobserveerde data. Een mogelijke keuze kan zijn dat alle schadejaren en afwikkeljaren worden

gebruikt, dit model zal dan vergelijkbaar zijn met het Chain Ladder model. Met behulp van Cross Validation is meer te zeggen over het optimale aantal te gebruiken parameters voor het model.

Door gebruik te maken van FSR hoeven we dus geen vaststaand model te kiezen, maar zetten we alle mogelijke modellen in een lange vector. FSR kiest dan uit de vector de meest geschikte parameters om het uiteindelijke model te vormen.

Het principe van methoden als FSR, Lasso (Tibshirani, 1996) of LAR (Least Angle Regression) is de volgende:

Als eerste werken we met de schatters  $X_{h,k}$  voor  $k = 1, \dots, p$  welke een vector is van schatters vergelijkbaar met  $x_i, y_j, r_j, f_j, p_i, h$ , etc. zoals deze gebruikt zijn in de modellen uit hoofdstuk 2, 4 en 5. Verder hebben we de uitkomsten uit geobserveerde data:  $C_h$  voor  $h = 1, \dots, N$ , welke een vector is van de data uit de matrix van  $C_{i,j}$  waarbij  $i + j \leq t + 1$ . Dit geeft het lineaire model:

$$f(x) = \beta_0 + \sum_{k=1}^p X_{h,k} \beta_k$$

met het criterium:

$$\min_{\beta} \sum_{i+j \leq t+1} \left( C_{i,j} - \beta_0 - \sum_k X_{h,k} \beta_k \right)^2 \quad \text{waarbij} \quad \sum_{k=1}^p |\beta_k| \leq s$$

Als  $s$  groot genoeg is, levert dit een gewone kleinste kwadraten schatter op. Hoe kleiner  $s$ , hoe kleinere  $\hat{\beta}$ , met veel componenten van  $\beta$  gelijk aan nul. De keuze van de grootte van  $s$  hangt hierdoor samen met de keuze van het aantal voorspellende parameters van  $\beta$  (ongelijk nul) welke mogen meedoen in het regressie model. Per stap wordt  $s$  vergroot zodat er per stap een  $\beta_k$  bijkomt, net zo lang totdat alle parameters in het model zitten. Er bestaan statistische computerprogramma's, zoals bijvoorbeeld  $R$ , die dit kunnen doorrekenen.

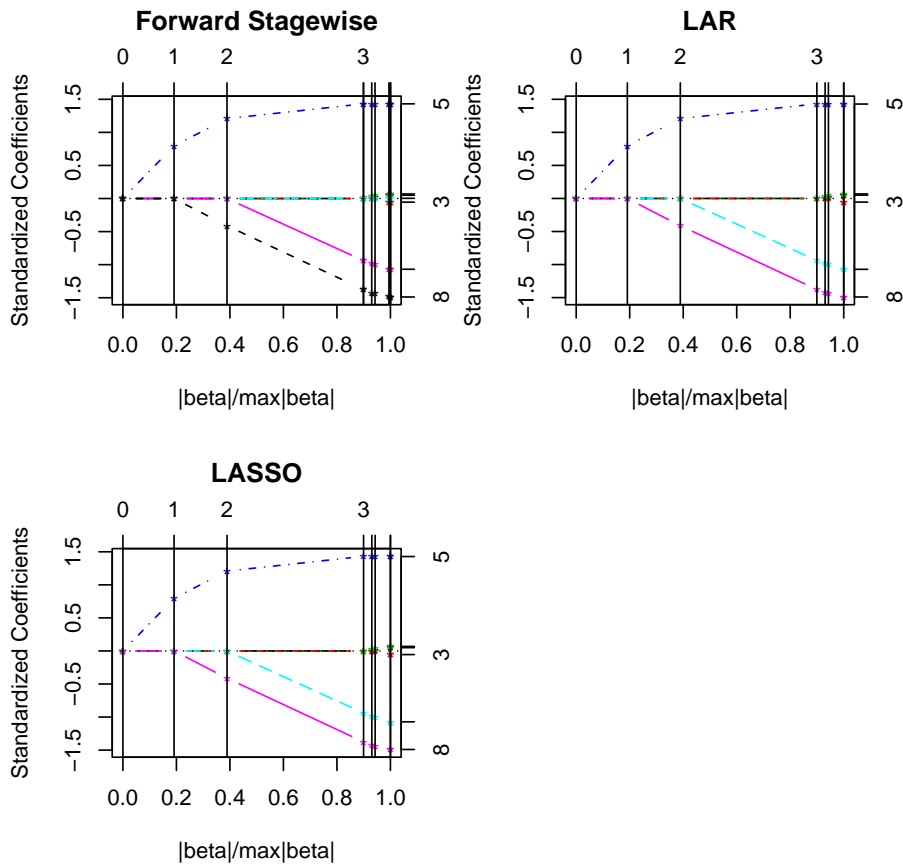
### Voorbeeld

In dit voorbeeld zullen we uitgaan van een multiplicatief verband tussen de data en gebruik maken van een loglineair model. Dit resulteert erin dat we de logaritme van de data invoeren. Bij het voorbeeld van de shadedriehoek uit figuur 1.5, zullen we laten zien hoe FSR werkt waarbij we gebruik maken van  $R$ . De functie in  $R$  heet `lars` en heeft als input een datavector  $C$  en een designmatrix  $X$ . Als eerste hebben we de vector van geobserveerde data  $C = (\log 60, \log 25, \log 10, \log 5, \log 63, \log 24, \log 11, \log 58, \log 23, \log 65)^T$ . Voor de schatters  $X$  maken we de keuze voor schadejaren (eerste vier kolommen), afwikkeljaren (vijfde tot en met de achtste kolom) en ten slotte een weegfactor per schadejaar in de laatste kolom. De weegfactor kan

geïnterpreteerd worden als aantallen verzekerden of inflatie. Dit levert de volgende designmatrix op:

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 4 \end{pmatrix}$$

De functie  $fit = lars(X, C, type = "for, lar of lasso")$  levert het volgende op:



Figuur 6.1

### Forward Stagewise Regression

R-squared: 0.999

Sequence of Forward Stagewise moves:

Var	5	8	7	4	-5	5	-5	2	3	6	-1	-9	-5
Step	1	2	3	4	4	5	5	6	7	8	9	9	10

### LAR

R-squared: 0.999

Sequence of LAR moves:

Var	5	8	7	4	2	3	-1	-6	-9
Step	1	2	3	4	5	6	7	7	7

### Lasso

R-squared: 1

Sequence of LASSO moves:

Var	5	8	7	4	2	3	-1	-6	-9
Step	1	2	3	4	5	6	7	7	7

Een negatieve waarde voor 'Var' geeft aan dat deze kolom lineair afhankelijk is van een al gebruikte kolom en daardoor niet mee zal doen in het model. In de parameter  $beta[i]$  zitten de waarden voor  $\beta$ , zoals in formule (25), waarbij  $i$  parameters aan het model meedoen. In dit voorbeeld valt het op dat het eerste schadejaar en de weegfactor per schadejaar belangrijk zijn om mee te nemen in het model.

De functie  $predict.lars(fit, X)$  genereert een matrix met in de  $i^e$  kolom, de schatters voor het model met  $i$  parameters. Per kolom staan schatters voor de waarden van de logaritme van  $C$ . Om schatters te krijgen voor alle data, en dus niet alleen voor de bovendriehoek, moeten we  $X$  aanpassen. De aangepaste designmatrix  $X$  gaat dan over alle data in plaats van alleen over de bovendriehoek en is op dezelfde manier in te vullen in de functie  $predict.lars$ . Met formule (25) kunnen we de steekproefvariantie  $\hat{\sigma}^2$  uitrekenen en vervolgens kunnen we met  $E(C_{i,j}) = e^{\eta_{i,j} + \frac{1}{2}\hat{\sigma}^2}$  een schatter bepalen voor de verwachte schadeclaims.

Wanneer we ervoor kiezen om 6 parameters mee te laten doen, dan krijgen we de volgende schadedriehoek:

schadejaar	afwikkeljaar			
	1	2	3	4
1	60	25	10	5
2	63	24	11	<b>5,3</b>
3	58	23	<b>10,9</b>	<b>5,3</b>
4	65	<b>24,5</b>	<b>11,3</b>	<b>5,5</b>

Figuur 6.2

De voorziening wordt  $R = 5,3 + 10,9 + 5,3 + 24,5 + 11,3 + 5,5 = 62,7$

Met behulp van de functie in R kunnen we de voorziening uitrekenen voor modellen met 1 tot en met 9 parameters:

aantal parameters:	1	2	3	4	5	6	7	8	9
IBNR voorziening:	200.4	123.7	62.6	62.6	62.7	62.7	62.2	62.2	62.2

### 6.3 Cross Validation

Wanneer er bij een bepaalde hoeveelheid data een model gemaakt is, dan kan de foutmarge bepaald worden. In het algemeen wordt hiervoor gekeken naar het verschil tussen de data en de modelschatters. Er wordt dus data vergeleken met een model dat verkregen is aan de hand van deze data. Het is dus niet objectief om op deze wijze een foutmarge te bepalen. Cross Validation is een methode om een objectieve foutmarge te bepalen. Bij Cross Validation wordt een deel van de data weg gelaten. Vervolgens wordt met de resterende data een model gemaakt. Tenslotte wordt de foutmarge bepaald aan de hand van het model en de weggelaten data. Nu hangen het model en de getoetste data niet van dezelfde data af en is derhalve objectiever. Omdat we in de schadedriehoeken maar een beperkte hoeveelheid data tot onze beschikking hebben door de geringe hoeveelheid afwikkeljaren, beperken we ons tot een 'Leave One Out Cross Validation'. Een algoritme voor dit proces kan als volgt beschreven worden:

#### Algoritme

Gegeven: data  $Y = (y_1, y_2, \dots, y_n)$

$k$  is het aantal parameters in het model,  $k = 1, \dots, m$

(1) Voor  $i = 1, \dots, n$ :

(2) Laat  $y_i$  weg, de overgebleven data is  $(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$

(3) Voor  $k = 1, \dots, m$ : het aantal parameters in het model:

(4) Op basis van de overgebleven data: schat  $\hat{y}_i$

(5) Bepaal de fout  $F(k, i) = (y_i - \hat{y}_i)^2$

Nu bevat  $P(k) = \frac{1}{n} \sum_{i=1}^n F(k, i)$  de totale kwadratische fout van een model met  $k$  parameters.

Naar verwachting ziet een plot van  $P$  eruit als een platgeslagen dalparabool. In het minimum vinden we de kleinste waarde voor de fout  $P$  bij een bepaalde hoeveelheid parameters  $k$ . Dit is het aantal dat we het beste voor het model kunnen gebruiken.

#### 6.4 F-toets

Een andere methode om het optimale aantal parameters in het model te onderzoeken, is gebruik te maken van de F-toets. Dit werkt als volgt:

(1) Bereken  $F$  met

$$F(k) = \frac{RSS(\hat{\beta}) - RSS(\tilde{\beta})}{RSS(\tilde{\beta})/(N - k - 2)} \quad (35)$$

Hierbij is  $N$  het aantal waarnemingen,

$k$  het aantal gebruikte parameters,

$RSS(\hat{\beta})$  is de Residual Sum of Squares van het geschatte model met  $k - 1$  parameters

$RSS(\tilde{\beta})$  de Residual Sum of Squares van het geschatte model met  $k$  parameters.

$F$  is een vector waarbij  $F(k)$  correspondeert met een F-toets met  $(N - k - 2)$  vrijheidsgraden bij het gebruik van een model met  $k$  parameters.

(2) We kunnen nu de overschrijdingskansen uitrekenen behorende bij een model met  $k$  parameters. In R kunnen we dit doen met de functie  $1 - pf(F, 1, (N - k - 2))$ . Als we een significantie niveau aanhouden van bijvoorbeeld 5%, dan mag het aantal parameters meedoen tot waar de overschrijdingskans onder de 5% blijft. Deze parameters hebben de grootste toegevoegde waarde.



## 7 Resultaten

Er zijn diverse methoden voor het schatten van voorzieningen besproken. In dit hoofdstuk laten we de resultaten zien van de Chain Ladder, Variant I op de Chain Ladder en het loglineaire model met FSR. Hiervan schatte we de bijbehorende standaardafwijking van  $R - \hat{R}$ . De standaardafwijking van de Chain Ladder is als eerste berekend met de parametrische bootstrap waarbij wordt uitgegaan van de Negatief Binomiale verdeling. Als tweede is de standaardafwijking voor de Chain Ladder berekend met de formules van Mack. Als derde hebben we de standaardafwijking berekend met de aanname dat de data normaal verdeeld is, zoals beschreven in 2.5. De standaardafwijking voor het loglineaire model is berekend met de bootstrap methode. In dit hoofdstuk vergelijken we steeds de geschatte standaard afwijkingen met de werkelijke fout. Als we uitgaan van reeds bekende data, kunnen we de werkelijke fout van een voorspelling bepalen

$$\sigma_{\text{werkelijk}} = R - \hat{R} = \sum_{i+j>t+1} C_{ij} - \hat{C}_{ij}$$

Vanwege de concurrentiegevoeligheid van de data, zal er in dit hoofdstuk alleen met fictieve data gewerkt worden. De data is representatief voor een zorgverzekeraar.

### 7.1 Alle data

De fictieve data die we hebben gebruikt, is fictieve data over de jaren 1999 tot en met 2003. Hiervan is alle fictieve data bekend en kunnen we onze schatters van de voorzieningen toetsen aan de gerealiseerde fictieve waarden. De dikgedrukte waarden zijn de fictieve gerealiseerde waarden, deze worden niet gebruikt bij de voorspellingen. In de praktijk wordt de data eerst gecorrigeerd op inflatie en aantallen verzekerden. Ook hier zullen we naar kijken en de resultaten vergelijken.

Gerealiseerde waarden:

schadejaar	afwikkeljaar		
	1	2	3
(1999) 1	298.559.344	74.641.462	1.396.226
(2000) 2	307.626.098	73.208.306	2.938.574
(2001) 3	359.991.621	112.090.931	2.937.355
(2002) 4	397.089.996	125.193.984	<b>2.899.879</b>
(2003) 5	374.414.630	<b>152.965.899</b>	<b>2.669.219</b>

Figuur 7.1

De IBNR voorziening is  $R = 2.899.879 + 152.965.899 + 2.669.219 = 158.534.997$ .

## 7.2 Chain Ladder schatter

Als eerste schatten we de voorzieningen met de Chain Ladder:

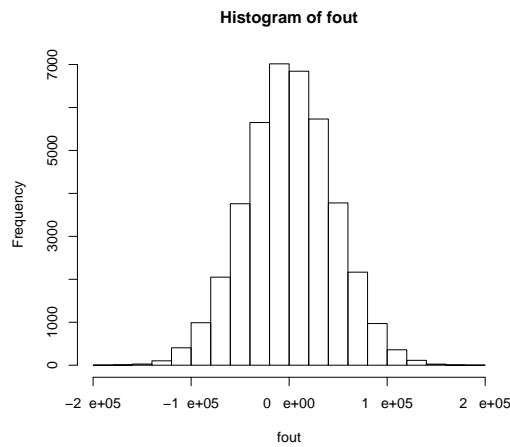
Chain Ladder, Jaardata

schadejaar	afwikkeljaar		
	1	2	3
(1999) 1	298.559.344	74.641.462	1.396.226
(2000) 2	307.626.098	73.208.306	2.938.574
(2001) 3	359.991.621	112.090.931	2.937.355
(2002) 4	397.089.996	125.193.984	<b>3.097.688</b>
(2003) 5	374.414.630	<b>105.775.357</b>	<b>2.848.027</b>

Figuur 7.2

De IBNR voorziening is hierbij  $R = 3.097.688 + 105.775.357 + 2.848.027 = 111.721.072$ .

Vervolgens bepalen we met de parametrische bootstrap de foutmarge van de voorziening met  $n = 40.000$  met de aanname dat  $C_{i,j} \sim NegBinom(x_i y_j, \omega)$ .



Figuur 7.3

Chain Ladder Fouten

type fout	$\hat{\sigma}$
Bootstrap, Negatief Binomiale verdeling	$\hat{\sigma} = 44.788$
Mack	$\hat{\sigma} = 15.987.729$
met de Normale verdeling	$\hat{\sigma} = 10.678.569$
werkelijke fout	$\sigma = 46.813.925$

Figuur 7.4

### Gecorrigeerde Jaardata

Vervolgens gaan we ook met de Chain Ladder de schatters berekenen op de gecorrigeerde data. We corrigeren op aantallen verzekerden en zorginflatie. Het resultaat vertalen we weer terug naar de gerealiseerde waarden, we krijgen dan:

Chain Ladder, Gecorrigeerde jaardata

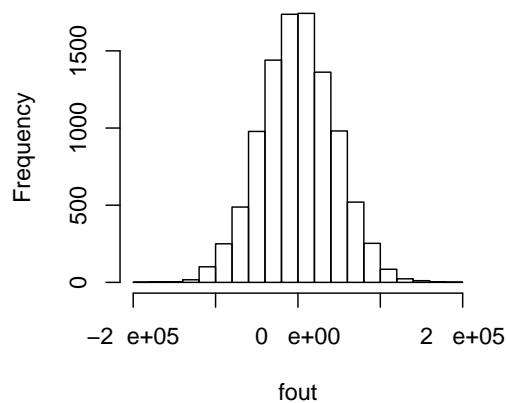
schadejaar	afwikkeljaar		
	1	2	3
(1999) 1	298.559.344	74.641.462	1.396.226
(2000) 2	307.626.098	73.208.306	2.938.574
(2001) 3	359.991.621	112.090.931	2.937.355
(2002) 4	397.089.996	125.193.984	<b>3.081.181</b>
(2003) 5	374.414.630	<b>103.912.973</b>	<b>2.821.863</b>

Figuur 7.5

De IBNR voorziening is hierbij  $R = 3.081.181 + 103.912.973 + 2.821.863 = 109.816.017$ .

Vervolgens bepalen we met de parametrische bootstrap de foutmarge van de voorziening met  $n = 40.000$  met de aanname dat  $C_{i,j} \sim NegBinom(x_i y_j, \omega)$ .

### Histogram of fout



Figuur 7.6

De standaardafwijking is  $\hat{\sigma} = 44.577$  op een totale voorziening van 109.816.017.

Het corrigeren van de data lijkt geen verbeteringen op te leveren ten opzichte van de ongecorrigeerde data. We zullen ons hierna beperken tot de Chain Ladder op de ongecorrigeerde data.

### 7.2.1 Chain Ladder met Variant I

Als toevoeging op de Chain Ladder om het aantal parameters te reduceren, hebben we in 5.1 Variant I geïntroduceerd, waarbij  $E(C_{i,j}) = xy_j$ . Om na te gaan of dit model bruikbaar is na correctie van jaareffecten, passen we dit model toe op de jaardata. Achtereenvolgens corrigeren we de jaardata, bepalen we de schatters met (27) en (28) en corrigeren we de schatters terug. We krijgen hierbij de schatters:

schadejaar	afwikkeljaar	
	2	3
(2002) 4		<b>2.966.283</b>
(2003) 5	<b>113.353.063</b>	<b>3.048.695</b>
	R=119.368.041	

Figuur 7.7

#### Chain Ladder met Variant Fouten

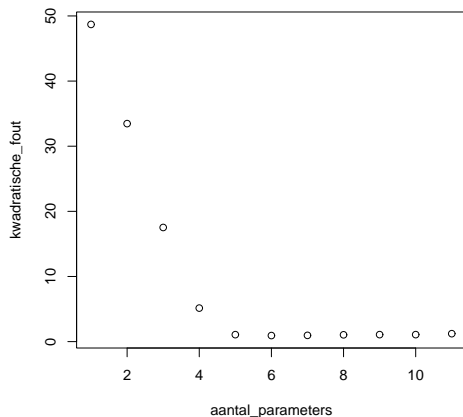
type fout	$\hat{\sigma}$
Bootstrap, Poisson verdeling	$\hat{\sigma} = 6.446.824$
met de Normale verdeling	$\hat{\sigma} = 14.319.472$
werkelijke fout	$\sigma = 39.166.956$

Figuur 7.8

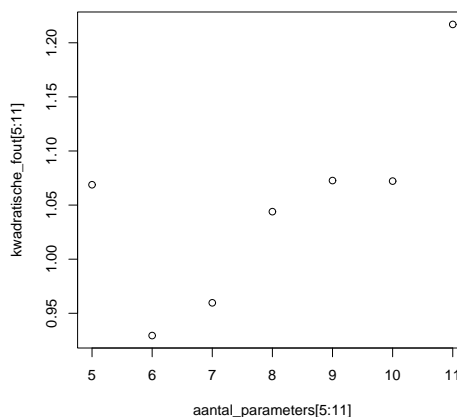
### 7.3 Loglineair model en FSR schatter

Een andere methode om schatters te bepalen voor de voorziening, is door gebruik te maken van een loglineair model. Hierbij zullen we Forward Stage-wise Regression gebruiken om te onderzoeken welke regressoren de meeste toegevoegde waarde hebben. Daarbij kunnen we gebruik maken van Leave One Out Cross Validation (LOOCV) en de F-toets om het meest geschikte aantal regressoren te bepalen.

Als eerste passen we LOOCV toe. In figuur 7.9 is de kwadratische fout te zien ten opzichte van het gebruikte aantal regressoren in het model. In afbeelding 7.9 is het nog onduidelijk of er sprake is van een uniek minimum. Hiervoor is in figuur 7.10 extra ingezoomd op het minimum, zodat we deze goed kunnen aflezen.



Figuur 7.9



Figuur 7.10

Het minimum correspondeert met het gebruik van 6 parameters. In vergelijking met de Chain Ladder is dit 4 regressoren efficiënter. De 6 parameters corresponderen achtereenvolgens met: • 3<sup>e</sup> afwikkeljaar, • 1<sup>e</sup> afwikkeljaar, • zorginflatie, • schadejaar 1999, • aantallen verzekerden en • schadejaar 2002.

Met de formules die gebruikelijk zijn bij loglineaire modellen, kunnen we vervolgens de voorziening inschatten met  $\widehat{C}_{i,j} = e^{\eta_{i,j} + \frac{1}{2}\widehat{\sigma}^2}$ , waarbij  $\sigma^2$  de steekproef variantie.

#### Loglineair model met FSR

schadejaar	afwikkeljaar		
	1	2	3
(1999) 1	298.559.344	74.641.462	1.396.226
(2000) 2	307.626.098	73.208.306	2.938.574
(2001) 3	359.991.621	112.090.931	2.937.355
(2002) 4	397.089.996	125.193.984	<b>3.143.974</b>
(2003) 5	374.414.630	<b>119.085.504</b>	<b>3.292.034</b>

Figuur 7.11

De IBNR voorziening is hierbij  $R = 3.143.974 + 119.085.504 + 3.292.034 = 125.521.512$ .

Wanneer we de F-toets toepassen, krijgen we de volgende overschrijdingskansen bij het toevoegen van de  $k^e$  parameter:

$k$	2	3	4	5	6	7	8	9	10
$p$	0,002	0,000	0,017	0,045	0,361	0,315	1,000	1,000	1,000

Wanneer we een significantie niveau hanteren van 5%, dan is het optimaal om tot en met 5 parameters toe te voegen. Dit is de helft van het aantal parameters dat de Chain Ladder gebruikt

Op dezelfde wijze als bij de LOOCV is gedaan, kunnen we nu de schadedriehoek volschatten. We krijgen dan als IBNR voorziening:  $R = 3.233.407 + 121.120.922 + 3.435.156 = 127.789.485$ . Vervolgens kunnen we met een bootstrap de standaardafwijking bepalen. Per bootstrap stap wordt er nieuwe data gegenereert en wordt er een optimaal aantal parameters bepaald met de F-toets. Gemiddeld worden er 5,3 parameters in het model gebruikt, het minimale aantal parameters is 4 en het maximale aantal is 7.

Fouten

type fout	$\hat{\sigma}$
Bootstrap	$\hat{\sigma} = 8.633.748$
werkelijke fout	$\sigma = 30.745.512$

Figuur 7.12

## 7.4 Per Verstrekking

Om trends per verstrekking eruit te kunnen lichten en mogelijk beter te kunnen voorspellen, zal hieronder dieper worden ingegaan op de diverse verstrekkingen. We zullen hiervoor kijken naar de Chain Ladder op ongecorrigeerde data en naar Loglineaire modellen. Opnieuw kijken we naar fictieve data.

### 7.4.1 Specialistische hulp

Gerealiseerde Specialistische hulp

schadejaar	afwikkeljaar		
	1	2	3
(1999) 1	22.060.176	4.386.875	165.622
(2000) 2	19.548.368	5.609.846	258.938
(2001) 3	28.839.126	9.522.099	314.421
(2002) 4	39.839.065	10.420.505	<b>323.460</b>
(2003) 5	38.756.115	<b>11.424.552</b>	<b>374.507</b>

Figuur 7.13

De IBNR voorziening is  $R = 323.460 + 11.424.552 + 374.507 = 12.122.519$ .

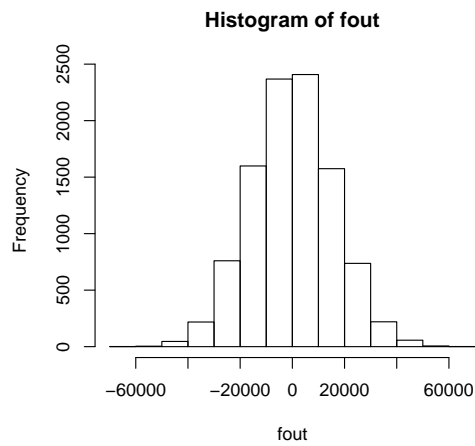
Achtereenvolgens bepalen we schatters met de Chain Ladder en het Logli-

neaire Model in combinatie met FSR en CV, waarbij 4 parameters optimaal blijken.

Schatters Specialistisch Hulp

schadejaar	2	3
	Chain Ladder schatter	
(2002) 4		<b>412.830</b>
(2003) 5	<b>10.521.047</b>	<b>404.761</b>
	IBNR=11.338.638	
	Chain Ladder + Variant	
(2002) 4		<b>299.784</b>
(2003) 5	<b>8.682.051</b>	<b>308.113</b>
	IBNR=9.289.948	
	Loglineair, FSR	
(2002) 4		<b>424.537</b>
(2003) 5	<b>11.864.290</b>	<b>498.587</b>
	IBNR=12.787.414	

Figuur 7.14



Figuur 7.15

### Fouten

<b>Chain Ladder Fout</b>	
Bootstrap, Negatief Binomiale verdeling	$\hat{\sigma} = 15.837$
Mack	$\hat{\sigma} = 1.930.508$
Normale verdeling	$\hat{\sigma} = 1.244.955$
Werkelijk	$\sigma = 783.881$
<b>Chain Ladder + Variant Fout</b>	
Bootstrap, Poisson verdeling	$\hat{\sigma} = 1.786.470$
Normale verdeling	$\hat{\sigma} = 3.040.288$
Werkelijk	$\sigma = 2.832.571$
<b>Loglineair, FSR Fout</b>	
Bootstrap	$\hat{\sigma} = 628.456$
Werkelijk	$\sigma = 664.895$

Figuur 7.16

### 7.4.2 Ziekenhuisverpleging

#### Gerealiseerde Ziekenhuisverpleging

schadejaar	afwikkeljaar		
	1	2	3
(1999) 1	174.972.382	36.498.952	916.868
(2000) 2	176.254.944	44.131.009	1.940.929
(2001) 3	209.913.550	55.576.558	2.200.486
(2002) 4	230.457.338	55.542.679	<b>2.004.823</b>
(2003) 5	224.311.443	<b>58.825.929</b>	<b>1.387.668</b>

Figuur 7.17

De IBNR voorziening is  $R = 62.218.420$ .

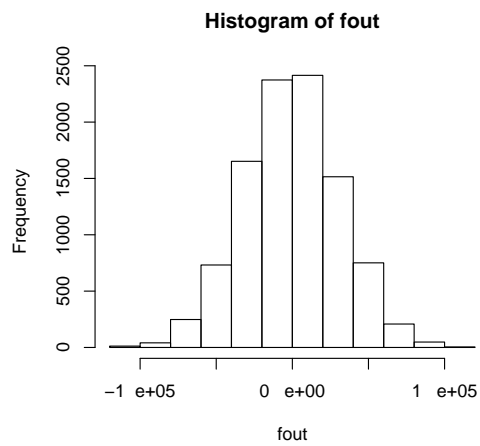
Achtereenvolgens bepalen we schatters met de Chain Ladder en het Loglineaire Model in combinatie met FSR en CV, waarbij 5 parameters optimaal blijken.



Schatters Ziekenhuisverpleging

schadejaar	2	3
	Chain Ladder schatter	
(2002) 4		<b>2.074.531</b>
(2003) 5	<b>54.335.064</b>	<b>2.021.192</b>
	IBNR=58.430.787	
	Chain Ladder + Variant	
(2002) 4		<b>2.046.536</b>
(2003) 5	<b>57.065.474</b>	<b>2.103.395</b>
	IBNR=61.215.406	
	Loglineair, FSR	
(2002) 4		<b>2.153.859</b>
(2003) 5	<b>58.365.104</b>	<b>2.241.667</b>
	IBNR=62.760.630	

Figuur 7.18



Figuur 7.19

## Fouten

<b>Chain Ladder Fout</b>	
Bootstrap, Negatief Binomiale verdeling	$\hat{\sigma} = 31.522$
Mack	$\hat{\sigma} = 5.505.179$
Normale verdeling	$\hat{\sigma} = 4.079.123$
Werkelijk	$\sigma = 3.787.633$
<b>Chain Ladder + Variant Fout</b>	
Bootstrap, Poisson verdeling	$\hat{\sigma} = 4.573.278$
Normale verdeling	$\hat{\sigma} = 5.779.959$
Werkelijk	$\sigma = 1.003.014$
<b>Loglineair, FSR Fout</b>	
Bootstrap	$\hat{\sigma} = 3.645.849$
Werkelijk	$\sigma = 542.210$

Figuur 7.20

### 7.4.3 CHA tapes

CHA tapes staat voor Clearing House Apothekers en zijn kosten van apotheken. Maandelijks worden er zogenaamde CHA tapes ingediend en verwerkt. Voor dit soort kosten wordt in het algemeen geen gebruik gemaakt van Chain Ladder schatters, omdat al met redelijk veel zekerheid te zeggen is hoeveel kosten er gemiddeld per persoon per maand gemaakt gaat worden. De totaal te verwachte kosten is het gemiddelde per persoon vermenigvuldigd met het totaal aantal verzekerden. Door de maandelijkse verwerking is niet te verwachten dat er nog kosten in het derde afwikkeljaar bij komen. Als er verschuivingen optreden, heeft dit waarschijnlijk te maken met verschuivingen in de verwerkingssnelheid.

#### Gerealiseerde CHA tapes

schadejaar	afwikkeljaar		
	1	2	3
(1999) 1	54.110.734	24.662.955	0
(2000) 2	66.161.156	15.090.125	0
(2001) 3	66.206.534	33.990.000	0
(2002) 4	62.945.000	44.875.000	<b>0</b>
(2003) 5	46.290.000	<b>66.338.000</b>	<b>0</b>

Figuur 7.21

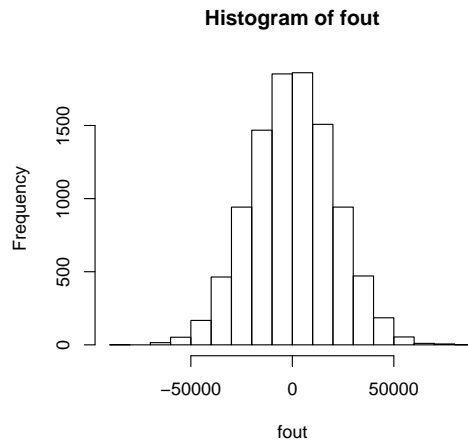
De IBNR voorziening is  $R = 66.338.000$ .

Achtereenvolgens bepalen we schatters met de Chain Ladder en het Logli-

neaire Model in combinatie met FSR en CV, waarbij 4 parameters optimaal blijken.

schadejaar	2	3
	Chain Ladder schatter	
(2002) 4		<b>0</b>
(2003) 5	<b>22.014.095</b>	<b>0</b>
	IBNR=22.014.095	
	Chain Ladder + Variant	
(2002) 4		<b>0</b>
(2003) 5	<b>34.388.060</b>	<b>0</b>
	IBNR=34.388.060	
	Loglineair, FSR	
(2002) 4		<b>0</b>
(2003) 5	<b>30.025.086</b>	<b>0</b>
	IBNR=30.025.086	

Figuur 7.22 Schatters CHA



Figuur 7.23

## Fouten

<b>Chain Ladder Fout</b>	
Bootstrap, Negatief Binomiale verdeling	$\hat{\sigma} = 20.229$
Mack	$\hat{\sigma} = 11.733.152$
Normale verdeling	$\hat{\sigma} = 7.101.379$
Werkelijk	$\sigma = 44.323.905$
<b>Chain Ladder + Variant Fout</b>	
Bootstrap, Poisson verdeling	$\hat{\sigma} = 3.469.798$
Normale verdeling	$\hat{\sigma} = 12.127.184$
Werkelijk	$\sigma = 31.949.940$
<b>Loglineair, FSR Fout</b>	
Bootstrap	$\hat{\sigma} = 5.236.853$
Werkelijk	$\sigma = 36.312.914$

Figuur 7.24

Alle schatters blijken niet erg passend. Dit is te wijten aan vertraagde verwerkingssnelheden in 2003. Wanneer de kosten geschat zouden worden met een ander model waarbij gekeken wordt naar gemiddelde jaartotalen, zal de schatter vermoedelijk beter worden.

### 7.4.4 Overige verstrekingen

#### Gerealiseerde Overige Verstrekingen

schadejaar	afwikkeljaar		
	1	2	3
(1999) 1	47.075.717	7.747.676	209.751
(2000) 2	45.661.631	8.447.325	668.708
(2001) 3	55.026.059	12.999.336	422.447
(2002) 4	62.439.811	14.355.801	<b>571.596</b>
(2003) 5	65.054.367	<b>16.377.417</b>	<b>907.044</b>

Figuur 7.25

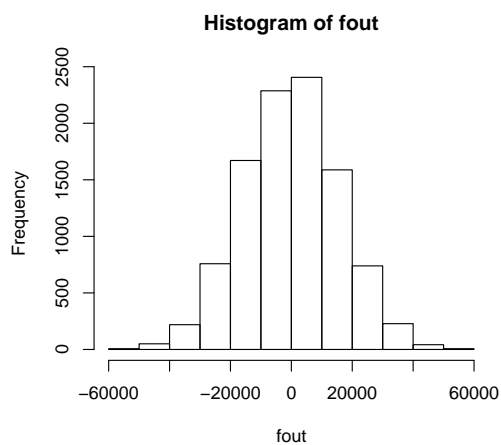
De IBNR voorziening is  $R = 17.856.057$ .

Achtereenvolgens bepalen we schatters met de Chain Ladder en het Loglineaire Model in combinatie met FSR en CV, waarbij 5 parameters optimaal blijken.

Schatters Overige Verstrekkingen

schadejaar	2	3
	Chain Ladder schatter	
(2002) 4		<b>564.563</b>
(2003) 5	<b>13.478.037</b>	<b>577.332</b>
	IBNR=14.619.932	
	Chain Ladder + Variant	
(2002) 4		<b>541.451</b>
(2003) 5	<b>12.772.439</b>	<b>556.495</b>
	IBNR=13.870.385	
	Loglineair, FSR	
(2002) 4		<b>555.059</b>
(2003) 5	<b>14.005.793</b>	<b>585.854</b>
	IBNR=15.146.706	

Figuur 7.26



Figuur 7.27

## Fouten

<b>Chain Ladder Fout</b>	
Bootstrap, Negatief Binomiale verdeling	$\hat{\sigma} = 15.623$
Mack	$\hat{\sigma} = 2.303.059$
Normale verdeling	$\hat{\sigma} = 1.744.892$
Werkelijk	$\sigma = 3.236.125$
<b>Chain Ladder + Variant Fout</b>	
Bootstrap, Poisson verdeling	$\hat{\sigma} = 2.183.609$
Normale verdeling	$\hat{\sigma} = 1.982.373$
Werkelijk	$\sigma = 3.985.672$
<b>Loglineair, FSR Fout</b>	
Bootstrap	$\hat{\sigma} = 1.602.458$
Werkelijk	$\sigma = 2.709.351$

Figuur 7.28

### 7.5 Samenvattend

We zijn begonnen met het bepalen van een schatter voor de totale data, vervolgens hebben we op verstrekkingsniveau schatter bepaald. Wanneer we nu de schatters voor de diverse verstrekkingen combineren, kunnen we weer een schatter genereren voor de totale data. In onderstaande tabel nogmaals een overzicht van de gerealiseerde waarden, de geschatte waarden op totaal niveau en de geschatte waarden op verstrekkingen niveau van  $\hat{C}_{4,3} + \hat{C}_{5,2} + \hat{C}_{5,3}$  en de IBNR voorzieningen. We maken hierbij onderscheid tussen het Chain Ladder model en tussen het Loglineair model. De standaardafwijking in het overzicht is berekend met de Bootstrap en de Negatief Binomiale verdeling, met Mack en met de bootstrap voor het loglineaire model. De formules van Mack hebben tot gevolg dat wanneer de data groter wordt, de variantie ook veel groter wordt. In de context van zorgverzekeringen is dit niet per se het geval en de waarde is hierdoor niet heel waarschijnlijk. De fout voor de Chain Ladder die met de bootstrap is bepaald, lijkt naast Mack onwaarschijnlijk klein. Hierdoor lijkt de fout die met de bootstrap is gegenereerd voor het loglineaire model en FSR het meest waarschijnlijk.

schadejaar	Chain Ladder	
	2	3
	Gerealiseerd	
(2002) 4		<b>2.899.879</b>
(2003) 5	<b>152.965.899</b>	<b>2.669.219</b>
	IBNR=158.534.997	
	Schatters op totaal	
(2002) 4		<b>3.097.688</b>
(2003) 5	<b>105.775.357</b>	<b>2.848.027</b>
	IBNR=111.721.072	
	$\hat{\sigma}_{bootstrap}=44.788$	
	$\hat{\sigma}_{Mack}=15.987.729$	
	$\hat{\sigma}_{normaal}=10.678.569$	
	$\sigma_{werkelijk}=46.813.925$	
	Schatters per verstrekking	
(2002) 4		<b>3.060.924</b>
(2003) 5	<b>100.348.243</b>	<b>3.003.285</b>
	IBNR=106.403.452	
	$\hat{\sigma}_{bootstrap}=43.563$	
	$\hat{\sigma}_{Mack}=13.304.314$	
	$\hat{\sigma}_{normaal}=8.465.423$	
	$\sigma_{werkelijk}=52.131.545$	

Figuur 7.29 Schatters

schadejaar	Chain Ladder met Variant I	
	2	3
	Gerealiseerd	
(2002) 4		<b>2.899.879</b>
(2003) 5	<b>152.965.899</b>	<b>2.669.219</b>
	IBNR=158.534.997	
	Schatters op totaal	
(2002) 4		<b>2.966.283</b>
(2003) 5	<b>113.353.063</b>	<b>3.048.695</b>
	IBNR=119.368.041	
	$\hat{\sigma}_{bootstrap}=6.446.824$	
	$\hat{\sigma}_{normaal}=14.319.472$	
	$\sigma_{werkelijk}=39.166.956$	
	Schatters per verstrekking	
(2002) 4		<b>2.887.771</b>
(2003) 5	<b>112.908.024</b>	<b>2.968.003</b>
	IBNR=118.763.798	
	$\hat{\sigma}_{bootstrap}=6.396.405$	
	$\hat{\sigma}_{normaal}=13.915.807$	
	$\sigma_{werkelijk}=39.771.199$	

Figuur 7.30 Schatters

schadejaar	Loglineair, FSR, F-toets	
	2	3
	Gerealiseerd	
(2002) 4		<b>2.899.879</b>
(2003) 5	<b>152.965.899</b>	<b>2.669.219</b>
	IBNR=158.534.997	
	Schatters op totaal	
(2002) 4		<b>3.233.407</b>
(2003) 5	<b>121.120.922</b>	<b>3.435.156</b>
	IBNR=127.789.485	
	$\hat{\sigma}_{bootstrap}=8.633.748$	
	$\sigma_{werkelijk}=30.745.512$	
	Schatters per verstrekking	
(2002) 4		<b>3.119.152</b>
(2003) 5	<b>112.777.651</b>	<b>3.229.185</b>
	IBNR=119.125.988	
	$\hat{\sigma}_{bootstrap}=6.609.060$	
	$\sigma_{werkelijk}=39.409.009$	

Figuur 7.31 Schatters



Uit bovenstaand overzicht zijn een aantal conclusies te trekken. Als eerste zijn de verschillende fouten bij het Chain Ladder model heel divers. Er is er geen die er echt als beste uitspringt. Als tweede heeft de schatter van de IBNR voorziening de kleinste werkelijke fout met het loglineaire model, vervolgens met Variant I op de Chain Ladder en tot slot met de Chain Ladder. Door een slimme keuze van de parameters in het loglineaire model, wordt overparameterisatie voorkomen en is dit model een stuk efficiënter dan de Chain Ladder. Ook Variant I is efficiënter met de hoeveelheid parameters en geeft een betere schatter dan de Chain Ladder. Als derde lijken de schatters op de totale schade een beter beeld van de werkelijkheid weer te geven dan de schatters op verstrekkingen niveau.

Toch wijkt de geschatte IBNR schade nog relatief veel af van de gerealiseerde IBNR schade. Dit is te wijten aan enerzijds een grote mutatie van aantallen verzekerden in het jaar 2001 door een fusie en anderzijds doordat de verstrekking CHA een afwijkende trend vertoont. De trend is afwijkend omdat er maar 2 afwikkellijaren zijn ten opzichte van 3 afwikkellijaren bij de andere verstrekkingen. Door het grote volume van het tweede afwikkelaar, wordt de fractie  $\hat{y}_j$  sterk beïnvloed. De grote werkelijke fout bij de schatters op verstrekkingen niveau, wordt bijna geheel veroorzaakt door de fout van de verstrekking CHA. Wanneer we de CHA eruit laten, wat gebruikelijk is vanwege andere voorspellingsmogelijkheden van de CHA, zal de fout op verstrekkingen niveau vele malen kleiner worden. Omdat het de moeite waard is om hier naar te kijken, berekenen we alle schatters en standaardafwijkingen op totaal en verstrekkingen niveau nogmaals voor de data waarbij CHA eruit gelaten is. We laten hierbij de bootstrap fout met de Negatief Binomiale verdeling weg, omdat de waarde minder goed past bij  $\sigma_{\text{werkelijk}}$ .

schadejaar	Chain Ladder	
	2	3
	Gerealiseerd	
(2002) 4		<b>2.899.879</b>
(2003) 5	<b>86.627.899</b>	<b>2.669.219</b>
	IBNR=92.196.997	
	Schatters op totaal	
(2002) 4		<b>3.120.466</b>
(2003) 5	<b>78.512.512</b>	<b>3.061.538</b>
	IBNR=84.694.516	
	$\hat{\sigma}_{Mack}=8.423.791$	
	$\hat{\sigma}_{normaal}=6.149.481$	
	$\sigma_{werkelijk}=7.502.481$	
	Schatters per verstrekking	
(2002) 4		<b>3.060.924</b>
(2003) 5	<b>78.334.148</b>	<b>3.003.285</b>
	IBNR=84.398.357	
	$\hat{\sigma}_{Mack}=6.271.996$	
	$\hat{\sigma}_{normaal}=4.608.015$	
	$\sigma_{werkelijk}=7.798.640$	

Figuur 7.32 Schatters zonder CHA

schadejaar	Chain Ladder met Variant I	
	2	3
	Gerealiseerd	
(2002) 4		<b>2.899.879</b>
(2003) 5	<b>86.627.899</b>	<b>2.669.219</b>
	IBNR=92.196.997	
	Schatters op totaal	
(2002) 4		<b>2.966.283</b>
(2003) 5	<b>78.965.003</b>	<b>3.048.695</b>
	IBNR= 84.979.981	
	$\hat{\sigma}_{bootstrap}=5.423.472$	
	$\hat{\sigma}_{normaal}=7.640.282$	
	$\sigma_{werkelijk}=7.217.016$	
	Schatters per verstrekking	
(2002) 4		<b>2.887.771</b>
(2003) 5	<b>78.519.964</b>	<b>2.968.003</b>
	IBNR= 84.375.738	
	$\hat{\sigma}_{bootstrap}=5.373.499$	
	$\hat{\sigma}_{normaal}=6.825.033$	
	$\sigma_{werkelijk}=7.821.259$	

Figuur 7.33 Schatters zonder CHA

schadejaar	Loglineair, FSR, F-toets	
	2	3
	Gerealiseerd	
(2002) 4		<b>2.899.879</b>
(2003) 5	<b>86.627.899</b>	<b>2.669.219</b>
	IBNR=92.196.997	
	Schatters op totaal	
(2002) 4		<b>3.129.705</b>
(2003) 5	<b>83.767.282</b>	<b>3.316.246</b>
	IBNR=90.213.233	
	$\hat{\sigma}_{bootstrap}=3.968.551$	
	$\sigma_{werkelijk}= 1.983.764$	
	Schatters per verstrekking	
(2002) 4		<b>3.119.152</b>
(2003) 5	<b>82.752.565</b>	<b>3.229.185</b>
	IBNR=89.100.902	
	$\hat{\sigma}_{bootstrap}=4.031.754$	
	$\sigma_{werkelijk}=3.096.095$	

Figuur 7.34 Schatters zonder CHA

Zonder de verstrekking CHA, zijn de schatters een stuk beter dan de schatters waarbij de verstrekking CHA wel is meegenomen. Bij de Chain Ladder zijn de fouten berekent met de methode van Mack en de fout met de normale verdeling beide acceptabel en geven ze een goed beeld van de fout omdat ze dichtbij de werkelijke fout liggen. Verder liggen de schatters en de fouten van de Chain Ladder en de Chain Ladder met Variant I heel erg dicht bij elkaar en geven ze een redelijk beeld van schatters en fouten. Het is dus heel goed mogelijk om efficiënter te zijn met de hoeveelheid parameters bij de Chain Ladder waarbij we toch nog een goede schatter krijgen. Opnieuw geeft het loglineaire model de beste schatter met de kleinste fout van de drie hierboven uitgewerkte methoden. Ook de schatter van de standaardafwijking is acceptabel in vergelijking met de werkelijke standaardafwijking. Kortom, het is aan te raden om bij het schatten de verstrekking CHA apart te schatten. De resultaten met het efficiënte loglineaire model komt als eerste in aanmerking met de parametrische bootstrap als methode om de fout te bepalen.

## 8 Samenvatting

In de wereld van verzekeraars worden ieder jaar schattingen gemaakt van de verwachte schadelast. Op deze schattingen worden de voorzieningen gebaseerd. Een voorziening is het bedrag dat de verzekeraar als actieve middelen beschikbaar houdt om toekomstige schadeclaims uit te betalen. Een groot deel van de schade wordt geclaimd in het jaar dat de schade is ontstaan. Er is ook een deel dat pas later wordt geclaimd, dit wordt IBNR schade genoemd. Hierbij staat IBNR voor 'Incurred But Not Reported'. Omdat de verzekeraar met bepaalde betrouwbaarheid in staat wil zijn om dit soort toekomstige claims te kunnen uitbetalen, moet IBNR schade in de schatter voor de verwachte schadelast worden meegenomen. Naast het genereren van een schatter is de verzekeraar geïnteresseerd in de betrouwbaarheid van de schatter, ofwel de standaardafwijking.

Tijdens het onderzoek zijn diverse modellen geanalyseerd die schatters voor de IBNR schade opleveren. Daarnaast is er gekeken naar methoden om een foutmarge te bepalen. Dit wordt gebruikt om de adequaatheid van de schatters weer te geven. Voor de foutmarge blijkt de Parametrische Bootstrap goed te werken.

Als eerste is er gekeken naar de populaire Chain Ladder methode. Dit is een methode die uitgaat van een constante verhouding tussen de verschillende afwikkeljaren. Bijvoorbeeld dat er in het eerste afwikkeljaar 75% wordt geclaimd, in het tweede afwikkeljaar nog 24% en in het laatste afwikkeljaar nog 1%. In het onderzoek is de Chain Ladder op diverse manieren benaderd, waarbij ook op diverse manieren de standaardafwijking is bepaald. De combinatie van schatter en standaardafwijking bepaalt hoe goed het model in de context van Zorgverzekeringen past. Bij de Chain Ladder methode levert de Normale verdeling de meest realistische standaardafwijking op. Dit geldt ook voor een variant op de Chain Ladder waarbij we gebruik maken van schadejaaronafhankelijke parameters. Bij deze variant corrigeren we de data eerst op jaareffecten. Deze aanpassing op de Chain Ladder zorgt ervoor dat er minder parameters in het model nodig zijn, hiermee kan overparameterisatie worden voorkomen.

Vervolgens is er gekeken naar loglineaire modellen. Deze modellen hebben als uitgangspunt dat de data uit multiplicatieve regressoren is opgebouwd. Omdat dit het geval is in de zorgverzekeringen-context, te denken aan zorginflatie en mutatie van verzekerden, past dit uitgangspunt goed bij de data. Voor de foutmarge is ook hier gebruik gemaakt van de bootstrap methode, waarbij de ruis normaal verdeeld is met verwachting nul en de variantie gelijk aan de steekproefvariantie.

Om uiteindelijk het meest geschikte model te bepalen, zijn we uitgegaan van een multiplicatieve opbouw van de data en hebben derhalve gekozen voor een loglineair model. Vervolgens hebben we een designmatrix gemaakt welke alle denkbare regressoren bevat die in een model kunnen zitten. Voorbeelden

hiervan zijn het schadejaar, afwikkeljaar, kalenderjaar, aantallen verzekerden, zorginflatie. Met behulp van Forward Stagewise Regression hebben we de computer laten uitrekenen wat de volgorde is van meest geschikte regressoren. De volgorde wordt berekend aan de hand van de kleinste kwadraten methode. Wanneer we een model willen maken met maar één regressor, moeten we de eerste kiezen van dit rijtje. Wanneer we een model met twee regressoren willen maken, dan kiezen we de eerste twee regressoren en zo verder. Om overparameterisatie te voorkomen, is er gekozen voor Cross Validation en de F-toets om het meest geschikte aantal regressoren te bepalen dat in een model mee mag doen. Wanneer we dit toepassen, valt er op dat dit model maar 50% van de regressoren gebruikt die in de Chain Ladder voorkomen en dus efficiënter is.

Tot slot zijn het Chain Ladder model, een variant op de Chain Ladder tegen overparameterisatie en het loglineaire model toegepast op de jaardata van 1999 tot en met 2003. De keuze om deze data te gebruiken is dat ook de resultaten hiervan bekend zijn waardoor te toetsen is of de schatters acceptabel zijn. Wanneer we de uitkomsten van deze drie modellen vergelijken met de gerealiseerde waarden, dan blijkt dat het loglineaire model de beste schatter oplevert met de kleinste werkelijke fout. Wanneer de Chain Ladder in de praktijk wordt gebruikt, is het gebruikelijk om de data eerst te corrigeren op aantallen verzekerden en op de zorginflatie. Een nadeel is dat er hierdoor twee parameters bij het model komen en overparameterisatie in de hand gewerkt wordt door de kleine hoeveelheid aan beschikbare data. Door dan gebruik te maken van de Variant op de Chain Ladder, wordt het aantal parameters weer gereduceerd en komt er een hele acceptabele schatter uit. Deze schatter en standaardafwijking zijn vergelijkbaar met de schatter en standaardafwijking van de gewone Chain Ladder.

Kortom, het loglineaire model in combinatie met Forward Stagewise Regression en de F-toets leveren het meest geschikte model door de kleinste werkelijke fout en goede mogelijkheden om de fout te bepalen. Hierbij zijn belangrijke aspecten dat er gelet is op de multiplicatieve opbouw van de data en het voorkomen van overparameterisatie.

## Appendix A

$\omega$  schatten voor over-dispersed Poisson proces.

```
C=jrdata # bevat jaardata
hoogte=dim(C)[1]
breedte=dim(C)[2]
x=rep(0,hoogte)
y=rep(0,breedte)
n=breedte

# voor willekeurige schadedriehoeken: x en y schatten
for (k in 1:(hoogte+1-n)){ # voor blok bekende data
  x[k]=sum(C[k,])
}
y[n]=sum(C[(1:(hoogte+1-n)),n])/(sum(x[1:(hoogte+1-n)]))

for (k in (hoogte+2-n):hoogte){ # voor de driehoek
  x[k]=sum(C[k,1:(hoogte+1-k)])/(1-(sum(y[(hoogte+2-k):n])))
  y[hoogte+1-k]=sum(C[1:k,(hoogte+1-k)])/(sum(x[1:k]))
}

# omega schatten
omega=1000 # starten met \omega^0
functie=matrix(0,hoogte, breedte)
afgeleide=matrix(0,hoogte,breedte)
fractie=10

while(abs(fractie)>0.0001){ # convergentie criterium

#functie uitrekenen
  for (i in 1:(hoogte+1-n)){ # voor blok bekende data
    for(j in 1:n){
      functie[i,j]=log(omega/(omega+x[i]*y[j])) +
        (C[i,j]+omega)/(x[i]*y[j]+omega)+1
    }
  }
  for (i in (hoogte+2-n):hoogte){ # voor de driehoek
    for(j in 1:(hoogte+1-i)){
      functie[i,j]=log(omega/(omega+x[i]*y[j])) +
        (C[i,j]+omega)/(x[i]*y[j]+omega)+1
    }
  }
}
```

```

# afgeleide functie uitrekenen

for (i in 1:(hoogte+1-n)){
    # voor blok bekende data
    for(j in 1:n){
        afgeleide[i,j]=(x[i]^2*y[j]^2+omega*C[i,j])/
            (omega*(x[i]*y[j]+omega)^2)
    }
}
for (i in (hoogte+2-n):hoogte){
    # voor de driehoek
    for(j in 1:(hoogte+1-i)){
        afgeleide[i,j]=(x[i]^2*y[j]^2+omega*C[i,j])/
            (omega*(x[i]*y[j]+omega)^2)
    }
}
fractie=sum(funcitie)/sum(afgeleide)
omega=omega-fractie

} # end while

```



## Appendix B

EM algortime

```
C=jrdata # matrix met data
hoogte=dim(C)[1]
breedte=dim(C)[2]

# x0 bepalen
x0=sum(C[1:(hoogte-breedte+1),1:breedte])/(hoogte-breedte+1)
x=rep(x0,hoogte)

# y0 bepalen
y=sum(C[1:(hoogte-breedte+1),1])/sum(C[1:(hoogte-breedte+1),])
for (i in 2:breedte){
  y=c(y,sum(C[1:(hoogte-breedte+1),i])/sum(C[1:(hoogte-breedte+1),]))
}
Cster=C # hier komt volgeschatte driehoek in
voorziening=0
marge=10

while(marge>.01){
# driehoek volschatten met laatst bekende x en y
  for(i in (hoogte-breedte+2):hoogte){
    for(j in (hoogte+2-i):breedte){
      Cster[i,j]=x[i]*y[j]
    }
  }
# bepalen van nieuwe x en y
  for (i in 1:hoogte){
    x[i]=sum(Cster[i,])
  }
  for(j in 1:breedte){
    y[j]=(sum(Cster[,j]))/(sum(Cster))
  }
# bepalen voorziening
  voorz=0
  for(i in (hoogte-breedte+2):hoogte){
    voorz=voorz+sum(Cster[i,((hoogte+breedte-1-i):breedte)])
  }
  voorziening=c(voorziening,voorz)
  marge=voorziening[length(voorziening)]-
    voorziening[length(voorziening)-1]
} # end while
```

## Appendix C

### Parametrische Bootstrap

```
FOUT=rep(0,40000)      # hierin komen alle bootstrap fouten te staan
for (h in 1:40000){    # steekproef grootte
  Cster=matrix(0,hoogte,breedte)
  for(i in 1:hoogte){  # steekproef trekken, matrix opvullen
    for (j in 1:breedte){
      Cster[i,j]=rpois(1,(xhat[i]*yhat[j]))  # of andere verdeling
    }
  }
  xster=1:hoogte      # xhat en yhat van de steekproef * schatten
  yster=1:breedte

  for (k in 1:(hoogte+1-breedte)){          # voor blok bekende data
    xster[k]=sum(Cster[k,])
  }
  yster[breedte]=sum(Cster[(1:(hoogte+1-breedte)),breedte])/
    (sum(xster[1:(hoogte+1-breedte)]))

  for (k in (hoogte+2-breedte):hoogte){     # voor de driehoek
    xster[k]=sum(Cster[k,1:(hoogte+1-k)])/
      (1-(sum(yster[(hoogte+2-k):breedte])))
    yster[hoogte+1-k]=sum(Cster[1:k,(hoogte+1-k)])/
      sum(xster[1:k])
  }

  fout=0                                     # bepalen van bootstrap fout
  for (i in (hoogte+2-breedte):hoogte){
    for (j in (hoogte+2-i):breedte){
      fout=fout+(Cster[i,j]-xster[i]*yster[j])
    }
  }
  FOUT[h]=fout
}

hist(fout)
mean(fout)
sqrt(var(fout))
```

## Appendix D

In David Skurnick, *A Survey of Loss Reserving Methods*, staat een uitgebreid overzicht van verschillende IBNR methoden inclusief een korte beschrijving van de methode.

Voor het uitgebreide overzicht verwijst ik u naar dit document, wat te vinden is op <http://www.casualtyactuarialsociety.com>. De beschreven methodes zijn:

- Individual case estimates
- Fast track reserves
- Tabular value
- Notice-average method
- Average value method
- Correct case reserve for bias
- Report year loss development
- Projection method
- Payment development method
- Runoff method
- IBNR reserve as a percentage of a base
- Modify last year's IBNR for growth
- Tarbell method
- Total loss reserves as a percentage of a base
- Runoff of cumulative incurred loss
- Accident year loss development
- Loss ratio method
- Accident year incurred loss development
- Accident year paid loss development
- Prospective test of reserves
- Lorah method
- Average value method
- Estimate loss reserves by a method which includes reopened claims
- Treat reopened claims like IBNR
- Balcarek method
- Salvage and subrogation
- Relationship between calendar year and accident year incurred loss
- Loss reserve method
- Ratio method
- Brian method
- Slifka method
- Projection method - Accident year basis
- Projection method - Policy year basis
- Dollar method

## Literatuur

- [1] Dalen, A. van (2006), *Een praktische methode voor het bepalen van de onzekerheid van uitloopschade*, De Actuaris, maart 2006
- [2] Dempster, A. P., Laird, N. M., Rubin, D. B. (1977), *Maximum likelihood from incomplete data via EM algorithm*, Journal of the Royal Statistical Society, Series B, 39:1–38.
- [3] Efron, B. (1979), *Bootstrap methods: another look at the jackknife*, The Annals of Statistics 7, 1-26
- [4] England, P.D. and Verrall R.J. (1998), *Standard Errors of Prediction in Claims Reserving: A Comparison of Methods*, ASTIN Colloquium, Glasgow, Scotland
- [5] Hastie, T., Taylor, J., Tibshirani, R., Walther, G. (2006), *Forward Stage-wise Regression and the Monotone Lasso*
- [6] Hoedemakers, T., Beirlant, J., Goovaerts, M.J., Dhaene, J. (2005), *Claims Reserving Using Generalized Linear Models*
- [7] Hoedemakers, T., Goovaerts, M. en Dhaene, J. (2003), *IBNR-Problematiek in historisch perspectief*, De Actuaris, november 2003
- [8] Hoedemakers, T., Goovaerts, M. en Dhaene, J. (2004) *De IBNR - Discussie, Lognormale Lineaire Regressiemodellen in Schadereservering*
- [9] Mack, T. (1993), *Distribution-free calculation of the standard error of Chain Ladder reserve estimates*, ASTIN Bulletin, Vol. 23, No. 2
- [10] Mack, T. (1994), *Which stochastic model is underlying the chain ladder method*, XXIV ASTIN Colloquium, Cambridge
- [11] Mack, T., Venter, G. (1999), *A Comparison of Stochastic Models that Reproduce Chain Ladder Reserve Estimates*, ASTIN Colloquium, Tokyo, Japan
- [12] McCullagh, P. and Nelder J.A. (1983), *Generalized Linear Models*
- [13] Piegorisch, W.W. (1990), *Maximum Likelihood Estimation for the Negative Binomial Dispersion Parameter*, Biometrics, Vol. 46, No. 3

- [14] Pinheiro, P.J.R., Andrade e Silva, J.M., Centeno, M.de L. (2002), *Bootstrap Methodology in Claim Reserving*
- [15] Renshaw, A.E. and Verrall R.J. (1994), *A Stochastic Model Underlying the Chain-Ladder Technique*, XXV ASTIN Colloquium, Cannes, France
- [16] Skurnick, D.(1973), *A Survey of Loss Reserving Methods*
- [17] Spiridon Penev, *An Introduction to the Bootstrap*, School of mathematics, Sidney
- [18] Verrall ,R.J. (1994), *A method for modelling varying run-off evolutions in claims reserving*, ASTIN Bulletin, Vol. 24, No. 2