



Matematisk statistik
Stockholms universitet

**En reservberäkningsmetodik baserad på enskilda
skador**

Elize Wästanfors

Examensarbete 2005:3

Postadress:

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm
Sverige

Internet:

<http://www.matematik.su.se/matstat>

En reservberäkningsmetodik baserad på enskilda skador

Elize Wästanfors *

Februari 2005

SAMMANFATTNING

För att kunna täcka framtida kostnader för skador som idag ännu inte anmälts samt för skador som rapporterats men inte slutreglerats, avsätter varje försäkringsbolag medel till en reserv. Det finns många tillvägagångssätt för att estimerar storleken på reserven. Den gemensamma utgångspunkten för de idag mest frekvent använda metoderna är en aggregation av data till triangelform. En välkänd metod som exemplifierar denna teknik och som använts i årtionden är Chain Ladder-metoden.

I denna studie utvecklas och utvärderas en idé till en metod för reservprediktion som istället grundas på individuella skador och försäkringar. Analysen delas upp i två delar, dels problemet att skatta antalet skador, dels svårigheten att uppskatta vad dessa skador kommer att kosta. Vad beträffande antalet skador baseras beräkningarna på en anpassad fördelningsfunktion för tid från skadetillfälle till inrapportering. Kostnaden skattas via en utjämnande kurvanpassning. Endast försiktiga slutsatser om denna individuella metodik kan dras på grund av få jämförelseperioder. För att metoden skall kunna fungera som ett användbart komplement är det självklart viktigt att både skattningarna av antalet skador och kostnaden för dessa är goda. Utifrån de data som studien grundar sig på har dock inte någon tillförlitlig metod för antalet skador kunnat fastställas.

* E-post: elize.wastanfors@trygghansa.se
Handledare: Esbjörn Ohlsson och Anders Lindström

A method based on individual claims in order to estimate the reserve of an insurance company

ABSTRACT

To be able to cover future expenses caused by incurred but not yet reported claims and reported but not settled claims, an insurance company must allocate capital to a reserve. There are many approaches to estimate the size of the reserve. The Chain Ladder method is one of the most used loss reserving techniques. Like other frequently used methods it is based on aggregated observations in the shape of a triangle.

In this paper we develop and evaluate an idea for a method based on individual claims and policies. The analysis is divided into two parts. In the first part an estimation of the *number* of IBNR claims is assessed. The calculations are based on a fitted distribution function for the time interval from occurrence to notification. In the second part, the claim *amounts* are estimated via an adjusting curve. Only vague conclusions about this new method can be made due to few periods of comparison. For the method to function, it is important that both claim counts and claim amounts are well estimated. However, on the basis of the data of this study, no reliable method in order to estimate the number of claims has been established.

FÖRORD

Denna rapport är utförd under hösten/vintern 2004/2005 på försäkringsbolaget Trygg-Hansa, aktuarieenheten i division Privat, Affärsområde Sak, i Stockholm. Uppgiften motsvarar 20 veckors arbete och utgör mitt examensarbete, omfattande 20 poäng för magisterexamen i matematisk statistik, från Matematik-ekonomi-linjen, Stockholms Universitet.

Jag vill med detta förord passa på att rikta ett tack till Trygg-Hansa som i och med erbjudandet av arbetsplats och egen dator givit mig möjligheten att få en inblick i arbetslivet.

Stort tack även till min handledare på Trygg-Hansa, Anders Lindström, för rådgivning, vägledning och för att han alltid funnits till hands.

Min handledare på Stockholms Universitet, Esbjörn Ohlsson, skall också ha ett varmt tack för stor hjälp med det teoretiska arbetet, uppmuntran och givande möten.

Till sist en stor eloge till hela aktuariegruppen för trevligt sällskap samt stöd och goda råd längs vägen.

INNEHÅLLSFÖRTECKNING

SAMMANFATTNING	1
ABSTRACT	2
FÖRORD	3
1 INLEDNING.....	5
1.1 RESERVSÄTTNING	5
1.2 SYFTE – EN ALTERNATIV METOD.....	5
1.3 AVGRÄNSNING.....	6
2 ANALYS	7
2.1 UTREDNING AV ANMÄLNINGSMÖNSTRET	7
2.2 FÖRDELNINGSANPASSNING AV ANMÄLNINGSTIDEN	9
2.3 SKATTNING AV ANTAL EFTERANMÄLDA SKADOR.....	12
2.3.1 <i>Dagsmetoden</i>	12
2.3.2 <i>Skadefrekvensmetoden</i>	14
2.3.3 <i>Chain Ladder</i>	16
2.3.4 <i>Jämförelse av metoderna</i>	17
2.4 SKATTNING AV KOSTNADEN FÖR DE EFTERANMÄLDA SKADORNA	17
2.4.1 <i>Utjämnande kurvanpassning</i>	19
2.5 IBNR-RESERVEN	21
3 RESULTAT OCH SLUTSATSER	23
4 KÄLLFÖRTECKNING	25
5 APPENDIX	26
5.1 GAMMAFÖRDELNINGEN	26
5.2 LOGNORMALFÖRDELNINGEN	26
5.3 PARETOFÖRDELNINGEN	27
5.4 GRUNDLÄGGANDE MODELLANTAGANDEN	28

1 INLEDNING

1.1 Reservsättning

Skadeförsäkringsbolag avsätter medel till den så kallade ersättningsreserven. Detta för att kunna täcka de framtida kostnaderna för skador inträffade under den bakomliggande perioden. Problemet är att den slutliga kostnaden för var och en av dessa skador inte är känd vid bokslutstillfället, dels för att alla skador ännu inte rapporterats av försäkringstagarna, dels för att det tar tid att slutreglera rapporterade skador. Man kan använda sig av följande uppdelning:

IBNR = *incurred but not reported* = inträffade men ej rapporterade skador

RBNS = *reported but not settled* = rapporterade men ej slutreglerade skador

För att kunna bestämma det totala reservbehovet måste alltså kostnaden för dessa två poster predikteras.

1.2 Syfte – en alternativ metod

Den mest använda metoden för att prediktera kostnaden för ovan nämnda IBNR- och RBNS-skador, och därmed kunna bestämma det totala reservbehovet, är den så kallade Chain Ladder-metoden. Denna metod utgår ifrån aggregerade data i triangelform.

Syftet med detta exjobb är att utifrån enskilda skador/försäkringar hitta och utvärdera ett alternativ till Chain Ladder-metoden och andra idag frekvent använda metoder. Vi önskar hitta en snabbare och enklare metod som dessutom kan tillämpas när som helst under året, inte bara vid årsskiftet. För att Chain Ladder skall kunna fungera vid godtycklig tidpunkt måste det finnas historiska data med denna uppdelning. Därmed inte sagt att denna indelning inte är möjlig, men beroende på vilken typ av försäkring man analyserar kommer för ändamålet uppdaterade data vara mer eller mindre lättillgängliga. Just i denna studie kommer vi dock att använda oss av sådana data som lätt går att aggregera både års- och månadsvis och därmed är tillräckliga för ett användande av Chain-Ladder.

Målet är att se om denna typ av individuell metodik kan fungera som ett bra komplement till befintliga modeller, om den är mer lätthanterlig och om den exempelvis ger mer exakta resultat. Därutöver vill vi som sagt kunna applicera den på kortare tidsintervall än år.

I utredningen behandlar vi endast de skador som har inträffat men ännu ej inrapporterats, IBNR. Inledningsvis anpassar vi en parametrisk fördelning för tid från skadetillfälle till anmälan. Utifrån denna fördelning skattar vi förväntat antal efteranmälda skador, dvs de skador för vilka vi måste ha avsatta medel, så att framtida kostnader kan täckas. Den förväntade skadekostnaden för dessa skador skall också beräknas, med hänsyn till beroendet på anmälningstiden.

Till sist skall vi utvärdera och bedöma resultatet av den nya metoden.

1.3 Avgränsning

För att kunna hålla detta arbete inom rimliga gränser måste vi göra vissa avgränsningar.

Den datamängd som har ställts till förfogande innehåller information om alla skador inom bransch Motor med skadetillfälle från 01 jan 1990 till 30 juli 2004. Vi har också vetskap om det totala antalet delkaskoförsäkringar från och med 1994 och framåt. För att kunna bedöma och analysera den nya metoden i förhållande till befintlig behövs ett färdigavvecklat datamaterial, så därför utesluter vi skador med skadedatum senare än 2002-12-31.

Vi väljer också att särskilja olika skadetyper, exempelvis stöld, vagnskada och glas, i det fall anmälningmönstret för dessa skadearter skiljer sig nämnvärt åt.

Vad beträffande anpassandet av en fördelning till anmälningstiden, så ska vi pröva tre olika fördelningar, nämligen gamma-, lognormal- samt Paretofördelningen. Parametrarna i fördelningarna kommer att skattas med hjälp av moment- och ML-metoden.

2 ANALYS

2.1 Utredning av anmälningmönstret

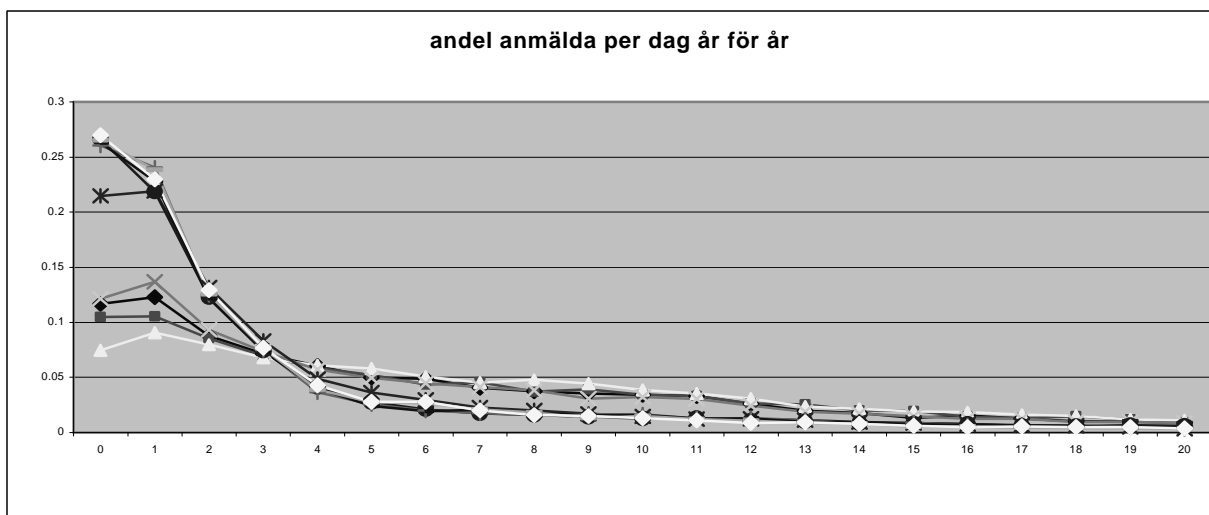
Inledningsvis undersöker vi anmälningmönstret mellan olika skadearter på privatägda personbilar. Vi vill bland annat se om medelvärdet för tid till anmälan skiljer sig mycket åt för varje typ av skada. Intuitivt känns det som att så är fallet. Rimligtvis borde ju en stulen bil locka till omedelbar skadeanmälan snarare än vad en spricka i bakrutan skulle göra. I tabell 2.1 syns medelvärdet och standardavvikelsen av tiden mellan skadedatum och rapporteringsdatum för några olika skador, uttryckt i dagar. Värdena är baserade på skador från alla åren i materialet.

Tabell 2.1

	MEDELVÄRDE	STANDARDVARIATION
Personbil - stöld	11,1	30,8
Personbil - vagnskada	14,7	37,4
Personbil - glas	57,4	69,2
Personbil - räddning	47,2	54,2
Personbil - totalt	29,3	55,5

Då anmälningstrukturen varierar anmärkningsvärt väljer vi att analysera de olika typerna av skador separat, för att få en bättre anpassad fördelningsfunktion. Till att börja med koncentrerar vi oss på stöldskador på personbil.

För att en parametrisk anpassning till anmälningstiden ska vara meningsfull krävs stabila utvecklingsmönster över åren. Därför måste vi undersöka hur anmälningstrukturen för stöldskador på personbil ser ut år för år. Om de skiljer sig nämnvärt faller hela idén om ett tidsberoende. Figur 2.1 visar anmälningmönstret för vart och ett av skadeåren 1993-2002 för de första tjugo dagarna.



Figur 2.1: andel anmälda per dag för åren 1993-2002, de fyra undre linjerna representerar 1993-1996 och den övre gruppen visar anmälningmönstret för åren 1997-2002

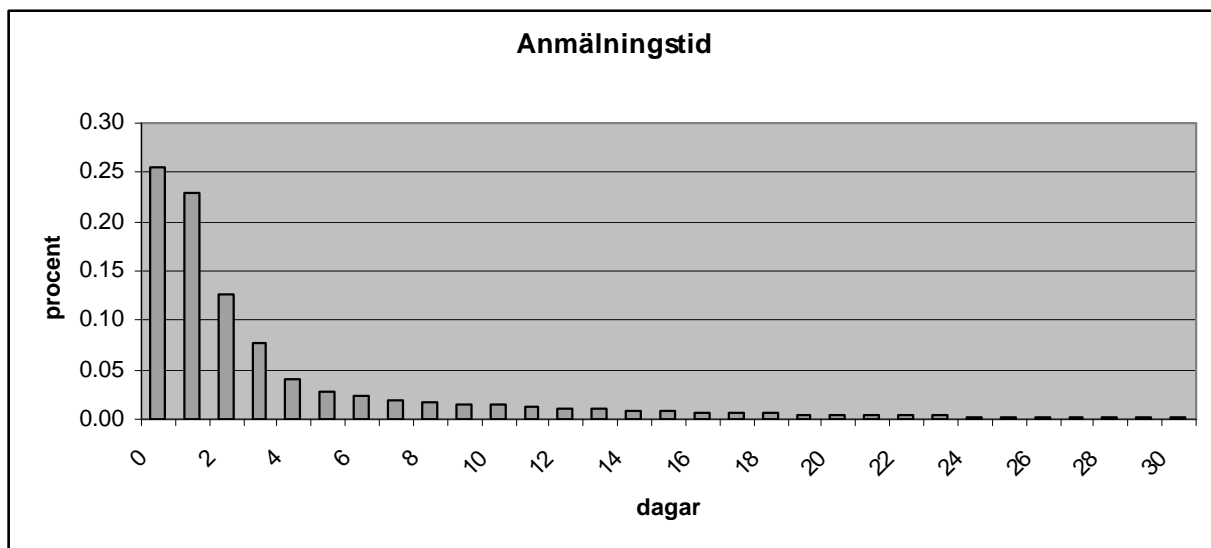
Vi ser en intressant struktur: de fyra linjerna som motsvarar åren 1993-96 ligger samlade, medan åren 1997-2002 har grupperat sig. Under de tidiga åren verkar det som om färre skador

anmäldes under de första dagarna jämfört med de senare årtalen. Det märkliga är att förändringen inte sker successivt med tiden. Skillnaden mellan de två grupperna är tydlig.

Finns det då någon förklaring till detta fenomen? Jo, under 1997 förändrades både arbetsätt och hanteringen av skador på Trygg-Hansa. Idag registreras allt direkt i telefon och det finns en strävan om att alla skador som anmäls ska registreras samma dag som de kommer in. Innan 1997 blev många av skadorna liggande några dagar innan de registrerades. Detta betyder att figur 2.1 ger oss en falsk bild. Trots att grafen visar att exempelvis drygt 10 % av skadorna år 1994 anmäldes samma dag, kanske det korrekta värdet är 25%. Vi kan ana en fara med att basera en metod på ett empiriskt mönster som den mänskliga faktorn är med och påverkar. Verkligheten följer kanske ett stabilt mönster, där arbetsättet inte alls behöver göra det!

För vårt anpassningsförfarande innebär detta att vi endast kan utgå ifrån skador från åren 1997 och framåt, för att undvika en felaktig anpassning. Skador med skadetillfälle efter 2002-12-31 tas bort ur datamängden, eftersom vi vill arbeta med ett material som idag är i det närmaste färdigavvecklat. Två skador, vilkas anmälningstid har passerat preskriptionstiden tre år för stöldskador på personbil, exkluderar vi från beräkningarna. Av de återstående 120 044 skadorna finns fortfarande 145 stycken öppna efter 2002-12-31. 81 stycken av dessa har reserv = 0, och de antas som slutreglerade. 64 skador återstår som alltjämt ligger öppna. Det faktum att de ligger öppna är dock inget problem för anpassningen av anmälningstiden för IBNR, så i första delen lämnas de oredigerade och kvarstår i analysen. Kostnaden för dessa antar vi sammanfalla med skadereglerarnas bedömning av den slutliga kostnaden.

Det vanligaste är att en stöldskada anmäls samma dag som skadan inträffat. 25,6 % av skadorna anmäls samma dag. Efter en månad är 95,3 % av skadorna inrapporterade. Medelvärde för tid till anmälan är nu 7,6 dagar (jämför med tabell 2.1) samtidigt som standardavvikelsen är 28,3 dagar. Observera att i och med borttagandet av skador inträffade tidigare än 1997 ur analysen, har medelvärdet för tid till anmälan sjunkit med hela 3,5 dagar, vilket motsvarar drygt 30 %.

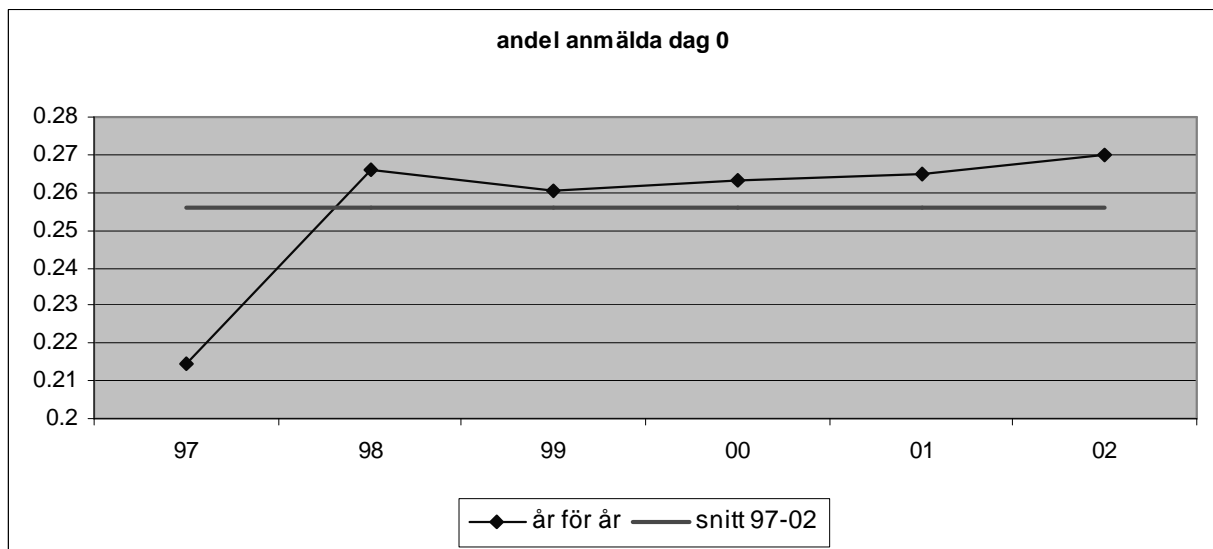


Figur 2.2: andel skador som anmäls på 0,1,...,30 dagar för skador inträffade under perioden 1997-2002

2.2 Fördelningsanpassning av anmälningstiden

Vi ska anpassa en parametrisk fördelning till den empiriska fördelningen för tiden från skadetillfälle till anmälan. Låt T vara denna anmälningstid. En parametrisk anpassning är mer lätthanterlig än den empiriska fördelningen när man vill göra förändringar i modellen, till exempel för att förutsäga något om framtiden. Den empiriska fördelningen kan vara besvärlig vid omfattande beräkningar och ger dessutom sannolikhetsmassan noll i intervall där det inte inträffat några observationer. En fördel med den empiriska fördelningen är dock att den är fri från antaganden. Därutöver kan självklart inte alla ojämnheter i den empiriska fördelningen fångas upp av en parametrisk. Men lätthanterligheten i kombination med att en parametrisk fördelning ger oss utjämning, gör att denna ändå är att föredra.

Som tidigare nämnts inrapporteras i detta fall en stor andel av skadorna redan samma dag. På grund av den stora skillnaden mellan sannolikheterna för dag 0 jämfört med dag 1, dag 2 osv kan då en bra anpassning bli svår att hitta om dessa skador bibehålls i datamaterialet under anpassningsförfarandet. Därför skattar vi sannolikheten $P(T=0)$ med den empiriska snittsannolikheten för detta utfall över åren 1997-2002, nämligen 0,256. Denna empiriska sannolikhet för andel anmälda dag 0 är stabil över åren, se figur 2.3. Övergångsåret 1997 avviker en aning, men inte så att det stör analysen.



Figur 2.3: andel anmälda samma dag som skadan inträffat för vart och ett av åren 1997-2002 jämfört med andel anmälda dag 0 under hela perioden (observera skalan på y-axeln).

Till resterande material anpassar vi $F_T(t)=P(T\leq t|t>0)$ genom att testa tre olika fördelningar, nämligen gammafördelningen, lognormalfördelningen och Paretofördelningen. Parametrarna skattas dels med momentmetoden, dels med ML-metoden. I momentmetoden bestämmer man parametrarna så att de första momenten i den empiriska fördelningen respektive den parametriska fördelningen överensstämmer med varandra. I ML-metoden får vi våra parameterskattningar genom att maximera likelihoodfunktionerna. (fördelningar och skattningar finns i appendix, för härledningar hänvisas till Johansson (1997)).

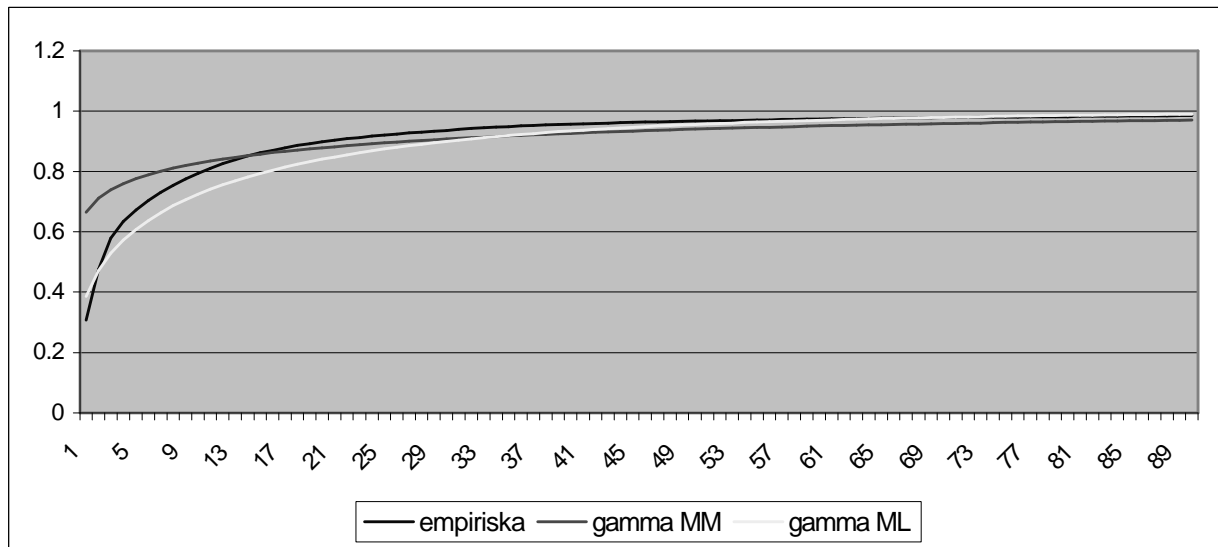
Tabell 2.2 : parameterskattningarna för de tre olika parametriska fördelningarna med de två olika skattningsmetoderna

FÖRDELNING	METOD	PARAMETERSKATTNINGAR	
gamma	Moment	$\alpha = 0,0985$	$\beta = 103,3125$
	ML	$\alpha = 0,2976$	$\beta = 34,1993$
lognormal	Moment	$\mu = 1,1144$	$\sigma = 2,3272$
	ML	$\mu = 1,2779$	$\sigma = 1,5354$
Pareto	Moment	$\alpha = 12,4019$	$\gamma = 2,2186$
	ML	$\alpha = 6,1149$	$\gamma = 1,5932$

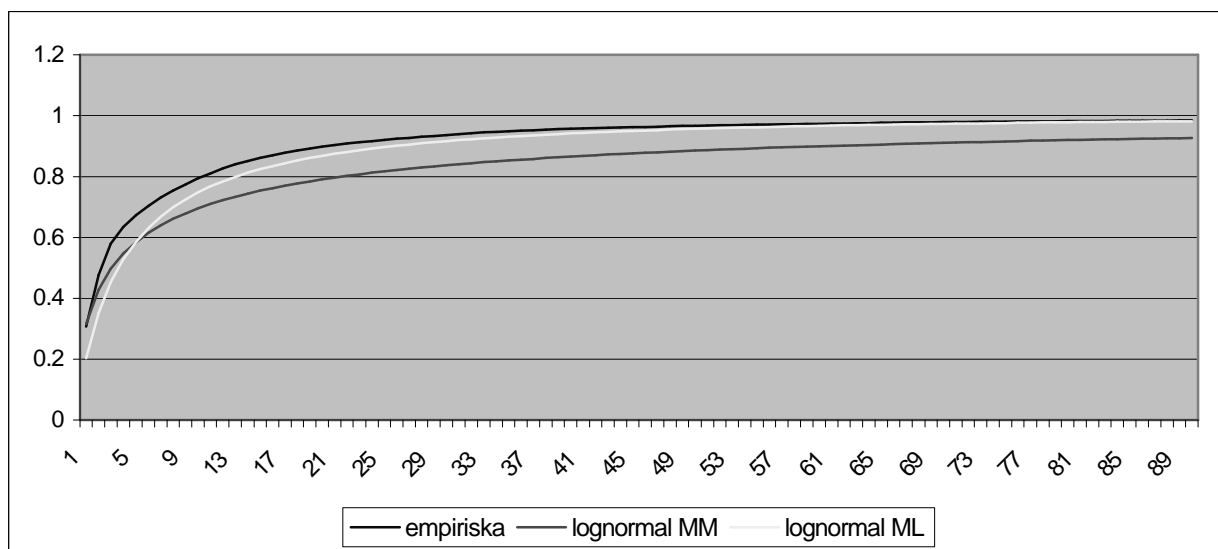
En jämförelse av värdena för de olika fördelningsfunktionerna med den empiriska fördelningsfunktionen (utan skador med $T = 0$) samt plottar över dessa ger båda en indikation på att Paretofördelningen med ML-skattade parametrar passar bäst. I beräkningarna hädanefter väljer vi därför den som vår anpassade fördelningsfunktion.

Tabell 2.3 : en jämförelse av den empiriska fördelningsfunktionens värden med de olika anpassade fördelningarnas värden för skador anmälda på 1,2,...,25 dagar

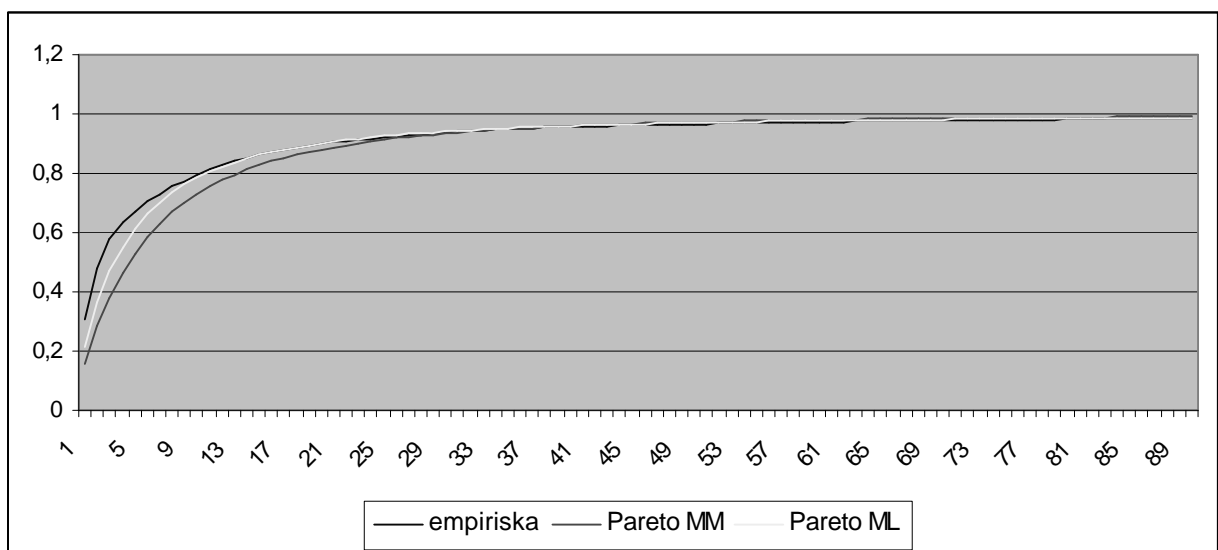
DAGAR	EMPIR	GAM MM	GAM ML	LOGN MM	LOGN ML	PAR MM	PAR ML
1	0,307	0,665	0,387	0,316	0,203	0,158	0,214
2	0,476	0,711	0,472	0,428	0,352	0,282	0,363
3	0,579	0,739	0,529	0,497	0,454	0,382	0,471
4	0,635	0,760	0,573	0,547	0,528	0,462	0,552
5	0,672	0,776	0,608	0,584	0,585	0,528	0,614
6	0,704	0,790	0,638	0,614	0,631	0,583	0,664
7	0,731	0,801	0,664	0,640	0,668	0,629	0,703
8	0,754	0,811	0,686	0,661	0,699	0,669	0,736
9	0,775	0,820	0,707	0,679	0,725	0,702	0,764
10	0,794	0,828	0,725	0,695	0,748	0,731	0,786
11	0,811	0,835	0,741	0,709	0,767	0,756	0,806
12	0,826	0,841	0,756	0,722	0,784	0,777	0,823
13	0,840	0,847	0,770	0,733	0,799	0,796	0,837
14	0,851	0,853	0,782	0,744	0,812	0,813	0,850
15	0,862	0,858	0,794	0,753	0,824	0,828	0,861
16	0,870	0,862	0,804	0,762	0,835	0,841	0,871
17	0,878	0,867	0,814	0,770	0,844	0,853	0,880
18	0,886	0,871	0,824	0,777	0,853	0,863	0,888
19	0,893	0,875	0,832	0,784	0,861	0,873	0,895
20	0,899	0,879	0,841	0,791	0,868	0,881	0,901
21	0,904	0,882	0,848	0,797	0,875	0,889	0,907
22	0,908	0,886	0,855	0,802	0,881	0,896	0,912
23	0,913	0,889	0,862	0,807	0,887	0,902	0,917
24	0,917	0,892	0,868	0,812	0,892	0,908	0,921
25	0,921	0,895	0,874	0,817	0,897	0,914	0,925



Figur 2.4: den empiriska fördelningsfunktionen för tid till anmälan jämfört med den anpassade gammafördelningen



Figur 2.5: den empiriska fördelningsfunktionen för tid till anmälan jämfört med den anpassade lognormalfördelningen



Figur 2.6: den empiriska fördelningsfunktionen för tid till anmälan jämfört med den anpassade Paretofördelningen

2.3 Skattning av antal efteranmälda skador

På kommande sidor ska vi angripa det viktiga problemet att beräkna det förväntade antalet skador som idag har inträffat, men som ännu inte rapporterats, IBNR.

Vi tänker oss att vi idag vill kalkylera hur mycket medel som måste avsättas för stöldskador inträffade på personbil någon gång under den bakomliggande treårsperioden. (Periodlängden tre år (1095 dagar) är vald eftersom preskriptionstiden för stöldskador är tre år, men man skulle i princip kunna välja hur lång eller kort period som helst. Bland alla de skador jag tittat på finns det dessutom bara två skador som inte anmälts på tre år.) Varje dag under denna treårsperiod har det inträffat ett visst antal skador. Några av dem har fram till idag hunnit inrapporteras. Dessa är kända. En del kan dock fortfarande komma att anmälas någon gång i framtiden. Dessa måste vi skatta.

2.3.1 Dagsmetoden

För överskådlighetens skull ställer vi upp problemet enligt tabell 2.4 nedan. Observera att dagarna till skillnad från den sedvanliga triangeluppställningen motsvarar kalenderdagar, inte utvecklingsdagar. Så om vi väljer *idag* som 31 dec 2001 svarar *dag 0* mot 1 jan 1999. Vi är då intresserade av de skador som har inträffat under perioden 01jan99-31dec01, men som idag, den 31 dec 2001, ännu inte anmälts.

Tabell 2.4

	Idag									
Dag	0	.	.	.	i	1095
N ₀										N _{0,1095}
N ₁₀₉₅										N _{1095,0}

Låt N_i = det totala antalet skador som inträffat dag i .

Idag känner vi $N_{0,1095}, \dots, N_{i,t}, \dots, N_{1095,0}$, där

$N_{i,t}$ = antal skador som inträffat dag i och som har rapporterats på $0 \leq \text{dagar} \leq 1095 - i$

Speciellt är:

$N_{0,1095}$ = antal skador som inträffat dag 0 och som har rapporterats på $0 \leq \text{dagar} \leq 1095$

$N_{1095,0}$ = antal skador som inträffat dag 1095 och som har rapporterats på 0 dagar, dvs samma dag som skadan inträffat

Det vi söker är $N_i - N_{i,t} = N^{\geq t}$, dvs antal skador som inträffat dag i , men som fortfarande kan komma att anmälas någon gång från imorgon och i dagar framåt.

Låt I_1 och I_2 vara två indikatorvariabler:

$$I_1\{j\} = \begin{cases} 1 \text{ om skada } j \text{ är rapporterad } 0 < \text{dagar} \leq 1095 - i & \text{med sh } (1 - p) \cdot F_T(t) \\ 0 \text{ annars} \end{cases}$$

$$I_2\{j\} = \begin{cases} 1 \text{ om skada } j \text{ är rapporterad på } t = 0 \text{ dagar} & \text{med sh } p = P(T = 0) \\ 0 \text{ annars} \end{cases}$$

där den anpassade fördelningsfunktionen $F_T(t) = P(T \leq t | T > 0)$ är sannolikheten att en skada anmäls på $\leq t$ dagar, givet att den inte anmälts samma dag.

Vi kan då skriva $N_{i,t}$ som:

$$N_{i,t} = \sum_{j=1}^{N_i} I_1\{j\} + \sum_{j=1}^{N_i} I_2\{j\} = \sum_{j=1}^{N_i} (I_1\{j\} + I_2\{j\})$$

Tar vi väntevärdet på båda sidor får vi:

$$\begin{aligned} E(N_{i,t}) &= E\left(\sum_j^{N_i} (I_1\{j\} + I_2\{j\})\right) = E(N_i) \cdot E(I_1\{j\} + I_2\{j\}) = \\ &= E(N_i) \cdot (F_T(t)(1 - p) + p) \end{aligned}$$

Detta leder oss till en skattning av N_i^{DM} (där DM står för DagsMetoden):

$$\hat{N}_i^{DM} = \frac{N_{i,t}}{(F_T(t)(1 - p) + p)} \quad i+t=1095 \quad (2.1)$$

Observera här att $N_{i,t}$ är känd samt att p skattas med den empiriska sannolikheten för att $T=0$, medan $F_T(t)$ skattas enligt avsnitt 2.2.

Det skattade totala antalet efteranmälda skador för perioden i fråga blir alltså:

$$\sum_{i=1}^{1095} (\hat{N}_i - N_{i,t}) = \sum_{t=1}^{1095} \hat{N}^{\geq t} \quad (2.2)$$

När vi applicerar (2.2) på några olika testperioder får vi nedanstående skattningar jämfört med vad det faktiska antalet efteranmälda skador blev. I tabellen visas också det totala antalet

skador, både det verkliga och det skattade, för att ge en känsla för hur stor del av alla skadorna som egentligen efteranmäls. ”Idag känt” N innebär exempelvis för 01jan98-31dec00 de skador som inträffat under just nämnda period och som också hunnit anmälas fram till ”idag”, den 31 dec 2000.

I beräkningarna har Paretofördelningen med ML-skattade parametrar använts som den anpassade $F_T(t)$.

Tabell 2.5

INTRÄFFANDE-PERIOD	TOTALT N	SKATTAT N	”IDAG KÄNT” N	EFTER-ANMÄLDA	SKATTADE EFTER-ANMÄLDA	DIFFERENS
01jan95-31dec97	56200	55969	55604	596	365	-231
01jan96-31dec98	62810	62709	62305	505	404	-101
01jan97-31dec99	68911	68945	68516	395	429	34
01jan98-31dec00	67573	67466	67165	408	301	-107
01jan99-31dec01	59250	59035	58811	439	224	-215
01jan00-31dec02	51133	51047	50768	365	279	-86

Vi får ett som synes ojämnt resultat. Observera dock att resultaten för de två första perioderna bör tolkas med försiktighet. Vi upptäckte ju tidigare att anmälningsstrukturen för skador inträffade före och efter 1997 i jämförelse ser annorlunda ut. Därför beror kanske differenserna delvis på att vår anpassade fördelningsfunktion inte lämpar sig till dessa tidiga år.

I tabellen kan vi utläsa att för period 01jan97-31dec99 blir skattningen nära det riktiga resultatet, men till exempel för perioden 01jan99-31dec01 fås en markant underskattning. Vid närmare granskning framgår det att ”dagsskattningarna” blir väldigt skakiga på slutet av testperioden. Med andra ord har vi en mycket stor varians då t är litet. Beror det på att den anpassade parametriska fördelningen är illa vald även till åren efter 1997? Om vi påminner oss sambandet mellan det totala antalet skador och de som vi idag har vetskap om så inser vi att det inte är den främsta anledningen. Vi förstår att det är antalet $N_{i,t}$, dvs de idag kända skadorna, som inverkar mest på skattningen.

Ponera att det för två dagar sedan inträffade sextio skador och att ingen av dessa idag hunnit anmälas. Sannolikheten för denna händelse är stor, trots att de flesta skador anmäls samma dag, så detta antagande känns rimligt. Enligt (2.1) blir då det förväntade totala antalet skador och därmed också det skattade antalet efteranmälda skador noll. Denna skattning skulle erhållits oavsett vilken fördelning vi använt oss av. Man förväntar sig alltså ingen skada överhuvudtaget och missar med andra ord alla de sextio skadorna. Skulle å andra sidan, säg 17 skador, ha hunnit inrapporteras idag, blir skattningen av de efteranmälda skadorna i det närmaste identisk med den verkliga summan.

Denna stora slumpmässiga variation i skattningarna av antal efteranmälda skador känns inte helt tillfredsställande. Hur ska vi då angripa problemet för att få mer tillförlitliga skattningar?

2.3.2 Skadefrekvensmetoden

Hur stora bestånden är vid olika tidpunkter, dvs hur stort det totala antalet försäkringar är (i det här fallet delkaskoförsäkringar), har vi kännedom om. Antalet skador som förväntas

inträffa kan vi uppskatta via skadefrekvensen. Ju fler skador som inträffar, desto fler skador borde rimligtvis också efteranmälas. Så, om vi konstruerar en modell för skadefrekvensen för stöldskador kan vi via (2.3) nedan skatta antalet förväntade skador per dag i (där SM står för SkadefrekvensMetoden):

$$\hat{N}_i^{SM} = \frac{\text{skadefrekvens}}{365} \cdot \text{bestånd} \quad (2.3)$$

Analysen av skadefrekvensen gör vi med generaliserade linjära modeller, GLM. GLM är en generalisering av den allmänna linjära modellen. Utifrån de modellantaganden som föreligger inom skadeförsäkring (se appendix) är antalet inträffade skador Poissonfördelat. Skadefrekvensen får då en relativ Poissonfördelning. Med hjälp av programpaketet Emblem ansätter vi en multiplikativ modell till skadefrekvensen och erhåller också våra parameterskattningar. Den multiplikativa modellen innebär att logaritmen av väntevärdet är linjär i ett antal förklarande variabler. Förklarande variabler kan i det här fallet exempelvis vara bilmärke och bilens ålder.

Om vi bara har en trovärdig modell för skadefrekvensen får vi genom sambandet (2.3) ”dagsskattningar” som inte påverkas av de idag kända skadorna. Skattningarna kommer att vara lika stabila alternativt icke-stabila oavsett om det gäller antal skador inträffade dag 1 eller dag 1000 i testperioden.

Viktigt är dock att inte bortse från det vi dag redan vet. Vi önskar en sammanvägning av de två ovan beskrivna skattningsmetoderna. Vi väljer att använda samma angreppssätt som Bornhuetter-Ferguson gör när de väger samman Chain Ladder med känd skadekostnad.

Låt oss därför skriva om vårt N_i , det totala antalet skador inträffade dag i , som:

$$N_i = \underbrace{N_{i,t}}_{\text{hitills anmälda - känt}} + \underbrace{N_i - N_{i,t}}_{\text{antal efteranmälda - okänt}} = N_{i,t} + \left(1 - \frac{N_{i,t}}{N_i}\right) N_i$$

”Vikten” $\frac{N_{i,t}}{N_i}$ skattas med $(F_T(t)(1-p) + p) = w$ enligt (2.1)

Detta ger oss vårt önskade resultat:

$$\hat{N}_i = \hat{N}_i^{DM} \cdot w + \hat{N}_i^{SM} \cdot (1-w) \quad (2.4)$$

Det vi söker är som bekant antal efteranmälda skador, $N_i^{\geq t}$:

$$\begin{aligned} &= \hat{N}_i - N_{i,t} = \hat{N}_i^{SM} \cdot (1-w) = \hat{N}_i^{SM} \cdot (1 - (F_T(t)(1-p) + p)) = \\ &= \hat{N}_i^{SM} \cdot (1-p)(1-F_T(t)) = \frac{\text{skadefrekvens}}{365} \cdot \text{bestånd} \cdot (1-p)(1-F_T(t)) \end{aligned} \quad (2.5)$$

Vi antar i denna studie att bestånden är konstanta halvårsvis och gör även samma antagande om skadefrekvensen. Dessa antaganden grundar vi på det faktum att bestånden inte förändras sig så pass mycket från den ena dagen till den andra att det kommer ha avgörande betydelse för våra skattningar.

Den sammanvägda skadefrekvensmetoden (2.5) ger oss nu de skattningar av antal efteranmälda skador som visas i tabell 2.6. I tabellen ser vi också det skattade antalet efteranmälda skador vi skulle erhålla om vi via skadefrekvensen skulle gissa exakt rätt antal skador per dag.

Tabell 2.6

INTRÄFFANDE PERIOD	EFTER-ANMÄLDA	SKATTADE EFTER-ANMÄLDA	DIFFERENS	SKATTADE EFTER-ANMÄLDA "EXAKT"	DIFFERENS
01jan95-31dec97	596	478	-118	444	-152
01jan96-31dec98	505	521	16	435	-70
01jan97-31dec99	395	479	84	435	40
01jan98-31dec00	408	362	-46	352	-56
01jan99-31dec01	439	354	-85	294	-145
01jan00-31dec02	365	348	-17	296	-69

Det är intressant att jämföra vårt skattade antal med de värden vi skulle ha fått vid korrekt antal skador per dag, eftersom vi på så sätt kan upptäcka var den främsta osäkerheten i våra skattningar ligger. Vi noterar att antalet efteranmälda skador beräknas bli konsekvent lägre vid "rätt" antal skador per dag än vid estimerat antal dagsskador. Endast i ett av fallen (period 01jan97-31dec99) innebär detta ett resultat närmare verkligheten. Med andra ord skulle en perfekt, dagsvis uppdaterad skadefrekvens inte ge oss bättre skattningar i just de här testperioderna. Intuitivt borde dock motsatsen vara fallet. Tyvärr har vi med så få data svårt att bedöma precisionen av vår modell, något som en simulering skulle kunnat hjälpa oss med.

2.3.3 Chain Ladder

Då Chain Ladder-metoden i denna uppsats endast ska användas att jämföra med, hänvisar vi exempelvis till Ohlsson & Johansson (2003) eller Dahl (2003) för detaljbeskrivning. Vi påminner oss här bara att Chain Ladder baseras uteslutande på tidigare erfarenheter. Metoden utgår ifrån aggregerade data i form av en triangel. Via utvecklingsfaktorer som bygger på denna historiska triangel predikterar man sedan det slutliga antalet rapporterade skador.

Nedan redovisas Chain Ladder-skattningarna av antal efteranmälda skador i denna studie.

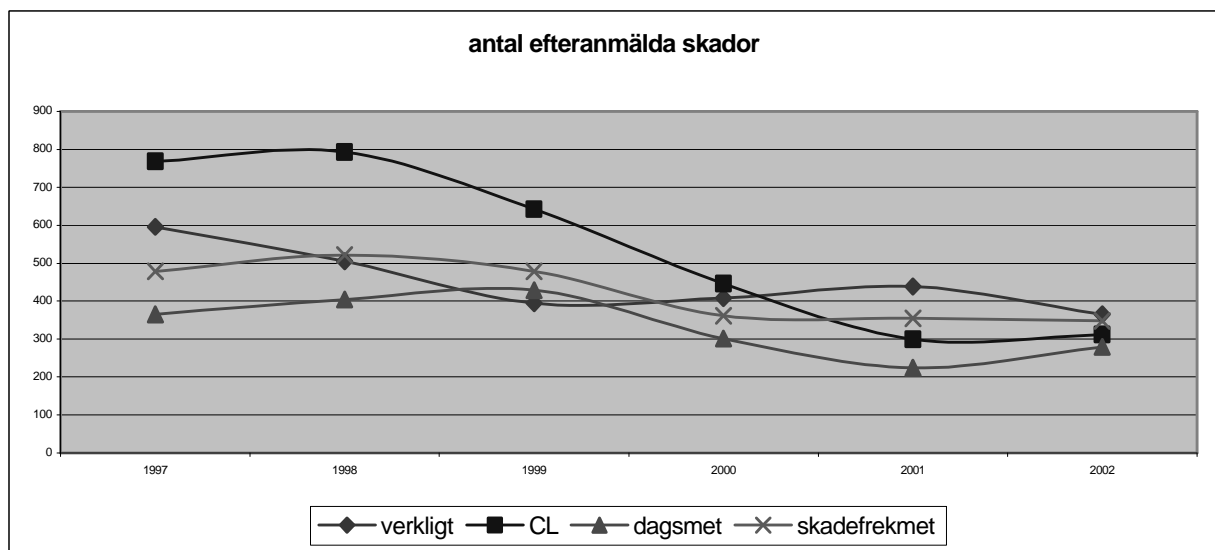
Tabell 2.7

INTRÄFFANDE PERIOD	EFTER-ANMÄLDA	SKATTADE EFTERANMÄLDA	DIFFERENS
01jan95-31dec97	596	769	173
01jan96-31dec98	505	793	288
01jan97-31dec99	395	643	248
01jan98-31dec00	408	447	39
01jan99-31dec01	439	299	-140
01jan00-31dec02	365	312	-53

Precis som för dagsmetoden och skadefrekvensmetoden ger Chain Ladder-metoden mycket varierande resultat. Överlag får vi dock fler överskattningar med Chain Ladder. En tänkbar förklaring till de ojämna skattningarna är att vi vid beräkningarna utgår från år, inte månader. Eftersom majoriteten av stöldskadorna på personbil anmäls inom bara några dagar får vi endast fyra skade- och utvecklingsår att jobba med. Utvecklingsfaktorerna blir då osäkra och fångar inte upp verkligheten på ett tillförlitligt sätt.

2.3.4 Jämförelse av metoderna

Låt oss jämföra våra skattningar med de som Chain-Ladder-metoden ger. Figur 2.7 visar de tre metodernas skattningar jämfört med det verkliga resultatet.



Figur 2.7: skattat antal efteranmälda skador jämfört med det egentliga antalet

Vi ser att den sammanvägda skadefrekvensmetoden är den metod som ger bäst resultat i fyra av de sex fallen.

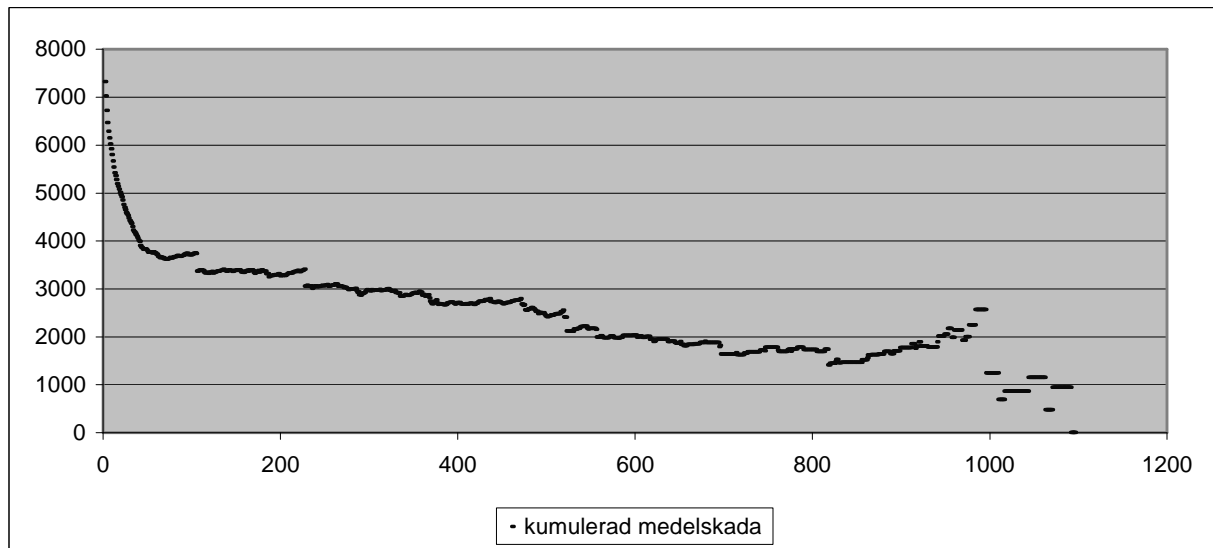
2.4 Skattning av kostnaden för de efteranmälda skadorna

Vi har nu det förväntade antalet efteranmälda skador. Men för att kunna bestämma IBNR-reserven behöver vi naturligtvis veta vad dessa skador kommer att kosta. Vi ställs dessvärre inför två problem vid beräkandet av kostnaden.

Det första problemet uppkommer i och med att då skadorna inte har anmälts ännu, har vi ingen information alls om deras egenskaper och premieargument. Vi vet inte om det rör sig om en stulen bil eller en stöld ur en bil, vi vet inte vilket bilmärke bilen i fråga har och vi vet inte hur gammal ägaren till bilen är. Det enda vi med säkerhet kan uttala oss om är att skadan kommer att anmälas vid någon godtycklig tidpunkt efter idag, vilket leder oss in på det andra problemet. Då vi inte vet exakt när i tiden inrapportering av skadan sker kan vi bara vara säkra på att de skador som inträffade för exempelvis fem dagar sedan och som inte anmälts idag, kommer att anmälas på ≥ 6 dagar. Detta gör att vi inte kan ansätta någon GLM med anmälningstiden som förklarande variabel.

Ett förslag som kan användas för att komma tillrätta med denna problematik är att ur vårt material av kända stölskador ta fram en medelskada, M_t , för skador anmälda på $\geq t$ dagar, $t:1-1095$, och se om vi kan skönja en regelbunden struktur. Om så är fallet skulle en utjämnande kurvanpassning vara en tänkbar ansats.

$$M_t = \frac{\text{sammanlagd skadekostnad för de skador med anmälningstid} \geq t \text{ dagar}}{\text{totalt antal skador med anmälningstid} \geq t \text{ dagar}} \quad (2.6)$$



Figur 2.8: medelskadan M_t (observera att M_t innebär medelskadan för skador anmälda på $\geq t$ dagar)

Som synes i figur 2.8 har vi en tydlig avtagande trend för denna ”kumulerade” medelskada. De skakiga observationerna på slutet kan vi bortse ifrån eftersom vi endast har enstaka skador där. Det stora hoppet vid 104 dagar är orsakat av en enskild skada med en slutlig skadekostnad på 913754 kr.

Självklart är även här stabila utvecklingsmönster nödvändiga för att en eventuell kurvanpassning ska vara möjlig. Låt oss därför undersöka M_t för vart och ett av åren och jämföra med M_t beräknad utifrån alla åren 1990-2002.

Tabell 2.8: medelskadan för skador anmälda på $\geq t$ dagar (observera den abnormala siffran för M_{100} för 2002, som beror på en enskild skadekostnad på 913754 kr för en skada med anmälningstid 104 dagar)

	Alla	90	91	92	93	94	95	96	97	98	99	00	01	02
M_1	7319	7054	7846	7605	7582	7166	7555	7374	7264	7585	6977	6717	7060	7025
M_2	7023	6774	7596	7479	7391	6952	7470	7123	6810	7004	6332	6314	6503	6503
M_3	6717	6468	7351	7232	7220	6736	7227	6906	6441	6483	5606	5752	5922	6020
M_{10}	5666	5439	6157	5913	6118	5634	6088	5986	5598	5330	4504	4532	5143	5440
M_{50}	3758	3513	3651	3739	3768	3385	3676	3610	3611	3409	3691	3497	3979	7529
M_{100}	3726	2994	3480	4081	3334	3419	3035	3514	3131	3321	3407	3469	3962	10650

Det verkar som om kostnaderna för skador från de tidiga åren överlag är högre än de senare. Vi drar oss till minnes den förändring i hantering av inrapporterade skador som skedde 1997 och som påverkade det mätbara anmälningsmönstret. Vi inser att denna utveckling av arbetssättet borde influera medelskadan likväl som rapporteringsstrukturen. Därför baserar vi istället skadekostnadsberäkningarna på skador med skadetillfälle från 1997-2002.

Tabell 2.9: medelskadan för skador med anmälningstid $\geq t$ dagar, beräknad utifrån åren 1997-2002

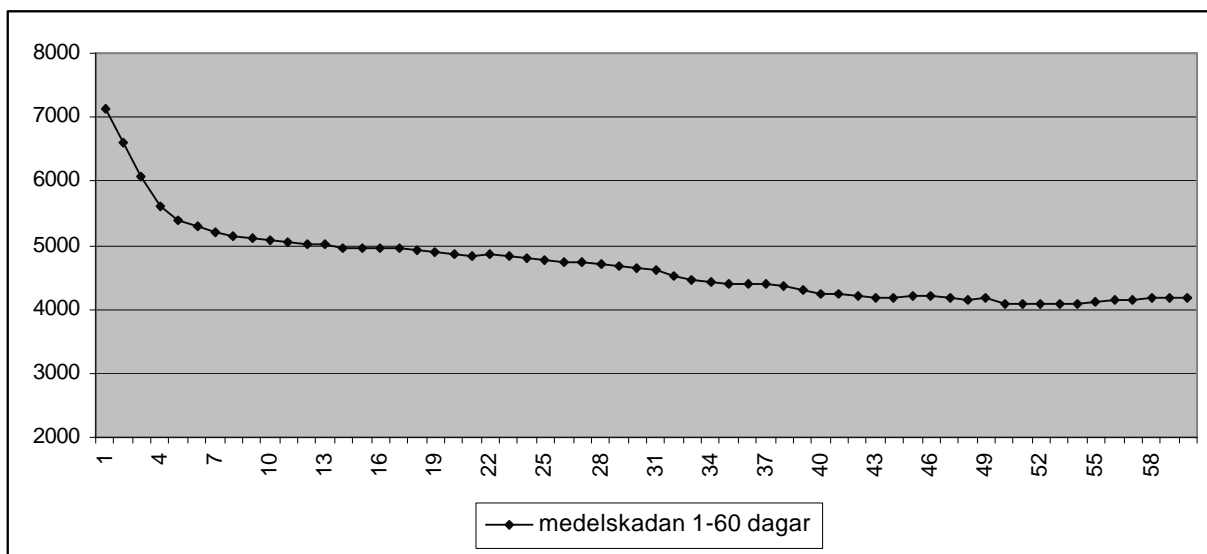
Medelskadan för skador anmälda på ≥ 1 dag, M_1	7124 kr
Medelskadan för skador anmälda på ≥ 2 dagar, M_2	6599 kr
Medelskadan för skador anmälda på ≥ 3 dagar, M_3	6062 kr
Medelskadan för skador anmälda på ≥ 10 dagar, M_{10}	5069 kr
Medelskadan för skador anmälda på ≥ 100 dagar, M_{100}	4406 kr
Medelskadan för skador anmälda på ≥ 500 dagar, M_{500}	2087 kr

Vi får ett liknande avtagande mönster för medelskadan vid 97-02-beräkningen som vid 90-02-beräkningen.

2.4.1 Utjämnande kurvanpassning

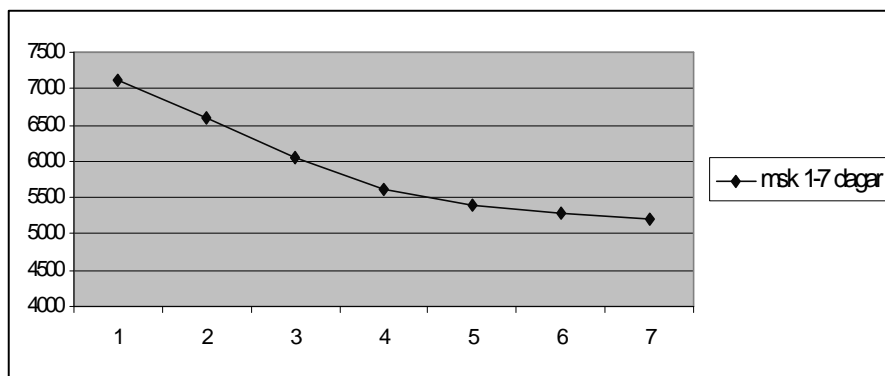
Precis som vid alla former av statistiskt modellbygge önskar vi hitta en förhållandevis enkel anpassning som samtidigt passar väl till data. Därigenom får vi en lättarbetad metod som förhoppningsvis ger bra utjämning av slumpmässiga effekter.

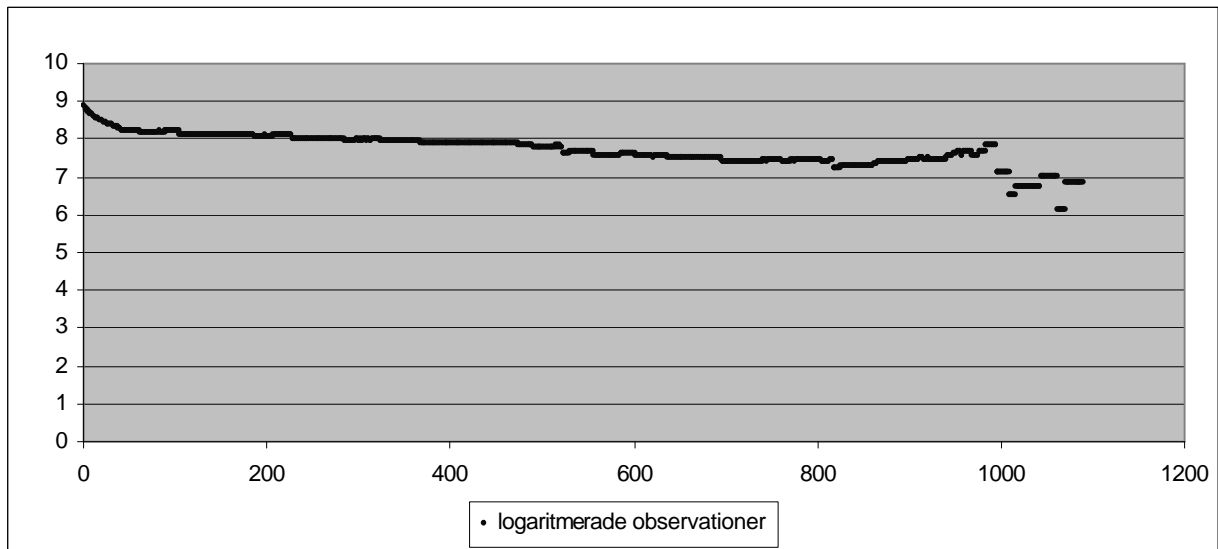
I vårt fall är det viktigast att vi har en god anpassning i början, eftersom det är där flest antal skador inträffar. Låt oss förstora kurvan och granska mönstret under de första två månaderna.



Figur 2.9: den kumulerade medelskadan för dag 1-60

De första sju dagarna ser ut att följa ett specifikt mönster, medan vi för resterande observationer troligtvis skulle tjäna på att logaritmera.





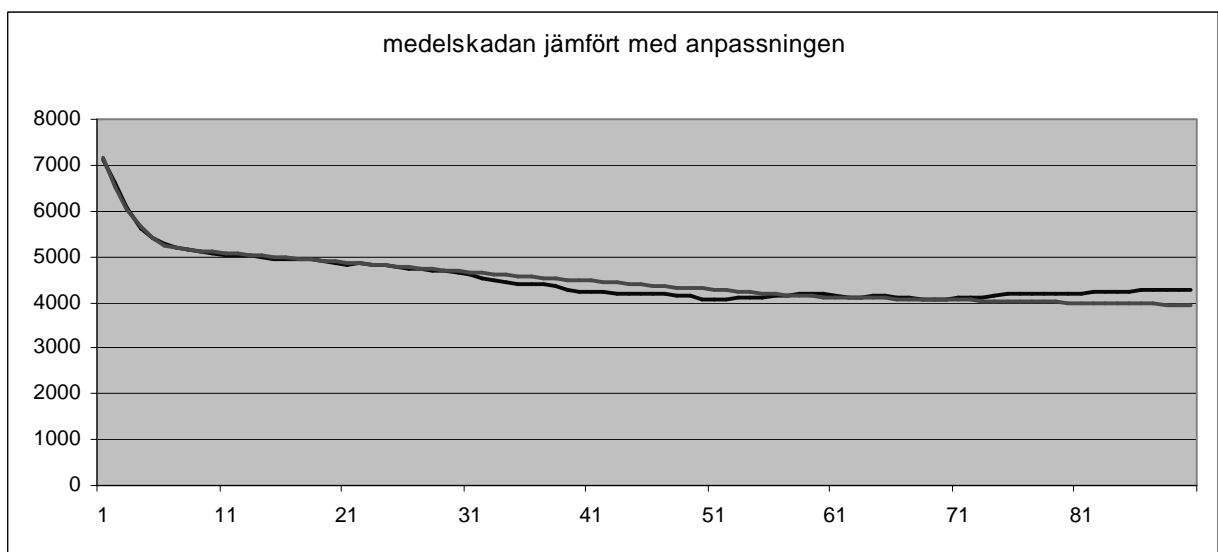
Figur 2.11: logaritmerad medelskada

Vi beslutar oss för att använda de empiriska värdena för kostnaden för de första sju dagarna. $M(x)$ står för medelskadan för skador med x eller fler dagar från skadetillfälle till anmälan.

M(1)	7124.3 kr
M(2)	6599.0 kr
M(3)	6062.3 kr
M(4)	5601.4 kr
M(5)	5400.0 kr
M(6)	5285.9 kr
M(7)	5191.6 kr

Efter att ha testat några olika tänkbara varianter väljer vi för 8 eller fler dagar nedanstående loglinjära modell:

$$M(x) = \begin{cases} \exp\{-0,0043x + 8,58\} & \text{för } 8 \leq x \leq 60 \\ \exp\{-0,0014x + 8,406\} & \text{för } 61 \leq x \leq 1095 \end{cases} \quad (2.7)$$



Figur 2.12: den empiriska kumulerade medelskadan jämfört med den utjämnande anpassningen $M(x)$. Observera skalan på x-axeln!

2.5 IBNR-reserven

För att beräkna reserven utnyttjar vi nu de skattningar vi har av antal efteranmälda skador per dag tillsammans med den förväntade kostnaden för dessa. Förfaringsättet blir som följer:

Tabell 2.10

t	ANTAL SKADOR MED ANMTID $\geq t$	KOSTNAD FÖR SKADOR MED ANMTID $\geq t$	FÖRVÄNTAD TOTALKOSTNAD
1	$N^{\geq 1}$	$M(1)$	$N^{\geq 1} * M(1)$
2	$N^{\geq 2}$	$M(2)$	$N^{\geq 2} * M(2)$
.	.	.	.
.	.	.	.
1095	$N^{\geq 1095}$	$M(1095)$	$N^{\geq 1095} * M(1095)$

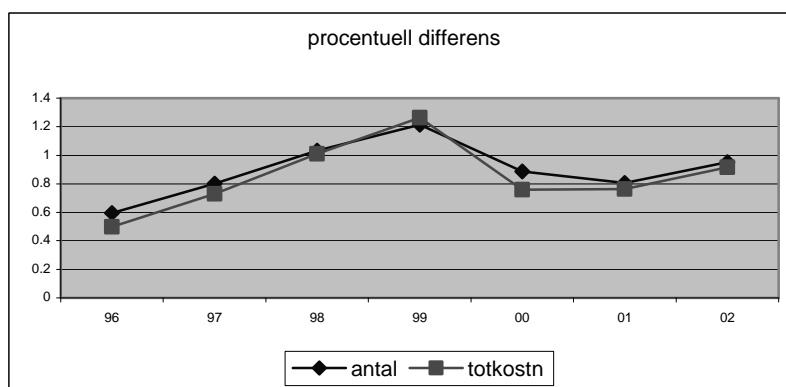
$$\text{RESERV} = \sum_{t=1}^{1095} N^{\geq t} \cdot M(t) \quad (2.8)$$

Låt oss nu enligt (2.8) beräkna den totala förväntade kostnaden för de efteranmälda skadorna, för de olika perioderna. Vi använder skadefrekvensmetodens skattningar av antalet skador. Eftersom vi hela tiden arbetar med ett slutreglerat material kan vi ju också kalkylera och jämföra med vad den faktiska kostnaden blev.

Tabell 2.11

INTRÄFFANDEPERIOD	VERKLIG KOSTNAD	UPPSKATTAD KOSTNAD	DIFFERENS
01jan95-31dec97	3 335 122 kr	2 435 483 kr	-27,0 %
01jan96-31dec98	2 619 078 kr	2 646 289 kr	+ 1,0 %
01jan97-31dec99	1 908 092 kr	2 412 140 kr	+26,4 %
01jan98-31dec00	2 384 658 kr	1 811 966 kr	-24,0 %
01jan99-31dec01	2 325 347 kr	1 777 222 kr	-23,6 %
01jan00-31dec02	1 912 435 kr	1 753 484 kr	-8,3 %

Den uppskattade kostnaden blir som synes i tabell 2.11 instabil. Beror fluktuationen då på kostnadsanpassningen eller skattningen av antal skador? Om vi granskar den procentuella differensen mellan verklig och uppskattad kostnad och jämför den med samma differens för inträffat vs skattat antal skador ser vi



att de följs åt. (se figur höger) Detta indikerar betryggande nog att kurvanpassningen $M(x)$ passar väl till var och en av de olika årgångarna. Det tyder också på att skattningarna av *antalet* skador spelar den centrala rollen för reservbestämningen. Exempelvis för perioden 96-98, då vi överskattade antal efteranmälda skador med endast 3,2%, tror vi också att den slutliga kostnaden bara blir marginellt större än den verkliga.

Vi gissar alltså att den förväntade kostnaden kommer att stämma väl överens med den verkliga om vi visste det korrekta antalet efteranmälda skador. I tabell 2.12 bekräftas våra aningar. Vi ser återigen att i denna metod med vår anpassning $M(x)$ är *antalet* skador avgörande för bestämmandet av reserven.

Tabell 2.12: den verkliga kostnaden för de efteranmälda skadorna jämfört med den estimerade kostnad vi skulle fått om antalet skador var känt

INTRÄFFANDEPERIOD	VERKLIG KOSTNAD	VERKLIGT ANTAL SKADOR*M(X)	DIFFERENS
01jan95-31dec97	3 335 122 kr	3 068 547 kr	-8,0 %
01jan96-31dec98	2 619 078 kr	2 538 006 kr	-3,1 %
01jan97-31dec99	1 908 092 kr	1 954 912 kr	+2,5 %
01jan98-31dec00	2 384 658 kr	2 037 331 kr	-14,6 %
01jan99-31dec01	2 325 347 kr	2 174 648 kr	-6,5 %
01jan00-31dec02	1 912 435 kr	1 726 567 kr	-9,7 %

3 RESULTAT OCH SLUTSATSER

Syftet med denna studie var att utvärdera en individuell metodik för skattningen av ett försäkringsbolags reserv. Huvudfokus har legat på de skador som idag ännu inte inrapporterats, IBNR. Det vi genomgående baserat våra beräkningar på är tiden från skadetillfälle till anmälan. I och med detta har vi stött på ett återkommande problem som genomsyrat hela analysen. Vid vilken tidpunkt vi än väljer att stanna och uppdatera reserven, vet vi bara *att* det kommer efteranmälas skador, inte *när*.

Tidigt i analysen insåg vi att det kan innebära en risk att låta ett empiriskt mönster, som delvis styrs av omständigheter kring arbetssätt och dylikt, ligga till grund för en metod. Förändringar i anmälningsmönstret kan ju då till viss grad vara orsakade av den mänskliga faktorn och inte av förändringar i den underliggande samhällsstrukturen. Detta måste hållas väl i minnet vid användandet av denna metod! Chain Ladder bygger visserligen också på ett utvecklingsmönster, men som i det här fallet då ändringar i skaderegleringsrutinerna främst påverkade fördelningen för de första dagarna, kommer inte det att ge utslag om Chain Ladder tillämpas på exempelvis månader.

Att arbetssättet kring hantering av inrapporterade skador förändrades markant under 1997 gjorde att vi var tvungna att begränsa vår datamängd drastiskt. Detta innebär tyvärr osäkrare skattningar och färre jämförelseperioder.

Vid anpassandet av antal efteranmälda skador, fann vi att Paretofördelningen med ML-skattade parametrar passade bäst. Men den empiriska fördelningsfunktionen, tillhörande det specifika valet av skadetyper (stölskadorna på personbil) vi avgränsat oss till i detta arbete, har ett mycket speciellt utseende. Som vi såg i avsnitt 2.1 anmäls hela 25,6% av skadorna samma dag som de inträffat och efter endast en vecka har 95,3% av skadorna inrapporterats. Denna struktur kan vara svår att fånga upp med en parametrisk fördelning. Visserligen har vi bara testat tre olika fördelningar, men kanske hade det varit lättare att hitta en bra anpassning om vi analyserat en skadefördelning med mer jämn spridning mellan antalet anmälningsdagar. Vi gör trots allt bedömningen att Paretofördelningen stämmer relativt väl överens med den empiriska fördelningsfunktionen.

Slutsatserna beträffande dagsmetoden blir att den inte är tillförlitlig. Vi upptäckte i avsnitt 2.3.1 att de idag kända skadorna inverkar mest på skattningen av de efteranmälda skadorna. En så skakig metod med denna stora slumpmässighet känns inte betryggande.

Skadefrekvensmetoden gav oss bättre skattningar än både dagsmetoden och Chain Ladder. Två omständigheter gör dock att vi har svårt att uttala oss om precisionen i denna modell:

- (i) vi utgår ifrån skador inträffade under en bakomliggande treårsperiod
- (ii) vår anpassade Paretofördelning passar bara till data från 1997 och framåt

Dessa två i kombination gör att vi egentligen bara har fyra testperioder där vi kan jämföra skattning med utfall. Vi skulle behöva simulera fram data för att kunna dra några säkra slutsatser.

Avseende kostnaden för IBNR-skadorna fick vi åter möta problemet med att vi bara vet *att*, inte *när*, en skada anmäls. Detta ledde till att alla medelskadorna M_t , för skador anmälda på $\geq t$ dagar, blev beroende av varandra och vi kunde inte ansätta någon generaliserad linjär modell.

En utjämnande kurvanpassning blev vårt förslag på lösning, se avsnitt 2.4.1. Den approximation vi valde visade sig ge tillfredsställande resultat rörande totalkostnaden, under förutsättning att skattningen av antalet skador var god.

Följden av den begränsade datamängd vi har blir att slutsatserna beträffande den individuella metodiken blir vaga. Utifrån de synliga resultat vi har kan vi ändå säga att metoden, med vissa förbehåll, verkar kunna fungera som ett komplement till dagens metoder. Som vi såg i avsnitt 2.5 förefaller antalet skador vara av betydande vikt för reservprediktionen, men i denna studie har ingen nöjaktig modell för dessa fastställts. Om så vore fallet skulle metoden dock kunna användas för att beräkna den förväntade totalkostnaden för de idag inträffade, men ej rapporterade skadorna. Fördelen med detta angreppssätt är att vi lätt kan utnyttja metoden vid godtycklig tidpunkt under året.

Till sist kan några aspekter som inte hänsyn tagits till nämnas:

- Vi har inte undersökt huruvida någon säsongseffekt föreligger. Anmälningmönstret kanske ser olika ut under året, vilket skulle innebära att separata fördelningsanpassningar skulle vara nödvändiga.
- Vi har begränsat oss till endast en speciell typ av skada. Vi kan dessvärre inte uttala oss alls om denna metodik fungerar bättre eller sämre på andra skadearter.
- Angående skadekostnaden har inflationen inte tagits i beaktande.

4 KÄLLFÖRTECKNING

Dahl, P. (2003): *Introduction to reserving, Corrected ed.* Handout.

Gut, A. (1995): *An intermediate course in probability.* Springer-Verlag New York, Inc.

Hogg, R. V. & Klugman, S. A. (1984): *Loss distributions.* John Wiley & Sons, Inc.

Johansson, B. (1997): *Matematiska modeller inom sakförsäkring.* Kompendium. Matematisk statistik, Stockholms universitet.

Ohlsson, E. & Johansson, B. (2003): *Prissättning inom sakförsäkring med Generaliserade linjära modeller.* Kompendium. Matematisk statistik, Stockholms universitet.

Reistad, A. (1999): *Fördelningsanpassning av anmälningstiden vid skadeförsäkring.* Forskningsrapport. Matematisk statistik, Stockholms universitet.

Ross, S. M. (2000): *Introduction to Probability Models.* Seventh edition. Academic Press, USA.

Tamhane, A. & Dunlop, D. (2000): *Statistics and data analysis – from elementary to intermediate.* Prentice Hall, Inc., Upper Saddle River, USA.

Taylor, G. (2000): *Loss Reserving: An Actuarial Perspective.* Kluwer Academic Publishers, USA.

5 APPENDIX

5.1 Gammafördelningen

Täthetsfunktion: $f(x) = \frac{1}{\Gamma(\mathbf{a})} x^{\mathbf{a}-1} \frac{1}{\mathbf{b}^{\mathbf{a}}} e^{-x/\mathbf{b}} \quad , \quad x \geq 0$

Fördelningsfunktion: $F(x) = \frac{1}{\Gamma(\mathbf{a})} \cdot \frac{1}{\mathbf{b}^{\mathbf{a}}} \int_0^x v^{\mathbf{a}-1} e^{-v/\mathbf{b}} dv \quad , \quad x \geq 0$

Väntevärde: $E(X) = \mathbf{a} \cdot \mathbf{b}$

Varians: $Var(X) = \mathbf{a} \cdot \mathbf{b}^2$

Momentskattning: $\hat{\mathbf{a}} = \frac{\bar{x}^2}{s^2}$
 $\hat{\mathbf{b}} = \frac{s^2}{\bar{x}}$

\mathbf{a} löses numeriskt ur :

ML-skattning: $\log(\mathbf{a}) - \log(\bar{x}) - \frac{\Gamma'(\mathbf{a})}{\Gamma(\mathbf{a})} + \frac{1}{n} \sum_{k=1}^n \log(x_k) = 0$

$$\mathbf{b} = \frac{\bar{x}}{\hat{\mathbf{a}}}$$

Här har sekantmetoden, se Johansson (1997), använts för att lösa ovanstående ekvation numeriskt.

5.2 Lognormalfördelningen

Täthetsfunktion: $f(x) = \frac{1}{\sqrt{2\pi s x}} e^{-\frac{1}{2}(\log x - m)^2 / s^2} \quad , \quad x > 0$

Fördelningsfunktion: $F(x) = \Phi\left(\frac{\log x - m}{s}\right) \quad , \quad x > 0$

Väntevärde: $E(X) = e^{m + \frac{1}{2}s^2}$

Varians: $Var(X) = e^{2m}(e^{2s^2} - e^{s^2})$

Momentskattning:
$$\hat{\mathbf{m}} = 2 \log(\bar{x}) - \frac{1}{2} \log(s^2 + \bar{x}^2)$$

$$\hat{\mathbf{s}}^2 = \log(s^2 + \bar{x}) - 2 \log(\bar{x})$$

ML-skattning:
$$\hat{\mathbf{m}} = \frac{1}{n} \sum_{k=1}^n \log(x_k)$$

$$\hat{\mathbf{s}}^2 = \frac{1}{n} \sum_{k=1}^n (\log(x_k))^2 - \hat{\mathbf{m}}^2$$

5.3 Paretofördelningen

Täthetsfunktion:
$$f(x) = \frac{\mathbf{g} \mathbf{a}^{\mathbf{g}}}{(\mathbf{a} + x)^{\mathbf{g}+1}}, \quad x > 0$$

Fördelningsfunktion:
$$F(x) = 1 - \left(\frac{\mathbf{a}}{\mathbf{a} + x} \right)^{\mathbf{g}}, \quad x > 0$$

Väntevärde:
$$E(X) = \frac{\mathbf{a}}{\mathbf{g} - 1}, \quad \mathbf{g} > 1$$

Varians:
$$\text{Var}(X) = \frac{\mathbf{g} \mathbf{a}^2}{(\mathbf{g} - 1)^2 (\mathbf{g} - 2)}, \quad \mathbf{g} > 2$$

Momentskattning:
$$\hat{\mathbf{a}} = \bar{x} \frac{s^2 + \bar{x}^2}{s^2 - \bar{x}^2}$$

$$\hat{\mathbf{g}} = \frac{2s^2}{s^2 - \bar{x}^2}$$

sätt
$$a(\mathbf{a}) := \frac{1}{n} \sum_{k=1}^n \frac{\mathbf{a}}{\mathbf{a} + x_k} \quad \text{och} \quad b(\mathbf{a}) := \frac{1}{n} \sum_{k=1}^n \log\left(\frac{\mathbf{a}}{\mathbf{a} + x_k}\right)$$

då löses \mathbf{a} numeriskt ur :

ML-skattning:
$$\frac{1}{a(\hat{\mathbf{a}})} + b(\hat{\mathbf{a}}) - 1 = 0$$

$$\hat{\mathbf{g}} = \frac{a(\hat{\mathbf{a}})}{1 - a(\hat{\mathbf{a}})}$$

Här har sekantmetoden, se Johansson (1997), använts för att lösa ovanstående ekvation numeriskt.

5.4 Grundläggande modellantaganden

1. Utfallen för olika försäkringsavtal är oberoende av varandra.
2. Utfallen i disjunkta tidsintervall är oberoende av varandra.
3. Inom en tariffcell har två försäkringsavtal med samma exponering samma fördelning för utfallen.