

Matematisk statistik
Stockholms universitet

**Stokastisk reservsättning med Tweedie-modeller
och bootstrap-simulering**

Totte Pikanen

Examensarbete 2005:4

Postadress:

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm
Sverige

Internet:

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

Stokastisk reservsättning med Tweedie-modeller och bootstrap-simulering

Totte Pikanen¹

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 ersättningsreserv. Det finns många tillvägagångssätt för att estimerar storleken på reserven. Historiskt sett har de flesta metoderna endast gett en punktskattning av det totala reservbehovet. På senare år har det kommit fram en rad olika metoder som förutom punktskattning även ger ett osäkerhetsmått för reserven. I denna studie undersöks tre olika generaliserade linjära modeller, närmare bestämt Tweedie-modellerna *överspridd Poisson*, *sammansatt Poisson* och *Gamma*, som genom bootstrap-simulering ger en simulerad fördelning för reserven. Metoderna jämförs med varandra och problem med algoritmerna belyses.

¹ E-post: Totte.pikanen@trygghansa.se
Handledare: Esbjörn Ohlsson, Anders Lindström

Stochastic claims reserving using Tweedie models and bootstrap simulation

ABSTRACT

To be able to cover future expenses caused by claims incurred but not reported and claims reported but not settled, a general insurance company must allocate capital to a reserve. There are many approaches to estimate the size of this reserve. Historically most of the methods have given just a point estimate of the required reserve. In recent years a number of methods has been introduced which in addition to a point estimate gives a measure of the uncertainty in the reserve. In this study we examine the use of three different generalized linear models, namely the Tweedie models *overdispersed Poisson*, *compound Poisson* and *Gamma*, that together with bootstrap simulation gives a simulated distribution for the reserve. The methods are compared to each other and problems with the algorithms are discussed.

FÖRORD

Huvuddelen av denna undersökning genomfördes under våren 2003 på försäkringsbolaget Trygg-Hansa, aktuarieenheten division Privat, Affärsområde Sak, i Stockholm. I januari 2005 återupptogs arbetet med kompletteringar och vissa förändringar som följd. Undersökningen utgör ett examensarbete, omfattande 20 poäng för magisterexamen i matematisk statistik på Matematisk-datalogiska 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 gav mig möjligheten att få en inblick i arbetslivet.

Stort tack även till min handledare på Trygg-Hansa, Anders Lindström, för rådgivning och vägledning.

Min handledare på Stockholms Universitet, Esbjörn Ohlsson, skall också ha ett stort tack.

Till sist en eloge till hela den dåvarande aktuariegruppen för trevligt sällskap samt stöd, råd och tips längs vägen.

1	BAKGRUND	5
1.1	UPPGIFT.....	6
2	METODER	6
2.1	ANTAGANDEN	6
2.2	BOOTSTRAP-SIMULERING	7
2.2.1	<i>Bootstrap-algoritmen steg för steg</i>	9
3	DATA	10
3.1	UTVECKLINGSTRIANGLAR	10
4	RESULTAT AV TILLÄMPNING PÅ DATA FRÅN TRYGG-HANSA	12
4.1	MODELLERNAS ANPASSNING TILL DATA	12
4.1.1	<i>Försäkringsklass civil</i>	13
4.1.2	<i>Försäkringsklass trafik</i>	14
4.1.3	<i>Försäkringsklass olycksfall</i>	15
4.2	SKATTAD RESERV UTAN BOOTSTRAP	16
4.3	PARAMETERN f	16
4.4	RESIDUALER	17
4.5	NEGATIVA VÄRDEN I PSEUDOTRIANGLARNA	17
4.6	SIMULERINGSTEGET	18
4.6.1	<i>Fördelningarnas utseende</i>	19
4.7	RESULTAT EFTER 1000 ITERATIONER	21
4.7.1	<i>Försäkringsklass civil</i>	22
4.7.2	<i>Försäkringsklass trafik</i>	24
4.7.3	<i>Försäkringsklass olycksfall</i>	26
5	SLUTSATSER	29
6	APPENDIX	31
6.1	PARAMETERBYTE I LOG-LINJÄRA MODELLEN	31
6.2	PARAMETERSKATTNING OCH ”HAT”-MATRISEN	31
6.3	FÖRDELNINGAR	32
6.3.1	<i>Överspridd Poisson, OdP</i>	32
6.3.2	<i>Sammansatt Poisson</i>	32
6.3.3	<i>Gamma</i>	33
	LITTERATUR	34

1 Bakgrund

Inom sakförsäkring händer det att det går lång tid från att en skada inträffar tills slutgiltig ersättning utgår från försäkringsbolaget. Det kan hända att det går en lång tid från skadetillfället tills skadan anmäls till bolaget eller att det tar lång tid från att skadan anmäls till bolaget tills det avgörs vilka belopp som skall utbetalas. I slutet av en försäkringsperiod (vanligen ett år eller ett kvartal) skall det i bolagets bokslut framgå hur mycket medel som avsätts för dessa inträffade men icke slutgiltigt reglerade skador. Denna avsättning av kapital kallas ofta för *ersättningsreserven*. Ersättningsreserven i sin tur brukar delas upp i reserv för inträffade men ej rapporterade skador, IBNR (Incurred But Not Reported), och reserv för rapporterade men ej slutreglerade skador, IBNER (Incurred But Not Enough Reported).

Att bestämma denna ersättningsreserv är ett statistiskt problem. Data brukar vara uppställda i en triangel kallad *utvecklingstriangel* där raderna representerar *skadeår* och kolumnerna *utvecklingsår*. Elementen i utvecklingstriangeln representerar utbetalningar gjorda under utvecklingsår j för skador inträffade under skadeår i och betecknas y_{ij} med motsvarande stokastiska variabel Y_{ij} , $i, j = 1, 2, \dots, m$. Utbetalningar gjorda under utvecklingsår j är gjorda $j - 1$ år efter det att skadan inträffade. Vanligen antar man att alla skador är slutreglerade efter m år. För (i, j) med $i + j \leq m + 1$ är y_{ij} kända och den statistiska uppgiften är att utvidga triangeln till en kvadrat innehållande prediktioner om framtida utbetalningar. Triangeln som utgör framtida utbetalningar kallas här för *framtidstriangeln*.

skadeår	utvecklingsår					
	1	2	3	...	m-1	m
1	y_{11}	y_{12}	y_{13}	...	$y_{1,m-1}$	$y_{1,m}$
2	y_{21}	y_{22}	y_{23}	...	$y_{2,m-1}$	
...		
m-1	$y_{m-1,1}$	$y_{m-1,2}$				
m	y_{m1}					

Figur A: Utvecklingstriangel

skadeår	utvecklingsår					
	1	2	3	...	m-1	m
1	y_{11}	y_{12}	y_{13}	...	$y_{1,m-1}$	$y_{1,m}$
2	y_{21}	y_{22}	y_{23}	...	$y_{2,m-1}$	$Y_{2,m}$
...
m-1	$y_{m-1,1}$	$y_{m-1,2}$	$Y_{m-1,3}$...	$Y_{m-1,m-1}$	$Y_{m-1,m}$
m	y_{m1}	Y_{m2}	Y_{m3}	...	$Y_{m,m-1}$	$Y_{m,m}$

Figur B: Utvecklingstriangel och Framtidstriangel

Den totala ersättningsreserven, R , består av summan av Y_{ij} i framtidstriangeln och skattas med $\hat{R} = \sum_{i+j>m+1} \hat{y}_{ij}$. Data är här alltså på inkrementell form så att utbetalt under år j verkligen betyder utbetalt *under* år j och inte *fram till och med* år j .

Den kanske vanligaste metoden för att bestämma ersättningsreserven är Chain Ladder-metoden. Metoden är en deterministisk metod som ger en punktskattning för reserven. På senare år har flera metoder som förutom punktskattning även ger prediktionsfel och i vissa fall fördelning för totala reserven kommit fram. I England & Verrall (2002) görs en sammanställning av en rad sådana så kallade *stokastiska reservsättningsmetoder*. England & Verrall (1999) föreslår en överspridd Poisson GLM-modell som med bootstrap-simulering ger en simulerad fördelning med percentiler för totala reserven. Pinheiro, Andrade e Silva & Centeno (2002) utvidgar metoden från England & Verrall (1999) med att även undersöka en Gamma GLM-modell samt med en korrigering av residualerna som används vid bootstrap-simuleringen. Mack (1993) föreslår en fördelningsfri metod för att beräkna prediktionsfelet hos Chain Ladder-skattningarna.

1.1 Uppgift

I detta arbete söker vi osäkerheten i den totala reserven genom dess fördelning. Arbetet är tänkt att kunna användas till att möta kommande EU-lagstiftning för försäkringsbolag. Lagstiftningen förväntas ställa kapitalkrav som bland annat kommer att bero på osäkerheten i reserven. De undersökta modellerna är tre generaliserade linjära modeller, GLM, som genom bootstrap-simulering ger en simulerad fördelning för totala ersättningsreserven. Modellerna har logaritmisk länkfunktion med den linjära strukturen $\log(\mathbf{m}_{ij}) = c + \mathbf{a}_i + \mathbf{b}_j$ och variansfunktionen $v(\mathbf{m}_{ij}) = \mathbf{m}_{ij}^p$ där p i detta arbete är 1, 1.5 eller 2.

Modeller med denna variansfunktion kallas för Tweedie-modeller. Modellerna svarar mot fördelningarna (överspridd) Poisson (då $p = 1$), sammansatt Poisson ($p = 1.5$) resp. Gamma ($p = 1.5$) varför modellerna i detta arbete genomgående kommer att kallas för OdP, SaPo och Gamma även om det egentligen bara är strukturen på variansfunktionen som är det viktiga i modellenanpassningen. Vilken aktuell fördelning som variansfunktionen svarar mot har betydelse endast i det sista steget, simuleringssteget, där det simuleras från dessa fördelningar i framtidstriangeln.

2 Metoder

Genom att ansätta en generaliserad linjär modell kan enskilda prediktioner göras för framtidstriangelns samtliga celler. Det är emellertid komplicerat eller ibland omöjligt att analytiskt beräkna prediktionsfelet hos skattningen för totala reserven som är summan av framtidstriangelns element. Även i de fall det går att beräkna prediktionsfelet för summan så är dessa beräkningar bara approximativa (Pinheiro, Andrade e Silva & Centeno (2002)). Det är därför bootstrap-simulering används i detta arbete för beräkning av prediktionsfelet. Förutom prediktionsfelet ger simuleringen även en empirisk fördelning ur vilken exempelvis percentiler kan utläsas.

2.1 Antaganden

Vi gör följande antaganden.

(A1) $\mathbf{m}_{ij} \equiv E[Y_{ij}] = \mathbf{g}_i \mathbf{d}_j$, där $\sum_{j=1}^m \mathbf{d}_j = 1$.

(A2) $Y_{ij}, Y_{i'j'}$, oberoende då $(i, j) \neq (i', j')$.

(A3) Fördelningen för Y_{ij} tillhör familjen EDM och har variansfunktion $v(\mathbf{m}) = \mathbf{m}^p$ där $p \in \{1, 1.5, 2\}$.

Exponentiella dispersionsmodeller, EDM, är familjen av fördelningar vars frekvensfunktion

kan skrivas på formen $f_Y(y) = \exp\left\{\frac{y\mathbf{q} - b(\mathbf{q})}{\mathbf{f}/w} + c(y; \mathbf{f}, w)\right\}$, se exempelvis Ohlsson &

Johansson (2003) för EDM:s användning inom GLM.

Antagande (A1) innebär att de utbetalda skadebeloppen modelleras multiplikativt där parametern \mathbf{g}_i tolkas som en nivåparameter, eller mer specifikt förväntade totala skadekostnaden, för skadeår i . Parametern \mathbf{d}_j tolkas som förväntade proportionen skadekostnad som uppstår under utvecklingsår j . Antagande (A2) innebär helt enkelt att utbetalningarna i olika celler antas vara oberoende och antagande (A3) att de har en varians proportionell mot väntevärdet upphöjt till p . Logaritmering av (A1) i bägge leden ger (A1') $\log(\mathbf{m}_{ij}) = \log(\mathbf{g}_i) + \log(\mathbf{d}_j)$, där $\log(\sum_{j=1}^m \mathbf{d}_j) = 0$, som genom ett lämpligt parameterbyte (se appendix för detaljer) och reducering av en parameter kan skrivas som (A1'') $\log(\mathbf{m}_{ij}) = c + \mathbf{a}_i + \mathbf{b}_j$, där $\mathbf{a}_1 = \mathbf{b}_1 = 0$.

Antagandena (A1), (A2) och (A3) definerar en generaliserad linjär modell, GLM, se McCullagh & Nelder (1989).

2.2 Bootstrap-simulering

Bootstrap-simulering som en datorbaserad statistisk metod introducerades av Efron 1979. Metoden ersätter analytiska beräkningar av fördelning och prediktionsfel hos parametrar och statistikor med en simuleringsalgoritm. Simuleringsalgoritmen går ut på att ur ett givet stickprov återskapa många nya stickprov genom slumpmässigt urval med återläggning. En introduktion till bootstrap-metoden finns i Efron & Tibshirani (1993).

Det finns olika sätt att gå till väga i bootstrap-algoritmen. Här används den typ som kallas för residual-bootstrap (Efron & Tibshirani, 1993) som innebär att man skapar nya pseudostickprov genom urval med återläggning bland residualerna i stället för urval bland själva observationerna. För detta krävs en residualdefinition som gör att residualerna kan betraktas som oberoende och lika fördelade. Här används den standardiserade Pearsonresidualen

$$(1) r_{ij}^{st} = \frac{r_{ij,pearson}}{\sqrt{\mathbf{f}(1-h_{ij})}}$$

där h_{ij} är diagonalelementet i "hat"-matrisen för observation y_{ij} (se appendix för definition

av "hat"-matrisen) och $r_{ij,pearson} = \frac{y_{ij} - \hat{\mathbf{m}}_{ij}}{\sqrt{v(\hat{\mathbf{m}}_{ij})}}$ är den "ostandardiserade" Pearsonresidualen.

Spridningsparametern \mathbf{f} skattas med Pearsons skattningen $\hat{\mathbf{f}} = \frac{\sum_{i+j \leq m+1} r_{ij,pearson}^2}{df}$ där

$df = n - p$ (= antal observationer - antal parametrar) är antalet frihetsgrader i modellen.

Ur (1) kan y_{ij} lösas ut vilket ger

$$(2) y_{ij}^* = r_{ij} \sqrt{\hat{\mathbf{f}}v(\hat{\mathbf{m}}_{ij})(1-h_{ij})} + \hat{\mathbf{m}}_{ij}.$$

Bootsrap-simuleringen börjar med att vi ur vårt stickprov \bar{y} (utvecklingstriangeln) först skapar n residualer. Sedan skapas genom slumpmässigt urval med återläggning bland residualerna en ny *pseudotriangel* \bar{y}^* genom relationen

$$(2^*) \quad y_{ij}^* = r_{i,j} \sqrt{\mathbf{fv}(\mathbf{m}_{ij})(1-h_{ij})} + \mathbf{m}_{ij}$$

där $r_{i,j}$ innebär slumpmässigt vald residual. På denna psuedotriangel appliceras modellen, parametrar skattas och prediktioner för varje cell i framtidstriangeln görs genom $\hat{y}_{ij}^b = \exp\{\hat{c} + \hat{\mathbf{a}}_i + \hat{\mathbf{b}}_j\}$, $i + j > m + 1$. Denna procedur upprepas B gånger och index b anger aktuellt iterationsnummer. När prediktioner i framtidstriangeln gjorts simuleras i varje cell i framtidstriangeln ur den aktuella modellens fördelning med prediktionerna \hat{y}_{ij}^b som väntevärden och parametern \mathbf{f} som skattas i varje iterationssteg utgör en parameter i den aktuella fördelningen som det simuleras ur (se appendix för exakt utseende hos fördelningarna). Sedan summeras dessa simulerade värden i framtidstriangeln och vi har efter B iterationer B stycken skattade reserver. Vi får även en empirisk fördelningsfunktion för totala reserven $\hat{F}(x) = \frac{\# \text{ obs} \leq x}{B}$ ur vilken percentiler kan utläsas.

Simuleringen sker alltså i två steg. I första steget simuleras en pseudotriangel fram genom slumpmässigt urval bland residualerna. I det andra steget sker simulering ur den aktuella fördelningen i framtidstriangelns alla celler. Tanken är att fånga upp dels variationen i parameterskattningarna och dels variationen i framtiden. England & Verrall (2002, kapitel 7) uttrycker detta som "Prediction variance = Estimation variance + Process variance". En bootstrap-algoritm som efter det första steget skattar den totala reserven genom summan av framtidstriangelns celler fångar bara felet i parameterskattningarna och missar därmed variationen kring de framtida väntevärdena.

2.2.1 Bootstrap-algoritmen steg för steg

I schemat nedan är bootstrap-algoritmen illustrerad steg för steg.

Steg 1 – Startprocedur

- Ansättning av modell med skattning av parametrarna $c, \mathbf{a}_i, \mathbf{b}_j$, ($i, j = 1, 2, \dots, m$) och \mathbf{f} . Parametrarna $c, \mathbf{a}_i, \mathbf{b}_j$ skattas med quasi-likelihood (se appendix) och \mathbf{f} med pearsonskattningen.
- Beräkning av anpassade värden \mathbf{m}_{ij} i utvecklingstriangeln ($i + j \leq m + 1$).
- Beräkning av de standardiserade residualerna enligt (1): $r_{ij}^{st} = \frac{y_{ij} - \mathbf{m}_{ij}}{\sqrt{\mathbf{f}v(\mathbf{m}_{ij})(1 - h_{ij})}}$.

Steg 2 – Simulering, upprepas B gånger

Substeg 2.1 - estimationsfelet

- Skapande av en pseudotriangel \bar{y}^* genom relationen $y_{ij}^* = r_{i,j} \cdot \sqrt{\mathbf{f}v(\mathbf{m}_{ij})(1 - h_{ij})} + \mathbf{m}_{ij}$ för varje plats i utvecklingstriangeln där residualerna $r_{i,j}$ väljs genom slumpmässigt urval med återläggning bland residualerna från steg 1.
- Ansättning av modell och skattning av parametrarna $c, \mathbf{a}_i, \mathbf{b}_j$, ($i, j = 1, 2, \dots, m$) och \mathbf{f} .
- Prediktioner för alla celler i framtidstriangeln genom $\hat{y}_{ij}^b = \exp\{c^b + \mathbf{a}_i^b + \mathbf{b}_j^b\}$, $i + j > m + 1$.

Substeg 2.2 - processfelet

- Skapande av simulerade värden $\hat{y}_{ij}^{b, sim}$, ($i + j > m + 1$), där b anger iterationssteg och sim att den är simulerad, i alla celler i framtidstriangeln. Simuleringen sker från aktuell fördelning med prediktionerna från substeg 2.1 som väntevärden.

- Beräkning av bootstrap-simulerad total reserv $\hat{R}^b = \sum_{i+j>m+1} y_{ij}^{b,sim}$.

Steg 3 – Resultat efter B iterationer

- Skattning av standardavvikelsen för reserven med $s_R = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{R}_b - \bar{R})^2}$.
- Skattning av fördelningen för reserven med den empiriska fördelningen $\hat{F}(x) = \frac{\# obs \leq x}{B}$.
- Beräkning av percentiler hos den framsimulerade fördelningen. Percentilerna definieras här genom $I_a = [aB]$ där a är en sannolikhet och funktionen $[x]$ betyder heltalsdelen av x . Percentilen är då observation nummer I_a i det ordnade stickprovet av simulerade reserver.

3 Data

Data som används i detta arbete är utvecklingstrianglar för tre försäkringsklasser hos Trygg-Hansa, civil, trafik, resp. olycksfall. För att skydda Trygg-Hansa är beloppen omräknade till fantasienheten *a-mark*. Trianglarna avser *utbetalt skadebelopp* och beloppen är avrundade till närmaste heltal.

Med utbetalt skadebelopp menas här bokade utbetalningar. I praktiken skulle en cell kunna vara negativ eftersom det ibland förekommer negativa utbetalningar det vill säga inbetalningar till bolaget. Detta sker exempelvis då bolaget betalar ut för en trafikskada som senare tillfaller ett annat försäkringsbolag varvid bolaget får tillbaks det utlagda beloppet. De negativa utbetalningarna är små i förhållande till de positiva varför negativa cellvärden uppstår mycket sällan och i förekommande fall hamnar bland de allra sista utvecklingsåren då utbetalningarna går mot noll. Aktuella data från Trygg-Hansa innehåller inga negativa värden.

Att det i praktiken skulle kunna hända att cellvärden är negativa är dock ett problem som måste övervägas noga. Skall modellerna tillåta negativa värden eller skall de bygga på att alla data är positiva och betrakta negativa värden, om de förekommer, som fel i data. I det senare fallet måste data justeras. Även om aktuella data från Trygg-Hansa enbart innehåller positiva värden blir detta problem aktuellt vid bootstrap-simuleringen då vissa pseudotrianglar innehåller negativa värden. Detta diskuteras vidare i avsnitt 4.5.

3.1 Utvecklingstrianglar

I datatriangel 1, 2 och 3 återfinns de tre utvecklingstrianglarna på inkrementell form. I cell (i,j) står utbetalt skadebelopp under utvecklingsår j för skadeår i .

Skade- år	Utvecklingsår											
	1	2	3	4	5	6	7	8	9	10	11	12
1991	455 026	239 004	37 780	12 100	12 725	11 748	3 800	2 709	656	1 204	3 535	2 758
1992	488 889	205 687	28 558	12 017	7 684	5 991	2 074	2 599	1 152	2 074	2 758	
1993	445 587	231 555	24 193	15 640	7 266	3 907	2 595	1 896	874	1 131		
1994	460 090	247 022	37 285	17 180	7 765	4 474	3 134	1 321	1 906			
1995	462 317	274 102	38 792	13 790	7 603	3 358	3 521	2 081				
1996	488 402	281 925	37 572	13 387	8 097	7 332	2 699					
1997	485 153	299 578	34 104	23 048	13 654	9 833						
1998	469 737	288 414	49 503	23 893	14 079							
1999	476 509	352 804	64 066	44 921								
2000	463 609	345 448	57 855									
2001	516 484	363 033										
2002	562 729											

Datatriangel 1: Utvecklingstriangel. Försäkringsklass civil. Enhet a-mark.

Skade- år	Utvecklingsår																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1985	156 063 230	111 838 130	21 142 392	12 241 537	8 410 511	14 679 382	9 828 106	14 210 733	15 454 940	15 861 540	6 923 365	21 738 979	10 329 201	13 453 250	10 437 657	9 579 207	17 715 929	5 419 239
1986	155 091 032	135 597 624	25 443 576	16 113 355	15 978 471	13 240 829	10 764 950	21 351 958	9 717 327	16 835 904	13 370 156	13 424 302	8 174 112	35 831 829	14 544 305	14 747 320	9 307 465	
1987	181 016 649	161 159 105	25 273 050	17 782 484	20 900 299	13 076 994	15 410 388	12 781 235	14 054 606	5 689 056	16 189 167	21 871 643	17 704 383	16 474 067	22 629 931	15 672 321		
1988	208 670 773	167 892 070	43 262 051	37 644 157	27 809 555	20 609 914	31 759 751	12 914 918	16 516 403	18 318 403	29 540 209	20 654 633	21 835 746	15 186 123	11 166 040			
1989	198 311 234	191 525 760	44 706 702	22 290 021	29 874 341	19 392 713	26 716 770	30 082 336	12 283 812	16 207 258	22 960 477	22 639 840	33 776 729	25 984 186				
1990	239 683 619	183 683 603	42 958 688	28 174 342	19 622 973	18 554 163	22 382 261	24 037 310	23 251 992	15 463 269	24 679 000	38 863 067	30 711 312					
1991	245 737 117	144 852 785	39 096 659	24 621 365	16 311 842	20 619 275	18 706 822	16 892 812	25 504 833	22 146 539	25 416 660	38 070 314						
1992	221 151 095	127 289 647	41 218 430	24 452 687	20 047 288	28 755 435	13 626 921	11 948 967	16 482 819	29 227 113	17 094 804							
1993	210 127 694	119 483 395	39 878 377	33 570 157	34 303 372	17 284 086	27 503 639	30 103 081	41 878 299	40 685 663								
1994	203 832 530	127 723 011	38 273 522	36 300 830	32 595 444	22 510 828	27 775 066	42 683 098	61 006 861									
1995	207 437 891	122 631 681	42 122 386	33 023 356	21 846 173	23 019 249	39 778 304	33 097 219										
1996	164 459 900	100 935 488	33 138 889	24 374 875	16 804 580	36 050 902	29 038 330											
1997	159 458 110	113 251 953	40 484 435	51 039 129	25 768 939	34 837 589												
1998	192 043 580	146 773 594	43 034 962	40 961 806	37 098 112													
1999	196 055 568	143 089 868	51 700 229	40 568 867														
2000	163 718 738	141 807 096	51 426 099															
2001	163 451 788	134 411 369																
2002	176 723 239																	

Datatriangel 2: Utvecklingstriangel. Försäkringsklass trafik. Enhet a-mark.

Skade- år	Utvecklingsår																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1986	23 630	25 020	6 950	11 472	12 519	5 713	4 244	5 873	4 158	3 472	6 028	3 165	2 119	1 773	2 475	2 930	292
1987	23 630	25 020	7 612	15 324	12 655	13 934	8 883	4 509	6 445	2 621	3 865	2 208	6 604	1 503	538	417	
1988	23 630	26 069	20 878	21 379	14 543	11 011	9 130	7 224	10 093	6 969	6 297	3 199	1 895	4 920	3 559		
1989	24 445	29 498	32 622	27 685	18 618	15 346	13 207	9 334	7 103	6 434	3 396	5 749	2 465	4 407			
1990	18 988	44 225	45 456	31 750	21 790	14 775	13 874	8 599	10 939	6 009	6 368	5 893	6 662				
1991	27 545	47 345	53 786	35 906	29 256	14 535	11 553	17 118	7 056	11 031	8 664	8 961					
1992	31 027	58 727	60 783	50 122	30 302	24 365	15 197	13 735	11 555	11 754	10 900						
1993	44 716	64 890	69 719	41 004	30 822	18 476	17 792	18 297	15 615	7 898							
1994	62 889	66 339	64 875	52 026	36 149	21 557	19 539	14 883	9 615								
1995	38 873	55 160	77 094	44 559	29 805	21 239	15 704	10 599									
1996	40 998	54 452	63 085	45 679	32 616	23 020	28 053										
1997	39 474	49 957	53 288	52 176	36 031	25 630											
1998	38 811	54 034	64 492	53 691	44 755												
1999	45 089	61 824	73 528	54 171													
2000	44 485	55 397	69 533														
2001	40 079	49 029															
2002	56 782																

Datatriangel 3: Utvecklingstriangel. Försäkringsklass olycksfall. Enhet a-mark.

De tre trianglarna har lite olika egenskaper. Trafiktriangeln består av betydligt större värden än de två andra och betraktas som ej avslutad med vilket menas att utbetalningar för de aktuella skadeåren förväntas fortsätta bortom det sista utvecklingsåret i triangeln. Trianglarna har också olika dimension där civil-triangeln har den minsta dimensionen.

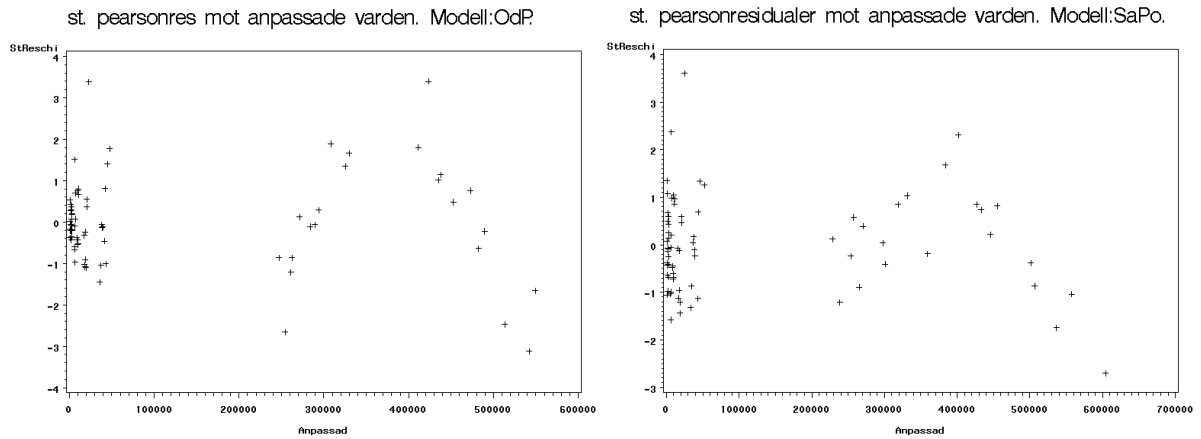
4 Resultat av tillämpning på data från Trygg-Hansa

Anpassning av modellerna och simulering sker med hjälp av programvaran SAS. Först undersöks modellernas anpassning till data där residualplottar för de olika modellerna redovisas och diskuteras. Sedan utförs själva bootstrap-simuleringen för de olika metoderna och resultaten redovisas.

4.1 Modellernas anpassning till data

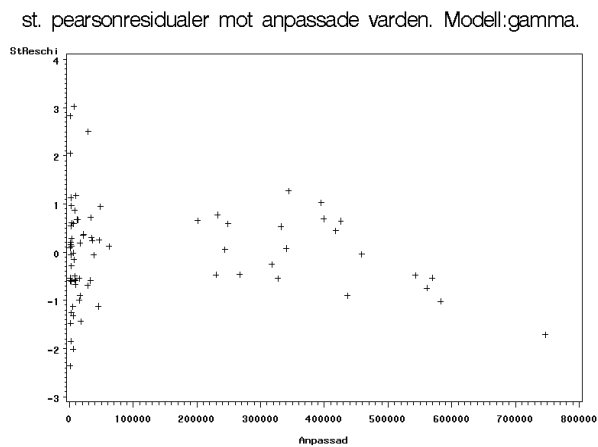
För att studera modellernas anpassning till data plottades de standardiserade pearsonresidualerna r_{ij}^{st} , enligt (1), mot sina anpassade värden \hat{m}_j . Enligt modellen kan dessa betraktas som oberoende och lika fördelade med väntevärde 0 och varians 1. För ett tillfredsställande resultat skall inga mönster finnas bland dem utan det skall se ut som ett jämt brus kring 0-axeln. Avvikelse från detta kan tyda på att modellen är felaktig på något sätt, se exempelvis McCullagh & Nelder (1989) kap. 12. På följande sidor är de standardiserade pearsonresidualerna plottade mot \hat{m}_j för alla modeller i alla försäkringsklasser.

4.1.1 Försäkringsklass civil



Figur 1 (till vänster): Standardiserade pearsonresidualer plottade mot \hat{m} . Modell OdP. Försäkringsklass civil.

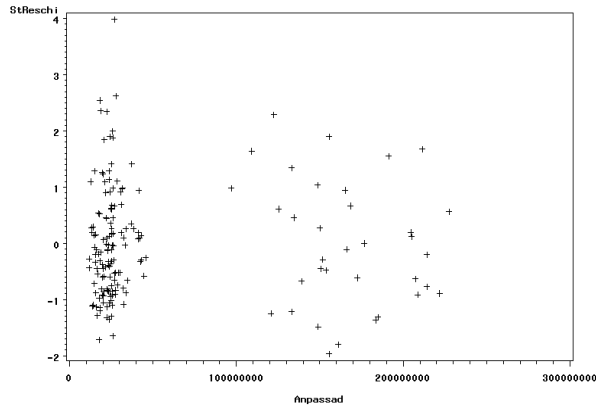
Figur 2 (till höger): Standardiserade pearsonresidualer plottade mot \hat{m} . Modell SaPo. Försäkringsklass civil.



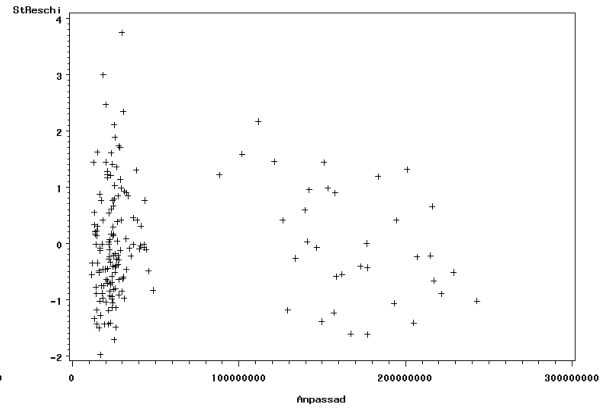
Figur 3: Standardiserade pearsonresidualer plottade mot \hat{m} . Modell Gamma. Försäkringsklass civil.

4.1.2 Försäkringsklass trafik

st. pearsonres mot anpassade varden. Modell:OdP.



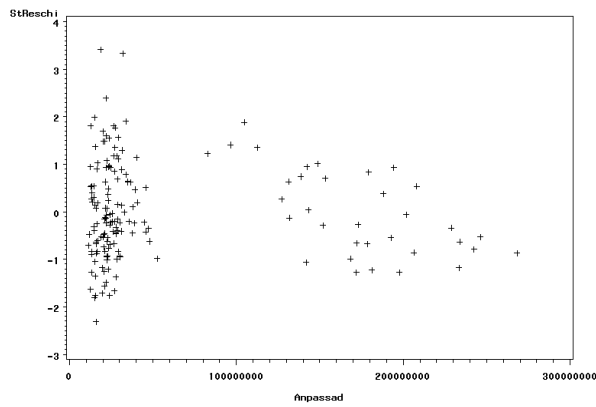
st. pearsonresidualer mot anpassade varden. Modell:SaPo.



Figur 4 (till vänster): Standardiserade pearsonresidualer plottade mot \hat{m} . Modell OdP. Försäkringsklass trafik.

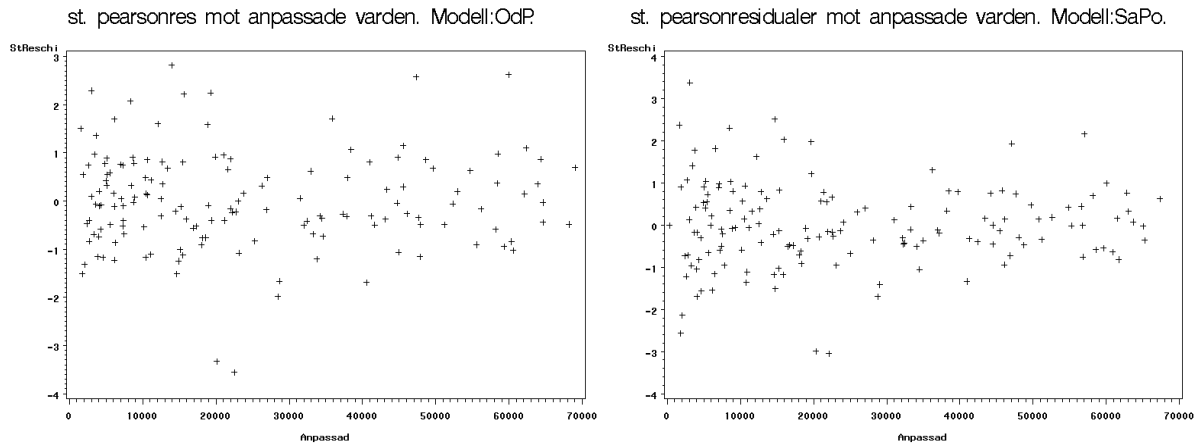
Figur 5 (till höger): Standardiserade pearsonresidualer plottade mot \hat{m} . Modell SaPo. Försäkringsklass trafik.

st. pearsonresidualer mot anpassade varden. Modell:gamma.



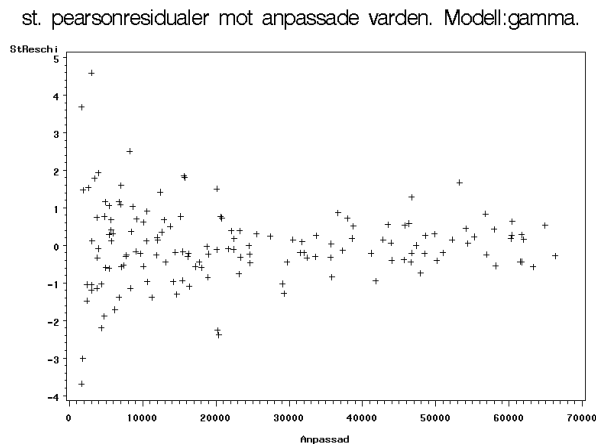
Figur 6: Standardiserade pearsonresidualer plottade mot \hat{m} . Modell Gamma. Försäkringsklass trafik.

4.1.3 Försäkringsklass olycksfall



Figur 7 (till vänster): Standardiserade pearsonresidualer plottade mot \hat{m} . Modell OdP. Försäkringsklass olycksfall.

Figur 8 (till höger): Standardiserade pearsonresidualer plottade mot \hat{m} . Modell SaPo. Försäkringsklass olycksfall.



Figur 9: Standardiserade pearsonresidualer plottade mot \hat{m} . Modell Gamma. Försäkringsklass olycksfall.

Vad det gäller försäkringsklass civil är det svårt att se något direkt mönster bland observationerna. Möjligen kan en parabel på formen $-x^2$ skönjas för OdP- och SaPo-modellen för stora \hat{m} där vi har få observationer. Tendensen är störst för OdP-modellen. Spännvidden för residualerna verkar öka med \hat{m} för OdP samtidigt som den ser ganska konstant ut för SaPo medan den för Gamma möjligen verkar sjunka.

På försäkringsklass trafik har vi en tät svärm med residualer för låga \hat{m} och en lite glesare svärm längre bort på \hat{m} -axeln. Residualerna är kraftigt konformade för Gamma-modellen. SaPo-modellens residualer ser ut att följa en linje på formen $-kx$ i den bortre glesare svärmen. OdP-residualerna ser bra ut.

På försäkringsklass olycksfall har vi ett jämnt residualspektrum för OdP-modellen varefter kon-utseendet ökar med ökande p i variansfunktionen.

För Gamma-modellen ser residualerna ut att avta med ökande \mathbf{m} för alla tre försäkringsklasser vilket kan tyda på en för hög potens i variansfunktionen. Tydligast är detta i olycksfall. Det som mest ser ut som ett ”skolboksexempel” av oberoende likafördelade observationer är OdP-residualerna för olycksfall. Där får man hjälp av att residualerna är jämnt fördelade med avseende på \mathbf{m} . Modellerna OdP och SaPo ser hyfsade ut i alla tre fallen och möjligen ser OdP-modellen bättre ut än SaPo. Kanske skulle $p = 1.25$ eller liknande ge bäst anpassning.

4.2 Skattad reserv utan bootstrap

Metoderna ger lite olika skattningar av den totala reserven och resultaten skiljer sig för de olika försäkringsklasserna. I tabell 10 ges skattningarna av den totala reserven för alla metoder. I tabell 11 uttrycks skattningarna som procent av OdP-skattningen (som alltid är samma som Chain Ladder-skattningen). Vi ser att i försäkringsklass civil ger metoderna i stort sett samma skattning av reserven, SaPo-skattningen är aningen större än OdP och Gamma-skattningen ännu lite större, dock bara cirka 1.7% större än OdP. I försäkringsklass trafik är skillnaderna större där SaPo ger en 8.6% större skattning än OdP och Gamma 18.2% större än OdP. I försäkringsklass civil är skillnaderna annorlunda. Där ger SaPo och Gamma aningen lägre skattningar än OdP, SaPo knappt 1% och Gamma drygt 2% lägre.

	OdP	SaPo	Gamma
Civil	726 869	731 649	739 210
Trafik	3 146 215 376	3 416 936 388	3 717 943 160
Olycksfall	1 664 686	1 652 360	1 627 902

Tabell 10: Skattade reserver för alla metoder.

	OdP	SaPo	Gamma
Civil	100	100.7	101.7
Trafik	100	108.6	118.2
Olycksfall	100	99.3	97.8

Tabell 11: Skattade reserver för alla metoder som procent av OdP.

Den stora skillnaden på försäkringsklass trafik är anmärkningsvärd. Skattningen av totala reservbehovet är nästan 20% större med Gamma än med OdP. Annars är det svårt att se några mönster, i två fall växer reservbehovet med växande p och i ett fall sjunker den.

4.3 Parametern f

Vid ansättning av modellerna skattades parametrarna c , \mathbf{a}_i , \mathbf{b}_j genom maximering av quasi-likelihooden (se appendix). Parametern f skattades genom pearsonskattningen

$$\hat{f} = \sum_{i+j \leq m+1} \frac{r_{ij,pearson}^2}{n-p} = \frac{1}{n-p} \sum_{i+j \leq m+1} \frac{(y_{ij} - \mathbf{m}_{ij})^2}{v(\mathbf{m}_{ij})}$$

där variansfunktionen $v(\mathbf{m})$ ju är

$v(\mathbf{m}) = \mathbf{m}$, $v(\mathbf{m}) = \mathbf{m}^{1.5}$ resp. $v(\mathbf{m}) = \mathbf{m}^2$ för de tre modellerna OdP, SaPo och Gamma.

Skattningen av \mathbf{f} för de tre modellerna i steg 1 i bootstrap-algoritmen finns i tabell 12 där vi ser mycket varierande värden. Skillnaderna beror på att variansfunktionen är nämnare i pearsonresidualen och ger därför mindre pearsonresidualer med ökande p . Följaktligen blir även $\hat{\mathbf{f}}$ mindre.

	OdP	SaPo	Gamma
Civil	2 156	9.58	0.066
Olycksfall	969	8.16	0.088
Trafik	3 174 359	534	0.093

Tabell 12: Skattade \mathbf{f} :n i steg 1.

4.4 Residualer

Vid skapandet av de standardiserade Pearsonresidualerna i steg 1 i bootstrap-algoritmen blir residualerna r_{1m}^{st} i nordöstra hörnet och r_{m1}^{st} i sydvästra hörnet lika med noll eftersom $\mathbf{m}_{1m} = y_{1m}$ och $\mathbf{m}_{m1} = y_{m1}$ per definition. Dessa residualer betraktas inte som observationer från den underliggande stokastiska variabeln och tas därför inte med bland residualerna ur vilka urval med återläggning sker. Vi tror helt enkelt inte på att vi har två observationer som passar perfekt varför vi skulle förlora variation om dessa noll-residualer togs med i simuleringen.

Statistiska data om de resulterande residualmängderna efter det inledande bootstrap-steget återfinns i Tabell 13. Medelvärde hamnade nära 0 och standardavvikelsen nära 1 överlag, som väntat.

		n-2	medelv.	st. avv.	min	max
Civil	OdP	76	-0.0091	1.099	-3.13	3.40
	SaPo	76	-0.0078	1.019	-2.70	3.61
	Gamma	76	0.0001	1.015	-2.36	3.02
Trafik	OdP	169	-0.0096	1.006	-1.96	3.99
	SaPo	169	-0.0062	1.007	-1.98	3.75
	Gamma	169	0.0001	1.011	-2.31	3.42
Olycksfall	OdP	151	0.0062	1.002	-3.56	2.82
	SaPo	151	-0.0027	1.011	-3.05	3.38
	Gamma	151	0.0000	1.047	-3.69	4.59

Tabell 13: Data om residualerna från steg 1 i bootstrap-algoritmen

4.5 Negativa värden i pseudotrianglarna

Även om aktuella data från Trygg-Hansa inte innehåller negativa värden är problemet med negativa värden i utvecklingstriangeln värt att diskutera. För några av modellerna uppstod relativt ofta negativa värden i pseudotrianglarna (se figur 14). Dessa kan uppstå om den slumpmässigt valda residualen är stor och negativ, se relation (2'). Vid en sådan situation kan parametrar skattas genom quasi-likelihood med det mindre stränga kravet att kolumnsummorna skall vara positiva i fallet OdP och Gamma. När negativa värden i

pseudotriangeln uppstår sker detta dock oftast någonstans i den nordöstra delen av triangeln varför kolumnsumman med stor sannolikhet blir negativ då detta sker. I fallet SaPo tillåts negativa värden men problem uppstår då SAS skall skatta parametrar. Den iterativa skattningsprocessen som SAS använder sig av (se appendix) konvergerar inte alltid varför en korrigering även i SaPo-fallet var nödvändig.

Här valdes den pragmatiska lösningen att inte tillåta några negativa värden över huvud taget. Vid ett negativt värde i pseudotriangeln ersattes den med ett annat värde. I fallet OdP och SaPo ersattes negativa värden i pseudotriangelarna med värdet noll. I Gammalfallet ersattes negativa värden i pseudotriangelarna med talet 100 i alla tre försäkringsklasserna. Hur England & Verrall (2001) och andra löst problemet framgår inte ur deras uppsatser.

Att ersätta negativa värden med positiva innebär att en del av den totala variationen underskattas vilket bör beaktas vid analyserna efter simuleringen. Speciellt om negativa värden uppstår ofta.

I tabell 14 nedan redovisas hur ofta ett negativt värde uppstod i pseudotriangelarna. I fallet OdP var detta ett relativt stort problem där det för försäkringsklass civil skedde 2728 gånger på olika ställen i triangeln och för försäkringsklass olycksfall 885 gånger. För modellen SaPo var det också ett problem där det i försäkringsklass civil var en residual ($r = -2.698$) som gav upphov till de negativa värdena. För Gamma-modellen uppstod inte negativa värden i civil eller trafik medan en mycket stor residual ($r = -3.687$) spökade i försäkringsklass olycksfall.

Att negativa värden inte uppstod så ofta i försäkringsklass trafik beror på att den ursprungliga triangeln har genomgående höga värden. Värdena för de sena utvecklingsåren har inte gått ner mot noll än.

	OdP	SaPo	Gamma
Civil	2728 gånger olika r olika ställen	253 gånger alltid $r = -2.697604$ utv.år 7 och framåt	inga negativa värden
Trafik	46 gånger $r \leq -1.310445$ alltid nordöstra hörnet	inga negativa värden	inga negativa värden
Olycksfall	885 gånger $r \leq -0.561069$ utv. år 6 och framåt	189 gånger $r \leq -1.504999$ utv.år 10 och framåt	357 gånger alltid $r = -3.687047$ utv. år 2 och framåt

Tabell 14: Förekomst av negativa värden i pseudotriangelarna under 1000 iterationer.

4.6 Simuleringssteget

Det är först vid simuleringen i framtidstriangeln som fördelningsantagandet gör sig gällande. Vid ansättning av modellen räckte det med specifikation av variansfunktion men här måste vi ha en fördelning att simulera ifrån. Simuleringen i framtidstriangeln med de predikterade värdena som väntevärden skedde på följande sätt. I fallet OdP; för att simulera från en överspridd Poissonfördelning med väntevärde y_{ij}^b simulerades från en poissonfördelad

variabel med väntevärde $\frac{y_{ij}^b}{\mathbf{f}^b}$ multiplicerat med \mathbf{f}^b . Indexet b anger aktuellt

iterationsnummer. För varje ny pseudotriangel skattas dispersionsparametern \mathbf{f} om på nytt. I fallet SaPo simulerades i två steg (se appendix om hur SaPo-variabeln är uppbyggd). Först

simulerades ett m fram från en poissonfördelning med väntevärde $\frac{(y_{ij}^b)^{2-p}}{\mathbf{f}^b(2-p)}$ varefter $y_{ij}^{b,sim}$

erhölls genom simulering ur en gammafördelning med parametrarna $\mathbf{a} = \frac{m(2-p)}{p-1}$ och

$\mathbf{b} = \mathbf{f}^b(p-1)(y_{ij}^b)^{p-1}$. I Gammalfallet erhölls $y_{ij}^{b,sim}$ genom simulering från en

gammafördelning med parametrarna $\mathbf{a} = \frac{1}{\mathbf{f}^b}$ och $\mathbf{b} = \mathbf{f}^b y_{ij}^b$ där gammafördelningen är sådan

att $E[Y] = \mathbf{ab}$ och $Var(Y) = \mathbf{ab}^2$.

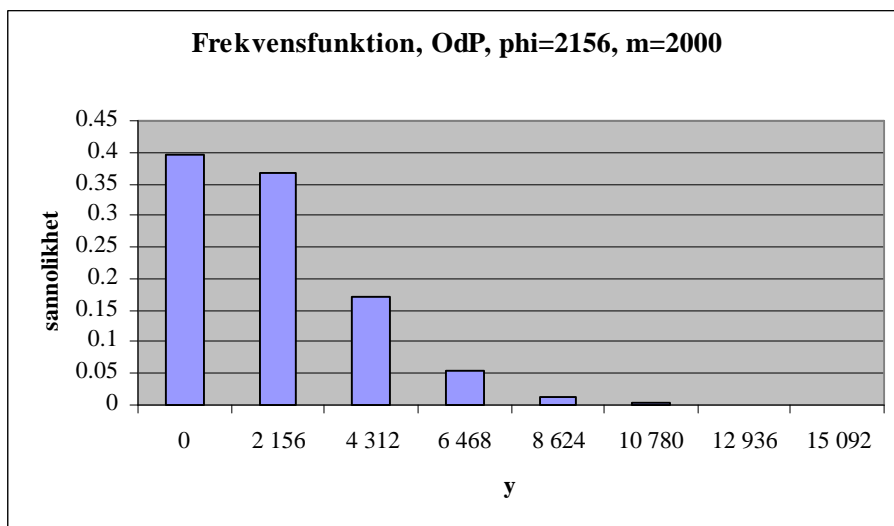
4.6.1 Fördelningarnas utseende

Modellen OdP är speciell genom att stödet för fördelningen beror på spridningsparametern \mathbf{f} . Givet parametern \mathbf{f} hamnar stödet på talen \mathbf{fn} , $n = 1, 2, 3, \dots$. Fördelningarna som det simuleras från i framtidstriangeln ser märkliga ut om \mathbf{f} är stort i förhållande till väntevärdet \mathbf{m} vilket är fallet i alla tre försäkringsklasserna. En typisk framtidstriangel i försäkringsklass civil för OdP-modellen är den som beräknas på den ursprungliga utvecklingstriangeln, se figur 15.

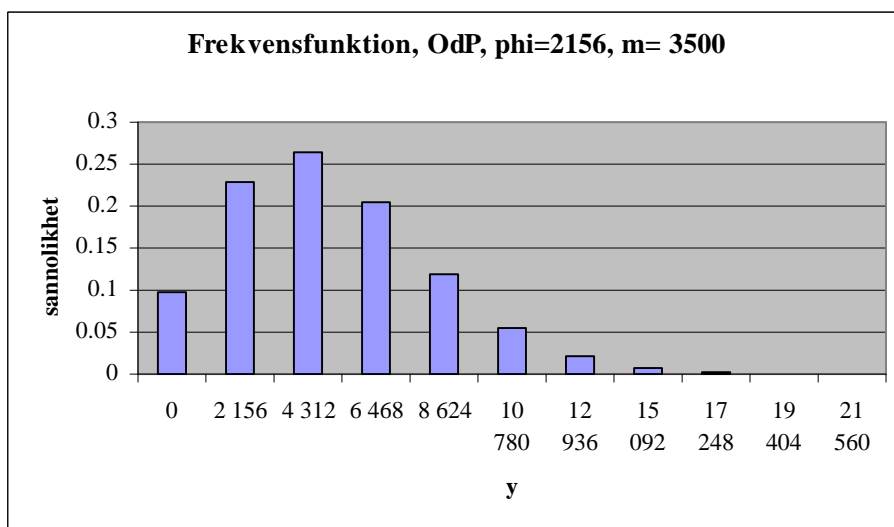
skadeår	utvecklingsår												
	1	2	3	4	5	6	7	8	9	10	11	12	
1													
2													2 684
3												3 015	2 607
4										1 520	3 208	2 774	
5									1 216	1 571	3 317	2 869	
6								2 322	1 271	1 642	3 466	2 997	
7							3 313	2 403	1 315	1 699	3 587	3 102	
8						7 205	3 265	2 368	1 296	1 675	3 535	3 057	
9					11 857	8 096	3 668	2 661	1 456	1 882	3 972	3 435	
10				21 782	11 231	7 668	3 475	2 520	1 379	1 782	3 762	3 253	
11			48 298	23 313	12 020	8 206	3 719	2 697	1 476	1 907	4 026	3 482	
12		337 798	49 452	23 869	12 307	8 402	3 807	2 762	1 511	1 953	4 122	3 565	

Figur 15: Framtidstriangel beräknad på ursprungsdata. Försäkringsklass civil, modell OdP.

Parametern \mathbf{f} skattades här till 2156 och vi ser att för utvecklingsår 9 och 10 är samtliga värden lägre än detta och att värdena i utvecklingsåren 7, 8, 11 och 12 bara är aningen större än 2156. Då värdena i framtidstriangeln, som fungerar som väntevärden i den fördelning som det simuleras ur, ligger i närheten av \mathbf{f} ser fördelningarna konstiga ut. I figur 16 och 17 finns två OdP-fördelningar plottade som illustrerar detta. Fördelningarna har stora punktmassor i ett fåtal punkter där sannolikhetsmassan är störst i $y = 0$ om \mathbf{m} är mindre än \mathbf{f} .



Figur 16: Frekvensfunktion för en OdP-fördelad s.v.



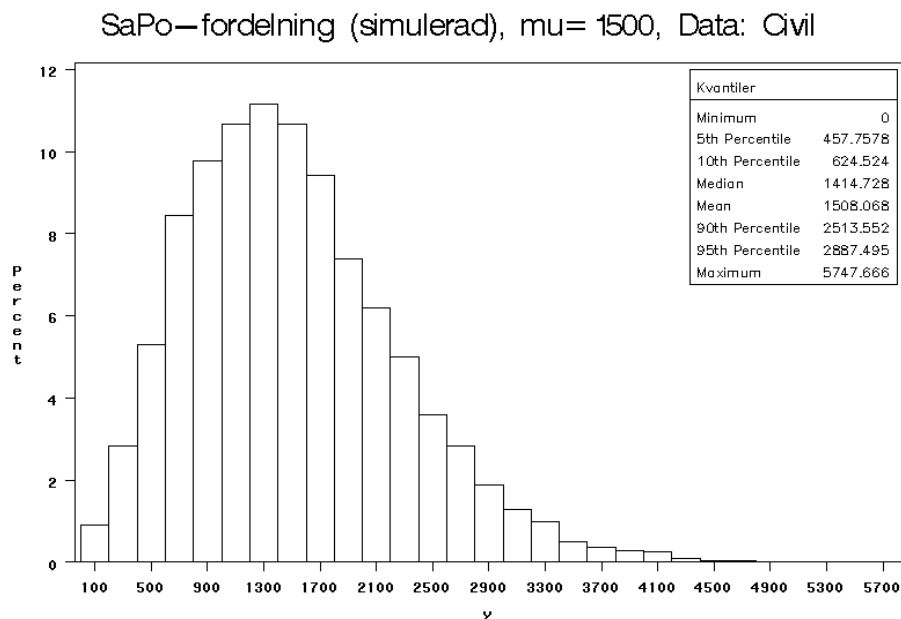
Figur 17: Frekvensfunktion för OdP-fördelad s.v.

Sannolikhetsmassan ligger på punkterna $n \cdot 2156$, $n = 0, 1, 2, \dots$ med en stor sannolikhet i nollan i figur 16 då $m = 2000$ vilket känns onaturligt då det som simuleras är framtida utbetalningar. Fördelningen borde vara kontinuerlig eller åtminstone approximativt kontinuerlig så att punkterna för stödet ligger tätt.

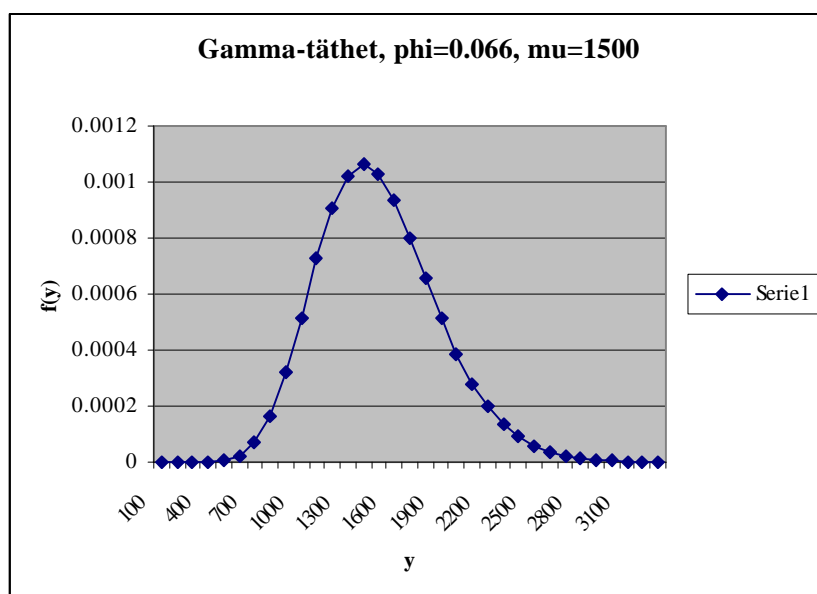
I försäkringsklass olycksfall är fenomenet densamma att f är stort i relation till framtidstriangelns värden men dock inte lika stort. I en typisk framtidstriangel beräknad på ursprungsdata där f skattades till 969 är det bara sista kolumnen i framtidstriangeln som innehåller värden mindre än det.

I fördelningarna SaPo och Gamma beror inte fördelningarnas stöd på parametern f . Där ser fördelningarna mer rimliga ut även för mindre värden i framtidstriangeln. Eftersom frekvensfunktionen för en SaPo-fördelad stokastisk variabel har ett komplicerat utseende (se appendix) återfinns i figur 18 en simulerad frekvensfunktion för en SaPo-fördelad variabel med väntevärde 1500 och spridningsparameter skattad till 9.58 då SaPo-modellen applicerades på den ursprungliga utvecklingstriangeln för försäkringsklass civil. Då Gamma-

modellen anpassades på samma utvecklingstriangel skattades f till 0.066 och en bild av frekvensfunktionen för en gammafördelad stokastisk variabel med väntevärde 1500 och spridningsparameter 0.066 finns i figur 19.



Figur 18: Simulerad frekvensfunktion för en SaPo-fördelad s.v.



Figur 19: Frekvensfunktion för Gamma-fördelad s.v.

4.7 Resultat efter 1000 iterationer

I denna undersökning valdes $B = 1000$ iterationer varvid 1000 simulerade totala reserver \hat{R}^b erhöles. I det följande undersöker vi prediktionsfelet, medelvärdet, fördelningarnas utseende

och percentiler i fördelningarna för respektive metod för de tre försäkringsklasserna. Prediktionsfelet som gäller som en skattning av estimationsfelet + processfelet för totala reserven skattades med standardavvikelsen av \hat{R}^b genom formeln

$$s = \sqrt{\frac{\sum_{b=1}^{1000} (R^b - \bar{R})^2}{1000 - 1}} \text{ där } \bar{R} = \frac{\sum_{b=1}^{1000} R^b}{1000} \text{ är medelvärdet av de simulerade skattade}$$

reserverna. Percentiler i den empiriska fördelningen erhöles genom observation nummer I_a i det ordnade stickprovet av simulerade reserver där $I_a = [a \cdot 1000]$ och a är procentnivån.

Den empiriska fördelningen erhöles genom $\hat{F}(x) = \frac{\# obs \leq x}{1000}$. Den empiriska

frekvensfunktionen för totala reserven erhöles genom att plotta ett histogram över \hat{R}^b .

4.7.1 Försäkringsklass civil

I tabell 20 ser vi percentiler hos den framsimulerade fördelningen samt medelvärdet, standardavvikelsen och variationskoefficienten, som är standardavvikelsen delat med medelvärdet, för försäkringsklass civil efter 1000 iterationer. På sista raden redovisas modellens ursprungligt predikterade totala reserv.

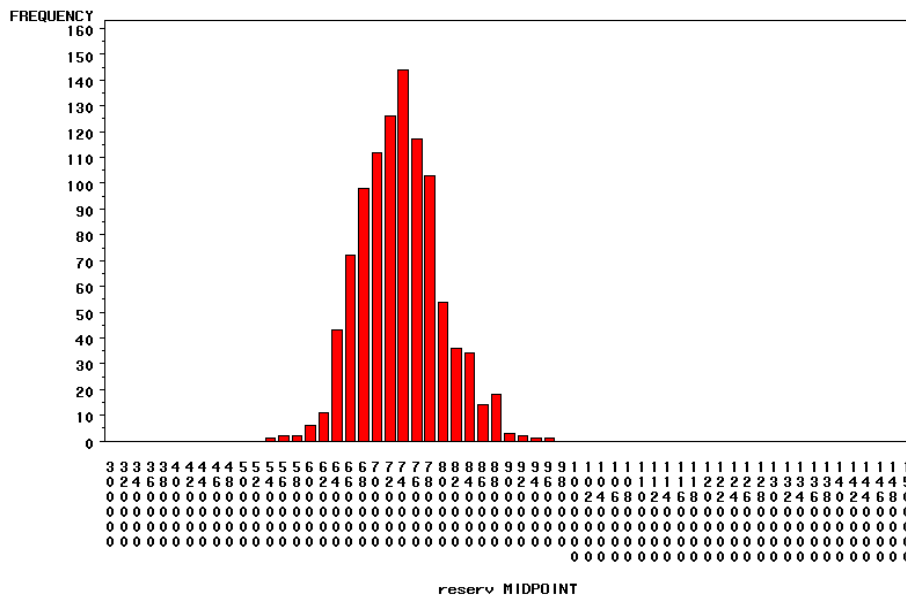
Percentiler	OdP	SaPo	Gamma
0.025	631 890	601 873	493 781
0.05	642 876	628 821	529 659
0.1	661 000	645 960	574 923
0.25	693 805	681 354	643 035
0.5	734 313	725 459	729 353
0.75	772 374	770 770	823 206
0.9	813 811	822 597	938 441
0.95	843 405	852 475	1 018 687
0.975	869 843	883 483	1 094 428
0.99	887 529	920 748	1 165 017
Medelv.	735 462	730 037	745 100
St.avv.	60 298	69 784	147 872
Std/Medelv	0.082	0.096	0.198
Modell	726 900	731 600	739 200

Tabell 20: Percentiler och standardavvikelse efter 1000 iterationer. Försäkringsklass civil.

Modellerna ger olika variation där OdP ger minst och SaPo aningen större variation. Gamma varierar mer än dubbelt så mycket som de övriga. Medelvärdena ligger ganska nära det ursprungligt skattade för alla tre modellerna vilket tyder på att bootstrap-simuleringen här inte verkar ha någon större bias vad det gäller skattning av väntevärdet. Att Gamma har en dubbelt så stor standardavvikelse som de andra är anmärkningsvärt.

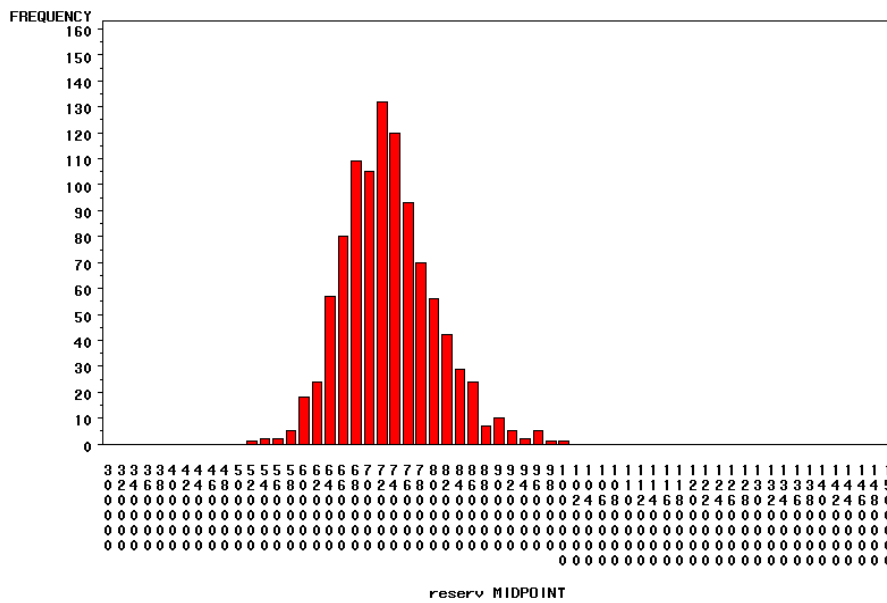
Histogram över simulerade reserver för respektive metod redovisas i figur 21, 22 och 23. Histogrammen är plottade i samma skala.

Simulerad total reserv, modell OdP, civil



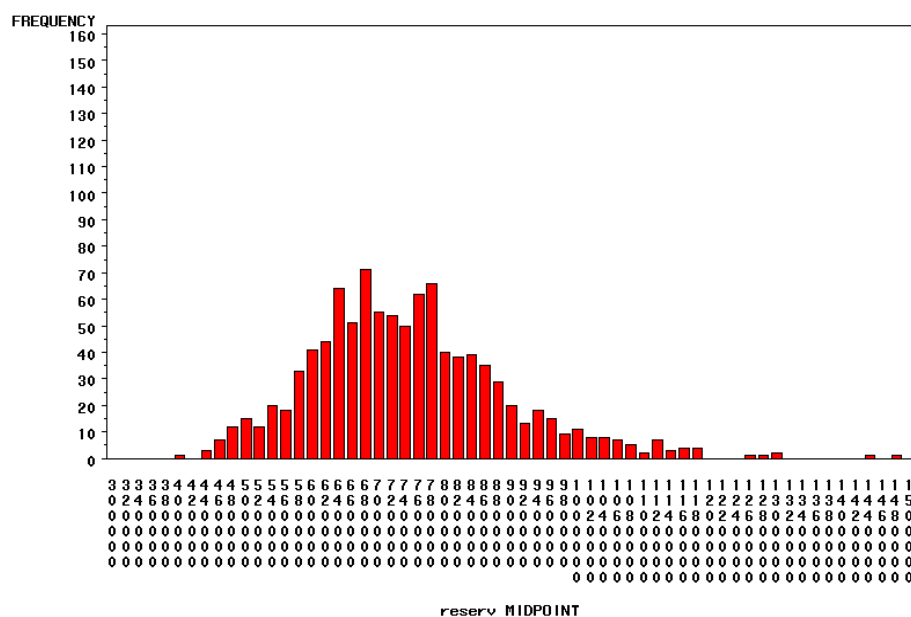
Figur 21: Histogram över simulerad totala reserver. Modell OdP, försäkringsklass civil.

Simulerad total reserv, modell SaPo, civil



Figur 22: Histogram över simulerad totala reserver. Modell SaPo, försäkringsklass civil.

Simulerad total reserv, modell Gamma, civil



Figur 23: Histogram över simulerad totala reserver. Modell Gamma, försäkringsklass civil.

4.7.2 Försäkringsklass trafik

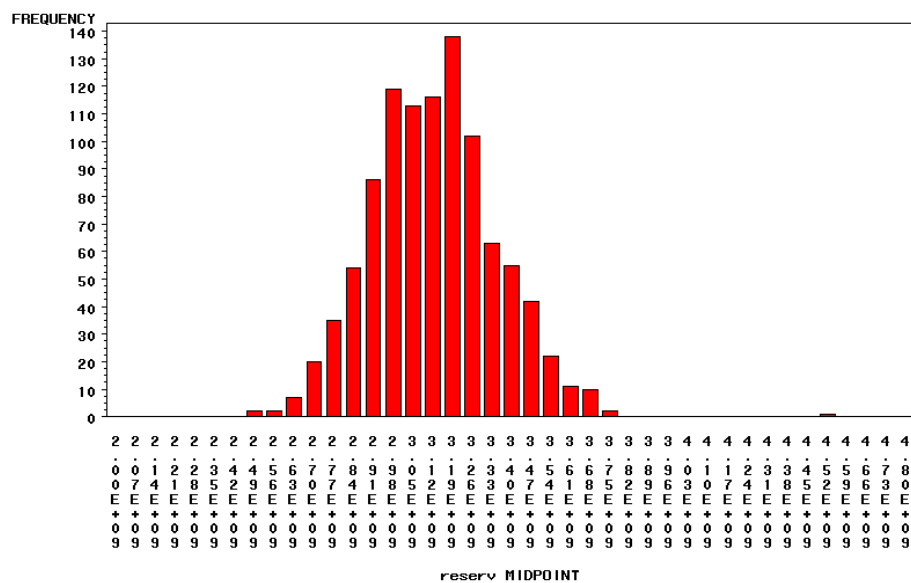
Percentiler, medelvärden och andra statistiska data efter 1000 bootstrap-iterationer på utvecklingstriangeln för försäkringsklass trafik återfinns i tabell 24. Variationen som del av väntevärdet är i samma storleksordning för alla tre modellerna, omkring 7%. För alla tre modellerna ligger det simulerade väntevärdet nära (inom $\pm 1\%$ av) det predikerade värde som modellen ursprungligen ger. Väntevärdena skiljer sig dock kraftigt åt såsom tidigare påpekats.

Percentiler	OdP	SaPo	Gamma
0.025	2 714 923 121	2 989 356 310	3 202 165 921
0.05	2 771 166 134	3 051 574 711	3 273 524 490
0.1	2 848 880 664	3 128 822 238	3 385 614 239
0.25	2 974 748 319	3 252 097 873	3 537 869 904
0.5	3 125 922 257	3 400 915 499	3 717 601 889
0.75	3 257 325 692	3 565 913 123	3 891 882 619
0.9	3 416 724 264	3 713 779 419	4 094 951 457
0.95	3 500 740 530	3 810 202 479	4 197 054 018
0.975	3 565 987 417	3 902 331 702	4 286 686 132
0.99	3 669 209 172	3 956 092 421	4 410 156 788
Medelv.	3 124 999 762	3 414 218 098	3 721 694 105
St.avv.	220 914 668	229 626 269	272 511 656
Std/Medelv	0.071	0.067	0.073
Modell	3 146 000 000	3 417 000 000	3 718 000 000

Tabell 24: Percentiler och andra data efter 1000 iterationer. Försäkringsklass trafik.

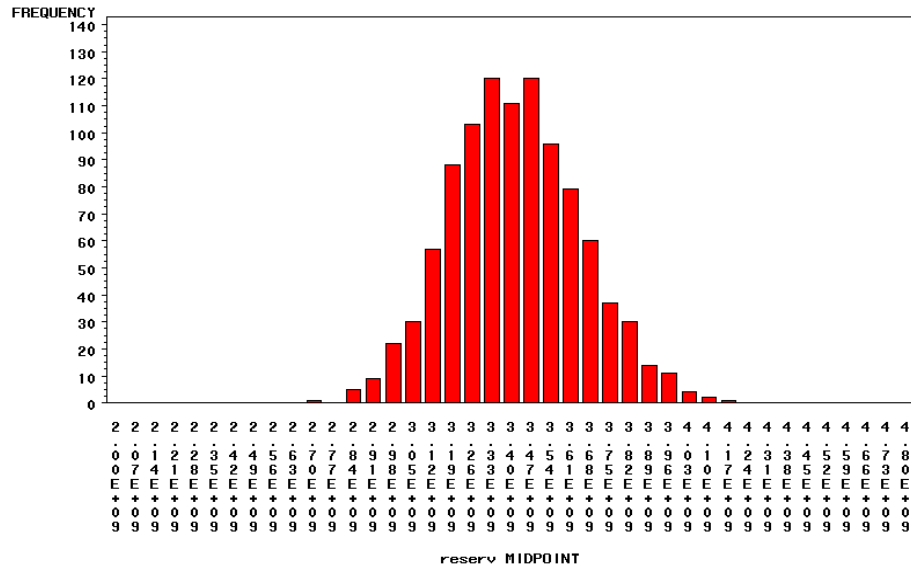
Resultatet av simuleringarna, för försäkringsklass trafik, i form av histogram återfinns i figur 25, 26 och 27.

Simulerad total reserv, modell OdP, trafik

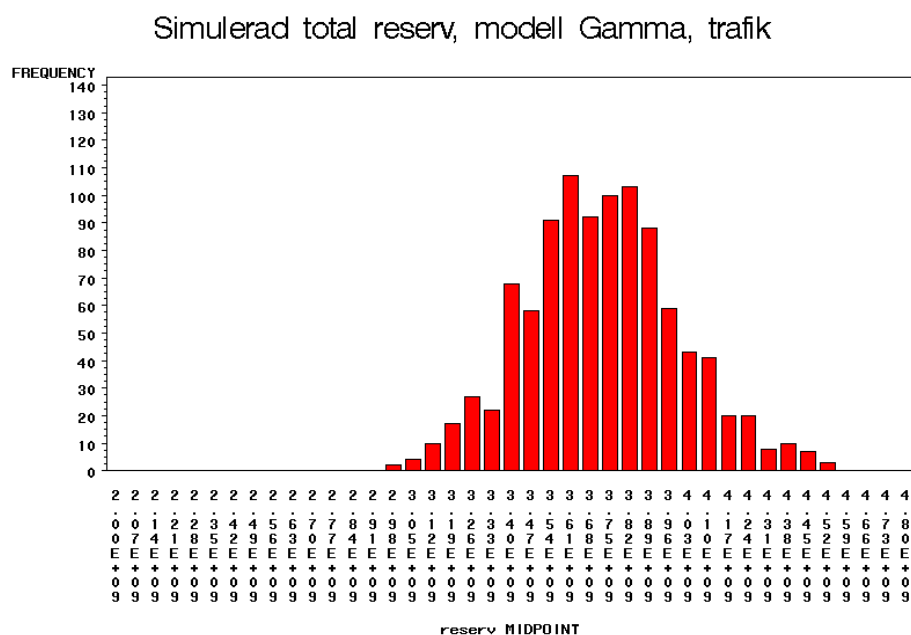


Figur 25: Histogram över simulerade totala reserver. Modell OdP, försäkringsklass trafik.

Simulerad total reserv, modell SaPo, trafik



Figur 26: Histogram över simulerade totala reserver. Modell SaPo, försäkringsklass trafik.



Figur 27: Histogram över simulerade totala reserver. Modell Gamma, försäkringsklass trafik.

4.7.3 Försäkringsklass olycksfall

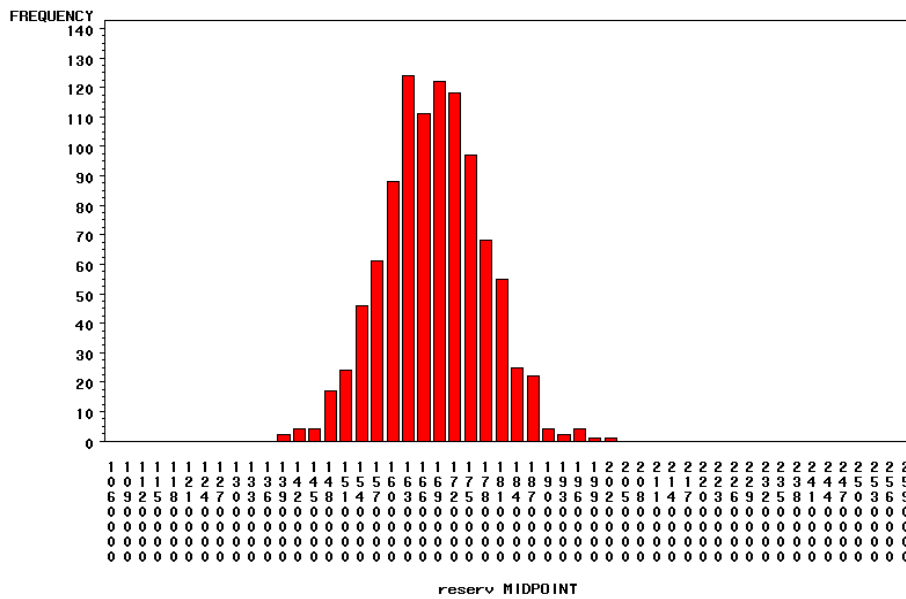
Percentiler och annat efter simuleringarna för försäkringsklass olycksfall finns i figur 28. OdP ger störst väntevärde, SaPo lite mindre och Gamma minst. Variationen är som i försäkringsklass civil klart störst för modell Gamma. OdP har minst varians och SaPo aningen större.

Percentiler	OdP	SaPo	Gamma
0.025	1 494 732	1 460 384	1 310 843
0.05	1 520 487	1 483 157	1 377 070
0.1	1 558 878	1 511 388	1 448 599
0.25	1 615 303	1 583 991	1 536 406
0.5	1 679 059	1 655 201	1 640 386
0.75	1 740 563	1 726 152	1 747 237
0.9	1 800 982	1 815 030	1 859 423
0.95	1 834 173	1 854 057	1 948 175
0.975	1 862 236	1 890 557	2 043 658
0.99	1 891 933	1 941 076	2 150 984
Medelv.	1 679 383	1 658 464	1 649 367
St.avv.	95 147	113 735	179 780
Std/Medelv	0.057	0.069	0.109
Modell	1 665 000	1 652 000	1 627 000

Figur 28: Percentiler och andra data efter 1000 iterationer. Försäkringsklass olycksfall.

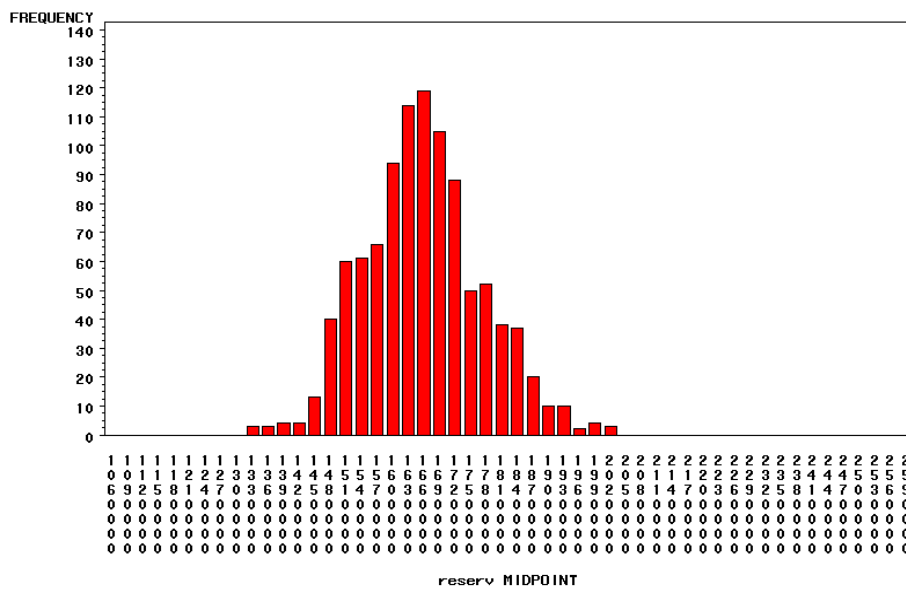
Resultatet av simuleringarna, för försäkringsklass olycksfall, i form av histogram återfinns i figur 29, 30 och 31.

Simulerad total reserv, modell OdP, olycksfall



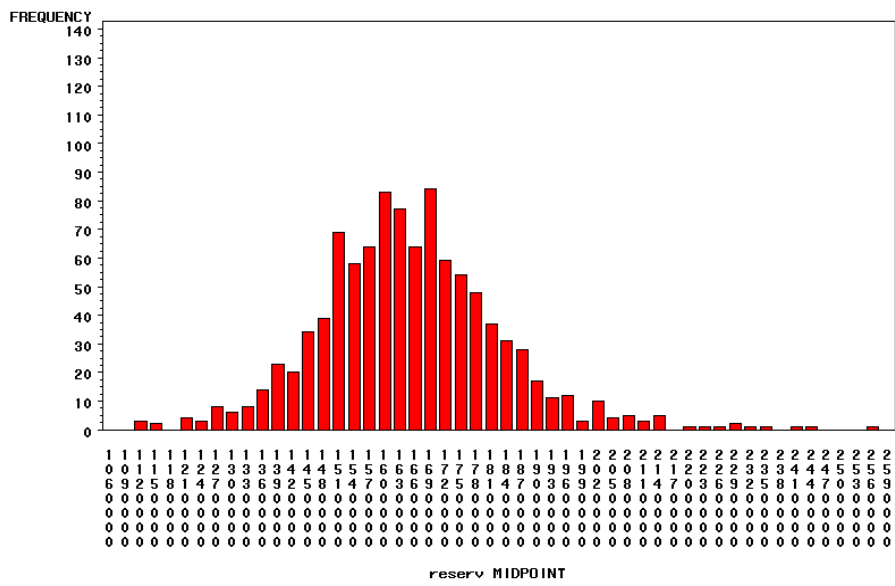
Figur 29: Histogram över simulerade totala reserver. Modell OdP, försäkringsklass olycksfall.

Simulerad total reserv, modell SaPo, olycksfall



Figur 30: Histogram över simulerade totala reserver. Modell SaPo, försäkringsklass olycksfall.

Simulerad total reserv, modell Gamma, olycksfall



Figur 31: Histogram över simulerade totala reserver. Modell Gamma, försäkringsklass olycksfall.

5 Slutsatser

I tabell 32 sammanfattas resultatet av simuleringarna. Där återfinns medelvärdet av simuleringarna samt medelvärdet delat med modellens predikterade värde och variationskoefficienten uttryckta i procent.

		OdP	SaPo	Gamma
Civil	Medelvärde	735 462	730 037	745 100
	Medelv./Modell	101.2	99.8	100.8
	Variationskoeff.	8.2	9.6	19.8
Trafik	Medelvärde	3 124 999 762	3 414 218 098	3 721 694 105
	Medelv./Modell	99.3	99.9	100.1
	Variationskoeff.	7.1	6.7	7.3
Olycksfall	Medelvärde	1 679 383	1 658 464	1 649 367
	Medelv./Modell	100.9	100.4	101.3
	Variationskoeff.	5.7	6.9	10.9

Tabell 32: Simuleringsresultat.

Simuleringarna för alla modeller i alla försäkringsklasser ger ett medelvärde som ligger ganska nära det som modellen ger ursprungligen. Den största biasen är för modell Gamma i försäkringsklass civil där simuleringarnas medelvärde är 1.3% större än modellens väntevärde. Detta indikerar att simuleringsproceduren är relativt bias-fri vad det gäller prediktion av modellens ursprungliga väntevärde trots problemen med negativa väntevärden i pseudotrianglarna (se kapitel 4.5).

Modellerna ger olika väntevärden utan något entydigt samband. I två fall av tre växer väntevärdet med växande p i variansfunktionen dvs. OdP, som ger samma skattning som Chain Ladder, ger minst reserv, SaPo lite större och Gamma störst. Detta var fallet för försäkringsklass civil och trafik. Skillnaderna är relativt små i försäkringsklass civil där Gamma ger 1.7% större skattad reserv än OdP medan de är större i försäkringsklass trafik där motsvarande skillnad är 18.2%. I det tredje fallet, för försäkringsklass olycksfall, sjunker väntevärdet med ökande p där Gamma ger minst reserv SaPo lite större och OdP störst där Gamma-modellens reserv är 2.2% mindre än OdP:s.

Variationskoefficienten varierar mellan metoderna och försäkringsklasserna. Gamma ger störst variationskoefficient i alla klasser. OdP har lägst variationskoefficient i klass civil och olycksfall där SaPo ligger ganska nära men något över. I försäkringsklass trafik är SaPo-modellens variationskoefficient aningen mindre än hos OdP-modellens. Största skillnaderna i variation är i försäkringsklass civil där variationskoefficienten är mer än dubbelt så stor för Gamma jämfört med OdP och SaPo och minsta skillnaderna finns försäkringsklass trafik där standardavvikelserna ligger nära varandra, variationskoefficienterna är 7.1, 6.7 resp. 7.3 procent.

Om man betraktar 90-, 95-, 97,5- och 99%-percentilen för alla tre metoderna i alla tre försäkringsklasser är resultatet entydigt. OdP ger minst percentil, SaPo lite större och Gamma störst.

Möjligen kan 1000 iterationer vara på gränsen till för lite. Tittar vi på Gamma-simuleringarna i figur 23 och 31 ser det ut som att den framsimulerade fördelningen inte riktigt har konvergerat.

Analysen av plottade residualer i kapitel 4.1 talar för OdP-modellen eller en modell med p nära ett. Problemen med många negativa värden i pseudotrianglarna däremot samt det märkliga utseendet hos fördelningarna det simuleras ur i framtidstriangeln talar emot OdP-modellen. De kraftigt konformade residualerna samt den stora variansen hos Gamma-modellen tyder på att variansantagandet med $p = 2$ kan vara för stort. Gammamodellen känns överlag aningen instabil. SaPo-modellens residual-plottar är inte helt tillfredsställande. Däremot känns fördelningarna det simuleras ur i framtidstriangeln rimliga och problemet med negativa värden i pseudotrianglarna är inte lika stort som i OdP.

I ett val mellan de tre modellerna OdP, SaPo med $p = 1.5$ och Gamma blir rekommendationen att välja modell SaPo. Den känns naturlig då den uppstår genom ett poissonfördelat antal gammafördelade variabler och vi här i en given cell har ett stokastiskt antal utbetalningar av varierande storlek. Möjligen skulle en modell med p någonstans mitt emellan 1 och 1.5 vara bäst. På senare tid har metoder och framförallt programvaror för att skatta p kommit som gör att en metod skulle kunna vara att först skatta p och sedan stoppa in den i en sådan här simuleringsalgoritm i SAS. I ett försök, egentligen utanför ramen för detta arbete, skattades p med hjälp av programvaran Emblem till 1.42, 1.42 och 1.65 för olycksfall, trafik respektive civil.

6 Appendix

6.1 Parameterbyte i log-linjära modellen

Parameterbytet för att gå från den log-linjära strukturen

(*) $\log(\mathbf{m}_j) = \log(\mathbf{g}_i) + \log(\mathbf{d}_j)$, där $\log(\sum_{j=1}^m \mathbf{d}_j) = 0$ till

(**) $\log(\mathbf{m}_j) = c + \mathbf{a}_i + \mathbf{b}_j$, där $\mathbf{a}_1 = \mathbf{b}_1 = 0$ sker genom att låta

$$c = \log(\mathbf{g}_1 \mathbf{d}_1), \mathbf{a}_i = \log(\mathbf{g}_i / \mathbf{g}_1), \mathbf{b}_j = \log(\mathbf{d}_j / \mathbf{d}_1) \Leftrightarrow \mathbf{g}_i = e^{c+\mathbf{a}_i} \left(\sum_{j=1}^m e^{\mathbf{b}_j} \right), \mathbf{d}_j = \frac{e^{\mathbf{b}_j}}{\sum_{j=1}^m e^{\mathbf{b}_j}}.$$

6.2 Parameterskattning och "hat"-matrisen

Vid parameterskattning sker först en övergång från triangelform till listform så att vi har observationerna y_{ij} i en lång vektor $\bar{y} = (y_1, y_2, \dots, y_n)^T$ där $n = \frac{m(m+1)}{2}$ är antalet observationer i utvecklingstriangeln. Observationerna y_{ij} ordnas så att de hamnar enligt $\bar{y} = (y_{11}, y_{12}, \dots, y_{1m}, y_{21}, \dots, y_{m1})^T$. Den generaliserade linjära modellen skrivs då på formen

$$g(\mathbf{m}) = X\bar{\mathbf{b}}$$

där funktionen g är log, X är den $\frac{m(m+1)}{2} \times (2m-1)$ -dimensionella designmatrisen

innehållande ettor och nollor, och $\bar{\mathbf{b}} = (\mathbf{b}_1, \dots, \mathbf{b}_{2m-1})^T$ är parametervektorn innehållande parametrarna c, \mathbf{a}_i och \mathbf{b}_j omdöpta och ordnade enligt $\bar{\mathbf{b}} = (c, \mathbf{a}_2, \dots, \mathbf{a}_m, \mathbf{b}_2, \dots, \mathbf{b}_m)^T$. Då exempelvis $m = 4$ har vi $g(\mathbf{m}) = X\bar{\mathbf{b}} \Leftrightarrow$

$$\begin{bmatrix} \log(\mathbf{m}_1) \\ \log(\mathbf{m}_2) \\ \log(\mathbf{m}_3) \\ \log(\mathbf{m}_4) \\ \log(\mathbf{m}_5) \\ \log(\mathbf{m}_6) \\ \log(\mathbf{m}_7) \\ \log(\mathbf{m}_8) \\ \log(\mathbf{m}_9) \\ \log(\mathbf{m}_{10}) \end{bmatrix} = \begin{bmatrix} 110001000 \\ 110000100 \\ 110000010 \\ 110000001 \\ 101001000 \\ 101000100 \\ 101000010 \\ 100101000 \\ 100100100 \\ 100011000 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \\ \mathbf{b}_4 \\ \mathbf{b}_5 \\ \mathbf{b}_6 \\ \mathbf{b}_7 \\ \mathbf{b}_8 \\ \mathbf{b}_9 \end{bmatrix}$$

Skattningen av parametrarna c, \mathbf{a}_i och \mathbf{b}_j genom vanlig maximum likelihood sker genom maximering av likelihoodfunktionen

$l(\bar{\mathbf{q}}; \mathbf{f}, \bar{y}) = \frac{1}{\mathbf{f}} \sum_i (y_i \mathbf{q}_i - b(\mathbf{q}_i)) + \sum_i c(y_i, \mathbf{f})$. Teorin nedan är hämtad ur Ohlsson & Johansson (2003).

Derivering av denna med avseende på \mathbf{b}_j blir efter upprepad användning av kedjeregeln

$$\frac{\partial l}{\partial \mathbf{b}_j} = \frac{1}{\mathbf{f}} \sum_i \frac{y_i - \mathbf{m}_i}{v(\mathbf{m}_i) g'(\mathbf{m}_i)} x_{ij} = \{g(\mathbf{m}) = \log(\mathbf{m}), v(\mathbf{m}) = \mathbf{m}^p\} = \frac{1}{\mathbf{f}} \sum_i \frac{y_i - \mathbf{m}_i}{\mathbf{m}_i^{p-1}} x_{ij}.$$

Nollstället till dessa $2m+1$ derivator ger SAS genom Newton-Raphson-metoden enligt

$$(*) \mathbf{b}_{k+1} = \mathbf{b}_k - H \cdot \bar{s} \text{ där } H \text{ (Hessianen) är matrisen av andraderivator } \frac{\partial^2 l}{\partial \mathbf{b}_j \partial \mathbf{b}_k} \text{ och } \bar{s} \text{ är}$$

vektorn av förstaderivator till likelihooden. H ges av $H = -X^T W_0 X$ där X är designmatrisen och W_0 är en diagonalmatris vars element på rad i ges av

$$w_{0i} = w_{ei} + (y_i - \mathbf{m}_i) \frac{v(\mathbf{m}_i) g''(\mathbf{m}_i) + v'(\mathbf{m}_i) g'(\mathbf{m}_i)}{\mathbf{f}(v(\mathbf{m}_i))^2 (g'(\mathbf{m}_i))^3} \text{ där } w_{ei} = \frac{1}{v(\mathbf{m}_i) (g'(\mathbf{m}_i))^2}.$$

Hatmatrisen defineras här, enligt SAS default, som $\text{Hatmatrix} = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2}$ där $W = W_0$ om algoritmen slutade i första steget och W_e annars.

Maximum quasi likelihood-skattning av parametrarna sker på samma sätt. Enligt teorin innebär Maximum quasi likelihood att parametrarna skattas genom maximering av quasi-

likelihood-funktionen $Q(\bar{\mathbf{m}}, \bar{y}) = \sum_i Q_i(\bar{\mathbf{m}}, \bar{y}) = \sum_i \int_{y_i}^{\mathbf{m}_i} \frac{y_i - t}{\mathbf{f}v(t)} dt = \sum_i \int_{y_i}^{\mathbf{m}_i} \frac{y_i - t}{\mathbf{f}t^p} dt$ i stället för

den vanliga log-likelihooden. Denna fungerar som en äkta log-likelihood vad det gäller estimation av parametrarna, se Olsson & Johansson (2003) kap 3, och ger samma parameterskattningar som maximum likelihood. Fördelen med quasi-likelihood är att fördelning inte behöver specificeras utan det räcker med specifikation av variansfunktionen.

6.3 Fördelningar

6.3.1 Överspridd Poisson, OdP

Variansfunktionen $v(\mathbf{m}) = \mathbf{m}$ svarar mot en överspridd poissonfördelning. Givet \mathbf{f} är frekvensfunktionen för en överspridd Poisson variabel

$$f_Y(y) = \exp \left\{ \frac{y \log(\mathbf{m}) - \mathbf{m}}{\mathbf{f}} - \frac{y}{\mathbf{f}} \log(\mathbf{f}) - \log \left(\binom{y}{\mathbf{f}} \right) \right\} = \exp \left\{ \frac{y \log(\mathbf{m}) - \mathbf{m}}{\mathbf{f}} + \log \left(\frac{\mathbf{f}^{-\frac{y}{\mathbf{f}}}}{\binom{y}{\mathbf{f}}} \right) \right\} \text{ där}$$

$y = \mathbf{f}n, n \in \{0,1,2,\dots\}$, och 0 annars. För dessa fördelningar gäller att $\text{Var}(Y) = \mathbf{f}\mathbf{m} \Rightarrow \sqrt{\text{Var}(Y)} = \sqrt{\mathbf{f}}\sqrt{\mathbf{m}}$ det vill säga standardavvikelsen är proportionell mot $\sqrt{\mathbf{m}}$.

6.3.2 Sammansatt Poisson

Då $1 < p < 2$ i $v(\mathbf{m}) = \mathbf{m}^p$ visar Jørgensen (1997) att fördelningen blir

$$f_Y(y) = \exp\left\{\frac{y\mathbf{q} - b(\mathbf{q})}{\mathbf{f}} + c(y; \mathbf{f}, p)\right\} \text{ där}$$

$$b(\mathbf{q}) = \frac{1}{2-p} (\mathbf{q}(1-p))^{(p-2)/(p-1)}, \mathbf{m} = b'(\mathbf{q}) = (\mathbf{q}(1-p))^{1/(1-p)} \Leftrightarrow \mathbf{q} = \frac{\mathbf{m}^{1-p}}{1-p} \text{ och}$$

$$c(y; \mathbf{f}, p) = \begin{cases} \log\left(\frac{1}{y} \sum_{n=1}^{\infty} \frac{\mathbf{f}^{-n} (b(-\mathbf{f}/y))^n}{\Gamma(n(2-p)/(p-1))n!}\right) & \text{då } y > 0 \\ 0 & \text{annars} \end{cases} .$$

Denna fördelning uppkommer på följande sätt:

Antag att $Z = \sum_{i=1}^N X_i$ där $X_i \sim iid Ga(\mathbf{a}, \mathbf{b})$ och $N \sim Po(m)$ samt N och X_i oberoende. Med $Ga(\mathbf{a}, \mathbf{b})$ menas här gammafördelning med $E[X] = \mathbf{a}\mathbf{b}$ och $Var(X) = \mathbf{a}\mathbf{b}^2$. Parametrarna \mathbf{a} och \mathbf{b} i gammafördelningen samt m i poissonfördelningen väljs

enligt $\mathbf{a} = \frac{2-p}{p-1}$, $\mathbf{b} = \mathbf{f}(p-1)\mathbf{m}^{p-1}$ och $m = \frac{\mathbf{m}^{2-p}}{\mathbf{f}(2-p)}$. Den stokastiska variabeln Z har då ovan

nämnda fördelning. För dessa fördelningar gäller att $Var(Y) = \mathbf{f}\mathbf{m}^p \Rightarrow \sqrt{Var(Y)} = \sqrt{\mathbf{f}\mathbf{m}^p}$.

6.3.3 Gamma

Tätheten för en gammafördelad stokastisk variabel är

$$f_Y(y; \mathbf{a}, \mathbf{b}) = \frac{1}{\Gamma(\mathbf{a})} y^{\mathbf{a}-1} \frac{1}{\mathbf{b}^{\mathbf{a}}} e^{-\frac{y}{\mathbf{b}}} = \exp\left\{(\mathbf{a}-1)\log(y) - \log(\Gamma(\mathbf{a})) - \mathbf{a}\log(\mathbf{b}) - \frac{y}{\mathbf{b}}\right\} =$$

$$= \{\mathbf{f} = 1/\mathbf{a}, \mathbf{m} = \mathbf{a}\mathbf{b}\} = \exp\left\{\frac{y(-1/\mathbf{m}) - \log(\mathbf{m})}{\mathbf{f}} + \log\left(\frac{y^{1-\mathbf{f}}}{\mathbf{f}(\Gamma(1/\mathbf{f}))^{\mathbf{f}}}\right)\right\} =$$

$$= \{\mathbf{q} = -1/\mathbf{m}, b(\mathbf{q}) = \log(-1/\mathbf{q})\} = \exp\left\{\frac{y\mathbf{q} - b(\mathbf{q})}{\mathbf{f}} + c(y; \mathbf{f})\right\} .$$

För denna fördelning gäller att $Var(Y) = \mathbf{f}\mathbf{m}^2 \Rightarrow \sqrt{Var(Y)} = \sqrt{\mathbf{f}\mathbf{m}}$ d v s standardavvikelsen är proportionell mot väntevärdet.

Litteratur

Efron, B. & Tibshirani, R. J., 1993. *An Introduction to the Bootstrap*. Chapman and Hall.

England, P. D. & Verrall, R. J., 2002. *Stochastic claims reserving in general insurance*. Presenterat inför Institute of actuaries, 28 januari 2002.
<http://www.actuaries.org.uk/files/pdf/sessional/sm0201.pdf>

England, P. D. & Verrall, R. J., 1999. *Analytic and bootstrap estimates of prediction errors in claims reserving*. Mathematics and Economics, Elsevier, vol. 25(3).

England, P. D., 2001. *Addendum to "Analytic and bootstrap estimates of prediction errors in claims reserving"*. Research report no. 138 City University London.

Jørgensen, Bent, 1997. *The theory of dispersion models*. Chapman and Hall, New York.

Mack, T., 1993. *Distribution-free calculation of the standard error of chain-ladder reserve estimates*. ASTIN Bulletin, 23, No.2.

McCullagh, P. & Nelder J.A., 1989. *Generalized linear models*, second edition. Chapman and Hall.

Ohlsson, E. & Johansson, B., 2003. *Prissättning inom sakförsäkring med Generaliserade linjära modeller*.

Olsson, U., 2002. *Generalized linear models, an applied approach*. Studentlitteratur.

Pinheiro, P. J. R., Andrade e Silva, J. M. & Centeno, M., 2002. *Bootstrap methodology in claims reserving*. <http://www.casact.org/coneduc/reinsure/astin/2000/andradeesilva1.doc>

SAS Institute, 1999. *SASOnlineDocument*.