DOOR TOM HOEDEMAKERS[1], MARC GOOVAERTS[2] EN JAN DHAENE[3]

IBNR

IBNR-PROBLEMATIEK IN HISTORISCH PERSPECTIEF

De bouwstenen van de tegenwoordig gebruikte actuariele methoden ter berekening van schadereserves zijn elementaire statistische technieken die moeten behoren tot het instrumentarium van iedere schadeactuaris. Het blijkt met de standaardprogrammatuur GLIM, SAS of S-PLUS mogelijk te zijn modellen te kiezen en de bijbehorende parameters te schatten. Een aanzet hiertoe wordt gegeven in het college Schade: Statistische technieken, dat aan de UvA gegeven wordt. Zulke technieken zijn al heel oud, en de IBNR-technieken zijn dus niet, zoals sommigen denken, in het laatste decennium van de 20e eeuw aan het brein van actuarissen ontsproten!

Verbeek (1972) ontwierp het eerste statistisch model voor de beschrijving van het IBNR-probleem. De schattingen werden uitgevoerd door de maximum likelihood-benadering toe te passen op onafhankelijk veronderstelde resultaten per cel in de afwikkelingsdriehoek. Taylor (1977) voerde de separatiemethode in, naar analogie met die van Verbeek, en gaf het speciale karakter aan van de 'derde dimensie', die van de kalenderjaren. De Vijlder en Goovaerts (1979) stelden als eersten de gelijktijdige behandeling van de driehoek in de drie richtingen voor: ontstaansjaren, kalenderjaren en afwikkelingsjaren.

Met IBNR-technieken worden schattingen gegeven voor de toekomstige betalingen van een portefeuille. Dit zijn dan puntschattingen voor de verwachtingen daarvan. Deze schattingen zijn alle afhankelijk van wat er in het verleden gebeurd is (de linker-bovendriehoek van de afwikkelingsdriehoek), en hangen dus sterk samen. Het bepalen van de precieze correlaties is ondoenlijk, maar door deze afhankelijkheden iets te versterken tot comonotonie (de comonotone copula) zijn Goovaerts en Redant (1999) erin geslaagd verdelingsfuncties te construeren die enerzijds dicht bij elkaar liggen, anderzijds een boven- en een ondergrens geven voor de 'echte' verdelingsfunctie van de nog te volgen betalingen, en wel in de zin van stop-loss ordening. Soms worden deze toekomstige betalingen beschouwd als een som van afhankelijke lognormale stochasten. Een verkeerde methode is deze te benaderen door middel van één lognormale stochast met als verwachting en variantie de sommen van de individuele verwachtingen en varianties per cel. Op het laatste AFIR-colloquium wees ook Albrecht op deze problematiek.

Verschillende alternatieven zijn ontwikkeld voor de statistische behandeling van de resultaten uit het verleden, dus de modellering van de gegevens in de bovendriehoek. Maar de kernvraag van het schadereserveringsprobleem is de specifieke behandeling van de derde dimensie, en de propagatie van de fouten rekening houdend met de zeer beperkte informatie. De benadering van Posthuma Partners is daarom belangrijk in de zin dat bijkomende informatie (gerechtvaardigd als een Bayesiaanse aanpak) wordt toegevoegd om de onmogelijkheid van de schatting van en de modelkeuze naar de toekomst beter op te lossen.

VERGELIJKEN

We vergelijken de resulaten van de methode van Posthuma en Ter Berg met een aantal modellen uit de actuariële literatuur. De schadeclaims in Tabel 2, weergegeven in incrementele vorm, vinden hun oorsprong in 1991 (Historical Loss Development Study (1991) - Automatic Facultative General Liability (excluding Asbestos and Environmental)) en werden nadien door verscheidene auteurs behandeld, zoals Mack (1994), en England en Verrall (2002). Kenmerkend aan de data is dat zij een aanzienlijke variabiliteit tussen de onstaansjaren vertonen. Merk ook de negatieve incrementele waarde in cel (2,7) op.

	1	2	3	4	5	6	7	8	9	10
1	5012	3257	2638	898	1734	2642	1828	599	54	172
2	106	4179	1111	5270	3116	1817	−103	673	535	
3	3410	5582	4881	2268	2594	3479	649	603		
4	5655	5900	4211	5500	2159	2658	984			
5	1092	8473	6271	6333	3786	225				
6	1513	4932	5257	1233	2917					
7	557	3463	6926	1368						
8	1351	5596	6165							
9	3133	2262								
10	2063									

TABEL 1 RUN-OFF DRIEHOEK MET NIET-CUMULATIEVE CLAIMGEGEVENS

Analyse met SAS/S-PLUS leert ons dat we hier het best gebruik kunnen maken van een quasi-gammamodel in het GLM-framework. We modelleren de incrementele claims Y_{ij}, (i,j=1,..,10) met een logaritmische linkfunctie om een multiplicatieve parametrische structuur te ver-

[1] Tom Hoedemakers is PhD student aan de K.U. in Leuven

[2] Marc Goovaerts is hoogleraar Actuariaat aan de K.U. in Leuven en aan de Universiteit van Amsterdam

[3] Jan Dhaene is hoogleraar Actuariaat aan de K.U. in Leuven en aan de Universiteit van Amsterdam

krijgen en we linken de verwachte waarde van de responsvariabele met een lineaire predictor

$$\eta_{ij} = \alpha + \left[I_{(j \leq 2)}(j-1) + I_{(j>2)} \right] \beta_1 + \left[I_{(3 \leq j \leq 6)}(j-2) + I_{(j>6)} 4 \right] \beta_2 + I_{(j>6)}(j-6)\ \beta_3 + (i+j-2)\,\gamma$$

met α, $(\beta_1, \beta_2, \beta_3)$, en γ parameters die de ontstaansjaren, de afwikkelingsjaren en de kalenderjaren respectievelijk modelleren en waarbij de indicator $I_{(j \leq x)}$ als j, 0 anders.

Tabel 2 geeft de totale reserve en die van de rijtotalen weer voor de voornaamste methoden. Merk op dat vooral de lognormale aanpak er duidelijk bovenuit schiet. Dit is te wijten aan het feit dat als men eerst logaritmen van de data neemt, daarvan de verwachtingen zuiver schat en daar vervolgens de exponent van neemt, men een naar boven onzuivere schatting krijgt. De cijfers in deze tabel zijn berekend met behulp van het standaard statistische pakket SAS. Voor een beschrijving van de Bayesiaanse modellen hierin verwijzen we naar de papers van England en Verrall (1999, 2001, 2002).

Het correct schatten van de reserves gaat voor een deel samen met een juiste modellering van de reeds geobserveerde schadebedragen. Vaak zal bij dit modelleren eerst de chain-laddertechniek toegepast worden, omdat dit een welbekende en gemakkelijk te gebruiken methode is. De hieruit voortvloeiende parameters kunnen dan glad gemaakt worden om zo met behulp van extra informatie en de opinie van actuarissen - voor wat deze waard is - een correcte interpretatie van de resultaten te verkrijgen. De voornaamste kritiek op het chain-laddermodel is het feit dat het overgeparametriseerd zou zijn; het kent immers een parameter voor elk ontstaansjaar en voor elk afwikkelingsjaar. Een mogelijke oplossing is het fitten van een parametrische curve door het afwikkelingspatroon. Een veel gebruikt voorbeeld hiervan is de Hoerlcurve. Een bijkomende kritiek op de chain-laddertechniek is dat zij vaak onvoldoende resultaten zal produceren indien er onevenwichtigheid bestaat tussen de verhouding van de afwikkeling in de beginjaren. De Bornhuetter-Fergusontechniek brengt hier dan weer soelaas. Het idee bestaat erin om de resultaten te stabiliseren door gebruik te maken van initiële schattingen. Deze schattingen worden bepaald aan de hand van externe informatie en interpretaties van ter zake kundigen. In het Bayesiaanse model onderscheiden we twee uitersten: de chain-laddertechniek (geen enkele voorgaande prior informatie over de rij parameters) en de Bornhuetter-Fergusontechniek.

Jaar	Pp	QG.	Boot.	LogN.	Bayes1	Bayes2	Bayes3	B-F	Hoerl1	Hoerl2	ODP
1	10	–	–	–	–	–	–	–	435	–	–
2	43	163	177	378	155	159	166	154	634	243	154
3	177	472	639	1.021	625	636	635	617	1.447	885	617
4	594	1.058	1.655	3.334	1.659	1.649	1.675	1.636	2.704	2.033	1.636
5	1.532	2.173	2.770	3.586	2.811	2.808	2.820	2.747	4.269	3.582	2.747
6	3.214	4.291	3.769	6.132	3.717	3.731	3.713	3.649	4.304	3.849	3.649
7	5.814	6.886	5.459	7.565	5.511	5.539	5.523	5.435	5.801	5.393	5.435
8	9.722	10.067	11.259	18.526	11.000	11.030	11.090	10.907	11.650	11.091	10.907
9	13.839	13.970	10.902	27.510	10.900	10.840	10.870	10.650	10.936	10.568	10.650
10	18.964	18.765	16.580	60.772	14.440	16.990	14.640	14.206	18.109	17.654	16.339
Tot.	**53.909**	**57.845**	**53.210**	**128.823**	**50.820**	**53.380**	**51.130**	**50.002**	**60.291**	**55.297**	**52.135**

TABEL 2 RESERVE RESULTATEN

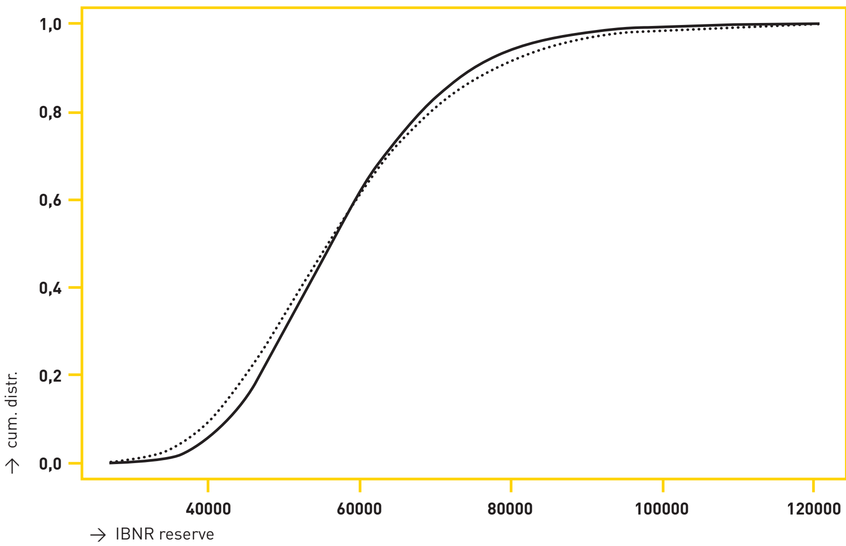
Pp: Het model van Posthuma en ter Berg, gebruikmakend van een constant tijdreeksmodel.
QG.: Quasi-gamma model als 'optimaal' aangemerkt door SAS.
Boot.: Bootstrap resultaten (1000 iteraties) in combinatie met het chain-laddermodel.
LogN.: De lognormale benadering met de chain-ladder lineaire predictor zonder correctie.
Bayes1/2/3: Het Bayesiaans chain-laddermodel met 'precise'/'vague'/'informative' priors (overdispersed Poisson model).
B-F: De Bornhuetter-Fergusontechniek.
Hoerl1/2: Hoerl-curveresultaten met / zonder extrapolatie (overdispersed Poisson model).
ODP: Het overdispersed Poisson model met de chain-ladder lineaire predictor.

VOORSPELLEN

Het bepalen van de reserve is een predictieproces. Gegeven de data poogt men een voorspelling te maken voor de toekomstige schadebedragen. Dit proces vergt een nauwkeurige statistische analyse. Een ander belangrijk punt in dit reserveringsproces is het schatten van de voorspelfout van de reserves, gegeven het onderliggend model. Deze voorspelfout, die we moeten beschouwen als een maat voor de precisie van de reserveschattingen, is belangrijk voor het opstellen van veilige reserves. Het probleem van het schatten van de voorspelfout van toekomstige betalingen en reserveschattingen kan worden teruggebracht tot het schatten van twee componenten: de procesfout en de schattingsfout. Naast varianties zijn andere maten, zoals scheefheidsmaten en risicomaten, ook zeer belangrijk in deze context. Het opstellen van de predictieverdeling is niet triviaal, aangezien de reserve een som van gecorreleerde stochastische veranderlijken is. Hierbij moet er zowel rekening worden gehouden met de variabiliteit van het onderliggend statistisch proces als met de variabiliteit van de parameterschatting. In de literatuur zijn er, naast de aantrekkelijke methode van Posthuma en Ter Berg, vier methoden bekend voor het opstellen van deze verdeling.

Bij de bootstraptechniek trekt men willekeurig (met teruglegging) één van de in de steekproef gerealiseerde residuen. Op die manier verkrijgt men een pseudo-dataset met een analoge onderliggende verdeling. Een tweede methode, die bekendstaat als het simuleren uit de parameters, is net als de bootstrapprocedure een tweeledig proces. Hierbij wordt de schattingsfout gesimuleerd door de parameters van een geschikte gezamenlijke verdeling te simuleren, gebruikmakend van informatie over het gefitte model. Op die manier kan de modelpredictor gevormd worden. Vervolgens kan, gegeven de gesimuleerde lineaire predictor, de fout die samenhangt met het onderliggend statistisch proces geïntegreerd worden door een willekeurige observatie uit de procesverdeling te trekken. Deze methode is alleen bruikbaar als de gezamenlijke verdeling van de parameters geïdentificeerd kan worden. Gebruik makend van Bayesiaanse methoden kan een posteriorverdeling van de parameters, conditioneel op de geobserveerde data, verkregen worden. Markov chain Monte Carlo (MCMC)-methoden leveren ten slotte de predictieverdeling. Deze methoden combineren schatting en predictie tegelijkertijd in hetzelfde kader. De vierde methode is gebaseerd op convexe orde-technieken, zodanig dat de benaderingen voor de verdelingsfunctie van de reserve kleiner of groter in stop-loss orde zijn dan de echte verdeling van de reserve. Deze aanpak, die vele actuariële en financiële toepassingen kent, berekent beneden- en bovengrenzen voor de reserve die een som is van afhankelijke stochastische variabelen. Zij maakt daarbij op een efficiënte wijze gebruik van de beschikbare informatie. We hebben deze vier methoden toegepast op bovenstaande dataset en vergeleken met de resultaten van Posthuma en Ter Berg. De resultaten vindt u terug in Tabel 3. Figuur 1 toont de verdelingsfunctie van de onder- en bovengrens. Beide liggen vrij dicht tegen elkaar, zodat zij samen betrouwbare informatie verschaffen over de echte verdeling van de totale reserve. Merk op dat deze laatste daar niet tussenin hoeft te liggen, maar de bijbehorende stop-loss premies moeten wel zo geordend zijn.

Een goede methode moet enerzijds aan alle elementaire statistische eigenschappen voldoen en anderzijds ook de mogelijkheid bieden om op een relatief eenvoudige wijze externe informatie te implementeren. In dat opzicht voldoet het model van Posthuma Partners zeer goed. Voor het merendeel van de kort beschreven methoden is een uitbreiding mogelijk om met externe informatie rekening te houden. Alle methoden zullen gebruikmaken van het verleden om een voorspelling voor de toekomst te maken. De toekomst is echter niet altijd een weerspiegeling van het verleden. Voorbeelden hiervan uit de praktijk zijn er voldoende, zoals optredende vertraging in de rechtspraak en het voorrang geven aan belangrijke zaken.



FIGUUR 1 DE CUMULATIEVE VERDELINGSFUNCTIES VAN DE ONDERGRENS (DOORGETROKKEN LIJN) EN VAN DE VERBETERDE BOVENGRENS (PUNTJES LIJN)

p	Bootstrap	Lognormal	Bayes2 ODP	Pp	ondergrens	verbeterde bovengrens
0,9	78.491	222.397	76.550	64.870	74.634	77.533
0,95	87.668	278.500	87.210	68.056	80.764	85.051
0,99	109.445	489.267	108.500	74.208	92.197	111.101

TABEL 3 VERGELIJKING VAN ENKELE HOGE KWANTIELEN ONDER DE VOORNAAMSTE BOVENGENOEMDE METHODEN

CONCLUSIE

Samenvattend kunnen we concluderen dat een reserveringsproces grosso modo tweeledig is. Enerzijds moet men tot een statistisch geschikt model komen, anderzijds is het van groot belang de verdeling van de reserve te bepalen, waaruit dan de nodige statistische maten afgeleid kunnen worden. De basis van heel het proces is natuurlijk het model. Er bestaat al een ruim gamma van modellen op de markt, maar steeds opnieuw moet men de modelassumpties nagaan. Het verkrijgen van de verdeling van de reserve is een zeer complex proces waarnaar de laatste tijd veel onderzoek is verricht. Uit ervaring kunnen we meegeven dat de benadering via de comonotone copula, zoals aangehaald in de introductie, zeer efficiënt is. Een bijkomend voordeel bij deze aanpak is dat men hier een stochastisch verdisconteringsproces kan inbouwen, zie Hoedemakers et al. (2003). Dit laatste is onder andere nuttig in een DFA-analyse. De methode van Posthuma Partners maakt gebruik van een deterministische verdiscontering, die toelaat stress testing uit te voeren. Ook de Bayesiaanse aanpak, die vrij eenvoudig te implementeren is, levert goede resultaten.

REFERENTIES

- Albrecht P. (1983). Parametric multiple regression risk models: Some connections with IBNR. *Insurance: Mathematics and Economics*, 2, 69-73.
- De Vijlder, F. en Goovaerts, M.J. (1979). Proceedings of the First Contact Group "Actuarial Sciences" Instituut voor Actuariële wetenschappen, K.U. Leuven, Belgium.
- England, P.D. en Verrall, R.J. (1999). Analytic and bootstrap estimates of prediction errors in claim reserving. *Insurance: Mathematics and Economics*, 25, 281-293.
- England, P.D. en Verrall, R.J. (2001). A flexible framework for stochastic claims reserving. *Proceedings of the Casualty Actuarial Society*, 88.
- England, P.D. en Verrall, R.J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3), 443-518.
- Goovaerts M.J. en Redant H. (1999). On the distribution of IBNR reserves. *Insurance: Mathematics and Economics*, 25, 1-9.
- Hoedemakers T., Beirlant J., Goovaerts M.J. en Dhaene J. (2003). Confidence Bounds for Discounted Loss Reserves. *Insurance: Mathematics and Economics*, To appear.
- Mack, T. (1994). Measuring the variability of chain-ladder reserve estimates. Casualty Actuarial Society, Spring Forum.
- Taylor G.C. (1977). Separation of Inflation and other Effects from the Distribution of Non-Life Insurance Claim Delays. *Astin Bulletin*, 9, 219-230.
- Verbeek H.G. (1972). An approach to the analysis of claims experience in motor liability excess of loss reinsurance.