

DE IBNR - DISCUSSIE

LOGNORMALE LINEAIRE REGRESSIEMODELLEN IN SCHADERESERVERING

Tom Hoedemakers¹ Marc Goovaerts^{1,2} Jan Dhaene^{1,2}

Men moet steeds in het achterhoofd houden dat een statistische analyse voor het schatten van de schadereserve slechts één component is in het hele reserveringsproces. Meestal beschikt de actuaire over een zekere voorkennis bij het analyseren van dergelijke datasets. Reserveren is immers een praktische aangelegenheid. We moeten de data proberen te begrijpen in plaats van in alle mogelijke situaties steeds hetzelfde model toe te passen. In die optiek is de bijdrage van Postuma en Ter Berg dan ook zeker aan te moedigen, aangezien zij ruimte laat voor het implementeren van extra informatie, iets wat in de toekomst zeker aan belang zal winnen.

Toch merken we op dat het lognormale lineaire regressiemodel een vaak toegepast model is, onder andere ook bij 'de Australische aanpak' die door H.J. Prins wordt gepromoot. Het gebruiken van commerciële slogans ter verpakking van een doordeweekse statistische techniek, met de bedoeling een zweem van wetenschappelijkheid te creëren is voor een actuaire een brug te ver lijkt ons. Het lognormale model werd immers als eerste stochastisch model gebruikt gelet op de eenvoud en dit op het moment dat het gebruik van modellering met een onderliggende GLM-benadering nog in de kinderschoenen stond. Anderzijds kan een lognormaal model eenvoudig geïmplementeerd worden - zonder gebruik te maken van gespecialiseerde software - met behulp van spreadsheets. De actuaries in Nederland waren trouwens al op de hoogte van het gebruik van het lognormale model ongeveer een kwarteeuw geleden door het NN-boekje: Loss Reserving Methods, van De Wit-Van Eeghen. Dit model nu bestempelen als innovierend is gebakken lucht.

Het is echter zeer belangrijk te vermelden dat het gebruik van lognormale lineaire regressiemodellen aanleiding kan geven tot totaal onbruikbare resultaten. Oorzaak hiervan is vaak dat de data niet vanzelfsprekend voldoen aan de modelassumpties die schuil gaan achter het lognormale model.

We bespreken achtereenvolgens kort de lognormaliteit, de afhankelijkheid en de homoscedasticiteit. Vervolgens gaan we even dieper in op de interpretatie van de logtransformatie en tenslotte wijzen we nog op het gevaar dat schuil gaat achter het simuleren uit een lognormaal model (explosierisico).

In een lognormaal lineair regressiemodel wordt eerst de logaritme genomen van de niet-cumulatieve claims Y_{ij} . De normale variabelen $Z_{ij} = \log(Y_{ij})$ worden dus verondersteld additief opgebouwd te zijn uit een deterministische component met gemiddelde η_{ij} en een normaal verdeelde storingsterm $\varepsilon_{ij} \sim N(0, \sigma^2)$.

Normaliteit kan eenvoudig getest worden aan de hand van allerlei statistische testen zoals o.a. de normale qq-plot, de Kolmogorov-Smirnov test, boxplots, histogrammen, ... De veronderstelling van homoscedasticiteit van de storingstermen zal niet voldaan zijn indien de fouten een ongelijke variantie bezitten. Dit probleem, dat in de praktijk vaak voorkomt, kan afgezwakt worden door gebruik te maken van gewogen regressie waarbij de gewichten empirisch bepaald worden. Bovendien komt het vaak voor dat de fouten gecorreleerd zijn van jaar tot jaar, waardoor de onafhankelijkheidsassumptie geschonden wordt.

¹ Department of Applied Economics, K.U.Leuven, Naamsestraat 69, 3000 Leuven, Belgium

² Institute of Actuarial Science, University of Amsterdam, Roeterstraat 11, 1018 WB Amsterdam, Netherlands

Door het werken met getransformeerde data wordt een correcte interpretatie van de dataset bemoeilijkt. Een bijkomend nadeel is dat de dataset moet bestaan uit positieve waarden. Indien er bijvoorbeeld een overschatting gebeurd is van het te betalen bedrag, kan bij het werken met niet-cumulatieve data een negatief bedrag ontstaan. Negatieve waarden kunnen ook optreden bij herverzekeringen of terugningen.

Steekproeven in het reële leven zijn nooit oneindig. Bij eindige steekproeven zijn de staarten echter vaak zwaarder dan bij oneindige het geval zou zijn. Ook bij het schatten van de IBNR reserve werken we met een eindig aantal jaren. Het behandelen van de steekproef als zijnde oneindig kan in vele gevallen echter leiden tot een onderschatting van de reserve.

Klassieke regressie-analyse in de lognormale context vereist volgende stappen

- 1) Neem de logaritme van de claims
- 2) Fit een model op de getransformeerde waarden gebruik makend van kleinste kwadraten regressie-analyse
- 3) Bereken schattingen voor de parameters
- 4) Transformeer correct terug!

Schattingen voor het gemiddelde worden niet eenvoudigweg bekomen door de lineaire predictor te exponentiëren. De gefitte waarden worden wel gegeven door

$$\hat{Y}_{ij} = \exp(\hat{\eta}_{ij} + 1/2\hat{\sigma}_{ij}^2) \quad (1)$$

met

$$\hat{\sigma}_{ij}^2 = \hat{\sigma}^2(B(X'X)^{-1}B')_{ij} + \hat{\sigma}^2 \quad (2)$$

en X, B de designmatrix corresponderend met de bovendriehoek resp. de benedendriehoek. Wanneer we nu gebruikmaken van een GLM-modellering met logaritmische linkfunctie (bv. een over-dispersed Poisson model), dan hebben we dat

$$\hat{Y}_{ij} = \exp(\hat{\eta}_{ij}) \quad (3)$$

De reden waarom er in (1) een variantie component aanwezig is, heeft te maken met het feit dat in een gegeneraliseerd lineair model de niet-cumulatieve data zelf als responsvariabelen gebruikt worden, terwijl men in de lognormale modellering werkt met de logaritme van de incrementele data als responsvariabelen.

Hieruit vloeit onmiddellijk voort dat een schatting van de verwachte reserves stijgt naarmate de onderliggende variantie component groter wordt. Deze vertekeningcorrectie heeft uiteraard ook zijn impact op een extensieve simulatie gebaseerd op dit lognormale model waarbij extreme observaties de verwachte waarde sterk beïnvloeden.

Concluderend kunnen we stellen dat een lognormale modellering correct kan zijn, maar het is zeker niet de modellering bij uitstek in een reserveringsproces.

In hetgeen volgt illustreren we het explosie-effect van het lognormale model aan de hand van een voorbeeld. Beschouw onderstaande driehoek bestaande uit niet-cumulatieve claims:

	1	2	3	4	5	6	7	8	9	10
1	100009	299968	414232	493342	685473	849023	636022	36606	104179	88370
2	6200	409596	284139	2513640	282809	204375	614953	210333	65252	
3	483439	156832	43447	382747	65722	66184	246965	10057		
4	13179	750669	894092	658117	229653	171645	319902			
5	349744	1995400	3488883	514386	103631	289768				
6	606697	700000	770310	1307200	1140218					
7	66023	880946	1756675	364362						
8	290301	1878413	3737107							
9	485011	439134								
10	180344									

Na een uitgebreide analyse m.b.v. het statistisch pakket SAS (door o.a. gebruik te maken van een parameter-selectietechniek) bekomen we een regressiemodel met als lineaire predictor gegeven door volgende formule:

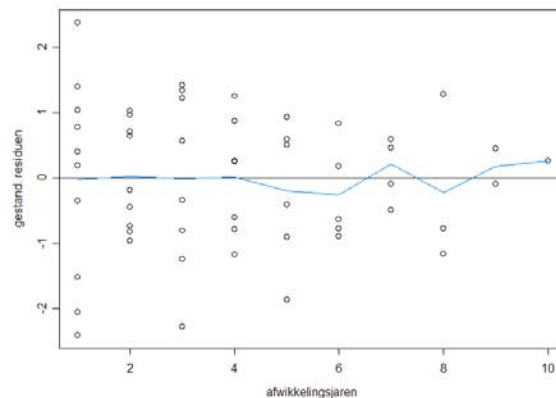
$$\eta_{ij} = I_{(i<3)}\alpha_1 + I_{(2<i<5)}\alpha_2 + I_{(i>4)}\alpha_3 + I_{(j>1)}\beta_1 + I_{(j>4)}\beta_2 + I_{(j>7)}\beta_3 + I_{(i+j-1>2)}\gamma_1, \quad (4)$$

met $(\alpha_1, \alpha_2, \alpha_3)$, $(\beta_1, \beta_2, \beta_3)$ en γ parameters die de ontstaansjaren, de afwikkelingsjaren en de kalenderjaren respectievelijk modelleren en waarbij bijvoorbeeld $I_{(j \leq 2)} = 1$ als $j \leq 2$, 0 anders.

Gebruikmakend van bovenstaande formule (4) en een lognormaal regressiemodel bekomen we volgende output in SAS

	estimate	s.e.	t-stat.	p-value
α_1	10,4689	0,5544	18,88	<0,0001
α_2	9,4446	0,6523	14,48	<0,0001
α_3	10,6876	0,6528	16,37	<0,0001
β_1	1,4500	0,3717	3,90	0,0003
β_2	-0,7258	0,325	-2,23	0,0302
β_3	-1,7482	0,4733	-3,69	0,0006
γ_1	1,7274	0,6612	2,61	0,0120

Niet minder dan 99,53% van de variantie in de historische claims wordt verklaard door de regressievergelijking. De aangepaste determinatiecoëfficiënt R_a^2 , dat rekening houdt met het aantal verklarende variabelen in het model en het aantal waarnemingen, heeft een waarde van 0,9946. De 'residual sum of squares', de stochastische component van de variatie van de afhankelijke variabele, bedraagt 42,0466. Akaike's Informatie Criterium (AIC) en het Bayesiaans Informatie Criterium (BIC) hebben respectievelijk een waarde van 56,05 en 70,1. Deze 'goodness-of-fit'-maten en de parameterschattingen met bijhorend p-waarden indiceren op het eerste zicht een adequate fit die de historische data statistisch correct beschrijft. Er is duidelijk geen sprake van overparameterisatie. Indien we echter naar de residuplots kijken, merken we op dat de Gauss-Markov condities niet voldaan zijn. Onderstaand hebben we de gestandaardiseerde residuen geplot tegenover de afwikkelingsjaren.



Het is duidelijk dat de residuen niet normaal verdeeld zijn met gemiddelde 0 en constante variantie σ^2 . De residuplot m.b.t. de kalenderjaarrichting vertoont een gelijkaardig patroon. De error-term moet onafhankelijk zijn van de plaats van observatie in de driehoek (homoscedasticiteit). Wanneer de storingsterm afhankelijk is van de ontstaansjaren, ontwikkelingsjaren of de kalenderjaren spreekt men van heteroscedasticiteit en betekent dit dat de vooropgestelde assumpties niet voldaan zijn. Uit bovenstaande residuplot blijkt duidelijk dat de spreiding van de gestandaardiseerde residuen in de richting van de ontwikkelingsjaren afneemt.

We kunnen dit laatste, het niet constant zijn van de variantie, eenvoudig testen, door de variantiefunctie te creëren in een bepaalde richting. Deze test is een variant van de wel bekende Breusch-Pagan test. Bijvoorbeeld in de ontwikkelingsjaren, kan men de volgende regressiefunctie schatten:

$$(\varepsilon_{ij}^*)^2 = \beta_0 + \beta_1 j,$$

met ε_{ij}^* de gestandaardiseerde residuen.

Nu kunnen we een t-test uitvoeren op β_1 om na te gaan of deze parameter significant is. Indien dit het geval is hebben we te maken met heteroscedasticiteit in die richting. De test in de twee andere richtingen is analoog. Onderstaand vind je de SAS-output voor de test in de drie richtingen, zijnde achtereenvolgens de ontstaansjaar-, de ontwikkelingsjaar- en de kalenderjaarrichting:

	estimate	s.e.	t-stat	p-value
intercept	1,2059	0,3276	3,68	0,0005
ontstaans	-0,0421	0,0698	-0,6	0,5489

	estimate	s.e.	t-stat	p-value
intercept	1,6856	0,3117	5,41	<0,0001
afwikkelings	-0,1621	0,0665	-2,44	0,0181

	estimate	s.e.	t-stat	p-value
intercept	2,4667	0,4763	5,18	<0,0001
kalender	-0,2042	0,0642	-3,18	0,0025

Hieruit blijkt dat er sprake is van heteroscedasticiteit in de ontwikkelingsjaar- en in de kalenderjaarrichting. Merk op dat een gewogen kleinste kwadraten methode hier geen soelaas biedt, mede aangezien de heteroscedasticiteit zich zowel manifesteert in de ontwikkelingsjaarrichting als in de kalenderjaarrichting. Merk ook op dat er enige sprake is van multicollineariteit. Dit vertaalt zich in het feit dat het maximum van de VIF's (Variance Inflation Factor) gelijk is aan 25,9545. Multicollineariteit kan leiden tot slechte, onstabiele voorspellingen.

Indien we tenslotte de verwachte totale reserve berekenen, bekommen we **25742041**.

Schadereservering is een predictieproces: gegeven de data, poogt men toekomstige claims te voorspellen. Wanneer men de predictiefout van toekomstige betalingen en reserveschattingen tracht te schatten aan de hand van klassieke statistische methoden, reduceert het probleem zich tot het schatten van 2 componenten: enerzijds de procesvariantie en anderzijds de schattingsvariantie. Merk op dat er een verschil bestaat tussen de predictiefout en de standaardfout. Deze laatste is de wortel van de schattingsvariantie. De predictiefout heeft te maken met de variabiliteit van een voorspelling, rekeninghoudend met de onzekerheid in de parameterschatting en met de intrinsieke, onderliggende variabiliteit in de voorspelde data. De RMSEP (root mean squared error of prediction) of predictiefout wordt in de literatuur ook wel soms verkeerdelijk de standaardfout genoemd. De MSEF wordt gegeven door volgende formule:

$$MSEP[\hat{Y}_{ij}] = E[(Y_{ij} - \hat{Y}_{ij})^2] \approx Var[Y_{ij}] + Var[\hat{Y}_{ij}]$$

De predictievariantie van toekomstige incrementele claims \hat{Y}_{ij} in het lognormale model wordt gegeven door:

$$MSEP[\hat{Y}_{ij}] = \hat{Y}_{ij}^2 (\exp(\hat{\sigma}_{ij}^2) - 1),$$

hetgeen overeenkomt met de standaardvorm van de variantie van een lognormale verdeling. Merk op dat $\hat{\sigma}_{ij}^2$ gegeven wordt door uitdrukking (2). De gekwadrateerde predictiefout van de totale reserve wordt in een lognormale setting gegeven door

$$MSEP[R] \approx \sum_{i,j} MSEP[\hat{Y}_{ij}] + 2 \sum_{i,j,k,l, ij \neq kl} \hat{Y}_{ij} \hat{Y}_{kl} (\exp(Cov[\hat{\eta}_{ij}, \hat{\eta}_{kl}]) - 1)$$

Indien we deze formule toepassen op bovenstaande dataset voor het lognormaal model, dan bekomen we een waarde van 9493554. Dit is ongeveer 37% van de totale reserve.

De gebreken aan bovenstaand model kunnen we verhelpen door over te gaan naar de familie van Gegeneraliseerde Lineaire Modellen (GLM's). Deze klasse veralgemeent op een natuurlijke wijze het klassieke lineaire regressiemodel. Heel wat auteurs onderstreepten reeds recentelijk de noodzaak van een dergelijke modellering in een schadereserveringstoepassing. Zie hiervoor onder andere England en Verrall (2002), Pinheiro, Andrade e Silva en Centeno (2002), Hoedemakers, Beirlant, Goovaerts en Dhaene (2003), Despeyroux, Lévi, Partrat en Vignancour (2003), Taylor, McGuire en Greenfield (2003),... In wat volgt geven we een kort overzicht van deze klasse lineaire modellen in de reserveringscontext.

De structuur van een dergelijk lineair model wordt gegeven door:

(1) $Y_{ij} \sim f(y; \mu_{ij}, \phi)$ met onafhankelijke Y_{ij} , $\mu_{ij} = E(Y_{ij})$ en waarbij $f(\cdot)$, de dichtheidsfunctie van Y_{ij} behoort tot de exponentiële familie van verdelingen. ϕ is een schaalparameter.

(2) $\eta_{ij} = g(\mu_{ij})$, $g(\cdot)$ wordt de linkfunctie genoemd.

(3) $\eta_{ij} = (X\vec{\beta})_{ij}$, X de regressiematrix corresponderend met de bovendriehoek.

De maximum likelihood schatters van de regressie parameters kunnen verkregen worden door de 'iteratively reweighted least squares' methode toe te passen, hetgeen de gewone least squares voor de normale lineaire regressiemodellen op een natuurlijke wijze uitbreidt.

Wanneer we een gegeneraliseerd lineair model definiëren, kunnen we de verdeling van Y_{ij} weglaten en enkel de variantiefunctie specificeren. De parameters worden dan geschat met behulp van maximum quasi-likelihood technieken in plaats van de traditionele maximum likelihood. De schatters blijven consistent. In dit geval vervangen we de verdelingsassumptie door $Var(Y_{ij}) = \phi V(\mu_{ij})$, waarbij $V(\cdot)$ de variantie functie is. Voor de normale verdeling is

$V(\mu_{ij}) = 1$, voor de Poisson verdeling (eventueel "over-dispersed", wanneer $\phi > 1$) is

$V(\mu_{ij}) = \mu_{ij}$ en voor de gamma verdeling hebben we dat $V(\mu_{ij}) = \mu_{ij}^2$. Merk op dat het

gebruik van de logaritmische linkfunctie in combinatie met een quasi-likelihood benadering impliceert dat de som van de incrementele claims in elke kolom én in elke rij positief moet zijn. Wanneer we met gamma (Poisson) modellen werken moet noodzakelijkerwijze elke incrementele waarde positief (niet-negatief) zijn. Kortom quasi-likelihood schatting in een GLM-kader laat ons toe de responsvariabele te modelleren in een regressie context zonder de verdeling ervan te specificeren. Deze eenvoudige en robuuste methode gebruikt enkel de meest elementaire informatie over de responsvariabele, nl. de gemiddelde-variantie relatie. Deze informatie alleen is veelal voldoende om consistente maximum likelihood schatters te verkrijgen.

Het meest aangewezen model in dit voorbeeld is een quasi-gamma model gegeven door volgende structuur

$$E[Y_{ij}] = \mu_{ij}$$

$$Var[Y_{ij}] = \phi \mu_{ij}^2$$

$$\log(\mu_{ij}) = \eta_{ij}$$

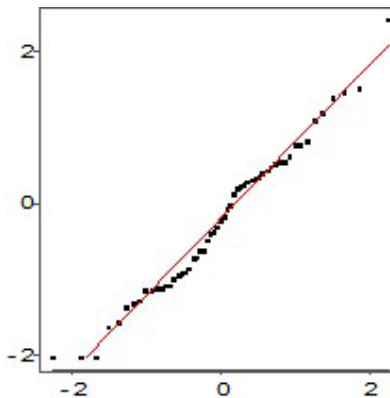
met η_{ij} , de lineaire predictor, gegeven door formule (4).

In dit geval krijgen we als SAS-output voor de parameterschattingen:

	estimate	s.e.	chi-sq	p-value
α_1	11,083	0,5234	448,2927	<0,0001
α_2	10,2105	0,6221	269,3822	<0,0001
α_3	11,2168	0,6296	317,3543	<0,0001
β_1	1,2155	0,3349	13,1763	0,0003
β_2	-0,8898	0,3136	8,0518	0,0045
β_3	-1,7014	0,4507	14,2501	0,0002
γ_1	1,6793	0,6147	7,4632	0,0063

Twee veel gebruikte goodness-of-fit maten in een GLM-context zijn de deviantie en de Pearson χ^2 -statistiek. Beide maten zijn in het normale geval gelijk aan de 'residual sum of squares'. De deviantie is hier gelijk aan 36,1011 en de Pearson χ^2 -statistiek heeft een waarde van 30,0704. Zowel de deviantie als de Pearson χ^2 -statistiek zijn, onder de aanname dat het model correct is, χ^2 -verdeeld met $n-p$ vrijheidsgraden (n =steekproefgrootte, p =aantal parameters). Aangezien beide statistieken in waarde kleiner zijn dan $\chi^2_{1-0.05;48} = 65,17077$ mogen we besluiten, met een betrouwbaarheid van 95%, dat het gefitte model zeer adequaat is. De criteria AIC en BIC hebben respectievelijk een waarde van 50,1011 en 64,1524.

Een normale qq-plot van de gestandaardiseerde deviantie residuen wordt onderstaand weergegeven. Merk op dat Pearson residuen de scheefheid van de onderliggende verdeling neigen weer te geven. Gestandaardiseerde deviantie residuen zijn benaderend normaal verdeeld wanneer het gefitte gegeneraliseerde lineaire model geschikt is.



De verwachte totale reserve is in dit geval gelijk aan **18449821**, hetgeen opmerkelijk minder is dan bij het lognormale model. Merk op dat de vertekeningcorrectie die bij een lognormale aanpak nodig is, wegvalt bij een GLM-modellering aangezien hier de niet-cumulatieve data zelf als responsvariabelen gebruikt worden, terwijl men in een lognormaal kader werkt met de logaritme van de incrementele data als responsvariabelen. Hieruit vloeit onmiddellijk voort dat een schatting van de verwachte reserves stijgt naarmate de onderliggende variantiecomponent groter wordt.

De gekwadrateerde predictiefout voor individuele predicties en voor de totale reserve wordt hier gegeven door

$$MSEP[Y_{ij}] \approx \phi \hat{\mu}_{ij}^{\rho} + \hat{\mu}_{ij}^2 Var[\hat{\eta}_{ij}]$$

$$MSEP[R] \approx \sum_{i,j} \phi \hat{\mu}_{ij}^2 + \sum_{i,j} \hat{\mu}_{ij}^2 Var[\hat{\eta}_{ij}] + 2 \sum_{i,j,k,l \text{ } ij \neq kl} \hat{\mu}_{ij} \hat{\mu}_{kl} Cov[\hat{\eta}_{ij}, \hat{\eta}_{kl}]$$

De RMSEP voor de totale reserve is in dit geval 5070881, hetgeen slechts 27% van de reserve is.

Een 97,5% betrouwbaarheidsgrens, gebaseerd op de normale verdeling en de predictiefout (als zijnde de standaardafwijking van de predictieverdeling), is in het lognormale geval gelijk aan **44349065** en in het quasi-gamma geval gelijk aan **28388565**. Tot slot vergelijken we de verwachte reserves onder het lognormale model en het quasi-gamma model voor de verschillende rijtotalen.

AY	Lognormaal	Quasi-gamma
1	0	0
2	120.089	88.134
3	90.072	73.665
4	135.108	110.498
5	1.319.015	854.561
6	2.163.442	1.406.850
7	3.007.869	1.959.138
8	4.700.287	3.303.737
9	6.392.705	4.648.336
10	8.085.123	6.391.701
Tot.	25.742.041	18.449.821

Opmerking:

Voor een goede beschrijving van de verscheidenheid aan modellen en toetsen wordt de lezer verwezen naar SAS ([8]), waar men de wetenschappelijke onderbouw vindt eerder dan een poging (verdienstelijk overigens) om elementaire statistische begrippen op commerciële wijze voor te stellen als verkoopsargument. Men dient de doorgedreven opleiding van de actuaris naar waarde te schatten

Besluit: Vaak voldoen de data niet rechtstreeks aan de vooropgestelde modelassumpties: het tegelijk gelden van symmetrie (normaliteit), constante variantie en additiviteit van systematische effecten is veelal geen realistische aanname. Om te proberen daar toch aan te voldoen worden vaak de data getransformeerd, maar in de schadeactuariële praktijk heeft dat niet altijd het beoogde effect. Transformaties hebben verder als nadeel dat een eenvoudige interpretatie van de parameters bemoeilijkt wordt. Mathematisch-statistische optimaliteitseigenschappen gaan bovendien verloren bij de terugtransformatie. De vertekeningcorrectie die zich, na een logaritmische transformatie van de responsvariabele, opdringt impliceert een extra variantiecomponent in de uitdrukking voor de predictie van een individuele claim. Het effect van deze correctie hangt samen met de onderliggende variabiliteitsgraad. Deze vertekeningcorrectie heeft uiteraard ook zijn impact op een extensieve simulatie gebaseerd op dit lognormale model waarbij extreme observaties de verwachte waarde

sterk beïnvloeden. Aan de hand van een voorbeeld hebben we in deze bijdrage getracht het explosierisico dat samengaat met het lognormale model uit te leggen. Simuleren uit een dergelijk model, als een methode om de predictieverdeling te bekomen, resulteert in totaal waardeloze en onbruikbare resultaten. Zoals uiteengezet in vorig nummer, kan men ondergrenzen en bovengrenzen bepalen aan de hand van de theorie van de comonotonie. De schatting van percentielen buiten de comonotone grenzen kan zeer in vraag gesteld worden zeker daar de boven- en de benedengrens kunnen uitgedrukt worden met behulp van comonotone vectoren. Raar, maar waar.

Referenties

- [1] Christofides, S. (1990). Regression models based on log-incremental payments. *Claims Reserving Manual*, 2, Institute of Actuaries, London.
- [2] Despeyroux, A., Lévi, C., Partrat, C. en Vignancour, J. (2003). Techniques for valuation of a general insurance company within the framework of IAS standards : some proposals. *Proceedings of the Seventh International Congress on Insurance: Mathematics and Economics, Lyon*.
- [3] England, P.D. en Verrall, R.J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3), 443-518.
- [4] Hoedemakers, T., Beirlant, J., Goovaerts M.J. en Dhaene J. (2003) Claims reserving using generalized linear models. *Onderzoeksrapport 0332 TEW, K.U.Leuven*.
- [5] Pinheiro, P.J.R., Andrade e Silva, J. en Centeno, M. (2002) Bootstrap methodology in claim reserving. *The Journal of Risk and Insurance*, 70(4), 701-720.
- [6] Taylor, G., McGuire, G. en Greenfield, A. (2003). Loss reserving: past, present and future. *Onderzoeksrapport 109, University of Melbourne*.
- [7] Verrall, R.J. (1991). On the estimation of reserves from loglinear models. *Insurance: Mathematics & Economics*, 10(1), 75-80.
- [8] Zie <http://v8doc.sas.com/sashtml/>