1989 General Insurance Convention

The application of additive and multiplicative Generale Linear Interactive Models (GLIM) in Health Insurance.

Paper for the XXI ASTIN Colloquium
New York, 1989

subject: Empirical Investigations.

Bob J.J. Alting von Geusau
June 1989.

1. Introduction and Summary.

For the private health insurance industry in the Netherlands a model is developped which can describe accurately the interaction between the various influencing factors which determine the size of health claims.
Most of these factors such as

- sex of the insured,
- age of the insured, or
- type of coverage,

can be considered as working multiplicatively on the base claim size.

For other factors this seems less true: it might be that for them an additive model will be more appropriate.

This paper provides a description of an example of a model which originally uses the multiplicative specifications but which is generalised easily to the use of additive factors.
Only the basic technique is given, which means that some major points are still open: for instance when to use the purely multiplicative model and when the mixed one.

The paper is concluded with some examples from real life.

2. The General Linear Interactive Model (GLIM): the multiplicative Gamma-case.

If the claimcost per insured k are symbolized by the variable z_k and the claimcost are supposed to be Gamma distributed, we can define as the density of z_k :

$$(1) \qquad \frac{z_k^{\phi_k - 1} \, e^{-\frac{\phi_k z_k}{\mu_k}} \, \phi_k^{\phi_k} \, \mu_k^{-\phi_k}}{\Gamma(\phi_k)}$$

The parameters μ_k and ϕ_k in (1) are also visible of course in

(2) the expectation μ_k

the variance $\dfrac{\mu_k^2}{\phi_k}$

In principle both sets of parameters can be of a very complicated structure, but in multiplicative GLIM we restrict ourself to the μ_k according to

$$(3) \qquad \mu_k = e^{\vec{x}_k^{\,t} \vec{\beta}}$$

In formula (3) $\vec{x}_k$ is a vector with zeros and ones in such a way that the correct linear combination of elements from the rating vector $\vec{\beta}$ is obtained.

The second restriction affects ϕ_k :

$$(4) \qquad \phi_k = \phi \text{ for all } k,$$

where Φ is the scale factor in the Gamma distribution: since at the moment we are only interested in the average value (3) we donot have to estimate the value of this parameter ϕ.

In order to estimate the values of the elements of $\vec{\beta}$ we use the sample $z_1, \ldots, z_k, \ldots, z_n$
and the log-likelihood function

$$(5) \qquad \log Л = \text{constant} + \sum_{k=1}^{n} \{\phi_k \log z_k - \log\Gamma(\phi_k)\} + \sum_{k=1}^{n} \phi_k \log\phi_k + \sum_{k=1}^{n} \{-\frac{\phi_k z_k}{\mu_k} - \phi_k \log\mu_k\} = \quad \text{since (3) and (4) holds}$$

$$= \text{constant} + \text{function}(\phi, z_k) + \phi\{\sum_{k=1}^{n} -z_k e^{-\vec{x}_k^t \vec{\beta}} - \sum_{k=1}^{n} \vec{x}_k^t \vec{\beta}\}$$

As a function of $\vec{\beta}$ (5) has its maximum when

$$(6) \qquad \frac{\partial \log Л}{\partial \beta_m} = \phi \{- \sum_{k=1}^{n} x_{km} + \sum_{k=1}^{n} x_{km} z_k e^{-\vec{x}_k^t \vec{\beta}}\} = 0$$

for $m=1,\ldots,M$, where x_{km} is element number m in the vector $\vec{x}_k$.

The consequence of this definition is that for each claim we cannot have more than M rating factors.

Formula (6) is a system of M equations with M unknowns $\beta_1, \beta_2, \ldots, \beta_M$ but since this system is not linear we need a little bit more than some basic knowledge of matrix algebra.

Van Eeghen, Greup and Nijssen in "Ratemaking" (1983) describe the elegant method of scoring as a way to solve (6) as follows:

Define the information matrix $I(\vec{\beta})$ by:

(7) $$I(\vec{\beta}) = -\text{Expectation}\{\frac{\partial^2 \log \Pi}{\partial\vec{\beta}\,\partial\vec{\beta}^t}\}$$

This matrix is negative definite for the values

(8) $$\vec{\beta} = \hat{\vec{\beta}}_n$$ where $\hat{\vec{\beta}}_n$ is the n^{th} approximation of the solution $\vec{\beta}$ of (6).

The method of scoring finally solves (6) by the iterative procedure

(9) $$\hat{\vec{\beta}}_{n+1} = \hat{\vec{\beta}}_n + I(\vec{\beta})^{-1}_{\vec{\beta}=\hat{\vec{\beta}}_n}\left(\frac{\partial \log \Pi}{\partial\vec{\beta}}\right)_{\vec{\beta}=\hat{\vec{\beta}}_n}$$

Formula (9) converges quite easily when the chosen initial value

(10) $$\hat{\vec{\beta}}_0$$

is not too unreasonable.

It makes the situation easier that in this specific case (7) boils down to

(11) $$I(\vec{\beta}) = \sum_{k=1}^{n} \vec{x}_k \, \vec{x}_k^{\,t}$$

for each $\vec{\beta}=\vec{\beta}_n$ which means that it is necessary to compute the inverse matrix of (11) only once.

3. GLIM generalised to the mixed additive-multiplicative case.

In health insurance it may be necessary to generalise (3) to

(12) $$\mu_k = \vec{y}_k^{\,t} \, \vec{\alpha} + e^{\vec{x}_k^{\,t} \vec{\beta}} = N_k \qquad \text{(say)}$$

for the observations z_k with k=1,..,n where:

- $\vec{y}_k$ the specific vector for observation k with zeros and ones for the additive factors, and
- $\vec{\alpha}$ the vector with the A additive factors.

As an example of such an additive factor the deductible per insured can be considered.

For the log-likelihood function this means:

(13) $$\log \Pi = \text{constant} + \sum_{k=1}^{n} \text{function}(\phi, z_k) +$$

$$+ \phi\{ \sum_{k=1}^{n} -z_k N_k^{-1} - \sum_{k=1}^{n} \log N_k \}$$

Therefore (6) becomes a system of (M+A) equations with (M+A) unknowns:

$$(14)\qquad \frac{\partial \log \Pi}{\partial \alpha_a} = \phi\{\sum_{k=1}^{n} z_k y_{ka} N_k^{-2} - \sum_{k=1}^{n} y_{ka} N_k^{-1}\} \qquad \text{for } a=1,..,A$$

$$\text{and}\quad \frac{\partial \log \Pi}{\partial \beta_m} = \phi\{\sum_{k=1}^{n} z_k x_{km} e^{\vec{x}_k^t \vec{\beta}} N_k^{-2} - \sum_{k=1}^{n} x_{km} e^{\vec{x}_k^t \vec{\beta}} N_k^{-1}\} \qquad \text{for } m=1,..,M.$$

which again cannot be solved directly.

With scoring it takes some tedious calculations but finally we find for the information matrix the composition of the submatrices:

for a=1,..,A and b=1,..,A and for m=1,..,M and p=1,..,M of the matrices given by:

$$(15)^{*}\qquad -\text{Expectation}\{\frac{\partial^2 \log \Pi}{\partial \alpha_a\, \partial \alpha_b}\} = \sum_{k=1}^{n} y_{ka} y_{kb} N_k^{-2}$$

$$(15)^{**}\qquad -\text{Expectation}\{\frac{\partial^2 \log \Pi}{\partial \alpha_a \partial \beta_m}\} = \sum_{k=1}^{n} x_{km} y_{ka} e^{\vec{x}_k^t \vec{\beta}} N_k^{-2}$$

$$(15)^{***}\qquad -\text{Expectation}\{\frac{\partial^2 \log \Pi}{\partial \beta_m \partial \beta_p}\} = \sum_{k=1}^{n} x_{km} x_{kp} e^{2\vec{x}_k^t \vec{\beta}} N_k^{-2}$$

which totalises into the information matrix by the recipe:

$$(16) \qquad I(\vec{\alpha},\vec{\beta}) = \left[\begin{array}{c|c} (15)^{***} & (15)^{**} \\ \hline (15)^{**} & (15)^{*} \end{array}\right]$$

and leads finally to the scoring formula:

$$(17) \qquad \begin{pmatrix} \vec{\beta}_{n+1} \\ \vec{\alpha}_{n+1} \end{pmatrix} = \begin{pmatrix} \vec{\beta}_{n} \\ \vec{\alpha}_{n} \end{pmatrix} + I(\vec{\alpha},\vec{\beta})^{-1} \begin{pmatrix} \dfrac{\partial \log \Pi}{\partial \vec{\beta}} \\ \dfrac{\partial \log \Pi}{\partial \vec{\alpha}} \end{pmatrix} \quad \text{with } \vec{\alpha}=\vec{\alpha}_n \text{ and } \vec{\beta}=\vec{\beta}_n$$

The convergence of (17) is much less convincing but in the cases I have analysed the solution came up after some time: the number of necessary iterations appeared to be about three times as high as in the purely multiplicative case.

This is the price we have to pay for the fact that (16) is not strictly negative definite in all cases.

4. Some results on real data.

The methods mentioned in the previous paragraphs have been applied to the 1987 and the 1988 data from private health insurance companies in the Netherlands.

This data has been collected by the Foundation KISG from 3.5 million insureds with regards to the following activities:

- the cost of hospitalisation
- the cost of a specialist in a clinic
- the cost of a specialist without hospitalisation
- the cost of paramedical aid
- the cost of other medical care, such as medical transportation, artificial legs and arms, dentures, obstetrical aid, home care, etcetera.

Of these activities the gross and the net amounts per insured per year have been collected plus the information related to the the individual situation of the insured with regards to:

- hospitalisation 3rd class or better
- sex of the insured
- area where the insured is living: the highly populated Western part of the Netherlands is much more expensive than the North and the South
- the age of the insured
- the level of the deductible
- the type of health insurance contract: in the Netherlands one can buy private health insurance through an individual contract or in some cases collectively via the employer in a group contract
- the fact whether the cost of the family doctor is insured yes or no: it seems that this factor indicates a totally different type of risk.

On this information I have applied the general GLIM-models in a score of variants of which I give three examples.

The first one is a purely multiplicative one with compound age factors:

			Year 1987	Year 1988
Base Claim per year net		Dfls	667	664
Multiplicative factors in %:				
Hospitalisation Class:	3		100	100
	1 & 2		114	112
Sex	male		100	100
	female		99	100
Area:	average		100	100
	cheap		91	96
	expensive		105	109
Age:	0- 4		109	112
	5- 9		51	51
	10-14		66	70
	15-19		64	65
	20-24 male		54	53
	idem female		87	82
	25-29 male		49	50
	idem female		158	156
	30-34 male		58	59
	idem female		161	161
	35-39 male		71	69
	idem female		121	123
	40-44		100	100
	45-49		120	120
	50-54		147	147
	55-59		179	185
	60-64		231	238
	65-69		326	333
	70-74		412	434
	75+		548	578
Deductible:	low		100	100
	average		95	88
	high		58	57
Contract:	individual		100	100
	collective		105	103
Family Doctor:	included		100	100
	not included		84	83

Table 1. Estimation of the base claim per year and the multiplicative factors in % on the KISG 1987 and 1988 deck, with the introduction of separate multiplicative factors to sex for some ages.

An example of how table 1 works:

let us take a gentleman of over 75:

base claim	fl. 667
multiplicative factors:	
* 2nd classe	1,14
* male	1
* expensive area	1,05
* age 75+	5,48
* low deductible	1
* individually insured	1
* family doctor included	1
	-----------
total nett claim per year	fl. 4375

In a high deductible scenario the table let this amount shrink to 0,58 * 4375 = 2538, which is around two thousand guilders less.

This difference seems to be too large.
So in table 2 we use the mixed GLIM where the deductibles compounded with some age factors are tagged as additive factors.

Deductible		Age for male and female:			
		0-19	20-39	40-59	60+
low: under fl. 250 per year					
	1987:	180	173	205	836
	1988:	189	22	249	729
average: fl. 250 - 750 per year					
	1987:	178	81	178	1068
	1988:	150	73-	155	547
high: over fl. 750 per year					
	1987:	0	22-	116-	324
	1988:	0	178-	91-	10

Table 2a. Additive constants in the mixed GLIM-model for the claimfactor 'Deductible' on KISG 1987 and 1988: these constants can only be used in connection with table 2b.

		Year 1987	Year 1988
Base claim per year	Dfls.	477	445
Multiplicative factors in %:			
Hospitalisation Class:	3	100	100
	1 & 2	118	115
Sex:	male	100	100
	female	101	101
Area:	average	100	100
	cheap	89	95
	expensive	106	108
Age:	0- 4	102	112
	5- 9	37	39
	10-14	51	59
	15-19	50	54
	20-24 male	46	83
	idem female	84	118
	25-29 male	39	77
	idem female	169	211
	30-34 male	50	88
	idem female	167	214
	35-39 male	64	100
	idem female	120	164
	40-44	100	100
	45-49	121	121
	50-54	154	157
	55-59	196	208
	60-64	131	195
	65-69	254	323
	70-74	369	464
	75+	560	674
Contract:	individual	100	100
	collective	106	104
Family Doctor:	included	100	100
	not included	79	80

Table 2b. Estimation of the base claim in guilders per year and the multiplicative factors in % on KISG 1987 and 1988: this table can only be used in connection with table 2a.

Back to our example to demonstrate how both tables are used together:

base claim 1987	fl. 477
multiplicative factors:	
* 2e class	1,18
* male	1
* expensive area	1,06
* age 75+	5,60
* individual cover	1
* family doctor included	1
sub result	fl. 3341
additive factor:	
+ low deductible, age 75+	836
total net claim in 1987	fl. 4177

With a high deductible the amount would have been fl. 3665 - that is 512 guilders lower - and that seems more plausible. This better fit is also indicated by the lower mean squared error: in table 2a/2b this reads 685, while in table 1 it is 734.

Finally I have listed table 3, where the multiplicative GLIM is demonstrated on a separate insurance activity: the cost of hospitalisation in 1987.

The introduction of separate age/sex factors leads to interesting results:

Base claim in 1987		283	
Multiplicative factors in %:			
Hospitalisation class:	3	100	
	1 & 2	113	
Sex:	male	100	
	female	90	
Area:	average	100	
	cheap	91	
	expensive	103	

		male	female
Age:	0 - 4	189	
	5 - 9	60	
	10 - 14	56	
	15 - 19	82	
	20 - 24	91	
	25 - 29	61	184
	30 - 34	66	191
	35 - 39	81	153
	40 - 44	100	164
	45 - 49	131	196
	50 - 54	184	221
	55 - 59	253	257
	60 - 64	358	347
	65 - 69	570	488
	70 - 74	739	668
	75+	1007	1024

Deductible:	low	100
	average	98
	high	65
Contract:	individual	100
	group	102
Family doctor:	included	100
	not included	82
Standaarddeviation MSE:		605

Table 3. Estimation of the base claim per year in guilders and the multiplicative factors in % for the cost of hospitalisation in 1987 with the introduction of separate multiplicative factors to sex for some ages.

5. Resume.

Pour la description d'un portefeuille Couts Medicaux la methode GLIM (General Linear Interactive Model) peut donner des instruments utilisables. Il parait possible de generaliser cette méthode avec des variantes mixtes du genre additif-multiplicatif.

Finallement le papier donne quelques exemples.

Bob J.J. Alting von Geusau,
Juin 1989 - June 1989.