ON THE IMPORTANCE OF DISPERSION MODELING FOR CLAIMS RESERVING: AN APPLICATION WITH THE TWEEDIE DISTRIBUTION

2012 CAS SPRING MEETING, Phoenix

Jean-Philippe Boucher (and Danaïl Davidov)

*Quantact / Département de mathématiques
Université du Québec à Montréal, Canada
http://www.quantact.uqam.ca/en/jpboucher.html

Modeling Run-off Triangles

We use classic run-off triangles such as:

- Chain-Ladder;
- Bornhuetter-Ferguson;
- London-Chain;
- etc.

Instead of using these deterministic methods, we used stochastic methods which allow us to compute the distribution of the reserve, and for example the variance.

A standard stochastic approach is Mack's model, where non-parametric modeling is used.

In our work, we prefered parametric model. In our case, it supposes than each payment $C_{i,j}$, with i the line number and j the colum number, is following a specific distribution. We used the following parametrization:

$$l_i(\text{line}) = \begin{cases} 1 & \text{if the payment occured in year } i \\ 0 & \text{if not} \end{cases}$$

$$c_j(\text{column}) = \begin{cases} 1 & \text{if the payment occured in development year } j \\ 0 & \text{if not} \end{cases}$$

Modeling Run-off Triangles (example)

In practice, it means that a run-off triangle will be modeled as:

Year	1	2	3	4
1	100	125	75	50
2	200	225	175	
3	325	335		
4	350			

$C_{i,j}$	Value	X_0	c_2	c_3	c_4	l_2	l_3	l_4
C_{11}	100	1	0	0	0	0	0	0
C_{12}	125	1	1	0	0	0	0	0
C_{13}	75	1	0	1	0	0	0	0
C_{14}	50	1	0	0	1	0	0	0
C_{21}	200	1	0	0	0	1	0	0
C_{22}	225	1	1	0	0	1	0	0
C_{22}	175	1	0	1	0	1	0	0
C_{31}	325	1	0	0	0	0	1	0
C_{32}	335	1	1	0	0	0	1	0
C_{41}	350	1	0	0	0	0	0	1

Poisson regression

If we suppose that each payment $C_{i,j}$ (i.e. each value in a single cell) is following a Poisson distribution, such as:

$$\Pr(C_{ij} = c_{ij}) = \frac{\mu_{ij}^{c_{ij}} e^{-\mu_{ij}}}{c_{ij}!}$$

with the mean parameter modeled as:

$$\mu_{i,j} = g(\gamma X_0 + \boldsymbol{\alpha}' \boldsymbol{L} + \boldsymbol{\beta}' \boldsymbol{C}),$$

it can be shown that MLE of the parameter $\boldsymbol{\theta} = \{\gamma, \boldsymbol{\alpha}, \boldsymbol{\beta}\}$ lead to the same reserves as the Chain-Ladder algorithm (Hachemeister and Stanard (1975), and published in english in Mack(1991)).

Using the same idea, we can then model $C_{i,j}$ with a negative binomial, normal, gamma, inverse-gaussian... or with a Tweedie distribution.

Tweedie Distribution

We use normalized increment payments ($Y_{i,j} = C_{i,j}/\omega_{i,j}$). The Tweedie distribution is based on assumptions:

- The number of payments $N_{i,j}$ is Poisson distributed (with mean $\lambda_{i,j}\omega_{i,j}$);
- Each payment $X_{i,j}^{(k)}$ is gamma distributed with mean $\tau_{i,j}$ and shape parameter ν;
- $C_{i,j} = \sum_{k=1}^{N_{i,j}} X_{i,j}^{(k)}$.

Then, as shown in Jorgensen(1997), the mean and variance of $Y_{i,j}$ are given by:

$$\begin{aligned} \mathrm{E}[Y_{i,j}] &= \mu_{i,j} \left(= \kappa_p'(\theta_{i,j}) = \frac{\partial(\kappa_p(\theta_{i,j}))}{\partial\theta_{i,j}} \right), \\ \mathrm{Var}[Y_{i,j}] &= \frac{\phi_{i,j}}{w_{i,j}} \mu_{i,j}^p \left(= \frac{\phi_{i,j}}{w_{i,j}} \kappa_p''(\theta_{i,j}) \right). \end{aligned}$$

with $p = (\nu + 2)/(\nu + 1)$ and $\mu_{i,j} = \exp(X_{i,j}\beta)$.

Tweedie Distribution

The loglikelihood can then be expressed as:

$$l = \sum_{i,j} \left(\log \left(c \left(y_{i,j}; \frac{w_{i,j}}{\phi_{i,j}}; p \right) \right) + \frac{w_{i,j}}{\phi_{i,j}} \left(y_{i,j} \frac{\mu_{i,j}^{1-p}}{1-p} - \frac{\mu_{i,j}^{2-p}}{2-p} \right) \right) .$$

The theory of GLM can be used for inference.

Tweedie Distribution

The dispersion parameter can be estimated in at least two ways. The first approach is the maximum likelihood estimator. The second approach uses the deviance principle. Both methods generate a constant $\phi_{i,j} \equiv \phi$.

$$\phi_{i,j} \equiv \phi \quad = \quad \frac{-\sum_{i,j} w_{i,j} \left(y_{i,j} \frac{\mu_{i,j}^{1-p}}{1-p} - \frac{\mu_{i,j}^{2-p}}{2-p} \right)}{(1+\nu)\sum_{i,j} r_{i,j}} ,$$

$$\phi_{i,j} \equiv \phi \quad = \quad \sum_{i,j} \frac{2}{N-Q} \left(y_{i,j} \frac{y_{i,j}^{1-p} - \mu_{i,j}^{1-p}}{1-p} - \frac{y_{i,j}^{2-p} - \mu_{i,j}^{2-p}}{2-p} \right) .$$

Other details can be found in our paper, or in Wutrich's paper (2003) who first applied the Tweedie distribution on run-off triangles.

Variance Modeling

The idea is to suppose a dispersion parameter that depends on i and j (lines and columns).

We can justify this modeling by a simple example. Suppose $C = \sum_{k=1}^{N} X_k$, meaning $E[C] = E[N]E[X_k]$ and $Var[C] = E[X_k]^2 Var[N] + E[N] Var[X_k]$, and

	Case 1	Case 2	Case 3
E[N]	10	20	10
Var[N]	10	20	10
E[X_k]	10	10	20
Var[X_k]	100	100	400
E[C]	100	200	200
Var[C]	2000	4000	8000

Dispersion Models

A dispersion model has a flexible variance structure denoted by:

$$\phi_{i,j} = \exp\{Z_{i,j}\gamma\} ,$$

where $\phi_{i,j}$ is the dispersion factor of cell (i,j), $Z_{i,j}$ is the $(i,j)^{th}$ row of the design matrix with the corresponding vector of parameters γ.

We use rows and/or columns to explain the dispersion just as we would for the means.

The maximum likelihood estimates are obtained through direct optimization of the likelihood function:

$$
\begin{aligned}
l &= \sum_{i,j} r_{i,j} \log\left(\frac{(w_{i,j}/\phi_{i,j})^{\nu+1} y_{i,j}^{\nu}}{(p-1)^{\nu}(2-p)}\right) - \log\left(r_{i,j}!\Gamma\left(r_{i,j}\nu\right) y_{i,j}\right) \\
&\quad + \frac{w_{i,j}}{\phi_{i,j}}\left(y_{i,j}\frac{\mu_{i,j}^{1-p}}{1-p} - \frac{\mu_{i,j}^{2-p}}{2-p}\right) .
\end{aligned}
$$

However, we can use other estimation techniques such as the double GLM.

Double GLM

The technique:

- The mean gets optimized using a Tweedie model with a fixed deviance and fixed p;
- Then the deviance-responses are optimized using the saddle-point approximation which supposes that the $d_{i,j}$ are approximately distributed as $\phi_{i,j}\chi_1^2$;
- Because this distribution is a particular case of the gamma distribution (with its own dispersion parameter equaling 2), we can therefore use the gamma model to find a good estimation of $\phi_{i,j}$;
- Finally, the dispersion-prior exposures are inserted back again in the mean sub-model for the next iteration of the algorithm.

Double GLM

On a mathematical point of view, we use iteratively the following equations

$$\beta^{k+1} = (X^T W X)^{-1} X^T W z \ ,$$

$$\gamma^{k+1} = (Z^T W_d Z)^{-1} Z^T W_d z_d \ ,$$

for some W, W_d and z, z_d, depending on γ, β.

Inference using the double GLM or direct optimization of the loglikelihood are equivalent.

REML

It is well known that the maximum likelihood variance estimators are biased downwards (ex: $S^2 = \frac{n}{n-1}\hat{\sigma}^2_{MLE}$)when the number of parameters used to estimate the fitted values is large compared with the number of observations.

In normal linear models, restricted maximum likelihood (REML) is usually used to estimate the variances, and this produces estimators which are approximately and sometimes exactly unbiased.

Note that this correction only targets the estimation of the variances, and thus has a residual effect on the means.

Quantact

REML with Tweedie

The REML technique can be applied on the Tweedie model, using the results of Smyth and Verbyla (1999) or Dunn (2001). Indeed, the DGLM method is still used but the weights W are modified, and the loglikelihood $l()$ of the Tweedie is replaced by $l^*()$:

$$l^*(y,\ \beta,\ \gamma,\ p) \quad = \quad l(y,\ \beta,\ \gamma,\ p) + \frac{1}{2}\log\left|X^T WX\right|,$$

where $l(y,\ \beta,\ \gamma,\ p)$ is the log-likelihood and $\frac{1}{2}\log\left|X^T WX\right|$ is the REML adjustment.

UQÀM
Université du Québec à Montréal

Applied Example

We consider Swiss Motor Industry data as analyzed in Wutrich(2003).

We have observations of incremental paid losses and the number of payments for 9 accident years on a horizon of up to 11 development years.

We also suppose that the exposure is the number of reported claims for each accident year (we suppose that it is sufficiently developed after 2 years).

Incremental payments, number of payments and weights used are:

AY	1	2	3	4	5	6	7	8	9	10	11
1	17841110	7442433	895413	407744	207130	61569	15978	24924	1236	15643	321
2	19519117	6656520	941458	155395	69458	37769	53832	111391	42263	25833	
3	19991172	6327483	1100177	279649	162654	70000	56878	9881	19656		
4	19305646	5889791	793020	309042	145921	97465	27523	61920			
5	18291478	5793282	689444	288626	345524	110585	115843				
6	18832520	5741214	581798	248563	106875	94212					
7	17152710	5908286	524806	230456	346904						
8	16615059	5111177	553277	252877							
9	16835453	5001897	489356								

AY	1	2	3	4	5	6	7	8	9	10	11	$w_{i,j}$
1	6229	3500	425	134	51	24	13	12	6	4	1	112953
2	6395	3342	402	108	31	14	12	5	6	5		110364
3	6406	2940	401	98	42	18	5	3	3			105400
4	6148	2898	301	92	41	23	12	10				102067
5	5952	2699	304	94	49	22	7					99124
6	5924	2692	300	91	32	23						101460
7	5545	2754	292	77	35							94753
8	5520	2459	267	81								92326
9	5390	2224	223									89545

Applied Example

We applied several models, all four with the use of the number of payments:

1. a constant dispersion model (Model I) ;
2. a model that directly optimizes the log-likelihood function (Model II);
3. a double generalized linear model (Model III) ;
4. a double generalized linear model with REML (Model IV).

For the constant dispersion model, (Model I), we replicate the procedure in Wutrich(2003), by using a direct maximum likelihood estimation for μ, ϕ and *p*.

For the variance models (Models II, III and IV) , we believe that the Swiss Motor data might have different trends for the frequency and severity over the development periods, but not in the accident year direction.

Hence, we suppose that only the columns have a direct effect on the dispersion.

For all three of these models, we estimated a variance parameter for each column except for the last one which was regrouped with the second to last column.

Results

AY(i)	R_i	Estimation	Process	MSEP$^{1/2}$
1	-	-	-	-
2	326	1 869	1 861	2 638
3	21 565	15 601	21 795	26 804
4	40 716	19 144	29 962	35 556
5	89 298	25 976	46 538	53 297
6	138 335	30 564	58 556	66 052
7	204 262	35 230	72 833	80 906
8	360 484	45 664	102 268	111 999
9	597 056	61 307	136 903	150 003
Total	1 452 042	180 126	203 658	271 886

Table: Reserve point estimates and MSEP decomposition for Model I

Results

AY (i)	R_i	Estimation	Process	MSEP$^{1/2}$
1	-	-	-	-
2	324	546	550	775
3	21 352	16 978	24 517	29 822
4	40 185	19 994	31 771	37 538
5	87 224	28 118	52 617	59 659
6	138 203	32 871	64 695	72 567
7	202 469	34 772	73 968	81 733
8	359 148	40 833	96 159	104 470
9	596 118	47 064	113 899	123 239
Total	1 445 023	183 285	190 409	264 289

Table: Reserve point estimates and MSEP decomposition for Models II and III

Results

AY (i)	R_i	Estimation	Process	MSEP$^{1/2}$
1	-	-	-	-
2	325	563	568	800
3	21 357	17 044	24 601	29 928
4	40 205	19 914	31 569	37 325
5	87 224	27 665	51 600	58 549
6	138 317	32 261	63 294	71 041
7	202 512	34 032	72 155	79 777
8	359 344	39 826	93 538	101 663
9	596 578	45 830	110 665	119 780
Total	1 445 862	180 470	185 670	258 926

Table: Reserve point estimates and MSEP decomposition for Model IV

Discussion

One can notice that Model IV (REML) produces generally somewhat lower estimates than Models II and III for this particular example.

This seems contrary to the fact that REML tends to correct the ML tendency to under-estimate dispersion.

It turns out that Model IV has also different mean estimates which slightly alter the variance parameters.

Had the mean parameters been the same, then the variance parameters would have been higher with the REML procedure.

Thus, it should be noted that the REML procedure might prove useful as it both corrects the mean parameters (slightly) and the variance parameters.

Conclusion

Allowing for a flexible variance structure does not guarantee that the overall variance in the model will be different.

It is suggested to consider variance modeling when the underlying tendency of the frequency is different from the tendency of the severity.