Integrating ordinary GLM with credibility in a Compound Poisson model

Part 1 claim frequency

Lecture in Indonesia november 2014

Stig Rosenlund

The methods of this talk are implemented in the programming language Rapp. To find Rapp on the internet, search **free actuarial language.**

About me: PhD in Mathematical Statistics at the University of Göteborg in 1983. Actuary at the Swedish insurance company Länsförsäkringar 1983-2010. After retirement in 2010, I have resumed my research, now focusing on insurance mathematics. So far I have in the 2010's had three papers published - two in tariff analysis and one in reserving.

Please interrupt me with questions at any time!

This talk is based on a paper submitted for publication, but not yet accepted. The credibility literature is large, but here I mention just the book by Ohlsson and Johansson - former colleagues of mine at Länsförsäkringar. It is enough to follow my reasoning. I also mention two papers on non-credibility tariff analysis by me, which have some bearing on my credibility results.

Ohlsson, E. & Johansson, B. (2010).
Non-Life Insurance Pricing with Generalized Linear Models.
174 pages + preface. Springer, Berlin.

Rosenlund, S. (2010).
Dispersion estimates for Poisson and Tweedie models.
ASTIN Bulletin **40**(1), 271-279.

Rosenlund, S. (2014).
Inference in multiplicative pricing.
Scandinavian Actuarial Journal **2014**(8), 690-713.

My papers are included in the manual as appendices.

The setup is that we have a number of arguments, each with a finite number of classes, the same as in my lecture on tariff analysis with Rapp, But one of the arguments has so many classes, that the uncertainty for risk premium is very large for many of the classes. And there is no natural ordering of the classes, such that risk premium can be assumed to be monotone, or have at most one minimum or maximum, in the ordering of the classes. So credibility must be used.

The focus of this lecture will be mainly on variance estimates, in particular between-groups variance components.

The argument with many classes will be called the credibility argument. The non-credibility arguments are of two kinds; those that depend functionally on the credibility argument and those that do not. The former are called Auxiliaries. For example, if the credibility argument is geographical parishes, median income and population density are Auxiliaries. Non-auxiliaries are e.g. policy holder age and object age, since a parish has policies in several classes of these.

Like Ohlsson & Johansson I treat non-auxiliaries by letting exposure be normalized duration, defined for an individual object as a number proportional to duration multiplied with estimated multiplicative risk premium for non-auxiliaries. After having stated that, I do not further treat non-auxiliaries, except in the section on iterations. This is as Ohlsson & Johansson do.

My claim is that my analysis gives better results than previous authors have given. For one Auxiliary my results are exact, while for several Auxiliaries they are approximations.

In order to describe the essentials of my work more clearly, I will here assume that there are no non-auxiliary ordinary arguments and that there is one single Auxiliary. You can think of geographical parishes as the credibility argument and population density as the Auxiliary. All mathematical intricacies are present in this setup. I treat claim freqency and mean claim separately in this lecture, concentrating on claim freqency.

Model

The Compound Poisson distribution is assumed for total claim cost for group j = 1, 2, ... J, conditional on stochastic variables

Θ_{Fj} for claim frequency

Θ_{Mj} for mean claim

Namely, conditional on the Θs, the claim numbers are Poisson and the claim severities are IID and independent of the claim numbers.

Furthermore, all Θs are assumed to be independent with expectation 1.

Their standard deviations are τ_F for claim frequency and τ_M for mean claim.

In my submitted paper I refer to my published paper Rosenlund (2010) and to an older reference to justify the Compound Poisson assumption, visàvi the non-distributional treatment of Ohlsson & Johansson.

I will here treat only claim frequency, since the topic would be too large for one lecture if also mean claim was treated. Furthermore I will leave out proof details. Nevertheless a lot of notation and many formulas will be presented.

Notation for observables

We define the following observables, where J, k_{j1}, w_j and w_{j1} are deterministic.

$$
\begin{aligned}
J &= \text{number of groups}, && (1)\\
k_{j1} &= \text{the class of group } j \text{ in its Auxiliary, } j \in \{1, \ldots, J\}, && (2)\\
w_j &= \text{duration group } j, && (3)\\
w_{j1} &= \sum_{i:k_{i1}=k_{j1}} w_i = \text{sum of duration over group } j\text{'s value in its Auxiliary}, && (4)\\
N_j &= \text{number of claims group } j, && (5)\\
N_{j1} &= \sum_{i:k_{i1}=k_{j1}} N_i = \text{total claim number over group } j\text{'s value in its Auxiliary}, && (6)\\
Y_{\mathsf{F}j} &= \frac{N_j}{w_j} = \text{observed claim frequency group } j, && (7)\\
\hat{\mu}_{\mathsf{F}j} &= \frac{N_{j1}}{w_{j1}} = \text{estimated collective claim frequency group } j \text{ for its Auxiliary.} && (8)
\end{aligned}
$$

All j with the same k_{j1} have the same w_{j1}, N_{j1} and $\hat{\mu}_{\mathsf{F}j}$. For example, if the Auxiliary is population density, with four classes Low, Medium Low, Medium High, High, and the parish j falls in the Medium High category, then $k_{j1} = 3$ and the parish has the same estimated collective claim frequency as other parishes with Medium High population density.

Assumptions

We formulate two assumptions.

A1F. Conditional on stochastic variables $\Theta_{\mathsf{F}j}$ $(j = 1, \ldots, J)$, with expectation $\mathrm{E}[\Theta_{\mathsf{F}j}] = 1$ and variance $\mathrm{Var}[\Theta_{\mathsf{F}j}] = \tau_{\mathsf{F}}^2$, N_j is Poisson distributed with expectation $\mu_{\mathsf{F}j}\Theta_{\mathsf{F}j}w_j$, where $\mu_{\mathsf{F}j}$ is determined by the Auxiliary of j.

A2F. $(\Theta_{\mathsf{F}1}, N_1), \ \ldots \ , (\Theta_{\mathsf{F}J}, N_J)$ are independent.

It follows that $\mathrm{E}[\hat{\mu}_{\mathsf{F}j}] = \mu_{\mathsf{F}j}$.

Objective. To predict $\mu_{\mathsf{F}j}\Theta_{\mathsf{F}j}$ as well as possible.

We use the words predict and predictor, by the convention to reserve the word estimator for guesses on non-stochastic parameters (the Bayesian formulation set aside). Our predictors are guesses on stochastic variable outcomes that occurred in the past and can never be observed.

We now recapitulate parameters and functionals given above and define new ones.

$$\mu_{\mathsf{F}j} = \mathrm{E}[Y_{\mathsf{F}j}] \quad \text{claim frequency group } j, \tag{9}$$

$$\Lambda_{\mathsf{F}j} = \mu_{\mathsf{F}j}\Theta_{\mathsf{F}j} \quad \text{claim frequency group } j \text{ conditional on } \Theta_{\mathsf{F}j}, \tag{10}$$

$$\tau_{\mathsf{F}}^2 = \mathrm{Var}[\Theta_{\mathsf{F}j}] \quad \text{assumed independent of } j, \tag{11}$$

$$\nu_{\mathsf{F}j}^2 = \frac{1}{{w_{j1}}^2} \sum_{i:k_{i1}=k_{j1}} w_i^2(\mu_{\mathsf{F}i}/w_i + \mu_{\mathsf{F}j}^2\tau_{\mathsf{F}}^2) \text{ equal to } \mathrm{Var}[\hat{\mu}_{\mathsf{F}j}]. \tag{12}$$

Our predictors and estimators are

$$\Lambda_{\mathsf{F}j}^* = z_{\mathsf{F}j}Y_{\mathsf{F}j} + (1 - z_{\mathsf{F}j})\mu_{\mathsf{F}j} \quad \text{non-observable predictor of } \Lambda_{\mathsf{F}j}, \text{ see (17)}, \tag{13}$$

$$\widehat{\Lambda}_{\mathsf{F}j} = \widehat{z}_{\mathsf{F}j}Y_{\mathsf{F}j} + (1 - \widehat{z}_{\mathsf{F}j})\hat{\mu}_{\mathsf{F}j} \quad \text{estimated predictor of } \Lambda_{\mathsf{F}j}, \text{ see (23)}, \tag{14}$$

$$\hat{\tau}_{\mathsf{F}}^2 = \text{pseudo-estimator of } \tau_{\mathsf{F}}^2, \text{ see (22)}, \tag{15}$$

$$\hat{\nu}_{\mathsf{F}j}^2 = \frac{1}{{w_{j1}}^2} \sum_{i:k_{i1}=k_{j1}} w_i^2(\hat{\mu}_{\mathsf{F}i}/w_i + \hat{\mu}_{\mathsf{F}j}^2\hat{\tau}_{\mathsf{F}}^2) \text{ estimator of } \mathrm{Var}[\hat{\mu}_{\mathsf{F}j}]. \tag{16}$$

Note that we could write $\mu_{\mathsf{F}j}$ instead of $\mu_{\mathsf{F}i}$ in (12) and $\hat{\mu}_{\mathsf{F}j}$ instead of $\hat{\mu}_{\mathsf{F}i}$ in (16), since they are equal, respectively.

Best linear predictor of the credibility factor

For pricing we wish to use a predictor of $\mu_{\mathsf{F}j}\Theta_{\mathsf{F}j}$ as a claim frequency factor for the value j of the credibility argument. It consists of the collective claim frequency $\mu_{\mathsf{F}j}$, the same for all groups with the same value of the Auxiliary, and the specific factor $\Theta_{\mathsf{F}j}$.

We seek the best linear predictor of $\mu_{\mathsf{F}j}\Theta_{\mathsf{F}j}$ in L^2-norm in the form (13), i.e. $z_{\mathsf{F}j}$ is to be determined so that $\mathrm{E}[(\Lambda^*_{\mathsf{F}j} - \mu_{\mathsf{F}j}\Theta_{\mathsf{F}j})^2]$ is minimized. Then $\Lambda^*_{\mathsf{F}j}$ is the BLP, provided the BLP has the form (13). It is shown that this form holds.

Theorem 1. *For the BLP in L^2-norm of $\mu_{\mathsf{F}j}\Theta_{\mathsf{F}j}$ it holds*

$$z_{\mathsf{F}j} = \frac{\mathrm{Var}[\hat{\mu}_{\mathsf{F}j}] + \mu^2_{\mathsf{F}j}\tau^2_{\mathsf{F}} - \mathrm{Cov}(\hat{\mu}_{\mathsf{F}j}, Y_{\mathsf{F}j}) - \mu_{\mathsf{F}j}\mathrm{Cov}(\hat{\mu}_{\mathsf{F}j}, \Theta_{\mathsf{F}j})}{\mu_{\mathsf{F}j}/w_j + \mathrm{Var}[\hat{\mu}_{\mathsf{F}j}] + \mu^2_{\mathsf{F}j}\tau^2_{\mathsf{F}} - 2\mathrm{Cov}(\hat{\mu}_{\mathsf{F}j}, Y_{\mathsf{F}j})}. \tag{17}$$

Here τ^2_{F} is a variance component for variation between the groups. And $\mathrm{Var}[\hat{\mu}_{\mathsf{F}j}]$ is a component that gives the estimation variance. We shall give estimators of these variance components which can be inserted in the expression (17). The variance component for variation within group j, which is $\sigma^2_{\mathsf{M}j}$ for mean claim, is for frequency under the Poisson assumption simply $\mu_{\mathsf{F}j}/w_j$.

With only one Auxiliary, where $\hat{\mu}_{\mathsf{F}j} = N_{j1}/w_{j1}$, we can show the following.

Corollary 1.

$$z_{\mathsf{F}j} = \frac{\nu_{\mathsf{F}j}^2 + \mu_{\mathsf{F}j}^2\tau_{\mathsf{F}}^2 - \dfrac{w_j}{w_{j1}}\mu_{\mathsf{F}j}/w_j - \dfrac{2w_j}{w_{j1}}\mu_{\mathsf{F}j}^2\tau_{\mathsf{F}}^2}{\mu_{\mathsf{F}j}/w_j + \nu_{\mathsf{F}j}^2 + \mu_{\mathsf{F}j}^2\tau_{\mathsf{F}}^2 - \dfrac{2w_j}{w_{j1}}\mu_{\mathsf{F}j}/w_j - \dfrac{2w_j}{w_{j1}}\mu_{\mathsf{F}j}^2\tau_{\mathsf{F}}^2}. \tag{18}$$

Pseudo-estimator of variance component between groups

The excess of a random variable is defined in the following way.

$$\mathrm{e}(Z) = \frac{\mathrm{E}[(Z-\mu)^4]}{\mathrm{E}[(Z-\mu)^2]^2} - 3 \quad \text{for a random variable } Z \text{ with } \mathrm{E}[Z] = \mu.$$

This gives

$$\mathrm{Var}[(Z-\mu)^2] = \mathrm{E}[(Z-\mu)^4] - \mathrm{E}[(Z-\mu)^2]^2 = [\mathrm{e}(Z)+2]\mathrm{E}[(Z-\mu)^2]^2 = [\mathrm{e}(Z)+2]\mathrm{Var}[Z]^2. \tag{19}$$

We give a pseudo-estimator of τ_{F}^2 under the conditions that the 3:rd central moment and the excess of $\Theta_{\mathsf{F}j}$ are 0. This is true if $\Theta_{\mathsf{F}j} \sim N(1, \tau_{\mathsf{F}}^2)$. Now $\Theta_{\mathsf{F}j}$ is non-negative, but the normal distribution can hold approximatively with a τ_{F}^2 giving a small probability to negative values in $N(1, \tau_{\mathsf{F}}^2)$.

We have not been able to prove that there is exactly one solution $\hat{\tau}_{\mathsf{F}}^2$ of (22), only that there is at least one, which can be 0. For definiteness we thus define the solution as the largest solution of possibly more than one. The equation is easily solved by a simple binary search.

We give long formulas in order to be completely unambiguous. The factor $[\cdots]^{-1}$ of (21) serves to make $c_{\mathsf{F}1} + \ldots + c_{\mathsf{F}J}$ equal to 1. Let

$$u(x, y) = \frac{1}{y^3} + \frac{7x+2}{y^2} + \frac{4x}{y} + 2x^2, \tag{20}$$

and

$$c_{\mathsf{F}j} = \left(\frac{1}{\hat{\mu}_{\mathsf{F}j} w_j} + \hat{\tau}_{\mathsf{F}}^2\right)^2 u(\hat{\tau}_{\mathsf{F}}^2, \hat{\mu}_{\mathsf{F}j} w_j)^{-1} \left[\sum_{i=1}^{J} \left(\frac{1}{\hat{\mu}_{\mathsf{F}i} w_i} + \hat{\tau}_{\mathsf{F}}^2\right)^2 u(\hat{\tau}_{\mathsf{F}}^2, \hat{\mu}_{\mathsf{F}i} w_i)^{-1}\right]^{-1}, \tag{21}$$

$$\hat{\tau}_{\mathsf{F}}^2 = \sum_{j=1}^{J} c_{\mathsf{F}j} \hat{\tau}_{\mathsf{F}}^2 \left(\frac{1}{\hat{\mu}_{\mathsf{F}j} w_j} + \hat{\tau}_{\mathsf{F}}^2\right)^{-1} \left(\frac{Y_{\mathsf{F}j}}{\hat{\mu}_{\mathsf{F}j}} - 1\right)^2. \tag{22}$$

Lemma 1. *Assume* **A1F** *and* **A2F***. For the pseudo-estimator $\hat{\tau}_{\mathsf{F}}^2$ we have $\mathrm{E}[\hat{\tau}_{\mathsf{F}}^2] \approx \tau_{\mathsf{F}}^2$. If $\mathrm{E}[(\Theta_{\mathsf{F}j} - 1)^3] = 0$ and $\mathrm{e}(\Theta_{\mathsf{F}j}) = 0$, then $\mathrm{Var}[(Y_{\mathsf{F}j}/\mu_{\mathsf{F}j} - 1)^2] = u(\tau_{\mathsf{F}}^2, \mu_{\mathsf{F}j} w_j)$ and $\hat{\tau}_{\mathsf{F}}^2$ is approximately optimal in the sense of having the smallest variance of estimators in the form (22) with $c_{\mathsf{F}1} + \ldots + c_{\mathsf{F}J} = 1$.*

Credibility factor estimator for predictor and correction for bias

We can now put estimators into the right side of equation (18).

Theorem 2. *An estimated credibility factor is*

$$\widehat{z}_{\mathsf{F}j} = \frac{\hat{\nu}^2_{\mathsf{F}j} + \hat{\mu}^2_{\mathsf{F}j}\left(1 - \dfrac{2w_j}{w_{j1}}\right)\hat{\tau}^2_{\mathsf{F}} - \dfrac{w_j}{w_{j1}}\hat{\mu}_{\mathsf{F}j}/w_j}{\hat{\mu}_{\mathsf{F}j}/w_j + \hat{\nu}^2_{\mathsf{F}j} + \hat{\mu}^2_{\mathsf{F}j}\left(1 - \dfrac{2w_j}{w_{j1}}\right)\hat{\tau}^2_{\mathsf{F}} - \dfrac{2w_j}{w_{j1}}\hat{\mu}_{\mathsf{F}j}/w_j}. \tag{23}$$

We have thus determined the predictor $\widehat{\Lambda}_{\mathsf{F}j} = \widehat{z}_{\mathsf{F}j}Y_{\mathsf{F}j} + (1 - \widehat{z}_{\mathsf{F}j})\hat{\mu}_{\mathsf{F}j}$ by (14) depending on variance estimates. The exposure-weighted total of the non-observable predictors is unbiased. Namely

$$\mathrm{E}[\sum_{j=1}^{J}\Lambda^*_{\mathsf{F}j}w_j] = \sum_{j=1}^{J}[z_j\mu_{\mathsf{F}j} + (1 - z_j)\mu_{\mathsf{F}j}]w_j = \sum_{j=1}^{J}\mu_{\mathsf{F}j}w_j = \mathrm{E}[\sum_{j=1}^{J}Y_{\mathsf{F}j}w_j] = \mathrm{E}[\sum_{j=1}^{J}N_j]. \tag{24}$$

The replacement of $z_{\mathsf{F}j}$ and $\mu_{\mathsf{F}j}$ with estimates $\widehat{z}_{\mathsf{F}j}$ and $\hat{\mu}_{\mathsf{F}j}$ can cause a bias. In examples the bias was found to be negligible. It can be corrected by a simple factor.

Combining claim frequency and mean claim

The claim frequency results are combined with corresponding mean claim results. if we wish to use these results for risk premium. We then define

$$\widehat{\Lambda}_{\mathsf{FM}j} = \widehat{\Lambda}_{\mathsf{F}j}\widehat{\Lambda}_{\mathsf{M}j} \tag{25}$$

and apply a correction factor

We have to assume that $\{\Theta_{\mathsf{F}j}\}_1^J$ and $\{\Theta_{\mathsf{M}j}\}_1^J$ are independent. Claim numbers are S-ancillary for the mean claim parameters in the Compound Poisson model, which motivates (25).

Simulation results

We have compared our pseudo-estimators with the classical type ones that are rendered in Ohlsson and Johansson (2010).

Claim frequency and mean claim were analyzed separately. The basic model assumptions were obeyed, but we did not let the distributions of $\Theta_{\mathsf{F}j}$ and $\Theta_{\mathsf{M}j}$ have zero excess or even mostly third central moment zero, as our pseudo-estimator theorems presuppose. From a practical viewpoint these assumptions are unrealistic, but they admit relatively simple and mathematically consistent estimators. The estimators have to be reasonably robust against departures from the assumptions in order to be useful, though. Instead we used some uniform and Γ-distributions for $\Theta_{\mathsf{F}j}$ and $\Theta_{\mathsf{M}j}$.

We found that the pseudo-estimator $\hat{\tau}_{\mathsf{F}}^2$ by (22) is the best one, except for some cases with small τ_{F}^2.

The mean claim pseudo-estimator $\hat{\tau}_{\mathsf{M}}^2$, not treated here, generally seemed to be best for light-tailed conditional claim distributions, while the non-pseudo-estimator seemed best for heavy-tailed ones.

Application

Credibility analysis according to this and the larger submitted paper can be done with Rapp, with a choice of choosing the classical estimators or the pseudo-estimators given in my paper.

A demo run of Taranmultiklass2.Rpp follows. It might take place in a separate session.