Credibility by Esbjörn Ohlsson
Lecture in Indonesia november 2014

Stig Rosenlund

I have been asked to give an overview of the credibility theory found in the book *Non-Life Insurance Pricing with Generalized Linear Models* by Ohlsson, E. & Johansson, B. (2010), Springer, Berlin. Esbjörn and Björn were my colleagues at Länsförsäkringar Alliance (its logo up left) for many years. Both are associate professors at Stockholm University. Esbjörn published the theory in 2008 in an article in Scandinavian Actuarial Journal. This material has been incorporated in the book, so the article is needed only for proper assignment of priority.

Please interrupt me with questions at any time!

The methods are implemented in Rapp alongside my own methods. I will give a demo in Rapp.

The theory builds on the Bühlmann-Straub credibility model. The aim is to find risk factors for an MLF = *multi-level factor*. The MLF has many classes, where many have too few claims to admit estimating a risk factor from the claim statistics of the class only. Think for example of 2000 geographical parishes, where many have just a couple or no claims.

Beside the MLF we have ordinary tariff arguments, such as owner age, vehicle age and motor effect. Claim frequency and mean claim are treated separately. The formulation of the theory is the same for both.

Now take one of these. For the ordinary arguments, a base factor and risk factors per argument and class are estimated with ordinary GLM analysis. Log link, i.e. a multplicative model, is used.

The combination with credibility consists in defining new exposures and risk statistics by multiplication with powers of the GLM estimates. The Bühlmann-Straub model is then applied to these.

Some notation. We call a class of an MLF a group. Exposure is duration for claim frequency and number of claims for mean claim. Key ratio denotes observed claim frequency = (number of claims)/duration and observed mean claim = (claim cost)/(number of claims), respectively.

The non-hierarchical model (p. 81–85)

$$
\begin{aligned}
J &= \text{number of groups}, \\
j &= \text{group index}, \\
i &= \text{tariff cell index } 1, 2, \ldots \text{ for ordinary arguments}, \\
t &= \text{observation index } 1, 2, \ldots \text{ in a group } j, \\
w_{ijt} &= \text{exposure, where } i \text{ is the tariff cell of } j, \\
Y_{ijt} &= \text{key ratio.}
\end{aligned}
$$

For a given j the number of different t is determined by some subdivision of j for claim frequency, for example disjoint time periods. For mean claim it is the number of claims.

Let μ be the base factor and γ_i be the factor product for tariff cell i, so that $\mu\gamma_i$ is the expected key ratio - claim frequency and mean claim, respectively. The method treats these as known in the sense that estimates of them are plugged into the formulas.

Now U_j is a random effect for the MLF group j with $\mathrm{E}[U_j] = 1$ that we wish to predict (in one terminology) or estimate (in another terminology). Following the author we use the words estimate and estimator. The treatment is in terms of $V_j = \mu U_j$, which the author finds more convenient.

We now formulate the model, where p is a number which in applications is 1 for claim frequency and 2 for mean claim.

Model assumptions

(1) $\mathrm{E}[Y_{ijt} \mid V_j] = \gamma_i V_j$.

(2) $\mathrm{E}[\mathrm{Var}[Y_{ijt} \mid V_j]] = \dfrac{\gamma_i^p \sigma^2}{w_{ijt}}$.

(3) The random vectors (Y_{ijt}, V_j); $j = 1, \ldots, J$; are independent.

(4) V_j are identically distributed with $\mathrm{E}[V_j] = \mu > 0$ and $\mathrm{Var}[V_j] = \tau^2$ for some $\tau^2 > 0$.

(5) For any j, conditional on V_j, the Y_{ijt} are independent.

We may call τ^2 a between-groups variance component and σ^2 a within-groups variance component.

New exposures and risk statistics are defined as

$$\widetilde{Y}_{ijt} = \frac{Y_{ijt}}{\gamma_i}, \qquad \widetilde{w}_{ijt} = w_{ijt}\gamma_i^{2-p}.$$

With time-honored actuarial terminology, we can say that we norm away the GLM estimates. Then we can use Bühlmann-Straub credibility on the normed statistics.

The goal is to find a linear estimator $\widehat{V}_j$ that minimizes the expected squared difference between it and the actual value V_j. It is called the credibility estimator.

The result is this. Define the normed exposures

$$\widetilde{w}_{.j.} = \sum_{i,t} \widetilde{w}_{ijt}$$

and the weighted normed statistics averages

$$\overline{\widetilde{Y}}_{\cdot j \cdot} = \frac{\displaystyle\sum_{i,t} \widetilde{w}_{ijt} \widetilde{Y}_{ijt}}{\widetilde{w}_{\cdot j \cdot}}.$$

Theorem 4.3. *Under the model assumptions, the credibility estimator of V_j is*

$$\widehat{V}_j = \tilde{z}_j \overline{\widetilde{Y}}_{\cdot j \cdot} + (1 - \tilde{z}_j)\mu$$

where

$$\tilde{z}_j = \frac{\widetilde{w}_{\cdot j \cdot}}{\widetilde{w}_{\cdot j \cdot} + \sigma^2/\tau^2}.$$

The estimated key ratio (claim frequency and mean claim, respectively) for policies with the MLF j will be $\gamma_i \widehat{V}_j$.

The main job is now to estimate the variance components τ^2 and σ^2. This follows from an adaptation of Bühlmann-Straub. The GLM estimates are plugged into the formulas, as if they were the true values.

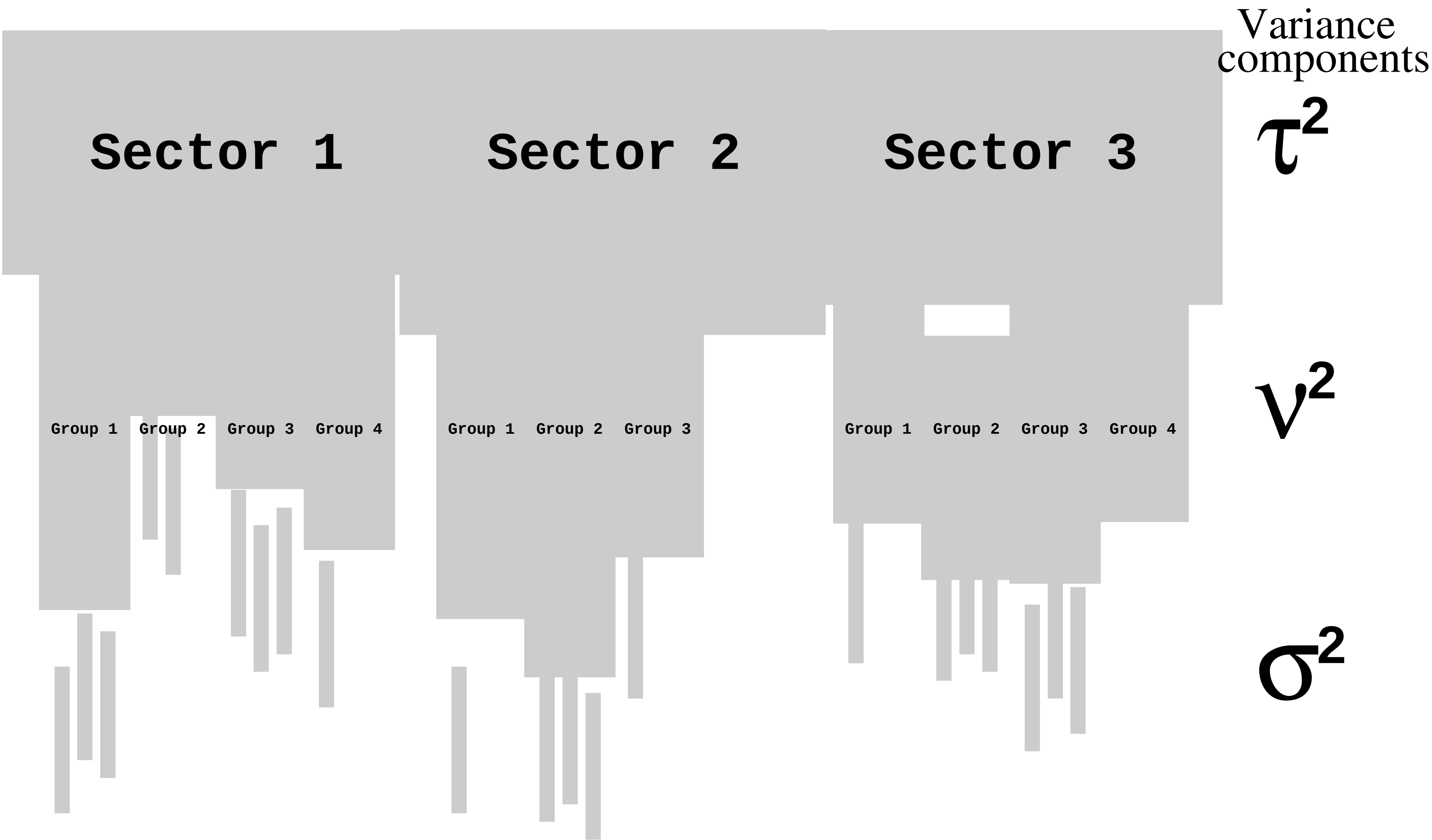
In addition the authors describe a backfitting iterative algorithm, going back and forth between GLM and credibility.

The hierarchical model (p. 90-94)

Credibility is made more complex in a hierarchical setting with two levels. Think of car brands with different car models. The model is then that we have J sectors, each with K_j groups. A sector can be all insured Skoda vehicles, and a group under the sector all insured Skoda can be Octavia Combi.

In this model a random effect for sector j is U_j with $\mathrm{E}[U_j] = 1$, like for MLFs in the the non-hierarchical model. For group k in sector j we have one more random effect U_{jk} with $\mathrm{E}[U_{jk} \mid U_j] = 1$. The total random effect is the product $U_j U_{jk}$.

A figure can illustrate the setting.



Examples of outcomes of risk factors depending hierarchically on sector and group.

There are more parameters and statistics, and the formulas are of course more involved, than in the non-hierarchical model.

The GLM parameters μ and γ_i are the same as before. Reformulated parameters $V_j = \mu U_j$, as before, and $V_{jk} = \mu U_j U_{jk} = V_j U_{jk}$ are introduced. The treatment is in terms of those.

Exposures and key ratios now have one more index k for group within sector.

Assumptions parallell to (1) - (5) above on expectations, variances, independence, conditional independence, and IID are stated. They are in full the following.

Model assumptions

(1) $\mathrm{E}[Y_{ijkt} \mid V_j, V_{jk}] = \gamma_i V_{jk}$.

(2) $\mathrm{E}[\mathrm{Var}[Y_{ijkt} \mid V_j, V_{jk}]] = \dfrac{\gamma_i^p \sigma^2}{w_{ijkt}}$.

(3) The vectors (Y_{ijkt}, V_j, V_{jk}); $j = 1, \ldots, J$; are independent.

(4) (V_j, V_{jk}) are identically distributed with

$$\mathrm{E}[V_j] = \mu > 0 \quad \text{and} \quad \mathrm{E}[V_{jk} \mid V_j] = V_j.$$

(5) For any j, conditional on V_j, the (Y_{ijkt}, V_{jk}) are independent.

(6) For any (j, k), conditional on (V_j, V_{jk}), the Y_{ijkt} are independent.

Define

$$\tau^2 = \mathrm{Var}[V_j] \quad \text{and} \quad \nu^2 = \mathrm{E}[\mathrm{Var}[V_{jk} \mid V_j]].$$

We can say that τ^2 is a between-sectors variance component, ν^2 a between-groups variance component and σ^2 a within-groups variance component.

The definitions of normed exposures, risk statistics and weighted normed statistics averages are here

$$\widetilde{Y}_{ijkt} = \frac{Y_{ijkt}}{\gamma_i}, \quad \widetilde{w}_{ijkt} = w_{ijkt}\gamma_i^{2-p}, \quad \text{and} \quad \overline{\widetilde{Y}}_{.jk.} = \frac{\sum_{i,t} \widetilde{w}_{ijkt}\widetilde{Y}_{ijkt}}{\widetilde{w}_{.jk.}}.$$

For the first level – the sector level – we have

Theorem 4.5. *Under the model assumptions, the credibility estimator of V_j is*

$$\widehat{V}_j = q_j \overline{\overline{\widetilde{Y}}}^z_{.j..} + (1 - q_j)\mu$$

where

$$z_{jk} = \frac{\widetilde{w}_{.jk.}}{\widetilde{w}_{.jk.} + \dfrac{\sigma^2}{\nu^2}}, \qquad \overline{\overline{\widetilde{Y}}}^z_{.j..} = \frac{\displaystyle\sum_k z_{jk} \overline{\widetilde{Y}}_{.jk.}}{\displaystyle\sum_k z_{jk}}, \qquad \text{and} \qquad q_j = \frac{z_{j.}}{z_{j.} + \dfrac{\nu^2}{\tau^2}}.$$

A point as index for the $\widetilde{w}$:s and z:s means summation over the values of the index.

For the second level – the group level – it holds

Theorem 4.6. *Under the model assumptions, the credibility estimator of V_{jk} is*

$$\widehat{V}_{jk} = z_{jk}\overline{\widetilde{Y}}_{.jk.} + (1 - z_{jk})\widehat{V}_{j}.$$

The estimated key ratio (claim frequency and mean claim, respectively) for policies with sector j and group k will be $\gamma_i \widehat{V}_{jk}$.

Estimates of the variance components τ^2, ν^2 and σ^2 are given, stated as extensions of estimates credited to Bühlmann & Gisler (2005) and Ohlsson (2005), independently derived. As before, GLM estimates are plugged into the formulas, as if they were the true values.

A demo run of Taranmultiklass3.Rpp follows. It might take place in a separate session.