An Insurance Model with Collective Seasonal Random Factors

By Bjørn Sundt, University of Oslo

1. Let m be an unknown random variable. We shall say that an estimator $m^{(1)}$ is a better estimator of m than another estimator $m^{(2)}$ if

$$E(m^{(1)}-m)^2 < E(m^{(2)}-m)^2,$$

that is, we use quadratic loss.

Let $y_1, \ldots, y_s$ be observable random variables, and let

$$\tilde{m} = \gamma_0 + \sum_{i=1}^{s} \gamma_i y_i$$

be the best linear credibility estimator of m based on $y_1, \ldots, y_s$, that is, the best estimator of m of the form $g_0 + \sum_{i=1}^{s} g_i y_i$, where $g_0, g_1, \ldots, g_s$ are non-random numbers. It is easy to show that $\gamma_0, \gamma_1, \ldots, \gamma_s$ must satisfy

$$\begin{aligned} \gamma_0 &= E(m) - \sum_{i=1}^{s} \gamma_i E(y_i) \\ \sum_{i=1}^{s} \gamma_i C(y_i, y_j) &= C(m, y_j) \quad j = 1, \ldots, s \end{aligned} \tag{1}$$

(see e.g. [1]). We silently assume that whenever we develop best linear credibility estimators in the present paper, then (1) gives a unique determination of $\gamma_0, \gamma_1, \ldots, \gamma_s$.

2. Assume that we have an insurance portfolio consisting of n policies that have been running for the same r insurance terms, and let x_{ij} be the total claim amount of i-th policy in j-th term.

Policy i is characterized by an unknown random parameter θ_i. We assume that $\theta_1, \ldots, \theta_n$ are independent and identically distributed, and that θ_i is independent of x_{kj} for $k \neq i$. Given θ_i $x_{i1}, x_{i2}, \ldots$ are conditionally independent and identically distributed with common cumulative distribution $G(\cdot \mid \theta_i)$. The function $G(\cdot \mid \cdot)$ is independent of i.

We introduce $b(\theta_1) = E(x_{11} \mid \theta_1)$, $\beta = E(x_{11})$, $\bar{x}_i = \frac{1}{r} \sum_{j=1}^{r} x_{ij}$, and $\bar{\bar{x}} = \frac{1}{n} \sum_{i=1}^{n} \bar{x}_i$.

3. We want to find $\tilde{x}_{k,r+1}$, the best linear credibility estimator of $x_{k,r+1}$ based on the observed x_{ij}'s. If claim amounts from different policies are independent, it is wellknown (see e.g. [1]) that

$$\tilde{x}_{k,r+1} = \frac{r}{r+\kappa}\bar{x}_k + \frac{\kappa}{r+\kappa}\beta \tag{2}$$

with

$$\kappa = \frac{E\,V(x_{11}\,|\,\theta_1)}{V(b(\theta_1))}.$$

4. The assumption that claim amounts from different policies are independent, gives specifically that claim amounts $x_{1j}, \ldots, x_{nj}$ from a term j are independent. However, this is not always the case in practice. In many situations there seem to be seasonal random factors influencing the claim amounts of the whole portfolio. We shall look at four examples.

i) *Motor insurance.* Suppose that one winter the weather has given extremely icy roads. This could lead to many car accidents.
ii) *Marine insurance.* In a stormy year there could be lots of shipwrecks.
iii) *Forest fire insurance.* A dry summer could lead to many forest fires.
iv) *Burglary insurance.* The electricity cut in New York in 1977 caused an enormous amount of burglaries.

In these examples it seems natural to assume that each insurance term j is characterized by an unknown random parameter η_j and that

i) $x_{1j}, \ldots, x_{nj}$ are independent given η_j;
ii) $x_{1j}, \ldots, x_{nj}$ are independent of η_l for $l \neq j$;
iii) $\eta_1, \eta_2, \ldots$ are independent and identically distributed;
iv) the η_j's are independent of the θ_i's;
v) x_{ij} and x_{kl} are independent if both $i \neq k$ and $j \neq l$.

We also assume that the conditional distribution of x_{ij} given θ_i and η_j is on the form $F(\cdot\,|\,\theta_i, \eta_j)$ with $F(\cdot\,|\,\cdot,\cdot)$ independent of i and j.
We introduce $c(\eta_1) = E(x_{11}\,|\,\eta_1)$.

5. From Example 5 in [3] follows that $\tilde{x}_{k,r+1}$ is the best linear credibility estimator of $x_{k,r+1}$ based on $\bar{x}_k$ and $\bar{\bar{x}}$. Hence

$$\tilde{x}_{k,r+1} = \gamma_0 + \gamma_1\bar{x}_k + \gamma_2\bar{\bar{x}} \tag{3}$$

with γ_0, γ_1, and γ_2 determined by

$$\gamma_0 = (1-\gamma_1-\gamma_2)\,\beta \tag{4}$$

$$\gamma_1\, V(\bar{x}_k) + \gamma_2\, C(\bar{\bar{x}}, \bar{x}_k) = C(x_{k,\,r+1}, \bar{x}_k) \tag{5}$$

$$\gamma_1\, C(\bar{x}_k, \bar{\bar{x}}) + \gamma_2\, V(\bar{\bar{x}}) = C(x_{k,\,r+1}, \bar{\bar{x}}) \tag{6}$$

The following relations are easily verified.

$$C(\bar{x}_k, \bar{\bar{x}}) = V(\bar{\bar{x}}) \tag{7}$$

$$C(x_{k,\,r+1}, \bar{x}_k) = V(b(\theta_1)) \tag{8}$$

$$C(x_{k,\,r+1}, \bar{\bar{x}}) = \frac{1}{n}\, V(b(\theta_1)) \tag{9}$$

$$V(\bar{\bar{x}}) = \frac{1}{n}\, V(\bar{x}_1) + \left(1-\frac{1}{n}\right) C(\bar{x}_1, \bar{x}_2)$$

$$V(\bar{x}_1) = V(b(\theta_1)) + \frac{1}{r}\, E\, V(x_{11} \mid \theta_1)$$

$$C(\bar{x}_1, \bar{x}_2) = \frac{1}{r}\, V(c(\eta_1)).$$

Using (7), (8), and (9), (5) and (6) can now be rewritten

$$\gamma_1\, V(\bar{x}_1) + \gamma_2\, V(\bar{\bar{x}}) = V(b(\theta_1)) \tag{10}$$

$$(\gamma_1 + \gamma_2)\, V(\bar{\bar{x}}) = \frac{1}{n}\, V(b(\theta_1)). \tag{11}$$

Subtracting (11) from (10) gives

$$\gamma_1 (V(\bar{x}_1) - V(\bar{\bar{x}})) = \left(1-\frac{1}{n}\right) V(b(\theta_1)).$$

From this follows

$$\gamma_1 = \left(1-\frac{1}{n}\right) \frac{V(b(\theta_1))}{V(\bar{x}_1) - V(\bar{\bar{x}})} =$$

$$\left(1-\frac{1}{n}\right) \frac{V(b(\theta_1))}{V(\bar{x}_1) - \frac{1}{n}\, V(\bar{x}_1) - \left(1-\frac{1}{n}\right) C(\bar{x}_1, \bar{x}_2)} = \frac{V(b(\theta_1))}{V(\bar{x}_1) - C(\bar{x}_1, \bar{x}_2)} =$$

$$\frac{V(b(\theta_1))}{V(b(\theta_1)) + \frac{1}{r}\, E\, V(x_{11} \mid \theta_1) - \frac{1}{r}\, V(c(\eta_1))},$$

that is,

$$\gamma_1 = \frac{r}{r + \kappa - \varrho} \tag{12}$$

with

$$\varrho = \frac{V(c(\eta_1))}{V(b(\theta_1))}.$$

From (11) we get

$$\gamma_1 + \gamma_2 = \frac{1}{n}\frac{V(b(\theta_1))}{V(\bar{\bar{x}})} = \frac{V(b(\theta_1))}{V(\bar{x}_1) + (n-1)\, C(\bar{x}_1, \bar{x}_2)} =$$

$$\frac{V(b(\theta_1))}{V(b(\theta_1)) + \frac{1}{r} E\, V(x_{11} | \theta_1) + (n-1)\frac{1}{r} V(c(\eta_1))}.$$

giving

$$\gamma_1 + \gamma_2 = \frac{r}{r + \kappa + (n-1)\,\varrho}. \tag{13}$$

From (3), (4), (12), and (13) we now get

$$\tilde{x}_{k,\, r+1} = \frac{r}{r + \kappa - \varrho}(\bar{x}_k - \bar{\bar{x}}) + \frac{r}{r + \kappa + (n-1)\,\varrho}\,\bar{\bar{x}} + \frac{\kappa + (n-1)\,\varrho}{r + \kappa + (n-1)\,\varrho}\,\beta. \tag{14}$$

6. We shall look at some asymptotic results.

When $r \to \infty$, we have $\frac{r}{r + \kappa - \varrho} \to 1$, $\frac{r}{r + \kappa + (n-1)\,\varrho} \to 1$, $\frac{\kappa + (n-1)\,\varrho}{r + \kappa + (n-1)\,\varrho} \to 0$, $\bar{x}_k \overset{a.s.}{\to} b(\theta_k)$, and $\bar{\bar{x}} \overset{a.s.}{\to} \frac{1}{n}\sum_{i=1}^{n} b(\theta_i)$.

From this follows that

$$\tilde{x}_{k,\, r+1} \overset{a.s.}{\to} b(\theta_k)$$

when $r \to \infty$, which is very satisfying.

If $\varrho \neq 0$ and $n \to \infty$, we have $\frac{r}{r + \kappa + (n-1)\,\varrho} \to 0$, $\frac{\kappa + (n-1)\,\varrho}{r + \kappa + (n-1)\,\varrho} \to 1$, and $\bar{\bar{x}} \overset{a.s.}{\to} \frac{1}{r}\sum_{j=1}^{r} c(\eta_j) = \bar{c}$. This gives

$$\tilde{x}_{k,\, r+1} \overset{a.s.}{\to} \frac{r}{r + \kappa - \varrho}(\bar{x}_k - \bar{c}) + \beta. \tag{15}$$

Rewriting (14) gives

$$\tilde{x}_{k,\,r+1} = \frac{r}{r+\kappa-\varrho}\left[\bar{x}_k + \frac{n\,\varrho}{r+\kappa+(n-1)\,\varrho}(\beta-\bar{\bar{x}})\right] + \frac{\kappa-\varrho}{r+\kappa-\varrho}\beta. \qquad (16)$$

In (2) $\tilde{x}_{k,\,r+1}$ was a weighted sum of β and $\bar{x}_k$, the latter a reasonable estimator of $E(x_{k,\,r+1}\,|\,\theta_k)$. In (16) we can interprete that $\bar{x}_k$ has been replaced by $\bar{x}_k + \frac{n\,\varrho}{r+\kappa+(n-1)\,\varrho}(\beta-\bar{\bar{x}})$. A correction term

$$h(\bar{\bar{x}}) = \frac{n\,\varrho}{r+\kappa+(n-1)\,\varrho}(\beta-\bar{\bar{x}})$$

has now entered to compensate for the collective random fluctuations caused by the η_j's. We observe that $\frac{n\,\varrho}{r+\kappa+(n-1)\,\varrho}$ increases when ϱ increases, that is, we need a greater correction when the variance of the $c(\eta_j)$'s increases, or less precisely, when the collective seasonal fluctuations increase. We also have that $\frac{n\,\varrho}{r+\kappa+(n-1)\,\varrho}$ decreases when r increases, that is, we need smaller correction when time increases.

Corresponding to (15) we have that

$$h(\bar{\bar{x}}) \overset{a.s.}{\to} \beta-\bar{c}$$

when $n\to\infty$.

The above reasoning perhaps becomes clearer if we for a moment assume an additive model:

$$E(x_{ij}\,|\,\theta_i,\,\eta_j) = b(\theta_i) + a(\eta_j)$$

$$E(a(\eta_j)) = 0.$$

Then

$$E(h(\bar{\bar{x}})\,|\,\eta_1,\ldots,\eta_r) = -\frac{n\,\varrho}{r+\kappa+(n-1)\,\varrho}\frac{1}{r}\sum_{j=1}^{r}a(\eta_j)$$

and

$$h(\bar{\bar{x}}) \overset{a.s.}{\to} -\frac{1}{r}\sum_{j=1}^{r}a(\eta_j)$$

when $n\to\infty$, and the nature of $h(\bar{\bar{x}})$ as a mean to reduce the influence of η_j's is clear.

7. In practice we have to use estimated values for ϱ, κ, and β. Estimators for ϱ, κ, and β are proposed in [2].

8. As a special case of our general model we briefly mentioned an additive model. Another special case that seems more apt in practice, is a multiplicative model where

$$E(x_{11} \mid \theta_1, \eta_1) = b(\theta_1)\, a(\eta_1)$$

$$E(a(\eta_1)) = 1.$$

Our linear credibility estimator seems natural in the additive model, not equally natural in the multiplicative case. And we could ask: Could we do better?
We shall develop an alternative approach.
For any estimator $x^*_{k,\,r+1}$ of $x_{k,\,r+1}$ based on the observed x_{ij}'s we have

$$E(x^*_{k,\,r+1} - x_{k,\,r+1})^2 = E\, V(x_{11} \mid \theta_1) + E(x^*_{k,\,r+1} - b(\theta_k))^2.$$

Hence an estimator of $x_{k,\,r+1}$ is optimal if and only if it is an optimal estimator of $b(\theta_k)$.
What is now an optimal estimator of $b(\theta_k)$?
If we could observe the random variables $a(\eta_1), \dots, a(\eta_r)$, it would be natural to base our estimator on $\frac{x_{k1}}{a(\eta_1)}, \dots, \frac{x_{kr}}{a(\eta_r)}$ since $E\left(\frac{x_{kj}}{a(\eta_j)} \mid \theta_k, \eta_j\right) = b(\theta_k)$. From section III. 5.1 in [1] follows that the best linear credibility estimator of $b(\theta_k)$ based on $\frac{x_{k1}}{a(\eta_1)}, \dots, \frac{x_{kr}}{a(\eta_r)}$ must be on the form

$$\delta_0 + \delta_1 \frac{1}{r} \sum_{j=1}^{r} \frac{x_{kj}}{a(\eta_j)}$$

where δ_0 and δ_1 are real constants.
Unfortunately we do not know $a(\eta_1), \dots, a(\eta_r)$. But they can be estimated. A natural estimator of $a(\eta_j)$ is the best linear credibility estimator of $a(\eta_j)$ based on $x_{1j}, \dots, x_{nj}$, that is,

$$\tilde{a}_j = \frac{n}{n+\nu} \frac{{}_j\bar{x}}{\beta} + \frac{\nu}{n+\nu}$$

with

$$\nu = \frac{E\, V(x_{11} \mid \eta_1)}{V(c(\eta_1))}$$

and

$${}_j\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_{ij}.$$

Letting now $x^*_{kj} = \frac{x_{kj}}{\tilde{a}_j}$ and $\bar{x}^*_k = \frac{1}{r} \sum_{j=1}^{r} x^*_{kj}$, we want the best linear credibility estimator of $x_{k,\,r+1}$ based on $\bar{x}^*_k$. From (1) follows that this estimator is

$$\tilde{\tilde{x}}_{k,\,r+1} = \gamma_0 + \gamma_1 \, \bar{x}^*_k$$

with

$$\gamma_0 = E(x_{11}) - \gamma_1 \, E(x^*_{11})$$

and

$$\gamma_1 = \frac{C(\bar{x}^*_k, x_{k,\,r+1})}{V(\bar{x}^*_k)}.$$

9. The multiplicative model has also been treated by Welten [4]. He uses the premium (2). If the claim amounts in j-th term are small, this could be due to small $a(\eta_j)$. This would lead to too small premiums to cover the expected future claims. However, the small claim amounts in j-th term have given the insurance company a profit, and Welten argues that a part of this profit should be paid into a "bonus reserve" to cover the future claims.

10. Both $\tilde{x}_{k,\,r+1}$ and $\tilde{\tilde{x}}_{k,\,r+1}$ (and Welten's approach) were developed under the very strict assumption that our portfolio consists of n policies that have been running for the same r terms. It is very unlikely that we shall meet this situation in reality, and we therefore need a less restrictive model.

Unfortunately, in section 5 we would then get a messy system of linear equations, solvable, but too complicated to be of any practical use. The present author believes that instead of developing the best linear credibility formula, one ought to examine the structure of (14) and try to develop a similar, not too complicated formula under the more general conditions.

It seems that such an approach would be somewhat easier in the approach of section 8.

References

[1] *De Vylder, F.*: «Introduction aux Théories Actuarielles de Crédibilité», Office des Assureurs de Belgique, 208 pp. (1975).

[2] *Sundt, B.*: "On Models and Methods of Credibility", Statistical Research Report No. 7, 1978, Institute of Mathematics, University of Oslo.

[3] *Sundt, B.*: "Credibility and Relative Exchangeability", Scandinavian Actuarial Journal, 1979, pp. 45–51.

[4] *Welten, C. P.*: "The Unearned No Claim Bonus", *ASTIN-Bulletin*, Vol. 5, pp. 25–32 (1968).

Bjørn Sundt
Institute of Mathematics University of Oslo
Blindern, Oslo 3, Norway

Abstract

A model for an insurance portfolio is developed, in which each insurance term is characterized by an unknown random parameter influencing the claim amounts of all the policies in the portfolio in that term. Two different ratemaking procedures are developed, and an approach by Welten is briefly summarized.

Zusammenfassung

Es wird ein Versicherungsportefeuillemodell beschrieben, in dem jede Versicherungsperiode durch einen unbekannten Zufallsparameter charakterisiert ist, der seinerseits den Schadenbetrag aller Policen dieser Periode beeinflusst. Anschliessend werden zwei verschiedene Prämienberechnungsregeln entwickelt und eine verwandte Methode von Welten kurz dargestellt.

Résumé

Un modèle d'un portefeuille d'assurance est développé où chaque période d'assurance est charactérisée par un paramètre aléatoire qu influence les coûts de tous les sinistres de cette période. Ensuite, deux différentes règles de calculer des primes sont développées et une methode similaire de Welten est mentionnée.

Riassunto

Viene presentato un modello di portafoglio nel quale, per ogni periodo di assicurazione, si introduce un parametro aleatorio che influenza il costo di tutti i sinistri che avvengono in questo periodo. In seguito vengono sviluppate due regole differenti di calcolo dei premi e comparate col metodo di Welten.