Simplified estimation of structure parameters in hierarchical credibility

Ohlsson, Esbjörn

Länsförsäkringar insurance and
Mathematical Statistics
Stockholm University
SE-106 91 Stockholm, Sweden
Phone +46 8 16 45 58
Fax +46 8 612 67 17
E-mail: esbjorn.ohlsson@math.su.se

Abstract

Some of the structure parameters in Jewell's hierarchical credibility model are commonly estimated by pseudo-estimators. In this paper we present alternative, unbiased estimators, similar to those of the Bühlmann-Straub model. The main advantage of our estimators is that they are given on closed form, while the pseudo-estimators require iterative solution.

Keywords

Credibility theory, Jewell's hierarchical model, structure parameter, pseudo-estimator.

1 Introduction

In credibility models there are so called *structure parameters* that must be estimated before the calculation of the credibility estimators themselves. In Jewell's hierarchical credibility model there is one overall mean parameter μ and three *variance* structure parameters, here called σ^2, a and b – see definitions in the next section. This paper is concerned with the estimation of the variance structure parameters in the hierarchical credibility model.

For the Bühlmann-Straub model, standard textbooks present unbiased estimators of all structure parameters. For Jewell's hierarchical credibility model, there is a simple unbiased estimator of σ^2, while for a and b only so called *pseudo-estimators* are provided, see, e.g., Goovaerts & Hoogstad (1987) or Dannenburg, Kaas & Goovaerts (1996). A pseudo-estimator is motivated by a formula $a = E[f(\boldsymbol{X}, a)]$ where f is a function and $\boldsymbol{X}$ is a random vector. The estimator is computed by iteratively solving the equation $a = f(\boldsymbol{x}, a)$, where $\boldsymbol{x}$ is an observation of $\boldsymbol{X}$. As noted by Sundt (1987), the fact that $a = E[f(\boldsymbol{X}, a)]$ does not imply that the resulting estimator of a is unbiased.

In the present paper we present alternative, unbiased estimators of the structure parameters a and b that are easier to apply than the pseudo-estimators, since they do not require iterative solution. In our experience, the new estimators and the pseudo-estimators give rather similar results. Hence, the main advantage of the new ones is their simplicity in application. There is also a (minor) pedagogical point in not having to introduce the concept of a pseudo-estimator in elementary texts.

2 Jewell's hierarchical credibility model

Here we present the classical hierarchical credibility model of Jewell. We call our observations Y_{jkt}, where, in the terminology of Dannenburg et al. (1996), j indicates a *sector*, k is called a *cell* and t is an *exposure unit* within the cell (j, k). In practise, (j, k, t) may of course represent any hierarchical structure of insurance contracts – for instance, j might be a county, k a parish and t an individual insurance taken by a resident of that parish. In motor insurance, j might be a car brand, k a specific car model and t an individual car.

We introduce *random effects* U_j for the sectors $j = 1, \dots, J$; and U_{jk} for the cells $k = 1, \dots, K_j$. The basic model is that

$$E(Y_{jkt}|U_j, U_{jk}) = U_{jk} \tag{2.1}$$

and

$$E(U_{jk}|U_j) = U_j \tag{2.2}$$

by which also

$$E(Y_{jkt}|U_j) = U_j \tag{2.3}$$

The overall expectation is called μ.

$$\mu \doteq E(U_j) = E(U_{jk}) = E(Y_{jkt}) \tag{2.4}$$

Note that U_{jk} is the (conditional) mean of all observations for cell (j, k), while U_j is the (conditional) mean of all observations in sector j and μ is the mean of the entire population.

Remark 1. Many texts, like Goovaerts & Hoogstad (1987), introduce abstract risk parameters Θ_j and Θ_{jk} and put $\mu(\Theta_j, \Theta_{jk}) \doteq E(Y_{jkt}|\Theta_j, \Theta_{jk})$ and $\nu(\Theta_j) \doteq E(\mu(\Theta_j, \Theta_{jk})|\Theta_j) = E(Y_{jkt}|\Theta_j)$. However, since inference is only made on μ and ν, and not on the Θ:s themselves, there is no loss in generality from using the random effects notation instead, where $\mu(\Theta_j, \Theta_{jk})$ is replaced by U_{jk} and $\nu(\Theta_j)$ is represented by U_j. This idea, in another notation, was introduced by Dannenburg et al. (1996). $\square$

Remark 2. Defining $V_j = U_j/\mu$ and $V_{jk} = U_{jk}/U_j$ we can interpret our model as a *multiplicative* random effects model $E(Y_{jkt}|U_j, U_{jk}) = \mu V_j V_{jk}$, where V_j and V_{jk} are uncorrelated. If instead we define $\Xi_j = U_j - \mu$ and $\Xi_{jk} = U_{jk} - U_j$ we get the *additive* model $E(Y_{jkt}|U_j, U_{jk}) = \mu + \Xi_j + \Xi_{jk}$ of Dannenburg et al. (1996). Both these models have nice interpretation, but we use the U-parametrisation here for technical simplicity. (By the end of the day, all these three models of course give the same credibility risk premiums.) $\square$

All second-order moments are supposed to be finite. Furthermore, we make the following assumptions, which are essentially the same as (J1)-(J5) in Goovaerts & Hoogstad (1987), rewritten in our notation.

Assumption 1 *(a) The sectors are independent,* i.e. *the vectors* (U_j, U_{jk}, Y_{jkt}) *and* $(U_{j'}, U_{j'k'}, Y_{j'k't'})$ *are conditionally independent as soon as* $j \neq j'$.

(b) For every j*, conditional on* U_j *the cells are independent,* i.e. (U_{jk}, Y_{jkt}) *and* $(U_{jk'}, Y_{jk't'})$ *are conditionally independent if* $k \neq k'$.

(c) Conditional on (U_j, U_{jk}) *the exposure units are independent,* i.e. Y_{jkt} *and* $Y_{jkt'}$ *are conditionally independent for* $t \neq t'$.

(d) All random effects pairs (U_j, U_{jk}) *are identically distributed.*

(e) For all j, k and t we have

$$E[\mathrm{Var}(Y_{jkt}|U_j, U_{jk})] = \frac{\sigma^2}{w_{jkt}} \tag{2.5}$$

for some parameter σ^2. Here w_{jkt} is the exposure weight.

By Assumption 1(d), the U_{jk} are identically distributed, and their common expected variance is the variance at the cell level

$$a \doteq E[\mathrm{Var}(U_{jk}|U_j)] \tag{2.6}$$

Our final structure parameter b is simply the sector variance

$$b \doteq \mathrm{Var}[U_j] \tag{2.7}$$

which does not depend on j since the U_j are identically distributed random variables.

Remark 3. By the standard laws for computing variances by conditioning we find the unconditional variance

$$\mathrm{Var}(Y_{jkt}) = a + b + \frac{\sigma^2}{w_{jkt}}$$

which exhibits the structure parameters σ^2, a and b as *variance components.* □

We further introduce the credibility factors

$$z_{jk} \doteq \frac{w_{jk\cdot}}{w_{jk\cdot} + \sigma^2/a} \tag{2.8}$$

$$q_j \doteq \frac{z_{j\cdot}}{z_{j\cdot} + a/b} \tag{2.9}$$

Here and in the future, the dot notation is used to indicate summation, as in $w_{jk\cdot} = \sum_t w_{jkt}$. We give weighted means a superindex to indicate which weights are used, as in

$$\overline{Y}^w_{jk\cdot} = \frac{\sum_t w_{jkt} Y_{jkt}}{\sum_t w_{jkt}} \tag{2.10}$$

and

$$\overline{Y}^{zw}_{j\cdot\cdot} = \frac{\sum_k z_{jk}\overline{Y}^{w}_{jk\cdot}}{\sum_k z_{jk}} \tag{2.11}$$

The well-known (inhomogeneous) credibility estimator at the sector level is

$$\hat{U}_j = q_j\overline{Y}^{zw}_{j\cdot\cdot} + (1-q_j)\mu \tag{2.12}$$

Further, the credibility estimator at the cell level is

$$\hat{U}_{jk} = z_{jk}\overline{Y}^{w}_{jk\cdot} + (1-z_{jk})\hat{U}_j \tag{2.13}$$

See for example Dannenburg et al. (1996, Theorems 3.2.2 and 3.2.3) for a derivation of these estimators. The overall expectation μ may be given by prior information or estimated by e.g.

$$\hat{\mu} = \overline{Y}^{qzw}_{\cdots} = \frac{\sum_j q_j\overline{Y}^{zw}_{j\cdot\cdot}}{\sum_j q_j} \tag{2.14}$$

In order to apply the credibility estimators in practice, we must first estimate the variance parameters σ^2, a and b.

2.1 Traditional estimators of variance parameters

The estimators in this section can be found in Goovaerts & Hoogstad (1987, p. 90) or Dannenburg et al. (1996, p. 54), to which we also refer for proofs. Firstly, an unbiased estimator of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{\sum_j\sum_k(T_{jk}-1)}\sum_{j=1}^{J}\sum_{k=1}^{K_j}\sum_{t=1}^{T_{jk}} w_{jkt}(Y_{jkt}-\overline{Y}^{w}_{jk\cdot})^2 \tag{2.15}$$

where J is the number of sectors, K_j is the number of cells in sector j, and T_{jk} is the number of observations for cell (j,k).

For a and b, pseudo-estimators are derived from the fact that

$$a = E\left[\frac{1}{\sum_j (K_j - 1)} \sum_{j=1}^{J} \sum_{k=1}^{K_j} z_{jk} (\overline{Y}^{w}_{jk\cdot} - \overline{Y}^{zw}_{j\cdot\cdot})^2\right] \quad (2.16)$$

$$b = E\left[\frac{1}{J-1} \sum_{j=1}^{J} q_j (\overline{Y}^{zw}_{j\cdot\cdot} - \overline{Y}^{qzw}_{\cdots})^2\right] \quad (2.17)$$

Since z_{jk} is a function of a and q_j is a function of b, we can not simply drop the expectation signs to get unbiased estimators. However, the equation resulting from omitting the expectation in (2.16) can be solved for a by iteration; the solution is called a pseudo-estimator of a. This value is then inserted into (2.17), the expectation sign is dropped, and the resulting equation is iterated to give a pseudo-estimator of b.

Note that (2.16) and (2.17) do not imply that the pseudo-estimators are unbiased.

3 The new estimators

Here we derive the alternative estimators of a and b that is the core of this paper. The trick is to replace the z- and q-weighted means in (2.16) and (2.17) by means with weights that are *known* (or at least already estimated) at that stage in the calculations. For a, instead of $\overline{Y}^{zw}_{j\cdot\cdot}$ we will use

$$\overline{Y}^{ww}_{j\cdot\cdot} = \frac{\sum_k w_{jk\cdot} \overline{Y}^{w}_{jk\cdot}}{\sum_k w_{jk\cdot}}$$

and we suggest the following unbiased estimator

$$\hat{a} = \frac{\sum_j \sum_k w_{jk\cdot} (\overline{Y}^{w}_{jk\cdot} - \overline{Y}^{ww}_{j\cdot\cdot})^2 - \hat{\sigma}^2 \sum_j (K_j - 1)}{w_{\cdots} - \sum_j (\sum_k w_{jk\cdot}^2)/w_{j\cdot\cdot}} \quad (3.1)$$

Remark 4. Suppose our hierarchical model had just one sector j, in which case our model was in fact non-hierarchical, i.e. we had the Bühlmann-Straub model. We could then omit the index j and (3.1) would reduce to the standard unbiased estimator in the Bühlmann-Straub model, see e.g. Dannenburg et al. (1996), Theorem 2.3.1. □

Once we know σ^2 and a, we get z_{jk} from (2.8), and thereby we can compute $\overline{Y}_{j\cdot\cdot}^{zw}$ in (2.11). In place of $\overline{Y}_{\cdots}^{qzw}$ that appears in the pseudo-estimator of b we use

$$\overline{Y}_{\cdots}^{zzw} = \frac{\sum_j z_{j\cdot} \overline{Y}_{j\cdot\cdot}^{zw}}{\sum_j z_{j\cdot}}$$

Our suggested unbiased estimator of b is now

$$\hat{b} = \frac{\sum_j z_{j\cdot} (\overline{Y}_{j\cdot\cdot}^{zw} - \overline{Y}_{\cdots}^{zzw})^2 - \hat{a}(J-1)}{z_{\cdot\cdot} - \sum_j z_{j\cdot}^2 / z_{\cdot\cdot}} \tag{3.2}$$

The estimators $\hat{a}$ and $\hat{b}$ are simpler than the pseudo-estimators in that they do not require iteration. The unbiasedness is proved in Theorem 3.1 below.

Note that even though $\hat{b}$ as it stands is unbiased, in practice we plug in $\hat{a}$ and $\hat{\sigma}^2$ into z_{jk} and then $\hat{b}$ is no longer strictly unbiased. A similar remark goes for the equation (2.17) that defines the pseudo-estimator – this equation is not strictly true when we plug in $\hat{a}$ and $\hat{\sigma}^2$ into z_{jk}. This kind of problem is of course common in statistics, and will not be discussed further here.

Remark 5. **Negative values.** As with the corresponding estimator in the Bühlmann-Straub model, the suggested estimators may take on negative values. This problem is discussed for the Bühlmann-Straub model in Example 2.3.2 of Dannenburg et al. (1996). The conclusion is that we can not reject the hypothesis that the parameter is equal to zero, and so the corresponding level of random effects should be removed from the model.

In our experience, in situations when our $\hat{a}$ or $\hat{b}$ is negative, the corresponding pseudo-estimator iterations converge (slowly) to zero. This is in accordance with the result for the Bühlmann-Straub model in Dubey & Gisler (1981, Theorem 2) which states that the pseudo-estimator equation has one and only one positive solution if and only if the unbiased estimator gives a strictly positive value; if not, the pseudo-estimator converges to zero. $\square$

Remark 6. **Multi-level models.** We believe that similar estimators could be found for hierarchical models with three or more levels. It remains to work out the details for these multi-level models, though. $\square$

Theorem 3.1 *(a) With $\hat{a}$ as in (3.1), we have $E[\hat{a}] = a$.*

(b) With $\hat{b}$ as in (3.2), we have $E[\hat{b}] = b$.

Proof. For part (a), we first note that

$$\mathrm{Var}(\overline{Y}^w_{jk\cdot}|U_j) = E[\mathrm{Var}(\overline{Y}^w_{jk\cdot}|U_j, U_{jk})|U_j] + \mathrm{Var}(E[\overline{Y}^w_{jk\cdot}|U_j, U_{jk}]|U_j)$$

By (2.1), (2.5) and (2.6)

$$E[\mathrm{Var}(\overline{Y}^w_{jk\cdot}|U_j)] = E[\mathrm{Var}(\overline{Y}^w_{jk\cdot}|U_j, U_{jk})] + E[\mathrm{Var}(U_{jk}|U_j)] = \frac{\sigma^2}{w_{jk\cdot}} + a \tag{3.3}$$

Further, by (2.3),

$$E[\overline{Y}^{w}_{jk\cdot}|U_j] = U_j \qquad \text{and} \qquad E[\overline{Y}^{ww}_{j\cdot\cdot}|U_j] = U_j$$

This justifies the second equality in the following derivation, in which we further use the fact from Assumption 1(b) that $\{\overline{Y}^{w}_{jk\cdot}; k = 1, \ldots, K_j\}$ are conditionally independent given U_j.

$$\begin{aligned}
E\left[(\overline{Y}^{w}_{jk\cdot} - \overline{Y}^{ww}_{j\cdot\cdot})^2\right] &= E\left[E\left[\left(\overline{Y}^{w}_{jk\cdot} - \overline{Y}^{ww}_{j\cdot\cdot}\right)^2 |U_j\right]\right] \\
&= E\left[\mathrm{Var}\left(\overline{Y}^{w}_{jk\cdot} - \overline{Y}^{ww}_{j\cdot\cdot}|U_j\right)\right] \\
&= E\left[\left(1 - \frac{w_{jk\cdot}}{w_{j\cdot\cdot}}\right)^2 \mathrm{Var}(\overline{Y}^{w}_{jk\cdot}|U_j) + \sum_{\ell \neq k} \left(\frac{w_{j\ell\cdot}}{w_{j\cdot\cdot}}\right)^2 \mathrm{Var}(\overline{Y}^{w}_{j\ell\cdot}|U_j)\right] \\
&= \left(1 - \frac{w_{jk\cdot}}{w_{j\cdot\cdot}}\right)^2 \left(\frac{\sigma^2}{w_{jk\cdot}} + a\right) + \sum_{\ell \neq k} \left(\frac{w_{j\ell\cdot}}{w_{j\cdot\cdot}}\right)^2 \left(\frac{\sigma^2}{w_{j\ell\cdot}} + a\right) \\
&= \frac{\sigma^2}{w_{jk\cdot}} \left(1 - \frac{w_{jk\cdot}}{w_{j\cdot\cdot}}\right) + a\left(1 - 2\frac{w_{jk\cdot}}{w_{j\cdot\cdot}} + \sum_{\ell} \left(\frac{w_{j\ell\cdot}}{w_{j\cdot\cdot}}\right)^2\right)
\end{aligned}$$

This yields

$$\begin{aligned}
&E\left[\sum_{k=1}^{K_j} w_{jk\cdot}(\overline{Y}^{w}_{jk\cdot} - \overline{Y}^{ww}_{j\cdot\cdot})^2\right] \\
&\qquad = \sigma^2(K_j - 1) \;+\; a\left(w_{j\cdot\cdot} - 2\sum_{k} \frac{w^2_{jk\cdot}}{w_{j\cdot\cdot}} + \sum_{\ell} \frac{w^2_{j\ell\cdot}}{w_{j\cdot\cdot}}\right)
\end{aligned}$$

and thus

$$\begin{aligned}
&E\left[\sum_{j=1}^{J}\sum_{k=1}^{K_j} w_{jk\cdot}(\overline{Y}^{w}_{jk\cdot} - \overline{Y}^{ww}_{j\cdot\cdot})^2\right] \\
&\qquad = \sigma^2 \sum_{j}(K_j - 1) \;+\; a\left(w_{\cdots} - \sum_{j}\sum_{k} \frac{w^2_{jk\cdot}}{w_{j\cdot\cdot}}\right)
\end{aligned}$$

which we solve for a. Since $\hat{\sigma}^2$ is unbiased for σ^2 we conclude that (3.1) gives an unbiased estimator of a.

We turn to part (b) of the theorem and first note that by (3.3)

$$E[\mathrm{Var}(\overline{Y}^{w}_{jk\cdot}|U_j)] = \frac{a}{z_{jk}} \qquad \text{and hence} \qquad E[\mathrm{Var}(\overline{Y}^{zw}_{j\cdot\cdot}|U_j)] = \frac{a}{z_{j\cdot}}\,,$$

by the conditional independence assumption in Assumption 1(b). Now, since $E(\overline{Y}^{zw}_{j\cdot\cdot}|U_j) = U_j$,

$$\begin{aligned}
\mathrm{Var}(\overline{Y}^{zw}_{j\cdot\cdot}) &= E[\mathrm{Var}(\overline{Y}^{zw}_{j\cdot\cdot}|U_j)] + \mathrm{Var}(E[\overline{Y}^{zw}_{j\cdot\cdot}|U_j]) \\
&= E[\mathrm{Var}(\overline{Y}^{zw}_{j\cdot\cdot}|U_j)] + \mathrm{Var}[U_j] \\
&= \frac{a}{z_{j\cdot}} + b
\end{aligned}$$

Note that since $E[Y_{jkt}] = \mu$ we have $E[\overline{Y}^{zw}_{j\cdot\cdot}] = E[\overline{Y}^{zzw}_{\cdot\cdot\cdot}] = \mu$, which justifies the first equality below, where we further use the independence of Assumption 1(a),

$$\begin{aligned}
&E\left[(\overline{Y}^{zw}_{j\cdot\cdot} - \overline{Y}^{zzw}_{\cdot\cdot\cdot})^2\right] = \mathrm{Var}\left[\overline{Y}^{zw}_{j\cdot\cdot} - \overline{Y}^{zzw}_{\cdot\cdot\cdot}\right] \\
&= \left(1 - \frac{z_{j\cdot}}{z_{\cdot\cdot}}\right)^2 \mathrm{Var}(\overline{Y}^{zw}_{j\cdot\cdot}) + \sum_{\ell\neq j}\left(\frac{z_{\ell\cdot}}{z_{\cdot\cdot}}\right)^2 \mathrm{Var}(\overline{Y}^{zw}_{\ell\cdot\cdot}) \\
&= \left(1 - \frac{z_{j\cdot}}{z_{\cdot\cdot}}\right)^2 \left(\frac{a}{z_{j\cdot}} + b\right) + \sum_{\ell\neq j}\left(\frac{z_{\ell\cdot}}{z_{\cdot\cdot}}\right)^2 \left(\frac{a}{z_{\ell\cdot}} + b\right) \\
&= \frac{a}{z_{j\cdot}}\left(1 - \frac{z_{j\cdot}}{z_{\cdot\cdot}}\right) + b\left(1 - 2\frac{z_{j\cdot}}{z_{\cdot\cdot}} + \sum_{\ell}\left(\frac{z_{\ell\cdot}}{z_{\cdot\cdot}}\right)^2\right)
\end{aligned}$$

Finally we get the result

$$E\left[\sum_{j=1}^{J} z_{j\cdot}(\overline{Y}^{zw}_{j\cdot\cdot} - \overline{Y}^{zzw}_{\cdot\cdot\cdot})^2\right] = a(J-1) + b\left(z_{\cdot\cdot} - \frac{\sum_j z_{j\cdot}^2}{z_{\cdot\cdot}}\right)$$

and since $\hat{a}$ is unbiased, this completes the proof of part (b) of the theorem. $\square$

4 Numerical example

In our applications of hierarchical credibility at Länsförsäkringar we have found that the differences between the pseudo and the unbiased estimators are usually rather small. For confidentiality reasons, we do not present the results here. Instead we use the artificial population in Section 3.3 of Dannenburg et al. (1996) for illustration. In their Table 3.1, the values of $\overline{Y}^{w}_{jk\cdot}$ are given, but not the original data Y_{jkt} themselves. This is, however, enough for our purposes, if we take the stated value $\hat{\sigma}^2 = 15.89$ as given and only estimate a and b. (Remember that our approach uses the same unbiased $\hat{\sigma}^2$ as Dannenburg et al.)

The results are given in Table 4.1. Supposedly due to round-off errors in the data, our pseudo-estimates differ slightly from those provided by Dannenburg et al. In parenthesis we give the values when all the weights w_{jkt} are set equal to one.

	True value	*Pseudo-est.*	*Unbiased est.*
a	1.000	1.152 (1.093)	1.209 (1.093)
b	25.000	25.309 (25.259)	25.300 (25.259)

Table 4.1: *Comparison of estimators for the data in Table 3.1 of Dannenburg et al. (1996). (Equal weights example in parenthesis.)*

The difference between the estimators is negligible for b and less than 5% for a. In this case, we know the true values since the population is artificially generated. The pseudo-estimator of a is somewhat closer to the true value here, but this might well be due to chance. In the case with equal weights and balanced design (all K_j equal and all T_{jk} equal) it is not hard to show that the pseudo-estimator equations have explicit solutions that are equal to our unbiased estimators – as verified here by the numbers given in parenthesis.

In their comparison of the unbiased and pseudo-estimator of the parameter a in the Bühlmann-Straub model, Dubey & Gisler (1981) found that "neither of the two estimators is universally better than the other" (in terms of variance). Since the estimators considered in the present paper are closely related to the quoted estimators, one might conjecture that the conclusion of Dubey & Gisler are valid in our hierarchical case, too.

Conditional on this conjecture being true, our choice in the hierarchical case falls on the unbiased estimators for the sake of simplicity in application.

5 References

Dannenburg, D.R., Kaas, R. & Goovaerts, M.J. (1996): *Practical actuarial credibility models.* IAE, Amsterdam.

Dubey, A. & Gisler, A. (1981): *On Parameter Estimators in Credibility.* MVSVM, 81:2, 187-212.

Goovaerts, M.J. & Hoogstad, W.J. (1987): *Credibility theory.* Survey of Actuarial Studies, Nationale-Nederlanden N.V., Rotterdam.

Sundt, B. (1987): *Book review of Goovaerts & Hoogstad (1987).* ASTIN Bulletin 17, No. 2.