Zeitschrift: Mitteilungen / Vereinigung Schweizerischer Versicherungsmathematiker = Bulletin / Association des Actuaires Suisses = Bulletin / Association of Swiss Actuaries

Band: - (1987)

Heft: 1

Artikel: Hierarchical credibility revisited

Autor: Bühlmann, Hans / Jewell, William S.

DOI: https://doi.org/10.5169/seals-967143

Hans Bühlmann, Zürich, and William S. Jewell, Berkeley

Hierarchical Credibility Revisited

0 Introduction

About ten years ago some very basic papers on hierarchical credibility appeared in the actuarial literature: *Taylor* [1974] and [1979], *de Vylder* [1976] and [1977], *Jewell* [1975]. The papers by *Sundt* [1979] and [1980] must also be mentioned.
During the visit of the first author at the University of California in spring 1986 we have revisited the problem of hierarchical credibility. We found it worthwhile to clarify some points and to give a presentation which unites the different working techniques. Our main result is the recursive procedure for evaluating hierarchical credibilities. This result is already contained in a much more general result by *Norberg* [1980].
The methods and principles described in this paper are applicable to quite general hierarchical models [multidimensional, regression models (after a suitable transformation)]. To keep our presentation simple we however stick exclusively to the *one dimensional case* in this paper.

1 The Model

The structure of our model is characterized by different levels

Level 0: At this level we have the data $\mathcal{D}$, consisting of the *observable* random variables

$$X_{ij}, \quad i=1,2,\dots,n_j \quad j=1,2,\dots,N$$

We interpret $X_{ij} \sim$ total claims produced by risk j in year i.

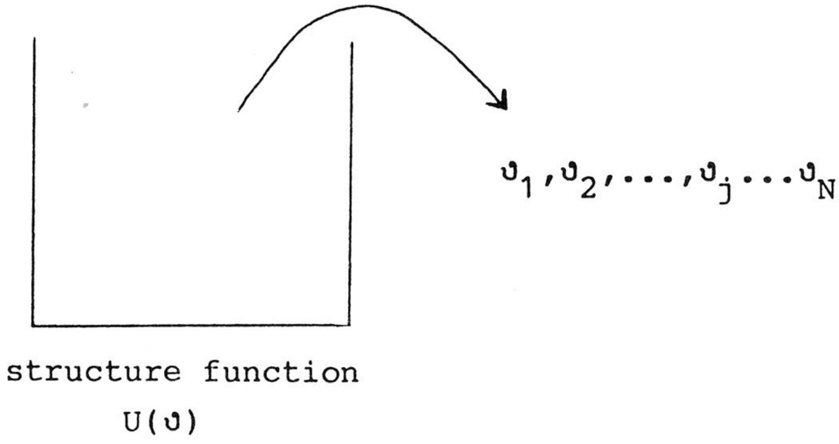
Level 1: At this level we have the *non observable* random variables

$$\vartheta_j; \quad j=1,2,\dots,N$$

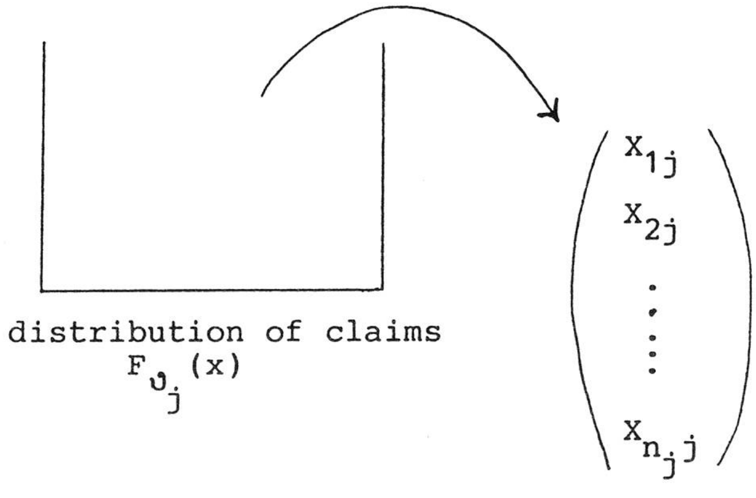
with the interpretation $\vartheta_j \sim$ risk quality for risk j.

The *standard credibility model* uses only these two levels by assuming that *all risk quality parameters* are drawn from *the same collective*. The complete drawing leading to the data is – in the standard credibility model – as follows:

Step 1: Draw (in an i.i.d. fashion) the risk qualities ϑ_j; $(j = 1, 2, \ldots, N)$ from the distribution $U(\vartheta)$



Step 2: For each risk j draw (in an i.i.d. fashion) the data X_{ij}; $(i = 1, 2, \ldots, n_j)$ from the distribution $F_{\vartheta_j}(x)$



The objection against this model is that we have created an intellectual framework, which permits *no* risk classification. Tariffs based on such a drastic simplification of the real world, have its merit and have also proved to be practicable. One of the most famous examples of such an application is the

1963 Swiss Automobile tariff. Nevertheless, one should learn from this paper that by passing over from the standard credibility model to a hierarchical credibility model one *can*, if one wishes, use credibility also *together* with risk classification schemes. The second message of this paper is that the practical handling of the hierarchical apparatus is not a hopeless affair. The key to keep the computational effort at a reasonable level is the *recursive calculation procedure* which we shall fully explain in this paper.
Let us then pass over from the standard credibility model to the hierarchical credibility model by *introducing more levels*. The basic idea is that

risks belong to cohorts,
cohorts belong to companies,
etc.

Of course, one is at liberty to introduce any number of levels as one pleases (and to use ones own names for the hierarchical grouping units – in fire insurance e.g. positions, books, classes, etc.). Again for simplicity of our presentation we stick here to two additional levels.

Level 2:

φ_k; $k = 1, 2, \ldots, s$
↑
quality for cohort k

Level 3: At this level we have the non observable *company qualities*

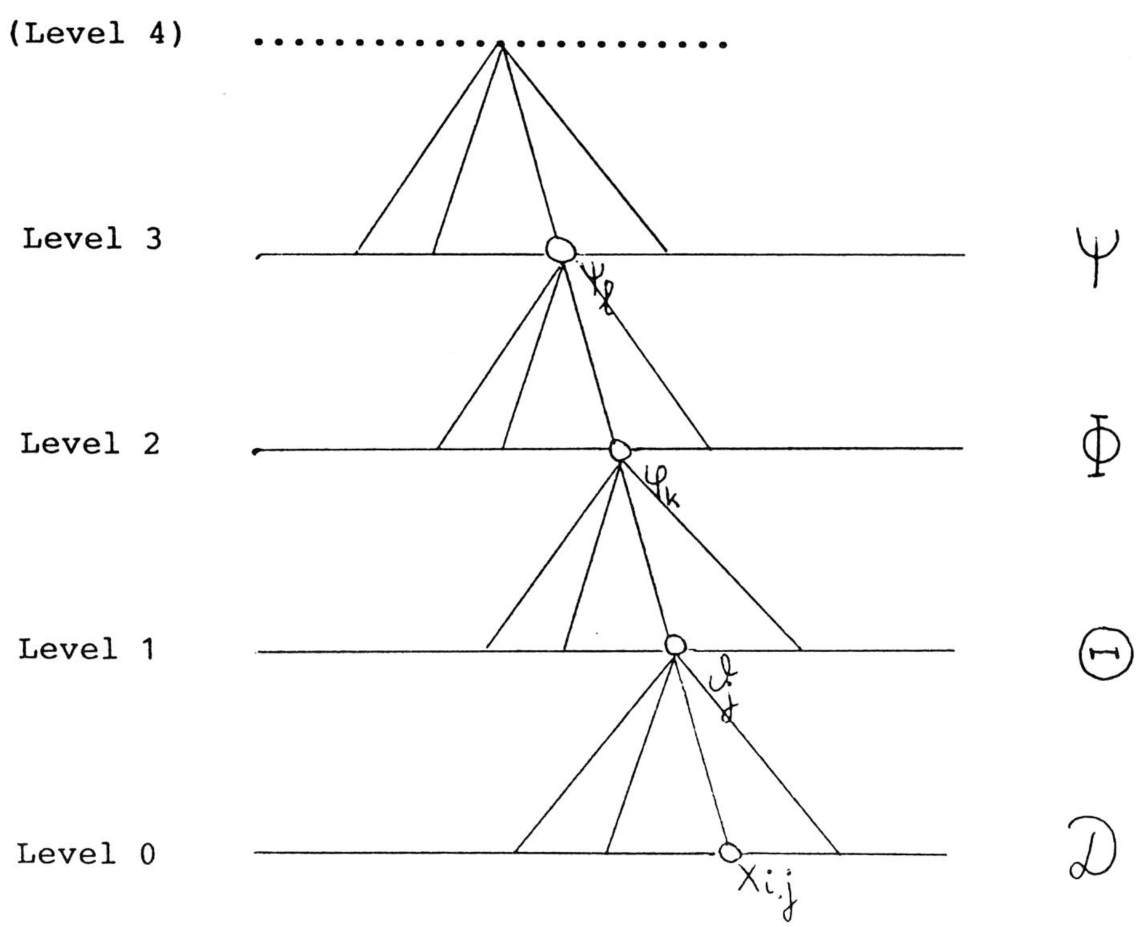
ψ_l; $l = 1, 2, \ldots, c$
↑
quality for company l

The whole structure becomes more understandable by the following graphical representation (next page)
The drawing shows only *one* data, *one* risk, *one* cohort, *one* company but it should be no problem to the reader to imagine the tree if all branches were drawn. It is also important that the reader notices our symbols (capital greek letters) for the *set of all variables* appearing at each level. We also use the notation

e.g. $\Phi(\psi_l)$ ~ all φ-variables deriving from company l
$\mathcal{D}(\varphi_k)$ ~ all data (X-variables) deriving from cohort k

for the quantities deriving from the ancestor in the bracket.

set of all variables



2 The Probability Structure in the Hierarchical Model

The probability structure in the hierarchical model is obtained by drawing the variables "from top down". This generates the complete probability distribution over the tree. In detail:

Level 3:

$\psi_l \quad (l = 1, 2, \ldots, c)$

are i.i.d. with *collective density* $r_3(\psi)$

Level 2: all

$\varphi_k \in \Phi(\psi_l)$

are *conditionally* i.i.d. with *company density* $r_2(\varphi/\psi_l)$

Level 1: all

$\vartheta_j \in \Theta(\varphi_k)$

are *conditionally* i.i.d. with *cohort density* $r_1(\vartheta/\varphi_k)$

Level 0: all

$X_{ij} \in \mathcal{D}(\vartheta_j)$

are *conditionally* i.i.d. with *risk density* $p(x/\vartheta_j)$

In the spirit of credibility theory we make no assumption on the parametric form of all the relevant densities and consider them as *unknown* to us.

Remarks:

1. On the top level the variables are i.i.d., on all other levels they are *conditionally* i.i.d.
2. Random variables in descending order in the tree

   $\psi_l, \quad \varphi_k, \quad \vartheta_j, \quad X_{ij}$ have the *Markov property* from left to right.

   Therefore in the conditional densities we indicate only the variable at the *lowest* known level.
3. It is clear how to construct a model with any number of levels.
   If we refer to this general model, we use

- the *same* notation as in this paper for levels 0 and 1,
- for levels $m \geqslant 2$ we then assume
  Level m: all

  $\varphi_k^{(m)} \in \Phi^{(m)}(\varphi_l^{(m+1)})$

  are conditionally i.i.d. with density $r_m(\varphi/\varphi_l^{(m+1)})$.

3 Relevant Random Variables and Constants at the different levels

Level 0 Data: $X_{ij} \sim$ total claims produced by risk j in year i

year \ risk	1	2	...	j	...	N
1	X_{11}	X_{12}	...	X_{1j}	...	X_{1N}
2	X_{21}	$\vdots$		X_{2j}		$\vdots$
$\vdots$	$\vdots$	$X_{n_2 2}$		$\vdots$		$\vdots$
$\vdots$	$\vdots$			$\vdots$		$X_{n_N N}$
$\vdots$	$X_{n_1 1}$			$\vdots$		
n_j				$X_{n_j j}$		

As – for fixed j – the X_{ij} are exchangeable random variables we can – without loss of information – summarize the risk data in *one* random variable for each risk

$$Y_1, Y_2, \ldots, Y_j, \ldots, Y_N$$

where

$$\boxed{Y_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij}.}$$

This reduction of the information into one variable is very basic. At the lowest level it is quite intuitive how this has to be done, namely just by taking the *average* over the observation. We call the resulting statistic the *linearly sufficient statistic* for the risk j and denote it by $B(\vartheta_j)$.

Hence

$$B(\vartheta_j) = Y_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij}$$

Level 1 (for each risk j)

The relevant quantities are

$$\mu(\vartheta_j) = E[X_{ij}/\vartheta_j] \qquad \sigma^2(\vartheta_j) = \mathrm{Var}[X_{ij}/\vartheta_j]$$

or relating to the summary statistic

$$E[Y_j/\vartheta_j] = \mu(\vartheta_j) \qquad \mathrm{Var}[Y_j/\vartheta_j] = \frac{\sigma^2(\vartheta_j)}{n_j}$$

Level 2 (for each cohort k)

$$M(\varphi_k) = E[\mu(\vartheta)/\varphi_k] = E[E[X_{ij}/\vartheta_j, \varphi_k]/\varphi_k]$$

$$F(\varphi_k) = E[\sigma^2(\vartheta)/\varphi_k] = E[\mathrm{Var}[X_{ij}/\vartheta_j, \varphi_k]/\varphi_k]$$

$$G(\varphi_k) = \mathrm{Var}[\mu(\vartheta)/\varphi_k] = \mathrm{Var}[E[X_{ij}/\vartheta_j, \varphi_k]/\varphi_k]$$

Observe that in the third expression one can choose any $\vartheta_j \in \Theta(\varphi_k)$. This third expression is anyway only written down to remind the reader of the Markov property of the sequence φ_k, ϑ_j, X_{ij} and will not be used in the following. On the higher levels we shall adhere to the simple conditioning by one variable which is more intuitive anyhow.

Level 3 (for each company l)

$$M(\psi_l) = E[M(\varphi)/\psi_l]$$

$$F(\psi_l) = E[F(\varphi)/\psi_l]$$

$$G(\psi_l) = E[G(\varphi)/\psi_l]$$

$$H(\psi_l) = \mathrm{Var}[M(\varphi)/\psi_l]$$

and finally at

Level 4 (collective, where we obtain the constants)

$M = E[M(\psi)]$	collective mean	
$F = E[F(\psi)]$	variation at level 0	(data)
$G = E[G(\psi)]$	variation at level 1	(risk)
$H = E[H(\psi)]$	variation at level 2	(cohorts)
$I = \mathrm{Var}[M(\psi)]$	variation at level 3	(companies)

Remark:
For the general model we would

$$\text{replace} \quad \begin{matrix} M \\ F \\ G \\ H \\ I \\ \\ \\ \end{matrix} \quad \text{by} \quad \begin{matrix} M \\ F_0 \\ F_1 \\ F_2 \\ \vdots \\ \\ F_L \end{matrix}$$

with

$$F_m = E[\mathrm{Var}[M(\varphi^{(m)})/\varphi^{(m+1)}]];$$

($\varphi^{(L+1)}$ is a degenerate random variable!)

4 The Problem

We want to find the credibility estimators for the
a) unknown risk quality:

$$\widehat{\mu(\vartheta_j)}, \quad \text{for all } j \quad \text{e.g. } \widehat{\mu(\vartheta_1)}$$

b) unknown cohort quality:

$$\widehat{M(\varphi_k)}, \quad \text{for all } k \quad \text{e.g. } \widehat{M(\varphi_1)}$$

c) unknown company quality:

$$\widehat{M(\psi_l)}, \quad \text{for all } l \quad \text{e.g. } \widehat{M(\psi_1)}$$

It will turn out that for solving a) one has automatically also to solve b) and c).
Hence, let us determine the inhomogeneous linear function

$$\widehat{\mu(\vartheta_1)} = \alpha_0 + \sum_{j=1}^{N} \alpha_j Y_j$$

$$\text{with} \quad E[(\mu(\vartheta_1) - \hat{\mu}(\vartheta_1))^2] = \min \; !$$

The brutal way, namely to solve a huge system of normal equations (*Jewell*, 1975) is not recommandable as a practicable solution. We rather rely on the Hilbert space technique advocated by *Taylor* (1974, 1979) and *de Vylder* (1976, 1977). According to Hilbert space terminology we write

$$\widehat{\mu(\vartheta_1)} = \text{Pro}[\mu(\vartheta_1)/Y_1, Y_2, \ldots, Y_N, 1]$$

and call it the

projection of $\mu(\vartheta_1)$ on the linear space spanned by the random variables Y_1, $Y_2, \ldots, Y_N$ and the constant 1.

Mathematical remark:
We work, of course, in the Hilbert space of square integrable random variables with scalar product $E[X_i \cdot X_j]$.

5 The Fundamental Principles for the Construction of Hierarchical Credibility Estimators

In the following

X, Y	are random variables
L, K	closed linear subspaces
a, b	scalars

Then we have

Principle I: (Linearity)

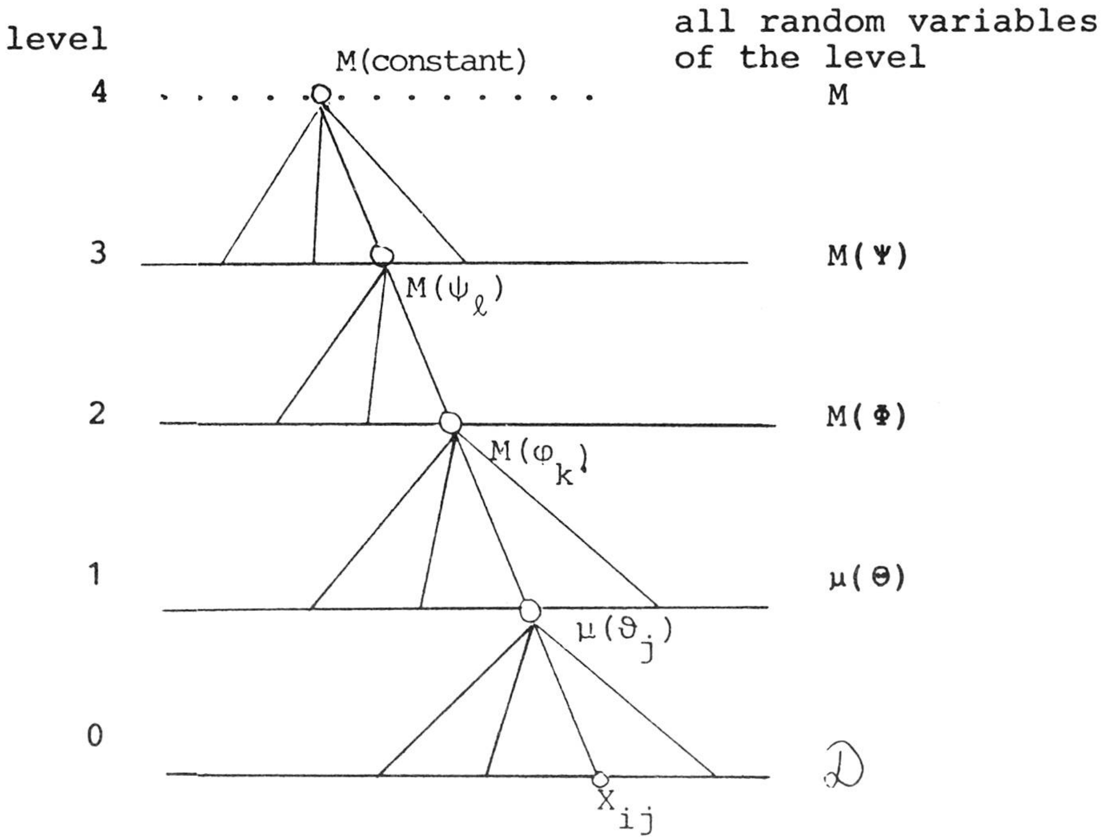
$$\mathrm{Pro}[aX + bY/L] = a\,\mathrm{Pro}[X/L] + b\,\mathrm{Pro}[Y/L]$$

Principle II: (Iterativity) Let $K \subset L$, then

$$\mathrm{Pro}[X/K] = \mathrm{Pro}[\mathrm{Pro}[X/L]/K]$$

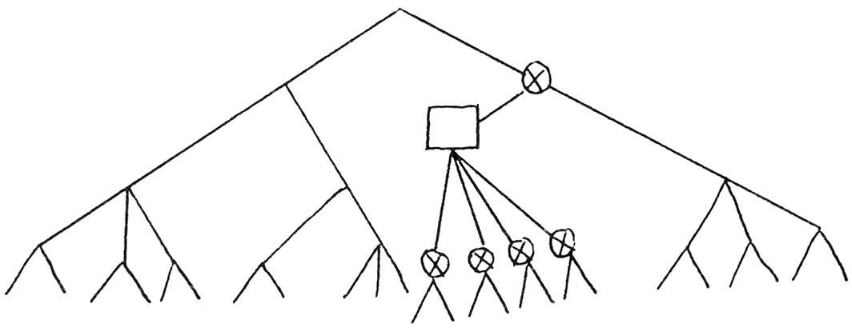
These two principles follow from the general properties of projection in any Hilbert space.

Principle III:
This needs more explanation. For this purpose let us draw again the hierarchical tree from Section 1. In this tree we, however, replace the random variables ϑ_j, φ_k, ψ_l by their corresponding means $\mu(\vartheta_j)$, $M(\varphi_k)$, $M(\psi_l)$ – which as functions of random variables are again random variables.



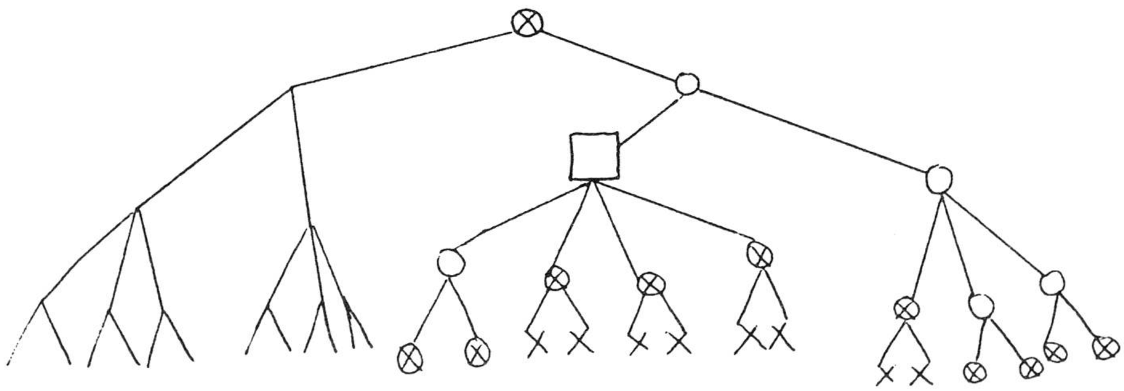
Principle III now says, that this tree has *the linear Markov property* [see *Witting*, 1987] i.e. projections of any element in the tree only depend on the *immediate known* neighbouring random variables in the tree.
The following examples may illustrate the idea of the linear Markov property.

i) ☐ random variable to be projected
all other random variables are known



ⓧ variables appearing in the projection

ii) ☐ random variable to be projected
x known random variables



ⓧ variables appearing in the projection
o "intermediate" projections

The proof of Principle III is given in the Appendix.

6 The Evaluation of the Hierarchical Credibility Estimator

The reader is reminded that the problem we want to solve is the

> projection of the mean variables of level 1 (e.g. $\mu(\vartheta_1)$) onto the levels 0 and 4 (data and constants)

in mathematical shorthand:
Find

$$\widehat{\mu(\vartheta_1)} = \mathrm{Pro}[\mu(\vartheta_1/\langle \mathcal{D}, M\rangle]$$

$\langle \ldots \rangle$ denotes the linear space spanned by the variables between "$\langle$" and "$\rangle$".

We also write

$L_{0,4}$ for $\langle \mathcal{D}, M\rangle$

and analogously

$L_{0,3,4}$ for $\langle \mathcal{D}, M(\Psi), M\rangle$
$L_{0,2,3,4}$ for $\langle \mathcal{D}, M(\Phi), M(\Psi), M\rangle$

Our program for the projection is *iterative*:

a) first project on $L_{0,2,3,4}$
b) then on $L_{0,3,4}$
c) then on $L_{0,4}$ which yields the result we want to obtain.

Execution of the program:

a) $\mathrm{Pro}[\mu(\vartheta_1)/\langle \mathcal{D}, M(\Phi), M(\Psi), M\rangle]$

$$= \mathrm{Pro}[\mu(\vartheta_1)/X_{11}, \ldots, X_{n_1 1}, M(\varphi_1)]$$

↑ by principle III

$$= \frac{n_1}{n_1 + F/G} Y_1 + \frac{F/G}{n_1 + F/G} M(\varphi_1)$$

↑ by classical credibility calculus

Observe that the formula above contains F and G and *not* $F(\varphi_1)$, $G(\varphi_1)$. This is a distinct difference to what we would obtain for a conditional expectation estimator. The projection estimator – due to its linear character – retains only the conditional means, whereas conditional variance components are averaged out over the collective.

Terminology. For a more intuitive background of the credibility estimator it is useful to adhere to the following terminology

i) $n_1 = \text{volume of risk 1} = V_1^{(1)}$

$n_j = V_j^{(1)}$

ii) $Z_1 = \dfrac{n_1}{n_1 + F/G} = Z_1^{(1)}$ credibility of risk 1

$Z_j = Z_j^{(1)}$

iii) $B(\vartheta_1) = Y_1$ linearly sufficient statistic for risk 1 (short: risk experience)

$B(\vartheta_j) = Y_j$

b) $\text{Pro}[\mu(\vartheta_1)/\langle \mathfrak{D}, M(\Psi), M\rangle]$

$$= \frac{n_1}{n_1 + F/G} Y_1 + \frac{F/G}{n_1 + F/G} \text{Pro}[M(\varphi_1)/\langle \mathfrak{D}, M(\Psi), M\rangle]$$

↑ by principles II and I

$$= Z_1^{(1)} B(\vartheta_1) + (1 - Z_1^{(1)}) \text{Pro}[M(\varphi_1)/\langle \mathfrak{D}, M(\Psi), M\rangle]$$

Through this part we are automatically lead to find the estimator of the *cohort quality*

The trick in evaluating *cohort quality* is to project first on $L_{0,1,3,4}$ and to use iterativity for obtaining the projection on $L_{0,3,4}$.

We then have

$$\mathrm{Pro}[M(\varphi_1)/\langle\mathcal{D}, \mu(\theta), M(\Psi), M\rangle] = \frac{r_1}{r_1 + G/H} \frac{\sum_{j=1}^{r_1} \mu(\vartheta_j)}{r_1} + \frac{G/H}{r_1 + G/H} M(\psi_1)$$

$r_1 \sim$ number of risks in cohort 1

By Principles II and I we get

$\mathrm{Pro}[M(\varphi_1)/\langle\mathcal{D}, M(\Psi), M\rangle]$

$$= \frac{1}{r_1 + G/H} \sum_{j=1}^{r_1} \left(Z_j^{(1)} Y_j + (1 - Z_j^{(1)}) \mathrm{Pro}[M(\varphi_1)/\langle\mathcal{D}, M(\Psi), M\rangle] \right)$$

$$+ \frac{G/H}{r_1 + G/H} M(\psi_1)$$

For abbreviation we write Pro for $\mathrm{Pro}[M(\varphi_1)\langle\mathcal{D}, M(\Psi), M\rangle]$ and collect terms. Hence,

$$\mathrm{Pro} \times \left(1 - \frac{1}{r_1 + G/H} \left(r_1 - \sum_{j=1}^{r_1} Z_j^{(1)} \right) \right) = \frac{\sum_{j=1}^{r_1} Z_j^{(1)} Y_j}{\sum_{j=1}^{r_1} Z_j^{(1)}} \frac{\sum_{j=1}^{r_1} Z_j^{(1)}}{r_1 + G/H} + \frac{G/H}{r_1 + G/H} M(\psi_1).$$

Multiplying both sides by $(r_1 + G/H)$ we obtain

$$\mathrm{Pro} \times \left(r_1 + \frac{G}{H} - r_1 + \sum_{j=1}^{r_1} Z_j^{(1)} \right) = \frac{\sum_{j=1}^{r_1} Z_j^{(1)} Y_j}{\sum_{j=1}^{r_2} Z_j^{(1)}} \cdot \sum_{j=1}^{r_1} Z_j^{(1)} + M(\psi_1) \cdot \frac{G}{H}.$$

This is where we can again adhere to our terminology of volume and linearly sufficient statistic.

We define

$$\underbrace{\sum_{j=1}^{r_1} Z_j^{(1)}}_{\text{sum of credibilities of risks}} = V_1^{(2)} = \text{volume of cohort 1}$$

$$\underbrace{\frac{\sum_{j=1}^{r_1} Z_j^{(1)} Y_j}{\sum_{j=1}^{r_2} Z_j^{(1)}}}_{\text{credibility weighted average of risk experiences}} = B(\varphi_1) = \text{linearly sufficient statistic for cohort 1 (short: cohort experience)}$$

Hence

$$\text{Pro}[M(\varphi_1)/\langle \mathcal{D}, M(\Psi), M\rangle] = \underbrace{\frac{V_1^{(2)}}{V_1^{(2)} + G/H}}_{\substack{Z_1^{(2)} \\ \text{credibility of cohort 1}}} B(\varphi_1) + \frac{G/H}{V_1^{(2)} + G/H} M(\psi_1)$$

c) The main step in the third projection – which can be performed in exactly the same way as under b) – is

$$\text{Pro}[M(\psi_1)/\langle \mathcal{D}, M\rangle] = \frac{V_1^{(3)}}{V_1^{(3)} + H/I} B(\psi_1) + \frac{H/I}{V_1^{(3)} + H/I} M$$

with

$$V_1^{(3)} = \sum_{\varphi_k \in \Phi(\psi_1)} V_k^{(2)} \qquad = \text{volume of company 1}$$

$$B(\psi_1) = \frac{1}{V_1^{(3)}} \sum_{\varphi_k \in \Phi(\psi_1)} V_k^{(2)} B(\varphi_k) = \text{linearly sufficient statistic of company 1}$$

7 Recursive Evaluation of Credibility Estimators in the Hierarchical Model

We have now all elements for the recursive procedure:

i) Evaluate $\widehat{M(\psi_l)}$ e.g. $\widehat{\mathrm{M}(\psi_1)}$

$$\widehat{M(\psi_1)} = Z_1^{(3)} B(\varphi_1) + (1 - Z_1^{(3)}) M;$$

where

$$Z_1^{(3)} = \frac{V_1^{(3)}}{V_1^{(3)} + H/I}.$$

ii) Evaluate $\widehat{M(\varphi_k)}$ e.g. $\widehat{\mathrm{M}(\varphi_1)}$

$$\widehat{M(\varphi_1)} = Z_1^{(2)} B(\varphi_1) + (1 - Z_1^{(2)}) \widehat{M(\psi_1)};$$

where

$$Z_1^{(2)} = \frac{V_1^{(2)}}{V_1^{(2)} + G/H}.$$

iii) Evaluate $\widehat{\mu(\vartheta_j)}$ e.g. $\widehat{\mu(\vartheta_1)}$

$$\widehat{\mu(\vartheta_1)} = Z_1^{(1)} Y_1 + (1 - Z_1^{(1)}) \widehat{M(\varphi_1)};$$

where

$$Z_1^{(1)} = \frac{n_1}{n_1 + F/G} = \frac{V_1^{(1)}}{V_1^{(1)} + F/G}.$$

It is worthwhile to summarize the two basic rules for the recursive evaluation.

Rule I: The volume measure at each level is obtained as the sum of the credibilities of the direct descendants at the next lower level.

Rule II: The linearly sufficient statistics (B-statistics) at each level are obtained by taking the credibility weighted average of the linearly sufficient statistics (B-statistics) at the next lower level.

8 Outlook

As indicated in the introduction this paper treats only the one dimensional case. It is very remarkable that the basic ideas explained here, carry over to the multidimensional and to the regression case (after a suitable transformation). As mentioned before, the results for the multidimensional and regression case can already be found in [*Norberg*, 1980]. Nevertheless, it would be recommendable to explicitely write up how the ideas presented here apply in the multidimensional case.
The problem of structural parameter estimation (estimation of M, F, G, H, I) is not presented here. We omit this important topic in order not to overload the present paper. However, we want to draw the attention of the reader to the note by *K.P. Mangold* appearing in the same number of the Bulletin. He basically uses the estimation technique proposed by *Sundt* in his 1986 paper.

Hans Bühlmann
Abt. Mathematik
ETH-Zentrum
8092 Zürich

William S. Jewell
Operations Research Center
University of California
Berkely, CA 94720
USA

Bibliography

Jewell, W.S. (1975): The use of collateral data in credibility theory: A hierarchical model, Giornale dell'Istituto Italiano degli Attuari 38, n° 1–2, pp. 1–16.
Mangold, K.P. (1987): Parameterschätzung im hierarchischen Credibility-Modell nach Sundt, BASA vol 87, 1.
Norberg, R. (1975): Credibility in the regression case, paper presented at the 1980 Oberwolfach Meeting on Risk Theory.
Sundt, Bjørn (1980): A multi-level hierarchical credibility regression model, SAJ 1980: 25–32 (No. 1).
Sundt, Bjørn (1979): A hierarchical credibility regression model, SAJ 1979: 107–113 (No. 8/3).
Sundt, Bjørn (1987): Some credibility regression models for the classification of individual passenger car models, to appear AB 17.1.
Taylor, G.C. (1977): Experience rating with credibility adjustment of the manual premium, AB 7, 323–336.
Taylor, G.C. (1979): Credibility analysis of general hierarchical models, SAJ 1979: 1–12 (No. 1).
de Vylder, Fl. (1976): Geometrical credibility, SAJ: 121–149 (No. 31).
de Vylder, Fl. (1977): Iterative credibility, BASA vol. 77, 1, 25–33.
Witting, Th. (1987): The linear Markov property in credibility theory, to appear in AB 17.1

Appendix

Proof of Principle III
(Linear Markov property of the hierarchical structure)

Remarks:

1. We only give the proof if the immediate neighbourhood of the projected variable is known. The general case then follows by iterativity of the projection.
2. Instead of projecting a very general variable we show that the linear Markov property holds for the projection of $M(\varphi_1)$. The reasoning can easily be transferred to the projection of any other variable. By working with the projection of $M(\varphi_1)$ we gain the advantage of a much more tractable notation.

Proof: By projecting $M(\varphi_1)$ on its immediate neighbours we obtain

$$\widehat{M(\varphi_1)} = \frac{r_1}{r_1 + G/H} \frac{1}{r_1} \sum_{j=1}^{r_1} \mu(\vartheta_j) + \frac{G/H}{r_1 + G/H} M(\psi_1)$$

we have to show that

$$\widehat{M(\varphi_1)} - M(\varphi_1)$$

is orthogonal to all random variables Y in the hierarchical tree.

Case a: Y is in direct ascending line of $M(\varphi_1)$

$$E[Y \cdot (\widehat{M(\varphi_1)} - M(\varphi_1))] = E[Y \cdot \underbrace{E[\widehat{M(\varphi_1)} - M(\varphi_1)/Y]}_{0} = 0$$

Case b: Y is in direct descending line of $M(\varphi_1)$

$$^{*}) \quad E[Y \cdot (\widehat{M(\varphi_1)} - M(\varphi_1))] = \frac{G/H}{r_1 + G/H} E[Y \cdot (M(\psi_1) - M(\varphi_1))]$$

$$+ \frac{1}{r_1 + G/H} \sum_{j=1}^{r_1} E[Y \cdot (\mu(\vartheta_1) - M(\varphi_1))]$$

Observe:

i) $E[Y\cdot(M(\psi_1)-M(\varphi_1))]=E[M(\varphi_1)\cdot(M(\psi_1)-M(\varphi_1))]=-H$

ii) $E[Y\cdot(\mu(\vartheta_j)-M(\varphi_1))]=0$

if Y is *not* directly descending from $\mu(\vartheta_j)$

$E[Y\cdot(\mu(\vartheta_j)-M(\varphi_1))]=G$

if Y is directly descending from $\mu(\vartheta_j)$ (which is only possible for *one* j)

Inserting i) and ii) into *) we obtain also 0.

Case c: Y is neither directly ascending nor descending from $M(\varphi_1)$. Then Y and $M(\varphi_1)$ have a common ancestor of *lowest level*. Call this common ancestor W. Observe that *given* W, Y and $M(\varphi_1)$ are conditionally independent. Hence,

$$E[\ Y\cdot(\widehat{M(\varphi_1)}-M(\varphi_1))]=E[E[Y/W]\cdot\underbrace{E[\widehat{M(\varphi_1)}-M(\varphi_1)/W]}_{0}]=0$$

This shows that in all cases $\widehat{M(\varphi_1)}-M(\varphi_1)$ is orthogonal on all variables Y. Consequently,

> $\widehat{M(\varphi_1)}$ is the *projection on the space spanned by all random variables* (and constants) Y q.e.d.

Summary

Based on some fundamental papers on hierarchical credibility that were published about 10 years ago, the problem in question is revisited by the authors.
The present survey on hierarchical credibility unites the different working techniques. To keep the presentation simple, the one-dimensional case is dealt with exclusivety although the methods and principles are applicable to quite general hierarchical models. The main result of the paper is the recursive procedure for evaluating hierarchical credibility.

Zusammenfassung

Gestützt auf einige grundlegende Arbeiten über hierarchische Kredibilität, welche vor etwa 10 Jahren publiziert worden sind, nehmen die beiden Autoren die Diskussion über dieses Thema wieder auf.
In der vorliegenden Übersicht geben sie eine Darstellung der hierarchischen Kredibilität, welche die unterschiedlichen Arbeitstechniken vereinheitlicht. Damit der Formelapparat nicht zu kompliziert wird, kommt nur der eindimensionale Fall zur Anwendung, obwohl die Methoden und Prinzipien auf allgemeine hierarchische Modelle anwendbar sind. Das Hauptresultat besteht in einem rekursiven Verfahren über die Abschätzung hierarchischer Kredibilität.

Résumé

Se basant sur quelques travaux fondamentaux publiés il y a environ 10 ans et se rapportant à la crédibilité hiérarchique, les auteurs reprennent le sujet.
Dans le présent article, ils proposent une présentation de la crédibilité hiérarchique qui unifie les différentes techniques de travail. Afin que les développements restent simples, il n'est traîté que le cas unidimensionnel bien que les méthodes et les principes soient applicables à des modèles hiérarchiques généraux. Le résultat principal consiste en un procédé récursif pour l'estimation de la crédibilité hiérarchique.