Multidimensional Credibility applied to estimating the frequency of big claims

Bühlmann Hans, Gisler Alois, Kollöffel Denise
ETH Zürich and Winterthur Insurance Company

Dedicated to Bill Jewell for his pathbreaking contributions to Credibility Theory.

Prof. Dr. H. Bühlmann
Nidelbadstrasse 22
CH - 8803 Rüschlikon

+41 1 724 10 84
hbuhl@math.ethz.ch

Prof. Dr. A. Gisler
Winterthur Insurance
P.O. Box 357
CH - 8401 Winterthur

+41 52 261 33 10
alois.gisler@winterthur.ch

D. Kollöffel
Buebenloostr. 1a
CH - 9500 Wil

denisek@student.ethz.ch

Abstract
First, the concept of multidimensional credibility is introduced and the corresponding credibility-estimators are derived. Then this methodology is applied to estimate the frequency of big claims. The multidimensional estimators are explicitely calculated and discussed. Next a simulation study is carried through and the behaviour of the multidimensional technique is studied and compared with the one-dimensional credibility estimators. Finally the theory is applied to a real data-set from practice.

Keywords: credibility, multidimensional, big claims, tariff making.

1 Introduction and Motivation

1.1 Historical Introduction

In his 1973 paper [7], Bill Jewell introduced the concept of multidimensional Credibility, which was a decisive contribution to the 1974 Berkeley Actuarial Research Conference on credibility theory [9]. In the same conference, Charly Hachemeister introduced his Credibility for Regression Models [6]. These two contributions must be seen as the pathbreaking contributions generalizing the credibility techniques into higher dimensions. Bill showed later in [8], how [6] could be derived from [7]. The other way could also have been taken.

1.2 Motivation

In many lines of business as for instance in motor insurance, the tariffs have become very refined in the sense that the premiums depend on many rating factors and co-variables. For calculating such tariffs it is standard to use multivariate statistical techniques like generalized linear modelling. These methods work well for the bulk of the data and for the big mass of "normal" claims. But what to do with the one or two percent of the biggest claims which, in some lines of business, are responsible for more than a half of the total claims amount and hence for more than a half of the pure risk premium? Often the tariff profile found by the multivariate statistical analysis is applied to the whole pure risk premium, which means that the risk profile for the big claims load is assumed to follow the same pattern as the risk profile of the claims load of the "normal" claims. Another approach sometimes found is to spread the big claims load equally over all risks, which is the same as assuming that all risks are equally exposed to big claims. The results coming out from these two somewhat extreme pure pragmatic ways of taking into account the big claims load might differ quite substantially. Hence whereas sophisticated multivariate statistical techniques are used for the bulk of the data, the methodology for taking into account the claims load of the big claims is still on a very simple actuarial level.

The present authors believe that multidimensional credibility is an appropriate answer in such situations, and this for two reasons: credibility is particularly suited in situations with scarce data and multidimensional credibility considers the observations of different categories (e.g. normal claims, big claims) simultaneously and let the data tell us what we can learn from the one category with respect to the other.

The treatment of the big claims load is only an example. There are many other situations where multidimensional credibility could be used. For instance in workmen's compensation in Switzerland we distinguish between accidents occurring 'at work' and those occurring 'not at work', and within each of these classifications, we differentiate between long-term benefits (for disability, or dependents) and short-term benefits (medical costs, wage payment interruptions). It is known that people who live 'dangerously' professionally, tend also to take more risks in their free time. There is therefore a connection between accidents at work and 'not at work'. This in turn means that we can learn something about the expected aggregate claim amount for accidents 'not at work' from the claims experience of accidents at work, and vice versa.

The essential feature of multidimensional credibility is to consider simultaneously the observations of different categories and to use this information in a methodologically consistent way.

To make this paper self-contained we summarize in the next section some main results from credibility theory. The specialists may skip section 2 and subsection 3.2.

2 Summary of some credibility results

We consider a random variable $\mu(\Theta)$ to be estimated on the basis of a random vector $\mathbf{X}$ with $\mathbf{X}' = (X_1, \ldots, X_J)$.

Definition 2.1 *An estimator* $\widehat{\mu(\Theta)}$ *of* $\mu(\Theta)$ *is said to be better than or equal to an estimator* $\widetilde{\mu(\Theta)}$, *if* $E\left[(\widehat{\mu(\Theta)} - \mu(\Theta))^2\right] \leq E\left[(\widetilde{\mu(\Theta)} - \mu(\Theta))^2\right]$, *that is we use quadratic loss.*

Definition 2.2 *The (inhomogeneous) credibility estimator* $\widehat{\widehat{\mu(\Theta)}}$ *of* $\mu(\Theta)$ *(based on* $\mathbf{X}$*) is the best estimator in the class*

$$L(\mathbf{X}, 1) := \left\{\widehat{\mu(\Theta)} : \ \widehat{\mu(\Theta)} = a_0 + \sum a_j X_j, \quad a_0, a_1, \cdots \in \mathbb{R}\right\}. \tag{2.1}$$

Definition 2.3 *The homogeneous credibility estimator* $\widehat{\widehat{\mu(\Theta)}}^{\text{hom}}$ *of* $\mu(\Theta)$ *(based on* $\mathbf{X}$*) is the best estimator in the class*

$$L_e(\mathbf{X}) := \left\{\widehat{\mu(\Theta)} : \ \widehat{\mu(\Theta)} = \sum a_j X_j, \ a_1, a_2, \cdots \in \mathbb{R}, \ \sum a_j \mu_{X_j} = E[\mu(\Theta)]\right\} \tag{2.2}$$

where $\mu_{X_j} = E[X_j]$.

The estimators $\widehat{\widehat{\mu(\Theta)}}$ and $\widehat{\widehat{\mu(\Theta)}}^{\text{hom}}$ defined in (2.1) and (2.2) as a solution to the least squares problem, are most elegantly understood as projections on the Hilbert space of all square integrable functions $\mathcal{L}^2(P)$. However we do not assume a special knowledge of the theory of Hilbert spaces in the following, so that the paper can easily be understood and followed by readers not familiar with this theory. The advantage of applying Hilbert space theory is that we can immediately apply our intuitive understanding of the properties of linear vector spaces and that certain features can be visualized and become much easier to understand.

Definition 2.4 (Hilbert space $\mathcal{L}^2(P)$)

1. $\mathcal{L}^2(P) := \left\{X : \ X \textit{ random variable with } E[X^2] = \int X^2 dP < \infty\right\}$
2. *If* X *and* Y *belong to* $\mathcal{L}^2(P)$ *then the inner product associated with the Hilbert space is given by*

$$< X, Y > := E[XY] \tag{2.3}$$

   and the corresponding norm is

$$\|X\| := < X, X >^{1/2} = E[X^2]^{1/2}. \tag{2.4}$$

In the following we assume that all considered random variables have finite second order moments and belong therefore to the Hilbert-space $\mathcal{L}^2(P)$.

Definition 2.5

1. *Two elements* X *and* $Y \in \mathcal{L}^2(P)$ *are* ***orthogonal*** $(X \perp Y)$, *if the inner product* $< X, Y >$ *equals* 0.
2. *For a closed subspace (or affine space)* $M \subset \mathcal{L}^2(P)$ *we define the orthogonal projection of* $Y \in \mathcal{L}^2(P)$ *on* M *as follows:* $Y^* \in M$ *is the* ***orthogonal projection of*** Y *on* M *(*$Y^* = \text{Pro}(Y \mid M)$ *) if* $Y - Y^* \perp Z_1 - Z_2$ *for all* $Z_1, Z_2 \in M$.

One can prove that the point Y^* always exists and is unique.

Theorem 2.6 *The following statements are equivalent:*

$$i) \quad Y^* = \mathrm{Pro}(Y \mid M). \tag{2.5}$$

$$ii) \quad Y^* \in M \text{ and } < Y - Y^*, Z - Y^* >= 0 \quad \text{for all } Z \in M. \tag{2.6}$$

$$iii) \quad Y^* \in M \text{ and } \|Y - Y^*\| \leq \|Y - Z\| \quad \text{for all } Z \in M. \tag{2.7}$$

Remarks:

- We call (2.6) the orthogonality condition.
- If M is a subspace, then the condition (2.6) can also be written as
$$< Y - Y^*, Z >= 0 \text{ for all } Z \in M. \tag{2.8}$$
- Note that $L(\mathbf{X}, 1)$ is a closed subspace and that $L_e(\mathbf{X})$ is a closed affine space of $\mathcal{L}^2(P)$.

Using this notation and these results, we can now reformulate the definition of the credibility estimators:

$$\widehat{\widehat{\mu(\Theta)}} := \mathrm{Pro}(\mu(\Theta) |\, L(\mathbf{X}, 1)). \tag{2.9}$$

$$\widehat{\widehat{\mu(\Theta)}}^{\mathrm{hom}} := \mathrm{Pro}(\mu(\Theta) |\, L_e(\mathbf{X})). \tag{2.10}$$

Looking at credibility estimators that way we can easily find the following

Theorem 2.7 (orthogonality conditions)

1. $\widehat{\widehat{\mu(\Theta)}}$ *is the (inhomogeneous) credibility estimator of* $\mu(\Theta)$, *iff*

$$i) \quad < \mu(\Theta) - \widehat{\widehat{\mu(\Theta)}}, 1 >= 0, \tag{2.11}$$

$$ii) \quad < \mu(\Theta) - \widehat{\widehat{\mu(\Theta)}}, X_j >= 0 \quad \text{for } j = 1, 2 \ldots, J. \tag{2.12}$$

2. $\widehat{\widehat{\mu(\Theta)}}^{\mathrm{hom}}$ *is the homogeneous credibility estimator of* $\mu(\Theta)$, *iff*

$$i) < \mu(\Theta) - \widehat{\widehat{\mu(\Theta)}}^{\mathrm{hom}}, 1 >= 0, \tag{2.13}$$

$$ii) < \mu(\Theta) - \widehat{\widehat{\mu(\Theta)}}^{\mathrm{hom}}, \widehat{\mu(\Theta)} - \widehat{\widehat{\mu(\Theta)}}^{\mathrm{hom}} >= 0 \quad \text{for all } \widehat{\mu(\Theta)} \in L_e(\mathbf{X}). \tag{2.14}$$

From the orthogonal conditions we easily obtain the well known normal equations.

Theorem 2.8 (normal equations)

$\widehat{\widehat{\mu(\Theta)}} = \widehat{a}_0 + \sum_j \widehat{a}_j X_j$ *is the (inhomogeneous) credibility estimator of* $\mu(\Theta)$, *iff the following normal equations are satisfied:*

$$i) \quad \sum_j \widehat{a}_j \mathrm{Cov}\,(X_j, X_k) = \mathrm{Cov}(\mu(\Theta), X_k), \quad k = 1, \ldots, n, \tag{2.15}$$

$$ii) \quad \widehat{a}_0 = \mu_0 - \sum_j \widehat{a}_j \, \mu_{X_j}. \tag{2.16}$$

3 Multidimensional Credibility

3.1 Starting point for multidimensional credibility

As an example let us consider the case, where we distinguish between "normal" and "big claims", and assume, that, in a first step, we are interested in the corresponding claim frequencies. For each risk i (the term risk is here and in the following also used as a synonym for risk category) and each year we have a two-dimensional vector of observations, namely the frequency of normal claims and the frequency of big claims. The expected value of the observation vector, given the risk characteristic Θ, is the vector

$$\boldsymbol{\mu}'(\Theta) = (\mu_1(\Theta), \mu_2(\Theta)). \tag{3.1}$$

Primarily we are interested in the pure risk premium

$$\boldsymbol{\mu}(\Theta) = \xi_1\mu_1(\Theta) + \xi_2\mu_2(\Theta), \tag{3.2}$$

where ξ_1 resp. ξ_2 are the expected values of the claim severities of normal resp. big claims, which might in more general models also differ between risks. For the purpose of simplicity of our presentation we omit the latter dependency in the applied part (section 4) of this paper. Because of the linearity property of the credibility estimator, we have

$$\widehat{\widehat{\mu(\Theta)}} = \xi_1\widehat{\widehat{\mu_1(\Theta)}} + \xi_2\widehat{\widehat{\mu_2(\Theta)}}. \tag{3.3}$$

Therefore, in order to find the credibility estimator for $\mu(\Theta)$, we need to determine the *credibility estimator of the vector* $\boldsymbol{\mu}(\Theta)$, i.e.

$$\widehat{\widehat{\boldsymbol{\mu}}}(\Theta)' = \left(\widehat{\widehat{\boldsymbol{\mu}}}_1(\Theta), \widehat{\widehat{\boldsymbol{\mu}}}_2(\Theta)\right) \tag{3.4}$$

Here, the vecotor $\boldsymbol{\mu}(\Theta)$ is two-dimensional. In the workmen's compensation example mentioned in section 1, $\boldsymbol{\mu}(\Theta)$ would be a vector with dimension four. The point is that the quantity to be estimated is *multi-dimensional* and hence the title *multi-dimensional credibility.*

We might first ask ourselves if it would make sense to estimate each of the components of the vector separately, on the basis of each of the associated observations (= corresponding components of the observation vector) and in this way reduce the problem to two one-dimensional problems. We could do that, if the components of the vector $\boldsymbol{\mu}(\Theta)$ were independent. However in most practical situations there is no a priori evidence that they are independent. On the contrary, there is mostly some evidence that there are dependences between the components and that we therefore can learn something from the observations of one component with respect to the others.

The essential point of *multidimensional credibility* is not only to *estimate a multidimensional vector* $\boldsymbol{\mu}(\Theta)$, but also *to use* for this estimation *simultaneously all available observations.*

3.2 The abstract multidimensional credibility model

In this subsection we concentrate on the essential mathematical structure for multidimensional credibility (Jewell[7]). It is very general, and we can use this structure also in other models like for instance in the regression model of Hachemeister [6]. In subsection 3.3 we will discuss a more concrete model.

3.2.1 Model

As already explained, the essential feature of multidimensional credibility is that the quantity to be estimated, $\boldsymbol{\mu}(\Theta)$, is multidimensional, i.e. a vector of length > 1.

In the abstract multidimensional credibility model, we assume that there is a random vector $\mathbf{X}$, having the same dimension as $\boldsymbol{\mu}(\Theta)$, which is individually unbiased, i.e. for which $E[\mathbf{X}|\,\Theta] = \boldsymbol{\mu}(\Theta)$. The goal is to find the credibility estimator based on $\mathbf{X}$.

The vector $\mathbf{X}$ considered here must not be the same as the observation vector, and as a rule this is not the case. As we will see, $\mathbf{X}$ is in general a compression of the original observations. For this reason the condition that $\boldsymbol{\mu}(\Theta)$ and $\mathbf{X}$ have the same length is no significant obstacle for practical applicatons. For the moment, we leave open the question of how the appropriate compression of the data is determined. The answer will be given in subsection 3.3 (see remark to (3.45)).

The vector $\boldsymbol{\mu}(\Theta)$, which we want to estimate is also not further specified here. However, in most concrete applications its meaning arises in a very natural way.

In the following, we want to derive the form of the credibility estimator of $\boldsymbol{\mu}(\Theta)$ based on this simple structure as just defined. For the homogeneous credibility estimator we need a corresponding collective. We will assume that we have a collective of contracts having the required structure. This leads us to the following

Model 3.1 (abstract multidimensional credibility)
We are given p-dimensional vectors $\{\mathbf{X}_i : i = 1, 2, \ldots, I\}$ *for which it holds that*

i) Conditionally, given Θ_i,

$$E[\mathbf{X}_i'|\,\Theta_i] = \boldsymbol{\mu}(\Theta_i)' = (\mu_1(\Theta_i), \mu_2(\Theta_i), \ldots, \mu_p(\Theta_i)), \tag{3.5}$$

$$\mathrm{Cov}(\mathbf{X}_i, \mathbf{X}_i'|\,\Theta_i) = \Sigma_i(\Theta_i). \tag{3.6}$$

ii) The pairs $(\Theta_1, \mathbf{X}_1)$, $(\Theta_2, \mathbf{X}_2), \ldots$ *are independent, and* Θ_1, $\Theta_2, \ldots$ *are independent and identically distributed.*

For each contract i, we are interested in estimating the associated vector $\boldsymbol{\mu}(\Theta_i)$.
This estimator will depend on the values of the structural parameters. The *structural parameters of the multidimensional credibility model are*

$$\boldsymbol{\mu} := E[\boldsymbol{\mu}(\Theta_i)], \tag{3.7}$$

$$S_i := E[\Sigma_i(\Theta_i)] = E[\mathrm{Cov}(\mathbf{X}_i, \mathbf{X}_i'|\,\Theta_i)], \tag{3.8}$$

$$T := \mathrm{Cov}\left(\boldsymbol{\mu}(\Theta_i), \boldsymbol{\mu}(\Theta_i)'\right). \tag{3.9}$$

Remarks:

- $\boldsymbol{\mu}$ is a vector of dimension p, S_i and T are $p \times p$ matrices.
- In insurance practice, it is almost always true that $S_i = w_i^{-1} \cdot S$, where w_i are suitable weights, respectively measures of volume. Therefore, in the cases which are relevant for practice, the matrices S_i have the same structure but are just weighted differently.

3.2.2 The (inhomogeneous) multidimensional credibility estimator

The multidimensional credibility estimator is defined componentwise. We thus have to find, for each component of the vector $\boldsymbol{\mu}(\Theta_i)$, the associated credibility estimator.

One can easily see that the credibility estimator for $\boldsymbol{\mu}(\Theta_i)$ depends only on the vector $\mathbf{X}_i$, and not on those $\mathbf{X}$-vectors associated with the other contracts. In order to simplify the notation, in this section we omit the subscript i.

The multidimensional credibility estimator is given by

$$\widehat{\widehat{\boldsymbol{\mu}(\Theta)}} := \text{Pro}(\boldsymbol{\mu}(\Theta) |\, L(\mathbf{X}, 1)), \tag{3.10}$$

where the projection operator should be understood componentwise and $L(\mathbf{X}, 1)$ is the linear subspace, spanned by 1 and the components of the vector $\mathbf{X}$, (cf. definition 2.2).

Now, for $k = 1, 2, \ldots, p$, let

$$\widehat{\widehat{\mu_k(\Theta)}} = a_{k0} + \sum_{j=1}^{p} a_{kj} X_j \tag{3.11}$$

be the credibility estimator for the components of $\boldsymbol{\mu}(\Theta)$. This estimator must satisfy the normal equations (see theorem 2.8):

$$\mu_k = a_{k0} + \sum_{j=1}^{p} a_{kj}\mu_j, \tag{3.12}$$

$$\text{Cov}(\mu_k(\Theta), X_m) = \sum_{j=1}^{p} a_{kj}\text{Cov}(X_j, X_m) \quad \text{for } m = 1, 2, \ldots, p. \tag{3.13}$$

More elegantly, using matrix notation we can write

$$\widehat{\widehat{\boldsymbol{\mu}(\Theta)}} = \mathbf{a} + A \cdot \mathbf{X}, \tag{3.14}$$

where

$$\mathbf{a} = \begin{pmatrix} a_{10} \\ \vdots \\ a_{p0} \end{pmatrix}, \qquad A = \begin{pmatrix} a_{11}, a_{12}, \ldots, a_{1p} \\ \vdots \\ a_{k1}, a_{k2}, \ldots, a_{kp} \\ \vdots \\ a_{p1}, a_{p2}, \ldots, a_{pp} \end{pmatrix}. \tag{3.15}$$

The normal equations (3.12) and (3.13) in matrix notation are given by

$$\begin{aligned} \boldsymbol{\mu} &= \mathbf{a} + A \cdot \boldsymbol{\mu}, &(3.16)\\ \operatorname{Cov}(\boldsymbol{\mu}(\Theta), \mathbf{X}') &= A \cdot \operatorname{Cov}(\mathbf{X}, \mathbf{X}'). &(3.17)\end{aligned}$$

Since

$$\begin{aligned} \operatorname{Cov}(\boldsymbol{\mu}(\Theta), \mathbf{X}') &= E\left[\operatorname{Cov}(\boldsymbol{\mu}(\Theta), \mathbf{X}' | \Theta)\right] + \operatorname{Cov}(\boldsymbol{\mu}(\Theta), E\left[\mathbf{X}' | \Theta\right]) \\ &= \operatorname{Cov}\left[\boldsymbol{\mu}(\Theta), \boldsymbol{\mu}(\Theta)'\right] &(3.18)\\ &= T, \\ \operatorname{Cov}(\mathbf{X}, \mathbf{X}') &= E\left[\operatorname{Cov}(\mathbf{X}, \mathbf{X}' | \Theta)\right] + \operatorname{Cov}(E\left[\mathbf{X} | \Theta\right], E\left[\mathbf{X}' | \Theta\right]) \\ &= S + T, &(3.19)\end{aligned}$$

we get from (3.17)

$$T = A \cdot (T + S) \tag{3.20}$$

and therefore

$$A = T \cdot (T + S)^{-1}. \tag{3.21}$$

Thus we have derived the following

Theorem 3.2 *The (inhomogeneous) multidimensional credibility estimator is given by*

$$\widehat{\widehat{\boldsymbol{\mu}(\Theta_i)}} = A_i \mathbf{X}_i + (I - A_i)\, \boldsymbol{\mu}, \tag{3.22}$$

where $\boldsymbol{\mu}$, T, S_i were defined in (3.7)-(3.9), and

$$A_i = T \cdot (T + S_i)^{-1}. \tag{3.23}$$

Remarks:

- The matrix $T + S_i$ is a covariance matrix and is therefore regular, for linearly independent components of the vector $\mathbf{X}$.
- Since T and S_i are symmetric, we also have that $A_i' = (T + S_i)^{-1} \cdot T$. Notice however that, in general $A_i \neq A_i'$, i.e. the credibility matrix A_i is not necessarily symmetric, even though T and S_i are symmetric.
- The estimator $\widehat{\widehat{\boldsymbol{\mu}(\Theta_i)}}$ is a weighted average of $\mathbf{X}_i$ and $\boldsymbol{\mu}$. The weighting results from the credibility matrix A_i and its complement $I - A_i$. It also holds that

$$\begin{aligned} A_i &= \left[T \cdot (T + S_i)^{-1} \cdot S_i\right] \cdot S_i^{-1} \\ &= \left(T^{-1} + S_i^{-1}\right)^{-1} \cdot S_i^{-1} \quad , &(3.24)\\ I - A_i &= \left(T^{-1} + S_i^{-1}\right)^{-1} \cdot T^{-1} \quad . &(3.25)\end{aligned}$$

  The weights are 'proportional' to their precision (inverse of the squared error matrix). In particular, we have that

$$\begin{aligned} T &= E\left[(\boldsymbol{\mu} - \boldsymbol{\mu}(\Theta_i)) \cdot (\boldsymbol{\mu} - \boldsymbol{\mu}(\Theta_i))'\right] &(3.26)\\ &= \text{squared error matrix of } \boldsymbol{\mu} \text{ with respect to } \boldsymbol{\mu}(\Theta_i), \\ S_i &= E\left[(\mathbf{X}_i - \boldsymbol{\mu}(\Theta_i)) \cdot (\mathbf{X}_i - \boldsymbol{\mu}(\Theta_i))'\right] &(3.27)\\ &= \text{squared error matrix of } \mathbf{X}_i \text{ with respect to } \boldsymbol{\mu}(\Theta_i). \end{aligned}$$

  The proportionality factor $\left(T^{-1} + S_i^{-1}\right)^{-1}$ is to be multiplied from the left.

3.2.3 The homogeneous multidimensional credibility estimator

In this section we will derive the homogeneous multidimensional credibility estimator. In contrast to the inhomogeneous estimator, this estimator has no constant term $\boldsymbol{\mu}$, i.e. the homogeneous credibility estimator contains an 'automatically built–in estimator' for the a priori expected value $\boldsymbol{\mu}$.

The homogeneous multidimensional credibility estimator is given by

$$\widehat{\widehat{\boldsymbol{\mu}(\Theta)}}^{\text{hom}} := \text{Pro}(\boldsymbol{\mu}(\Theta)|\, L_e(\mathbf{X}_1, \cdots, \mathbf{X}_I)), \tag{3.28}$$

where the projection operator is to be understood componentwise, and $L_e(\mathbf{X}_1, \ldots, \mathbf{X}_I)$ is again the affine subspace, spanned by the components of the vectors $\mathbf{X}_i$, $i = 1, 2, \ldots, I$, with the corresponding side constraints (componentwise unbiasedness).

Theorem 3.3 *The homogeneous multidimensional credibility estimator is given by*

$$\widehat{\widehat{\boldsymbol{\mu}(\Theta_i)}}^{\text{hom}} = A_i \mathbf{X}_i + (I - A_i)\, \widehat{\boldsymbol{\mu}}, \tag{3.29}$$

where T, S_i were defined in (3.7)-(3.9) and

$$A_i = T \cdot (T + S_i)^{-1}, \tag{3.30}$$

$$\widehat{\boldsymbol{\mu}} = \left(\sum_{i=1}^{I} A_i \right)^{-1} \sum_{i=1}^{I} A_i \mathbf{X}_i. \tag{3.31}$$

Remark:
(3.29) is obtained by replacing $\boldsymbol{\mu}$ in (3.22) by the credibility-weighted mean (3.31).

Proof of theorem 3.3: Since $L_e(\mathbf{X}_1, \ldots, \mathbf{X}_I) \subset L(\mathbf{X}_1, \ldots, \mathbf{X}_I, 1)$, because of the iterative property of projection operators we have that

$$\begin{aligned}
\widehat{\widehat{\boldsymbol{\mu}(\Theta_i)}}^{\text{hom}} &= \text{Pro}\,(\boldsymbol{\mu}(\Theta_i)|\, L_e(\mathbf{X}_1, \ldots, \mathbf{X}_I)) \\
&= \text{Pro}\,(\text{Pro}\,(\boldsymbol{\mu}(\Theta_i)|\, L(\mathbf{X}_1, \ldots, \mathbf{X}_I, 1))|\, L_e(\mathbf{X}_1, \ldots, \mathbf{X}_I)) \\
&= \text{Pro}\,(A_i \mathbf{X}_i + (I - A_i)\, \boldsymbol{\mu}|\, L_e(\mathbf{X}_1, \ldots, \mathbf{X}_I)) \\
&= A_i \mathbf{X}_i + (I - A_i)\, \text{Pro}\,(\boldsymbol{\mu}|\, L_e(\mathbf{X}_1, \ldots, \mathbf{X}_I))\,.
\end{aligned} \tag{3.32}$$

It remains then to show that

$$\widehat{\boldsymbol{\mu}} := \text{Pro}\,(\boldsymbol{\mu}|\, L_e(\mathbf{X}_1, \ldots, \mathbf{X}_I)) = \left(\sum_{i=1}^{I} A_i \right)^{-1} \sum_{i=1}^{I} A_i \mathbf{X}_i. \tag{3.33}$$

We want to check the orthogonality conditions (2.6), i.e. we must show that for all $\sum_{i=1}^{I} B_i \mathbf{X}_i$ with $\sum_{i=1}^{I} B_i = I$ it holds that

$$\text{Cov}\left(\widehat{\boldsymbol{\mu}}, \left(\sum_{i=1}^{I} B_i \mathbf{X}_i \right)' \right) = \text{Cov}\left(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\mu}}' \right)\ . \tag{3.34}$$

Because $\text{Cov}(\mathbf{X}_i, \mathbf{X}_i') = T + S_i$ and because of the independence of the $\mathbf{X}_i$ we have that

$$\begin{aligned}
\text{Cov}\left(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\mu}}'\right) &= \left(\sum_{i=1}^{I} A_i\right)^{-1} \cdot \left(\sum_{i=1}^{I} A_i \,\text{Cov}\left(\mathbf{X}_i, \mathbf{X}_i'\right)\, A_i'\right) \cdot \left(\sum_{i=1}^{I} A_i'\right)^{-1} \\
&= \left(\sum_{i=1}^{I} A_i\right)^{-1} \cdot \left(\sum_{i=1}^{I} T\, A_i'\right) \cdot \left(\sum_{i=1}^{I} A_i'\right)^{-1} \qquad (3.35) \\
&= \left(\sum_{i=1}^{I} A_i\right)^{-1} \cdot T.
\end{aligned}$$

Analogously (because $\sum_{i=1}^{I} B_i = I$)

$$\begin{aligned}
&\text{Cov}\left(\widehat{\boldsymbol{\mu}}_i, \left(\sum_{i=1}^{I} B_i \mathbf{X}_i\right)'\right) \\
&= \left(\sum_{j=1}^{I} A_j\right)^{-1} \left(\sum_{i=1}^{I} A_i \,\text{Cov}\left(\mathbf{X}_i, \mathbf{X}_i'\right)\, B_i'\right) \\
&= \left(\sum_{j=1}^{I} A_j\right)^{-1} \sum_{i=1}^{I} T \cdot B_i' \qquad (3.36) \\
&= \left(\sum_{j=1}^{I} A_j\right)^{-1} T\ , \qquad (3.37)
\end{aligned}$$

with which we have shown (3.34). This ends the proof of theorem 3.3 . □

3.3 The multidimensional Bühlmann-Straub model

3.3.1 Motivation and interpretation

In the Bühlmann-Straub model, there is associated with each risk i and for each year j a one-dimensional observation X_{ij} $(j = 1, 2, \ldots, n)$ satisfying the model assumptions of the well known Bühlmann-Straub model [2]. In the multidimensional *Bühlmann-Straub model,* these observations are multidimensional, i.e. the observations associated with the risk i in the year j are themselves vectors

$$\mathbf{X}_{ij}' = (X_{ij}^{(1)}, X_{ij}^{(2)}, \ldots, X_{ij}^{(p)})\ . \qquad (3.38)$$

Example 1: Workmen's Compensation (see section 1)

$$\mathbf{X}_{ij}' = (X_{ij}^{(1)}, X_{ij}^{(2)}, X_{ij}^{(3)}, X_{ij}^{(4)}), \qquad (3.39)$$

where
$\mathbf{X}_{ij}'$ = observation vector of the risk i in year j,
$X_{ij}^{(1)}, X_{ij}^{(2)}$ = observed claim ratios for short-term benefits and long-term benefits for accidents occurring at work and
$X_{ij}^{(3)}, X_{ij}^{(4)}$ = the corresponding observed claim ratios for accidents not occurring at work.

Example 2: Layering

$$\mathbf{X}'_{ij} = (X_{ij}^{(1)}, X_{ij}^{(2)}, \ldots, X_{ij}^{(p)}), \tag{3.40}$$

where
$X_{ij}^{(1)}$ = the claim ratio (or the claim frequency) associated with the claims in layer 1 (e.g. claims less than CHF 1'000),
$X_{ij}^{(2)}$ = the claim ratio (or the claim frequency) associated with the claims in layer 2 (e.g. claims between CHF 1'000 and CHF 50'000),
and so on.
A special case of this example would be the simple differentiation between normal claims and big claims, which is equivalent to a two layer system (claims under and above a specified threshold).

Example 3: Differentiation according to claim type, e.g. physical damage versus bodily injury

$$\mathbf{X}'_{ij} = (X_{ij}^{(1)}, X_{ij}^{(2)}), \tag{3.41}$$

where
$X_{ij}^{(1)}$ = the claim ratio (or the claim frequency) associated with physical damage and
$X_{ij}^{(2)}$ = the claim ratio (or the claim frequency) associated with bodily injury claims.

3.3.2 Definition of the model

We are given a collective of I risks or "risk categories", and let
$\mathbf{X}_{ij}$ = observation vector (length p) of risk i in year j ,
w_{ij} = associated known weight.

The weights w_{ij} can depend on the risk i as well as on the year j, but we assume that they are constant within the observation vector $\mathbf{X}_{ij}$.

Model 3.4 *The risk i is characterized by its individual risk characteristics ϑ_i, which is itself the realization of a random variable Θ_i, and it holds that:*

A1: Conditionally, given Θ_i, the $\mathbf{X}_{ij}$, $j = 1, 2, \ldots, n$, are independent with

$$E\left[\mathbf{X}_{ij} \mid \Theta_i, w_{ij}\right] = \boldsymbol{\mu}(\Theta_i) \quad \textit{(vector of length p)}, \tag{3.42}$$

$$\operatorname{Cov}\left(\mathbf{X}_{ij}, \mathbf{X}'_{ij} \mid \Theta_i, w_{ij}\right) = \frac{1}{w_{ij}} \cdot \Sigma(\Theta_i) \quad \textit{(}p \times p \textit{ matrix).} \tag{3.43}$$

A2: The pairs $\{(\Theta_i, \mathbf{X}_i) : i = 1, 2, ...I\}$ are independent, and Θ_1, $\Theta_2, \ldots$ are independent and identically distributed, where $\mathbf{X}'_i = \left(\mathbf{X}'_{i1}, \mathbf{X}'_{i2}, ..., \mathbf{X}'_{in}\right)$.

Remarks:

- For the case of simplicity we assume a common volume measure for all components (model assumption (3.43)). However the model can be easily extended to cover the situation of different volumes for each component. In this case with volume measure $w_{ij}^{(k)}$ for component k, the conditional covariance matrix would have the form $\left(\frac{\sigma_{kl}(\Theta_i)}{(w_{ij}^{(k)} w_{ij}^{(l)})^{1/2}}\right)_{k,l=1,\ldots,p}$.

- The functions $\boldsymbol{\mu}(\cdot)$ and $\Sigma(\cdot)$ are independent of i and j. The dependence of the conditional covariance on the individual risk is modelled using the volume measures w_{ij}.

3.3.3 Multidimensional credibility formula in the Bühlmann-Straub model

The multidimensional Bühlmann-Straub model does not fully fit into the framework of the abstract multidimensional credibility model described in subsection 3.2, in that not just one, but many vectors of observations are available for the ith risk. In order that we can use the results derived in section 3.2, we must first compress the sequence of multidimensional one-period claim ratios $\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in}$ into one single observation vector $\mathbf{B}_i$. We explicitly choose

$$\mathbf{B}_i = \left(\sum\nolimits_j \frac{w_{ij}}{w_{i\bullet}} X_{ij}^{(1)}, \cdots, \sum\nolimits_j \frac{w_{ij}}{w_{i\bullet}} X_{ij}^{(p)}\right)', \tag{3.44}$$

the average over time of the observations. This choice is intuitive. Also, it is easy to check that

$$\mathbf{B}_i = \operatorname{Pro}(\boldsymbol{\mu}(\Theta_i) | \, L_e^{ind}(\mathbf{X}_i)) \,, \tag{3.45}$$

$$\text{where } L_e^{ind}(\mathbf{X}_i) = \left\{\widehat{\boldsymbol{\mu}(\Theta_i)} : \widehat{\boldsymbol{\mu}(\Theta_i)} = \sum\nolimits_j a_{ij} X_{ij} \,; E\left[\widehat{\boldsymbol{\mu}(\Theta_i)} \middle| \Theta_i\right] = \boldsymbol{\mu}(\Theta_i)\right\}. \tag{3.46}$$

Remark:
One can show that the right hand side of (3.45) is the appropriate compression in all multidimensional credibility models. More details can be found in [1].

As we will now see, $\mathbf{B}_i$ is indeed an allowable compression, i.e. the credibility estimator depends on the data only via $\mathbf{B}_i$. For the compressed data $\mathbf{B}_i$ we have that

$$E\left[\mathbf{B}_i | \, \Theta_i\right] = \boldsymbol{\mu}(\Theta_i), \tag{3.47}$$

$$\begin{aligned} \operatorname{Cov}\left(\mathbf{B}_i, \mathbf{B}_i' | \, \Theta_i\right) &= \sum_{j=1}^{n} \frac{w_{ij}^2}{w_{i\bullet}^2} \operatorname{Cov}\left(\mathbf{X}_{ij}, \mathbf{X}_{ij}' \middle| \, \Theta_i\right) \\ &= \sum_{j=1}^{n} \frac{w_{ij}}{w_{i\bullet}^2} \Sigma(\Theta_i) = \Sigma(\Theta_i)/w_{i\bullet}. \end{aligned} \tag{3.48}$$

We can now apply theorem 3.2 to $\mathbf{B}_i$.

Theorem 3.5 *The (inhomogeneous)* ***credibility estimator*** *in the multidimensional Bühlmann-Straub model (model 3.4) is given by*

$$\widehat{\widehat{\boldsymbol{\mu}(\Theta_i)}} = A_i \mathbf{B}_i + (I - A_i)\, \boldsymbol{\mu}, \tag{3.49}$$

$$\textit{where} \quad A_i = T \cdot (T + S/w_{i\bullet})^{-1}. \tag{3.50}$$

Proof of theorem 3.5: From theorem 3.2 we know that $\widehat{\widehat{\boldsymbol{\mu}(\Theta_i)}}$ is the credibility estimator based on $\mathbf{B}_i$. It remains to show that $\widehat{\widehat{\boldsymbol{\mu}(\Theta_i)}}$ is also the credibility estimator based on all data (admissibility of the compression $\mathbf{B}_i$). To do this we must check the orthogonality conditions

$$\boldsymbol{\mu}(\Theta_i) - \widehat{\widehat{\boldsymbol{\mu}(\Theta_i)}} \perp \mathbf{X}_{kj} \quad \text{for all } k, j. \tag{3.51}$$

For $k \neq i$ we have that (3.51) holds because of the independence of the risks. It remains then to show (3.51) for $k = i$.

$$\begin{aligned} & E\left[\left(\boldsymbol{\mu}(\Theta_i) - A_i\mathbf{B}_i - (I - A_i)\,\boldsymbol{\mu}\right) \cdot \mathbf{X}'_{ij}\right] \\ = \;& A_i E\left[\left(\boldsymbol{\mu}(\Theta_i) - \mathbf{B}_i\right) \cdot \mathbf{X}'_{ij}\right] + (I - A_i) E\left[\left(\boldsymbol{\mu}(\Theta_i) - \boldsymbol{\mu}\right) \cdot \mathbf{X}'_{ij}\right] \end{aligned} \tag{3.52}$$

From

$$\mathbf{B}_i = \operatorname{Pro}(\,\boldsymbol{\mu}(\Theta_i) |\, L_e^{ind}\,(\mathbf{X}_i))$$

it follows that

$$\begin{aligned} E\left[\left(\boldsymbol{\mu}(\Theta_i) - \mathbf{B}_i\right) \cdot \mathbf{X}'_{ij}\right] &= E\left[\left(\boldsymbol{\mu}(\Theta_i) - \mathbf{B}_i\right) \cdot \mathbf{B}'_i\right] \\ &= E\left[\left(\boldsymbol{\mu}(\Theta_i) - \mathbf{B}_i\right) \cdot \left(\mathbf{B}_i - \boldsymbol{\mu}(\Theta_i)\right)'\right] \\ &= -E\left[\operatorname{Cov}\left(\mathbf{B}_i, \mathbf{B}'_i |\, \Theta_i\right)\right] \\ &= -S/w_{i\bullet}\;. \end{aligned} \tag{3.53}$$

For the second summand in (3.52), we get by conditioning on Θ_i

$$\begin{aligned} & E\left[\left(\boldsymbol{\mu}(\Theta_i) - \boldsymbol{\mu}\right) \cdot \mathbf{X}'_{ij}\right] \\ = \;& E\left[\left(\boldsymbol{\mu}(\Theta_i) - \boldsymbol{\mu}\right) \cdot \boldsymbol{\mu}(\Theta_i)'\right] \\ = \;& E\left[\left(\boldsymbol{\mu}(\Theta_i) - \boldsymbol{\mu}\right) \cdot \left(\boldsymbol{\mu}(\Theta_i) - \boldsymbol{\mu}\right)'\right] \\ = \;& T\;. \end{aligned} \tag{3.54}$$

Plugging $A_i = T \cdot (T + S/w_{i\bullet})^{-1}$ as well as (3.53) and (3.54) in (3.51) we see that

$$\begin{aligned} & E\left[\left(\boldsymbol{\mu}(\Theta_i) - A_i\mathbf{B}_i - (I - A_i)\,\boldsymbol{\mu}\right) \cdot \mathbf{X}'_{ij}\right] \\ = \;& -A_i\,(S/w_{i\bullet} + T) + T = 0\;. \end{aligned} \tag{3.55}$$

This completes the proof of theorem 3.5. □

4 Application to estimating the big claim frequency

4.1 The problem

Als already mentioned earlier, in many lines of business the premiums depend on several risk factors and covariates. To calculate such refined tariffs, there are used sophisticated multivariate statistical techniques like general linear modelling. Usually, the big claims are truncated in such a multivariate analysis. These big claims are very small in number, often only 1% - 2%, however the loss burden originating from this category of claims is often more than a half of the total claims load. Of course one does not forget or neglect this "big claims" burden in the tariff. But it is usually taken into account very pragmatically (cf. subsection 1.2). Hence there is still the problem of how to assess the pure risk premium originating from big claims. This paper aims to make a contribution to this problem on the basis of multidimensional credibility theory.

The problem of big claims in tariff-making has already been considered in the literature (see [4], [5], [10]). The new idea in this paper is, that there are simultaneoulsy considered pairs of data, namely the one belonging to "normal" claims and the one attributed to big claims. The problem fits perfectly into the framework of multidimensional credibility, and in this case, the dimension is only two. This has the advantage that we can calculate the entries of the credibility matrix explicitly and that we can interpret them. Moreover we will show how the structural parameters can be estimated from the data.

In this paper we will concentrate on the claim numbers. But the methodology can also be used for claim severities or for the observed pure risk premiums (= observed average total claim amount per risk). Indeed, D. Kollöffel investigated in [11] both, the credibility estimators of the claim frequencies as well as the one of the average claim severities.

4.2 Model-assumptions and Notations

We consider a portfolio of risks $i = 1, 2, ..., I$ each of them with observation periods $j = 1, 2, ..., J$. Let us denote by

$w_{ij} =$ number of year risks (weights),
$N_{ij} =$ number of claims.

We further assume that there are two categories of claims: "normal" claims and "big" claims and that each claim belongs to one of the two categories. The most common classification is to use a threshold criterion: claims with claim amount above the threshold are classified as big claims and the others as normal claims. But also other criteria like for instance the distinction between pure physical damage and bodily injury claims are possible.

Hence we can define the following two-dimensional observation vectors:

$$\mathbf{N}'_{ij} = (N_{ij}^{(1)}, N_{ij}^{(2)}), \text{ where}$$
$$N_{ij}^{(1)} = \text{number of normal claims,}$$
$$N_{ij}^{(2)} = \text{number of big claims,}$$

and

$$\mathbf{F}'_{ij} := \frac{\mathbf{N}'_{ij}}{w_{ij}} = (F_{ij}^{(1)}, F_{ij}^{(2)}), \text{ where}$$
$$F_{ij}^{(1)} = \text{frequency of normal claims,}$$
$$F_{ij}^{(2)} = \text{frequency of big claims.}$$

Model 4.1 *The following assumptions hold true:*

A1: $N_{ij}^{(1)}$ *and* $N_{ij}^{(2)}$ *are conditionally, given* Θ_i*, independent;*

A2: The observed frequency-vectors $\mathbf{F}_{ij}$ fulfill the conditions of the multidimensional Bühlmann-Straub model 3.4.

In particular the following relations are fulfilled in model 4.1:

$$
\begin{aligned}
E\left[\mathbf{F}_{ij}\middle|\,\Theta_i, w_{ij}\right]' &= \left(\mu_1\left(\Theta_i\right), \mu_2\left(\Theta_i\right)\right), && (4.1)\\
\operatorname{Cov}\left(\mathbf{F}_{ij}, \mathbf{F}'_{ij}\middle|\,\Theta_i, w_{ij}\right) &= \frac{1}{w_{ij}} \cdot \Sigma(\Theta_i), && (4.2)\\
\text{where } \Sigma\left(\Theta_i\right) &= \begin{pmatrix} \sigma_1^2\left(\Theta_i\right) & 0 \\ 0 & \sigma_2^2\left(\Theta_i\right) \end{pmatrix}. && (4.3)
\end{aligned}
$$

Claim numbers are very often assumed to be Poisson-distributed in practice. In this case, the variance is equal to the expected value. Therefore we will also consider the following model which includes the Poisson-case but which is more general.

Model 4.2 *In addition to the assumptions of model 4.1 we assume that there exists a constant $a \in \mathbb{R}$, such that*

$$
\begin{aligned}
\sigma_1^2\left(\Theta_i\right) &= a \cdot \mu_1\left(\Theta_i\right),\\
\sigma_2^2\left(\Theta_i\right) &= a \cdot \mu_2\left(\Theta_i\right).
\end{aligned}
$$

4.3 Credibility estimators

We want to estimate for each risk i the claim frequency of normal claims as well as the one of big claims, i.e. the vector

$$
\boldsymbol{\mu}\left(\Theta_i\right)' = \left(\mu_1\left(\Theta_i\right), \mu_2\left(\Theta_i\right)\right). \tag{4.4}
$$

With the assumptions of model 4.1 we find the following structural parameters

$$
\begin{aligned}
S &= \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}, && (4.5)\\
T &= \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix}, && (4.6)\\
\text{where } \rho &= \frac{\tau_{12}}{\tau_1\tau_2} = \frac{\operatorname{Cov}\left(\mu_1\left(\Theta_i\right), \mu_2\left(\Theta_i\right)\right)}{\sqrt{\operatorname{Var}\mu_1\left(\Theta_i\right) \cdot \operatorname{Var}\mu_1\left(\Theta_i\right)}}. && (4.7)
\end{aligned}
$$

The compressed observation vector $\mathbf{B}_i$ as defined in (3.44) is

$$
\begin{aligned}
\mathbf{B}_i &= (F_i^{(1)}, F_i^{(2)})', && (4.8)\\
\text{where } F_i^{(1)} &:= \sum_{j=1}^{J} \frac{w_{ij}}{w_{i\bullet}} F_{ij}^{(1)}, && (4.9)\\
F_i^{(2)} &:= \sum_{j=1}^{J} \frac{w_{ij}}{w_{i\bullet}} F_{ij}^{(2)}. && (4.10)
\end{aligned}
$$

We can now calculate the credibility matrix by applying theorem 3.5. To simplify notation we drop the index i and we write w instead of $w_{i\bullet}$.

We obtain

$$
\begin{aligned}
A &= T \cdot \left(T + \frac{S}{w}\right)^{-1} = \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix} \begin{pmatrix} \tau_1^2 + \frac{\sigma_1^2}{w} & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 + \frac{\sigma_2^2}{w} \end{pmatrix}^{-1} \qquad (4.11) \\
&= \frac{1}{\left(\tau_1^2 + \frac{\sigma_1^2}{w}\right)\left(\tau_2^2 + \frac{\sigma_2^2}{w}\right) - \rho^2\tau_1^2\tau_2^2} \cdot \begin{pmatrix} \tau_1^2\left(\tau_2^2 + \frac{\sigma_2^2}{w}\right) - \rho^2\tau_1^2\tau_2^2 & \frac{1}{w}\rho\tau_1\tau_2\sigma_1^2 \\ \frac{1}{w}\rho\tau_1\tau_2\sigma_2^2 & -\rho^2\tau_1^2\tau_2^2 + \tau_2^2\left(\tau_1^2 + \frac{\sigma_1^2}{w}\right) \end{pmatrix}.
\end{aligned}
$$

(4.11) does not look very appealing. In order to find simpler expressions for the credibility weights which are easier to interpret, we introduce the credibility coefficients

$$
\kappa_l = \frac{\sigma_l^2}{\tau_l^2}, \; l = 1, 2 \quad . \qquad (4.12)
$$

After some straightforward calculations we obtain for the entries of the credibility matrix A the following formulae

$$
\begin{aligned}
a_{11} &= \frac{w}{w + \kappa_1\left(1 + \frac{\rho^2}{1-\rho^2+\frac{\kappa_2}{w}}\right)}, \qquad (4.13) \\
a_{22} &= \frac{w}{w + \kappa_2\left(1 + \frac{\rho^2}{1-\rho^2+\frac{\kappa_1}{w}}\right)}, \qquad (4.14) \\
a_{12} &= \frac{\rho}{w} \cdot \frac{\sigma_1}{\sigma_2} \cdot \frac{\sqrt{\kappa_1\kappa_2}}{\left(1 + \frac{\kappa_1}{w}\right)\left(1 + \frac{\kappa_2}{w}\right) - \rho^2}, \qquad (4.15) \\
a_{21} &= \frac{\rho}{w} \cdot \frac{\sigma_2}{\sigma_1} \cdot \frac{\sqrt{\kappa_1\kappa_2}}{\left(1 + \frac{\kappa_1}{w}\right)\left(1 + \frac{\kappa_2}{w}\right) - \rho^2}. \qquad (4.16)
\end{aligned}
$$

Thus we have found the following

Theorem 4.3 (cred. estimator for the claim frequency) *The credibility estimators in model 4.1 for the claim frequencies are*

$$
\begin{aligned}
\widehat{\widehat{\mu_1(\Theta_i)}} &= \mu_1 + a_{11}^{(i)}\left(F_i^{(1)} - \mu_1\right) + a_{12}^{(i)}\left(F_i^{(2)} - \mu_2\right), \qquad (4.17) \\
\widehat{\widehat{\mu_2(\Theta_i)}} &= \mu_2 + a_{21}^{(i)}\left(F_i^{(1)} - \mu_1\right) + a_{22}^{(i)}\left(F_i^{(2)} - \mu_2\right),
\end{aligned}
$$

where $a_{11}^{(i)}, \ldots, a_{22}^{(i)}$ *are given by (4.13) - (4.16), if* w *is replaced there by* $w_{i\bullet}$.

Remarks

- With the additional assumption of model 4.2 including the Poisson-case, (4.15) and (4.16) become

$$a_{12} = \frac{\rho}{w} \cdot \sqrt{\frac{\mu_1}{\mu_2}} \cdot \frac{\sqrt{\kappa_1 \kappa_2}}{\left(1+\frac{\kappa_1}{w}\right)\left(1+\frac{\kappa_2}{w}\right)-\rho^2}, \tag{4.18}$$

$$a_{21} = \frac{\rho}{w} \cdot \sqrt{\frac{\mu_2}{\mu_1}} \cdot \frac{\sqrt{\kappa_1 \kappa_2}}{\left(1+\frac{\kappa_1}{w}\right)\left(1+\frac{\kappa_2}{w}\right)-\rho^2}. \tag{4.19}$$

It is convenient to bring the elements of the observation vector on the same scale, otherwise the credibility weights are not directly comparable. We therefore consider the "standardized" frequencies

$$\widetilde{F}_i^{(1)} := \frac{F_i^{(1)}}{\mu_1}, \qquad \widetilde{F}_i^{(2)} := \frac{F_i^{(2)}}{\mu_2}. \tag{4.20}$$

In practice, one encounters very often a multiplicative tariff structure, which has become even more popular with the GLM-modelling. Here one is interested in the relativities of the risk premiums between risks, i.e. in the ratio of the risk premium of risk i to the overall premium. We therefore want to find the estimators of

$$\widetilde{\boldsymbol{\mu}}\left(\Theta_i\right)' = \left(\widetilde{\mu}_1\left(\Theta_i\right), \widetilde{\mu}_2\left(\Theta_i\right)\right) := \left(\frac{\mu_1\left(\Theta_i\right)}{\mu_1}, \frac{\mu_2\left(\Theta_i\right)}{\mu_2}\right). \tag{4.21}$$

Of course, by using the linear property of credibility estimators, they can be immediately derived from theorem 4.3. For this purpose it is useful to introduce

$$\begin{aligned} \nu_i^{(1)} &:= w_{i\bullet}\mu_1, & \nu_i^{(2)} &:= w_{i\bullet}\mu_2, \\ \widetilde{\kappa}_1 &:= \tfrac{\mu_1^2}{\tau_1^2} = \mu_1\kappa_1, & \widetilde{\kappa}_2 &:= \tfrac{\mu_2^2}{\tau_2^2} = \mu_2\kappa_2. \end{aligned} \tag{4.22}$$

Then we obtain

Corollary 4.4 (credibility estimator for the standardized frequency)

$$\widehat{\widehat{\widetilde{\mu}_1(\Theta_i)}} = 1 + \widetilde{a}_{11}^{(i)}\left(\widetilde{F}_i^{(1)} - 1\right) + \widetilde{a}_{12}^{(i)}\left(\widetilde{F}_i^{(2)} - 1\right), \tag{4.23}$$

$$\widehat{\widehat{\widetilde{\mu}_2(\Theta_i)}} = 1 + \widetilde{a}_{21}^{(i)}\left(\widetilde{F}_i^{(1)} - 1\right) + \widetilde{a}_{22}^{(i)}\left(\widetilde{F}_i^{(2)} - 1\right) \tag{4.24}$$

where

$$\widetilde{a}_{11}^{(i)} = \frac{\nu_1^{(i)}}{\nu_1^{(i)} + \widetilde{\kappa}_1 \left(1 + \frac{\rho^2}{1-\rho^2+\frac{\widetilde{\kappa}_2}{\nu_2^{(i)}}}\right)}, \tag{4.25}$$

$$\widetilde{a}_{12}^{(i)} = \rho \cdot \frac{\mu_2}{\mu_1} \cdot \frac{\sigma_1}{\sigma_2} \cdot \frac{\sqrt{\frac{\widetilde{\kappa}_1}{\nu_1^{(i)}} \frac{\widetilde{\kappa}_2}{\nu_2^{(i)}}}}{\left(1 + \frac{\widetilde{\kappa}_1}{\nu_1^{(i)}}\right)\left(1 + \frac{\widetilde{\kappa}_2}{\nu_2^{(i)}}\right) - \rho^2}, \tag{4.26}$$

$$\widetilde{a}_{21}^{(i)} = \rho \cdot \frac{\mu_1}{\mu_2} \cdot \frac{\sigma_2}{\sigma_1} \cdot \frac{\sqrt{\frac{\widetilde{\kappa}_1}{\nu_1^{(i)}} \frac{\widetilde{\kappa}_2}{\nu_2^{(i)}}}}{\left(1 + \frac{\widetilde{\kappa}_1}{\nu_1^{(i)}}\right)\left(1 + \frac{\widetilde{\kappa}_2}{\nu_2^{(i)}}\right) - \rho^2}, \tag{4.27}$$

$$\widetilde{a}_{22}^{(i)} = \frac{\nu_2^{(i)}}{\nu_2^{(i)} + \widetilde{\kappa}_2 \left(1 + \frac{\rho^2}{1-\rho^2+\frac{\widetilde{\kappa}_1}{\nu_1^{(i)}}}\right)}. \tag{4.28}$$

Remarks

- With the additional assumption in model 4.2 , (4.26) and (4.27) become

$$\widetilde{a}_{12}^{(i)} = \rho \cdot \sqrt{\frac{\mu_2}{\mu_1}} \cdot \frac{\sqrt{\frac{\widetilde{\kappa}_1}{\nu_1^{(i)}} \frac{\widetilde{\kappa}_2}{\nu_2^{(i)}}}}{\left(1 + \frac{\widetilde{\kappa}_1}{\nu_1^{(i)}}\right)\left(1 + \frac{\widetilde{\kappa}_2}{\nu_2^{(i)}}\right) - \rho^2}, \tag{4.29}$$

$$\widetilde{a}_{21}^{(i)} = \rho \cdot \sqrt{\frac{\mu_1}{\mu_2}} \cdot \frac{\sqrt{\frac{\widetilde{\kappa}_1}{\nu_1^{(i)}} \frac{\widetilde{\kappa}_2}{\nu_2^{(i)}}}}{\left(1 + \frac{\widetilde{\kappa}_1}{\nu_1^{(i)}}\right)\left(1 + \frac{\widetilde{\kappa}_2}{\nu_2^{(i)}}\right) - \rho^2}. \tag{4.30}$$

- Note that the weights occurring in the above formulae are now the a priori expected claim numbers $\nu_1^{(i)}$ and $\nu_2^{(i)}$.

- Note also that the credibility coefficients are now

$$\widetilde{\kappa}_l = \mathrm{Var}\left(\widetilde{\mu}_l\left(\Theta_i\right)\right) = \mathrm{cv}\left(\mu_l\left(\Theta_i\right)\right)^{-2} \qquad l = 1,2 \quad , \tag{4.31}$$

$$\text{where } \mathrm{cv}\left(x\right) = \frac{\sqrt{Var\left[X\right]}}{E\left[X\right]} \text{ (coefficient of variance) .} \tag{4.32}$$

  Hence the credibility coefficients have a direct meaning and can easily be interpreted.

Since the standardized credibility estimators (cf. theorem 4.4) are just the non-standardized credibility estimators (cf. theorem 4.3) divided by the corresponding a priori expected values μ_1 resp. μ_2, we immediately find that the following relations between the credibility weights are satisfied.

Proposition 4.5

$$
\begin{aligned}
\widetilde{a}_{11} &= a_{11}, && (4.33)\\
\widetilde{a}_{22} &= a_{22}, && (4.34)\\
\widetilde{a}_{12} &= \frac{\mu_2}{\mu_1} a_{12}, && (4.35)\\
\widetilde{a}_{21} &= \frac{\mu_1}{\mu_2} a_{21}. && (4.36)
\end{aligned}
$$

4.4 Discussion of the credibility estimators

In the following we discuss the credibility estimator of the "standardized" frequencies as given by corollary 4.4 .

First we compare the multidimensional with the one-dimensional credibility estimators (each component estimated only based on the observations of that component). Let us denote by

$$
\begin{aligned}
\widetilde{\alpha}_1 &= \frac{\nu_1}{\nu_1 + \widetilde{\kappa}_1}, && (4.37)\\
\widetilde{\alpha}_2 &= \frac{\nu_2}{\nu_2 + \widetilde{\kappa}_2} && (4.38)
\end{aligned}
$$

the one-dimensional credibility weights.

Of course for $\rho = 0$ it holds that

$$\widetilde{a}_{11} = \widetilde{\alpha}_1, \quad \widetilde{a}_{22} = \widetilde{\alpha}_2, \quad \widetilde{a}_{12} = \widetilde{a}_{21} = 0 \ . \qquad (4.39)$$

i.e. the multidimensional credibility estimators coincide with the one-dimensional ones.

For $\rho \neq 0$ a comparison of (4.25) and (4.28) with (4.37) and (4.38) shows that

$$\widetilde{a}_{11} < \widetilde{\alpha}_1 \ , \quad \widetilde{a}_{22} < \widetilde{\alpha}_2 \ , \qquad (4.40)$$

hence the diagonal elements of the multidimensional credibility-matrix are smaller than the corresponding one-dimensional credibility weights, which is an interesting and not obvious result. The explanation is that this smaller weight given to the diagonal observation is compensated by giving some weight to the observation of the off-diagonal component.

As to the off-diagonal elements of the multidimensional credibility-matrix, we can conclude from (4.26) and (4.27) that

$$\frac{\widetilde{a}_{21}}{\widetilde{a}_{12}} = \frac{\mu_1^2}{\mu_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}. \qquad (4.41)$$

With the additional assumption of model 4.2 we have

$$\frac{\widetilde{a}_{21}}{\widetilde{a}_{12}} = \frac{\mu_1}{\mu_2}. \qquad (4.42)$$

(4.42) is an interesting relationship too. Since the number of big claims is in many lines of business in the range of 1% - 2%, it follows that $\widetilde{a}_{21}$ is 50 - 100 times bigger than $\widetilde{a}_{12}$. Hence we can learn much more from the observed standardized normal claim frequency with respect to the expected big claim frequency than vice versa, which is of course not very surprising.

Next we consider what happens when the considered risk group is very big, i.e. when the weights w, or as a consequence, the expected number of normal claims tends to infinity.

We then obtain for $|\rho| < 1$

$$\lim_{w\to\infty} \widetilde{a}_{11} = \lim_{w\to\infty} \widetilde{a}_{22} = 1, \tag{4.43}$$

$$\lim_{w\to\infty} \widetilde{a}_{12} = \lim_{w\to\infty} \widetilde{a}_{21} = 0. \tag{4.44}$$

Hence the credibility estimators tend to the observed mean of each of the component for $w \to \infty$, a result which is not surprising.

For $\rho = \pm 1$ we get

$$\lim_{w\to\infty} \widetilde{a}_{11} = \lim_{w\to\infty} \frac{w\mu_1}{w\mu_1 + \widetilde{\kappa}_1 + \frac{\widetilde{\kappa}_1}{\widetilde{\kappa}_2} w\mu_2} \tag{4.45}$$

$$= \frac{\widetilde{\kappa}_2\mu_1}{\widetilde{\kappa}_2\mu_1 + \widetilde{\kappa}_1\mu_2}, \tag{4.46}$$

and analogously

$$\lim_{w\to\infty} \widetilde{a}_{22} = \frac{\widetilde{\kappa}_1\mu_2}{\widetilde{\kappa}_2\mu_1 + \widetilde{\kappa}_1\mu_2}. \tag{4.47}$$

It is interesting to note that for $w \to \infty$ the sum of the diagonal credibility weights (right hand side of (4.46) and (4.47)) is equal to one.

For the off-diagonal elements of the credibility matrix we obtain for $\rho = \pm 1$

$$\lim_{w\to\infty} \widetilde{a}_{12} = \pm \lim_{w\to\infty} \frac{\mu_2}{\mu_1} \cdot \frac{\sigma_1}{\sigma_2} \cdot \frac{\sqrt{\frac{\widetilde{\kappa}_1}{w\mu_1} \cdot \frac{\widetilde{\kappa}_2}{w\mu_2}}}{\frac{\widetilde{\kappa}_1}{w\mu_1} + \frac{\widetilde{\kappa}_2}{w\mu_2} + \frac{\widetilde{\kappa}_1\widetilde{\kappa}_2}{w^2\mu_1\mu_2}} \tag{4.48}$$

$$= \pm\frac{\mu_2}{\mu_1} \cdot \frac{\sigma_1}{\sigma_2} \cdot \sqrt{\frac{\widetilde{\kappa}_1}{\mu_1} \cdot \frac{\widetilde{\kappa}_2}{\mu_2} \cdot} \frac{\mu_1\mu_2}{\widetilde{\kappa}_1\mu_1 + \widetilde{\kappa}_2\mu_2}$$

$$= \pm\frac{\mu_2}{\mu_1} \cdot \frac{\sigma_1}{\sigma_2} \cdot \sqrt{\frac{\widetilde{\kappa}_1}{\widetilde{\kappa}_2}} \cdot a_{22}, \tag{4.49}$$

and analogously

$$\lim_{w\to\infty} \widetilde{a}_{21} = \pm\frac{\mu_1}{\mu_2} \cdot \frac{\sigma_2}{\sigma_1} \cdot \sqrt{\frac{\widetilde{\kappa}_2}{\widetilde{\kappa}_1}} \cdot a_{11}. \tag{4.50}$$

With the additional assumption of model 4.2 we get

$$\lim_{w\to\infty} \widetilde{a}_{12} = \sqrt{\frac{\widetilde{\kappa}_1}{\widetilde{\kappa}_2}} \cdot a_{22} \ , \tag{4.51}$$

$$\lim_{w\to\infty} \widetilde{a}_{21} = \sqrt{\frac{\widetilde{\kappa}_2}{\widetilde{\kappa}_1}} \cdot a_{11} \ . \tag{4.52}$$

A case of special interest is the situation of an equal profile between normal and big claim frequency, i.e. the assumption, that

$$\mu_2(\Theta_i) = \frac{\mu_2}{\mu_1} \cdot \mu_1(\Theta_i). \tag{4.53}$$

Then $\rho = 1$ and $\widetilde{\kappa}_1 = \widetilde{\kappa}_2$, and we get for $w \to \infty$

$$\lim_{w\to\infty} \widetilde{a}_{11} = \lim_{w\to\infty} \widetilde{a}_{21} = \frac{\mu_1}{\mu_1+\mu_2} \ , \tag{4.54}$$

$$\lim_{w\to\infty} \widetilde{a}_{22} = \lim_{w\to\infty} \widetilde{a}_{12} = \frac{\mu_2}{\mu_1+\mu_2} \ . \tag{4.55}$$

and hence

$$\widehat{\widehat{\widetilde{\mu}_1(\Theta_i)}} = \widehat{\widehat{\widetilde{\mu}_2(\Theta_i)}} = 1 + \frac{\mu_1}{\mu_1+\mu_2}\left(\widetilde{F}_i^{(1)} - 1\right) + \frac{\mu_2}{\mu_1+\mu_2}\left(\widetilde{F}_i^{(2)} - 1\right). \tag{4.56}$$

(4.56) is an intuitively very plausible result. Note that the assumption (4.53) is equivalent to

$$\widetilde{\mu}_2(\Theta_i) = \widetilde{\mu}_1(\Theta_i). \tag{4.57}$$

Hence the estimator for the standardized frequency must be the same. Of course, this must hold true not only for $w \to \infty$, but in general. Indeed one finds after some calculation that

$$\widetilde{a}_{11} = \widetilde{a}_{21} \qquad \text{and} \qquad \widetilde{a}_{22} = \widetilde{a}_{12}, \tag{4.58}$$

and hence

$$\widehat{\widehat{\widetilde{\mu}_1(\Theta_i)}} = \widehat{\widehat{\widetilde{\mu}_2(\Theta_i)}} = 1 + \widetilde{a}_{11}^{(i)} \cdot \left(\widetilde{F}_i^{(1)} - 1\right) + \widetilde{a}_{22}^{(i)} \cdot \left(\widetilde{F}_i^{(2)} - 1\right). \tag{4.59}$$

The following figures 1-3 show the shape of the credibility weights (y-axis) in dependence of ν_1 (x-axis) for $\widetilde{\kappa}_1 = \widetilde{\kappa}_2 = 11$ $(\mathrm{cv}(\mu_l(\Theta_i) \simeq 30\%)$ and for the case, where the overall frequency of big claims is 50 times smaller than the one of normal claims. For technical reasons, the weights are denoted in the legend of the graphics by akl (k,l = 1,2) instead of $\widetilde{a}_{kl}$.

Figure 1 shows the credibility weights for $\rho = 0$. In this case, the off-diagonal elements are zero and are therefore not plotted in figure 1. From figure 1 we can see, that the credibility weight $\widetilde{a}_{11}$ takes very soon a value near to one, which means, that the credibility estimators deviate only little from the observed normal claim frequency (except for small risk groups with an expected number of normal claims less than 100). Hence in practice one can

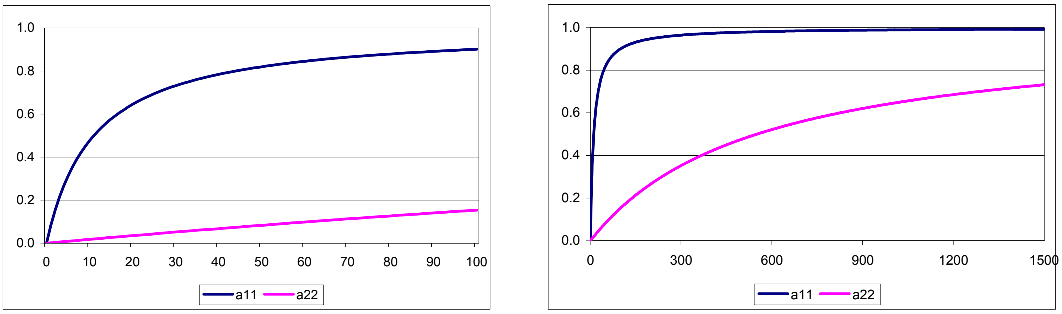

Figure 1: *credibility weights for* $\rho = 0$.

usually simply rely on the observations when estimating the normal claim frequency. However this is no longer the case for the big claim frequency.

Also for $\rho \neq 0$ one finds, that the credibility esimators of the normal claim frequencies deviate only little from the observed claim frequencies ($\widetilde{a}_{11}$ near to one and $\widetilde{a}_{12}$ close to zero) and that one can usually simply rely on the observations when estimating the expected normal claim frequency (except for small risk groups). The intersting thing is again the estimation of the big claim frequency. Therefore only the credibility weights relevant for the estimator of the big claim frequency, namely s $\widetilde{a}_{21}$ and $\widetilde{a}_{22}$, will be plotted in the figures 2 and 3,.

Figure 2 shows these credibility weights for $\rho = 0.5$. It can be seen from figure 2, that the weight $\widetilde{a}_{21}$ given to the observed normal claim frequency is not at all neglectible and, for quite a big range of ν_1,even higher than the weight given to the observed big claim frequency. This means that we can learn quite a lot from the observed normal claim frequency with respect to the expected big claim frequency.

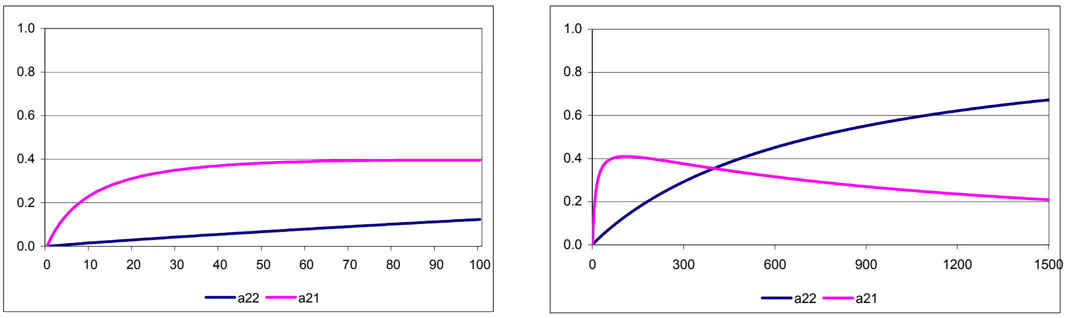

Figure 2: *credibility weights for* $\rho = 0.5$.

Finally, figure 3 shows the credibility weights for $\rho = 1$. Note, that together with $\widetilde{\kappa}_1 = \widetilde{\kappa}_2$, we consider the situation where normal and big claim frequencies have the same profile. In this case, the estimators of the standardized big and normal claim frequency coincide, i.e. estimating the "standardized" big claim frequency is the same as estimating the "standardized" normal claim frequency. Hence it is not surprising, that $\widetilde{a}_{21}$ is soon close to one and $\widetilde{a}_{22}$ close to zero.

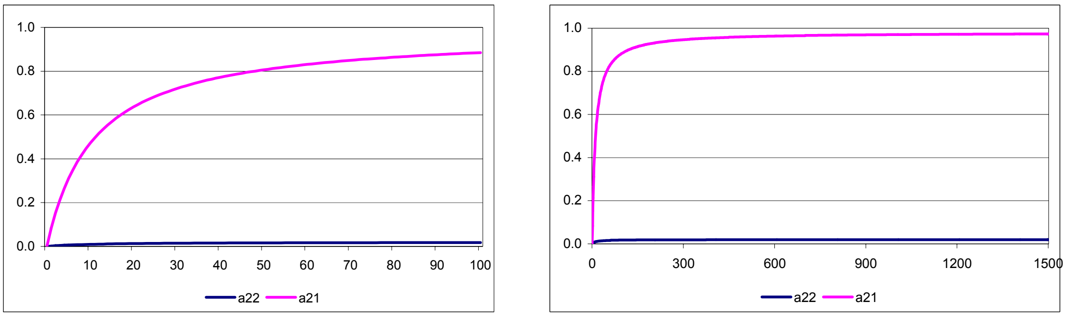

Figure 3: *credibility weights for* $\rho = 1$

4.5 Simulation study

The following results are taken from a simulation study carried through by D. Kollöffel in her diploma thesis [11].

4.5.1 Simulation models

The following model distributions were considered:

normal claims $N_i^{(1)} \sim Poisson(\Theta_i^{(1)})$,
$\Theta_i^{(1)} \sim 500 \cdot \Gamma(h, h)$
$h = 0.3^{-2}$.

big claims $N_i^{(2)} \sim Poisson(\Theta_i^{(2)})$.

version 1): $\Theta_i^{(2)}$ independent of $\Theta_i^{(1)}$ and
$\Theta_i^{(2)} \sim 10 \cdot \Gamma(h, h)$ with $h = 0.3^{-2}$;

version 2): $\Theta_i^{(2)} = \frac{\Theta_i^{(1)}}{100} + \widetilde{\Theta}_i^{(2)}$,
where $\widetilde{\Theta}_i^{(2)}$ is independent of $\Theta_i^{(1)}$ and
$\widetilde{\Theta}_i^{(2)} \sim 5 \cdot \Gamma(h, h)$ with $h = 0.3^{-2}$.

version 3): $\Theta_i^{(2)} = \frac{\Theta_i^{(1)}}{50}$.

Here the claim numbers may be interpreted as the observations of one single accident year or as the sum over several accident years.

For the structural parameters we find

structural parameters	$\mu_1 = \sigma_1^2$	$\mu_2 = \sigma_2^2$	τ_1^2	τ_2^2	τ_{12}	ρ
version 1	500	10	22'500	9	0	0
version 2	500	10	22'500	4.5	225	0.7071
version 3	500	10	22'500	9	450	1

(4.60)

Note, that model version 1 represents the case, where normal and big claim frequencies are independent and that, in this case, we can't learn anything from the observed number of normal claims with respect to the expected number of big claims. The other extreme case is model version 3, where the profile of the big claim frequency is the same as the profile of the normal claim frequency corresponding to model 4.2. Model version 2 is something in between, which will probably be the most common situation in practice.

For each of the model versions there were simulated 100 times a portfolio with 10 risks of equal weight, and for each of these risks there were simulated $\Theta_i^{(1)}, \Theta_i^{(2)}, N_i^{(1)}, N_i^{(2)}$. Since the model versions differ only in the modelling of the big claim frequency from each other, the same simulated data of $\Theta_i^{(1)}$ and $N_i^{(1)}$ were used in the simulation of the three model versions.

Then the multidimensional as well as the one-dimensional, the inhomogeneous as well as the homogeneous credibility estimators of $\Theta_i^{(1)}$and $\Theta_i^{(2)}$ were calculated and compared with $\Theta_i^{(1)}$and $\Theta_i^{(2)}$. For the inhomogeneous credibility estimators, the true structural parameter were used. For the homogeneous credibility estimators, the structural parameters had to be estimated too from the data of each of the simulated portfolios. The estimators used can be found in appendix A. Of course the precision of the estimators of the structural parameters depend on the size (number of risks (or risk categories)) of the portfolio. Our simulated portfolios contain 10 risks (or risk categories). For bigger portfolios the precision of the homogenous credibility estimators would be nearer to the one of the inhomogeneous estimator.

4.5.2 Simulation results

model version 1 The following table shows the structural parameters for the inhomogeneous credibility estimator known from the model as well as the average values of the parameter estimates for the homogeneous credibility estimates from the 100 simulated portfolios.

structural parameters	$\mu_1 = \sigma_1^2$	$\mu_2 = \sigma_2^2$	τ_1^2	τ_2^2	τ_{12}	ρ
inhomogeneous cred. est.	500.00	10.00	22'500	9.00	0	0
homogeneous cred. est.	504.04	9.98	23'607	8.07	-1.10	-0.0025

(4.61)

Analogously you find in the next tables the credibility weights of the inhomogeneous credibility estimators obtained with the above structural parameters as well as the average value of the homogeneous credility weights, averaged over the 100 simulated portfolios.

credibility weights	α_1	$\widetilde{a}_{11}$	$\widetilde{a}_{12}$	α_2	$\widetilde{a}_{22}$	$\widetilde{a}_{21}$
inhomog. cred.est.	97.83%	97.83%	0	47.37%	47.37%	0
homog. cred. est.	97.18%	96.88%	0.01%	35.66%	29.91%	-0.59%

(4.62)

Finally the next table shows the square root of the average quadratic loss.

$\sqrt{\text{quadratic loss of big claim frequency}}$	one-dim.	multi-dim.
inhomogeneous credibility estimators	24.7%	24.7%
homogeneous credibility estimators	26.1%	26.9%

(4.63)

Discusson
First we can see from (4.62), that for the inhomogeneous as well as for the homogeneous credibility estimators α_1 and $\widetilde{a}_{11}$ are close to 1 and that $\widetilde{a}_{12}$ is very small. Hence the credibility estimators of $\Theta_i^{(1)}$ (frequency of normal claims) will only little deviate from the observed frequencies, such that in practice it will be sufficiently accurate to rely on the observed individual frequency of normal claims and that there is no need for credibility for estimating the normal claim frequency.

From (4.62) it is also seen, that the credibility weights α_2 and $\widetilde{a}_{22}$ relating to the big claim frequency are smaller than 50% but substantially higher than 0, such that credibility does make sense here. Of course, in model version 1, the multidimensional inhomogeneous credibility estimators coincide with the one dimensional ones. But this is not necessarily the case for the homogeneous credibility estimators, since here the structural parameters are unknown and have to be estimated from the data and since the estimator for ρ might be different from zero. It is therefore not surprising that the quadratic loss of the homogeneous estimator is higher for the multidimensional model than for the one-dimensional one, because we fit an "over-parametrised" model with possibly $\rho \neq 0$. However, the difference and the loss of precision is very small when using the "over-parametrised" multidimensional model. Figure 4 shows the square root of the avarage squared loss of the homogeneous credibility estimators of the big claim frequency for each of the 100 simulated portfolios. The figure confirms, that the multidimensional estimators are only slightly worse than the one-dimensional ones.

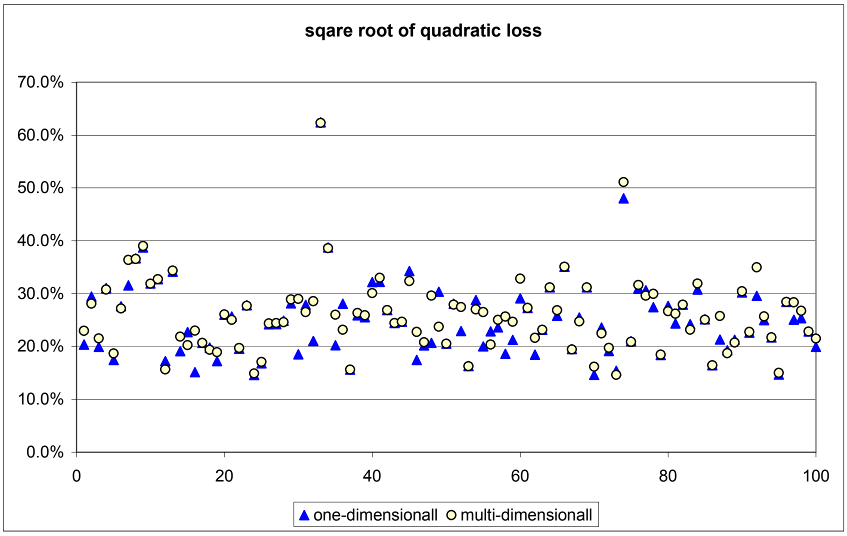

Figure 4: homogeneous: model version 1

model version 2 The following results were obtained:

structural parameters	$\mu_1 = \sigma_1^2$	$\mu_2 = \sigma_2^2$	τ_1^2	τ_2^2	τ_{12}	ρ
inhomogeneous cred. est.	500.00	10.00	22'500	4.5	225.00	0.707
homogeneous cred. est.	504.04	10.07	23'607	5.0	177.28	0.516

(4.64)

credibility weights	α_1	$\widetilde{a}_{11}$	$\widetilde{a}_{12}$	α_2	$\widetilde{a}_{22}$	$\widetilde{a}_{21}$
inhomog. cred.est.	97.83%	97.44%	0.80%	31.03%	18.69%	39.77%
homog. cred. est.	97.18%	96.71%	0.55%	13.41%	0.1341	27.79%

(4.65)

$\sqrt{\text{quadratic loss of big claim frequency}}$	one-dim.	multi-dim.
inhomogeneous credibility estimators	19.0%	13.4%
homogeneous credibility estimators	19.1%	17.8%

(4.66)

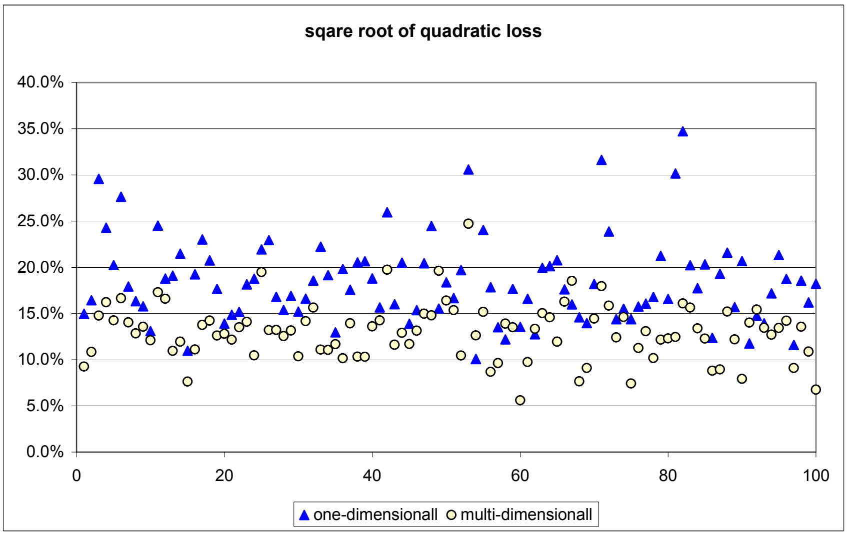

Figure 5: inhomogeneous, model version 2

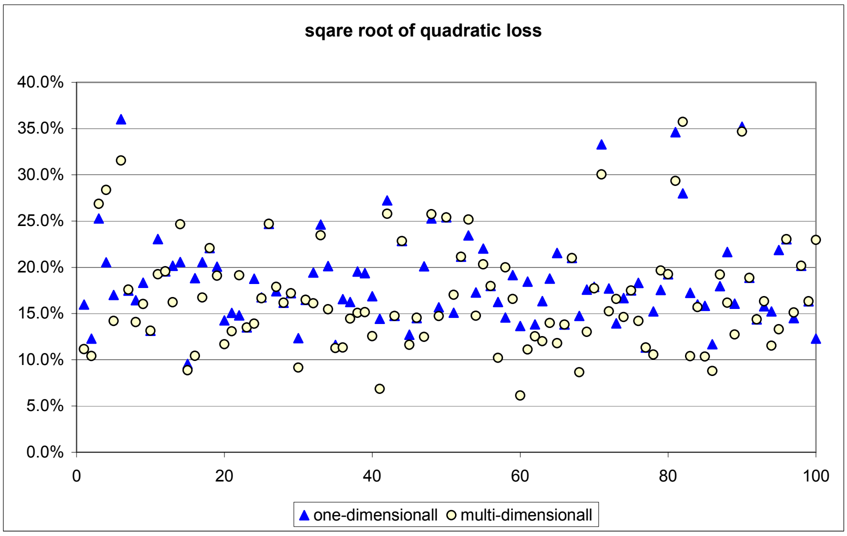

Figure 6: homogeneous: model version 2

Discussion
The quadratic loss of the inhomogeneous credibility estimator is reduced by a factor of 50%. Hence there is a substantial gain of precision when using the multi-dimensional instead of the one-dimensional credibility estimator. This can also been seen from figure 5.

The gain in precision is not so big for the homogeneous credibility estimator, but, as can be seen from figure 6, the multidimensional estimates are still in most of the 100 simulations clearly mor precise than the one-dimensional ones.

model version 3 The following results were obtained:

structural parameters	$\mu_1 = \sigma_1^2$	$\mu_2 = \sigma_2^2$	τ_1^2	τ_2^2	τ_{12}	ρ
inhomogeneous cred. est.	500	10.00	22'500	9.0	450.00	1
homogeneous cred. est.	504.04	10.26	23'607	9.0	378.28	0.821

(4.67)

credibility weights	α_1	$\widetilde{a}_{11}$	$\widetilde{a}_{12}$	α_2	$\widetilde{a}_{22}$	$\widetilde{a}_{21}$
inhomog. cred.est.	97.83%	95.95%	1.92%	47.37%	1.92%	95.95%
homog. cred. est.	97.18%	95.70%	1.40%	37.44%	7.77%	68.70%

(4.68)

$\sqrt{\text{quadratic loss of big claim frequency}}$	one-dim.	multi-dim.
inhomogeneous credibility estimators	21.5%	9.8%
homogeneous credibility estimators	23.0%	15.8%

(4.69)

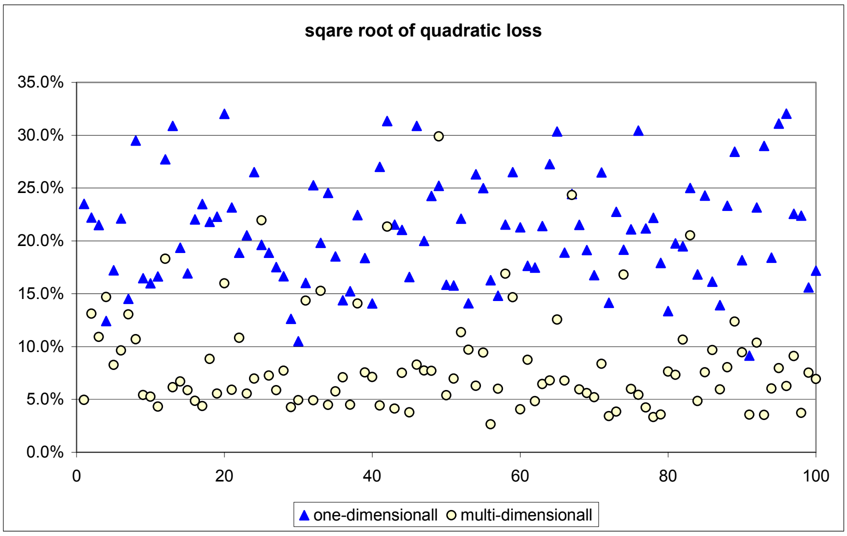

Figure 7: inhomogeneous: model version 3

Discussion
The quadratic loss of the inhomogeneous credibility estimator is reduced tremendously, which is also confirmed by figure 7. The gain of precision is also very substantial for the homogeneous credibility estimator, which can also be seen in figure 8.

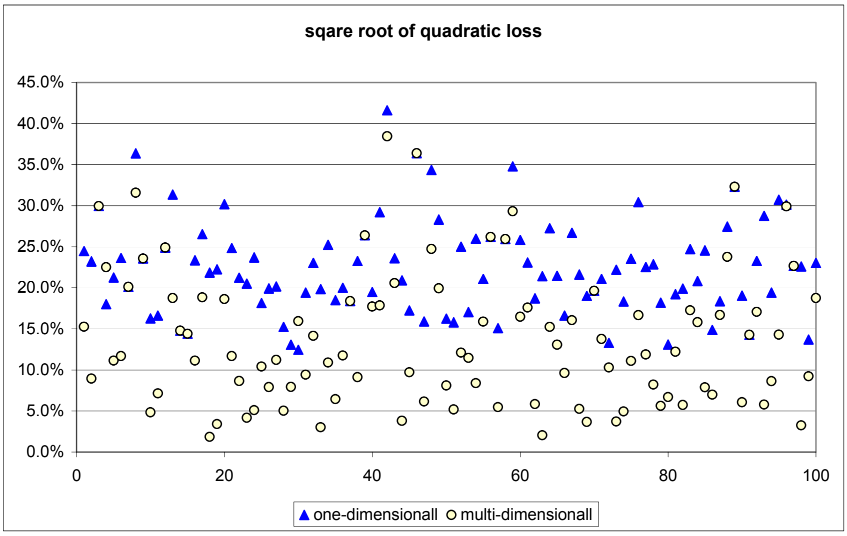

Figure 8: homogeneous: model version 3

summary conclusion from the simulation study Regarding the problem of estimating the big claim frequency, we can draw the following conclusions from the simulation study:

- We don't loose much of precision when using the homogeneous multidimensional credibility estimator in the case where the underlying processes of producing normal and big claims are independant of each other.
- The gain of precision is substantial when using multidimensional credibility in the case where there is a correlation in the underlying process between normal and big claim frequency. The gain of precision is the bigger, the bigger this correlation.
- In practice, the structural parameters are mostly unknown and have to be estimated from the data (homogeneous credibility estimators). The data should tell us, whether and how much normal and big claim frequencies are correlatet and how much we can learn from the observed frequency of normal claims with respect to the big claim frequency. The results of the simulation study show, that this objective is achieved to a great extent.

4.6 Application to a real data set from practice.

Table 1 shows the number of year risks as well as the observed number of "normal claims" (claim amount $<$ CHF 50'000) and "big claims" (claim amount $\geq$ CHF 50'000) of a major Swiss insurance company together with the corresponding claim frequencies for several regions in the line motor liability.

From these data we first we have to estimate the structural parameters by use of the the estimators in appendix A. The result can be found in table 2.

region	# of risks	# of norm. claims	# of big claims	norm. claim freq. in %o	big cl. freq. in %o	standardized norm.cl.fr.	big cl. fr.
	a	b	c	d=b/a	e=c/a	f=d/89.7	g=e/0.902
1	50'061	3'838	42	76.7	0.839	0.86	0.93
2	10'135	789	5	77.8	0.493	0.87	0.55
3	121'310	8'836	105	72.8	0.866	0.81	0.96
4	35'045	3'416	32	97.5	0.913	1.09	1.01
5	19'720	1'640	32	83.2	1.623	0.93	1.80
6	39'092	5'149	37	131.7	0.946	1.47	1.05
7	4'192	307	7	73.2	1.670	0.82	1.85
8	19'635	1'918	16	97.7	0.815	1.09	0.90
9	21'618	2'267	18	104.9	0.833	1.17	0.92
10	34'332	2'659	30	77.4	0.874	0.86	0.97
11	11'105	654	7	58.9	0.630	0.66	0.70
12	56'590	4'826	52	85.3	0.919	0.95	1.02
13	13'551	1'195	10	88.2	0.738	0.98	0.82
14	19'139	1'631	15	85.2	0.784	0.95	0.87
15	10'242	836	14	81.6	1.367	0.91	1.51
16	28'137	2'209	20	78.5	0.711	0.88	0.79
17	33'846	3'372	17	99.6	0.502	1.11	0.56
18	61'573	5'900	37	95.8	0.601	1.07	0.67
19	17'067	1'509	21	88.4	1.230	0.99	1.36
20	8'263	662	9	80.1	1.089	0.89	1.21
21	148'872	14'851	163	99.8	1.095	1.11	1.21
Total	763'525	68'464	689	89.7	0.902	1.00	1.00

Table 1: observed data in motor liability

$\widehat{\sigma_1^2}$	$\widehat{\sigma_2^2}$	$\widehat{\tau_1^2}$	$\widehat{\tau_2^2}$	$\widehat{\tau_{12}}$	$\widehat{\varrho}$
8.967E-02	9.024E-04	2.383E-04	2.956E-08	3.085E-07	0.116
			$\widehat{\mu_1}$	$\widehat{\mu_2}$	
multi-dim. cred.			87.5%o	0.892%o	
one-dim. cred.			87.5%o	0.895%o	

Table 2: Estimates of structural parameters

With these estimates of the structural parameters we obtain for the credibility coefficients

$$\widehat{\kappa_1} = \frac{\widehat{\sigma_1^2}}{\widehat{\tau_1^2}} = 376,$$

$$\widehat{\kappa_2} = \frac{\widehat{\sigma_2^2}}{\widehat{\tau_2^2}} = 3'053.$$

It is also interesting to look at the coefficient of variation of $\mu(\Theta_1)$ (normal claim frequency) and of $\mu(\Theta_2)$ (big claim frequency). We obtain

$$\begin{aligned} cv_1 &:= cv(\widehat{\mu(\Theta_1)}) = \frac{\widehat{\tau}_1}{\overline{F_1}} = 17.2\%, \\ cv_2 &:= cv(\widehat{\mu(\Theta_2)}) = \frac{\widehat{\tau}_2}{\overline{F_2}} = 19.1\%, \end{aligned}$$

where $\overline{F_1}$ and $\overline{F_2}$ are the observed normal resp. big claim frequency of the total over all regions (see formulae (A.11) and (A.13) in appendix A). The two coefficients of variation are very close to each other, something, which is in some way also quite natural.

We define the standardized homogeneous credibility estimators by

$$\widetilde{\widehat{\mu(\Theta_1)}}^{\text{hom}} := \frac{\widehat{\widehat{\mu(\Theta_1)}}^{\text{hom}}}{\overline{F_1}}, \qquad \widetilde{\widehat{\mu(\Theta_2)}}^{\text{hom}} := \frac{\widehat{\widehat{\mu(\Theta_2)}}^{\text{hom}}}{\overline{F_2}}.$$

Table 3 shows the credibility weights for these standardized homogeneous credibility estimators obtained with the estimated structural parameters according to table 2.

region	multidimensional				one-dimensional	
	$\widetilde{a}_{11}$	$\widetilde{a}_{12}$	$\widetilde{a}_{21}$	$\widetilde{a}_{22}$	$\widetilde{\alpha_1}$	$\widetilde{\alpha_2}$
1	99.2%	0.05%	4.9%	618%	99.3%	62.1%
2	96.4%	0.09%	9.3%	24.7%	96.4%	24.9%
3	99.7%	0.03%	2.6%	79.7%	99.7%	79.9%
4	98.9%	0.06%	6.0%	53.1%	98.9%	53.4%
5	98.1%	0.08%	7.7%	38.9%	98.1%	39.2%
6	99.0%	0.06%	5.6%	55.8%	99.0%	56.2%
7	91.8%	0.10%	10.4%	11.9%	91.8%	12.1%
8	98.1%	0.08%	7.7%	38.8%	98.1%	39.1%
9	98.3%	0.07%	7.4%	41.1%	98.3%	41.5%
10	98.9%	0.06%	6.0%	52.6%	98.9%	52.9%
11	96.7%	0.09%	9.2%	26.4%	96.7%	26.7%
12	99.3%	0.05%	4.5%	64.7%	99.3%	65.0%
13	97.3%	0.09%	8.7%	30.5%	97.3%	30.7%
14	98.1%	0.08%	7.8%	38.2%	98.1%	38.5%
15	96.4%	0.09%	9.3%	24.9%	96.5%	25.1%
16	98.7%	0.07%	6.6%	47.6%	98.7%	48.0%
17	98.9%	0.06%	6.1%	52.2%	98.9%	52.6%
18	99.4%	0.04%	4.3%	66.6%	99.4%	66.9%
19	97.8%	0.08%	8.1%	35.6%	97.8%	35.9%
20	95.6%	0.10%	9.7%	21.1%	95.6%	21.3%
21	99.7%	0.02%	2.2%	82.8%	99.7%	83.0%

Table 3: credibility weights for the standardized cred. est.

Finally, table 4 shows the credibility estimates of the standardized claim frequencies together with the corresponding observed frequencies.

region	# of risks	standandardized norm. cl. fr.			standandized big cl. fr.		
		observed	cred.estimates		observed	cred.estimates	
			multi-dim.	1-dim.		multi-dim.	1-dim.
1	50'061	0.86	0.86	0.86	0.93	0.95	0.95
2	10'135	0.87	0.87	0.87	0.55	0.87	0.88
3	121'310	0.81	0.81	0.81	0.96	0.96	0.97
4	35'045	1.09	1.09	1.09	1.01	1.01	1.00
5	19'720	0.93	0.93	0.93	1.80	1.30	1.31
6	39'092	1.47	1.46	1.46	1.05	1.05	1.02
7	4'192	0.82	0.83	0.83	1.85	1.08	1.10
8	19'635	1.09	1.09	1.09	0.90	0.96	0.96
9	21'618	1.17	1.17	1.17	0.92	0.98	0.96
10	34'332	0.86	0.86	0.86	0.97	0.97	0.98
11	11'105	0.66	0.67	0.67	0.70	0.88	1.91
12	56'590	0.95	0.95	0.95	1.02	1.01	1.01
13	13'551	0.98	0.98	0.98	0.82	0.94	0.94
14	19'139	0.95	0.95	0.95	0.87	0.94	0.94
15	10'242	0.91	0.91	0.91	1.51	1.11	1.12
16	28'137	0.88	0.88	0.88	0.79	0.89	0.89
17	33'846	1.11	1.11	1.11	0.56	0.77	0.76
18	61'573	1.07	1.07	1.07	0.67	0.78	0.77
19	17'067	0.99	0.99	0.99	1.36	1.12	1.13
20	8'263	0.89	0.90	0.90	1.21	1.03	1.04
21	148'872	1.11	1.11	1.11	1.21	1.18	1.18
Total	763'525	1.00	1.00	1.00	1.00	1.00	1.00

Table 4: credibility estimates of the standardized claim frequencies

Discussion

The credibility estimates of the normal claim frequencies do nearly fully coincide with the observed frequencies. This is not surprising and confirms our previous statements and findings.

One often used pragmatic approach would be to use the same profile resp. the same standardized estimates also for the big claims (cf. subsection 1). However, the observed data and the credibility estimates in table 4 tell us, that this pragmatic approach wouldn't be appropriate at all in this situation.

The other often encountered approach of spreading the claims laod of the big claims equally over all risks wouldn't be adequate either. This would mean that the big-claim frequency is assumed to be be the same for all regions. However the credibility estimates for the standardized big claim frequency speak another language and vary quite substantially between the regions.

The estimated correlation $\widehat{\rho}$ (see table 2) between the normal and the big claim frequency is pretty small. Therefore, the multidimensional credibility estimate deviates only little from the one dimensional one. However, this might be different in other practical situations. By using empirical multidimensional credibility, it is the data, which tell us, how much we can learn from the normal claims with respect to the big claims.

Acknowledgment
The authors would like to express their thanks to Mario Wüthrich, Winterthur Insurance Company, who has supported Mrs. Kollöffel in her diploma thesis, in particular for the simulation part.

A Appendix: Estimation of structural parameters

Since for Poisson-distributed claim numbers

$$\sigma_1^2 = \mu_1 \quad , \quad \sigma_2^2 = \mu_2 \tag{A.1}$$

we use as estimator of S

$$\widehat{S} := \begin{pmatrix} \widehat{\sigma_1^2} & 0 \\ 0 & \widehat{\sigma_2^2} \end{pmatrix}, \text{ where} \tag{A.2}$$

$$\widehat{\sigma_1^2} := \sum_{i=1}^{I} \frac{w_{i\bullet}}{w_{\bullet\bullet}} F_i^{(1)}, \tag{A.3}$$

$$\widehat{\sigma_2^2} := \sum_{i=1}^{I} \frac{w_{i\bullet}}{w_{\bullet\bullet}} F_i^{(2)}. \tag{A.4}$$

As estimator for the "between risks" covariance matrix we use

$$\widehat{T} := \begin{pmatrix} \widehat{\tau_1^2} & \widehat{\tau_{12}} \\ \widehat{\tau_{12}} & \widehat{\tau_2^2} \end{pmatrix}, \text{ where} \tag{A.5}$$

$$\widehat{\tau_1^2} := \max\left(\text{const} \cdot \left(T_1 - \frac{I \cdot \widehat{\sigma_1^2}}{w_{\bullet\bullet}}\right), 0\right), \tag{A.6}$$

$$\widehat{\tau_2^2} := \max\left(\text{const} \cdot \left(T_2 - \frac{I \cdot \widehat{\sigma_2^2}}{w_{\bullet\bullet}}\right), 0\right), \tag{A.7}$$

$$\widehat{\tau_{12}} := \text{signum } (T_{12}) \cdot \sqrt{\min\left\{(\text{const} \cdot T_{12})^2, \widehat{\tau_1^2} \cdot \widehat{\tau_2^2}\right\}}, \tag{A.8}$$

and where

$$\text{const} \quad := \frac{I-1}{I}\left\{\sum_{i=1}^{I} \frac{w_{i\bullet}}{w_{\bullet\bullet}}\left(1 - \frac{w_{i\bullet}}{w_{\bullet\bullet}}\right)\right\}^{-1}, \tag{A.9}$$

$$T_1 \quad := \frac{I-1}{I}\sum_{i=1}^{I} \frac{w_{i\bullet}}{w_{\bullet\bullet}}\left(F_i^{(1)} - \overline{F}^{(1)}\right)^2, \tag{A.10}$$

$$\overline{F}^{(1)} := \sum_{i=1}^{I} \frac{w_{i\bullet}}{w_{\bullet\bullet}} F_i^{(1)}, \tag{A.11}$$

$$T_2 := \frac{I-1}{I} \sum_{i=1}^{I} \frac{w_{i\bullet}}{w_{\bullet\bullet}} \left(F_i^{(2)} - \overline{F}^{(2)}\right)^2, \tag{A.12}$$

$$\overline{F}^{(2)} := \sum_{i=1}^{I} \frac{w_{i\bullet}}{w_{\bullet\bullet}} F_i^{(2)}, \tag{A.13}$$

$$T_{12} := \frac{I-1}{I} \sum_{i=1}^{I} \frac{w_{i\bullet}}{w_{\bullet\bullet}} \left(F_i^{(1)} - \overline{F}^{(1)}\right)^2 \left(F_i^{(2)} - \overline{F}^{(2)}\right). \tag{A.14}$$

References

[1] Bühlmann H. and Gisler A., *book on credibility,* forthcoming.

[2] Bühlmann H. and Straub E. (1970), *Glaubwürdigkeit für Schadensätze,* Bulletin of the SAA, 111-133 (or (1972), *Credibility for loss ratios,* translated by C.E. Brooks, ARCH).

[3] Bühlmann H., Gisler A., Jewell W.S. (1982), *Excess claims and data trimming in the context of credibility rating procedures,* Bulletin of the SAA

[4] Gisler A. (1980), *Optimum trimming of data in the credibility model,* Bulletin of SAA, 117-147.

[5] Gisler A. and Reinhard P. (1993), *Robust Credibility,* ASTIN Bulletin.

[6] Hachemeister C.A. (1975), *Credibility for regression models with application to trend;* in *"Credibility, theory and application",* Proceedings of the Berkeley Actuarial Research Conference on credibility; Academic Press, New York, 129-163.

[7] Jewell, W.S. (1973), *multidimensional Credibility,* Operations Research Center 73-7, Berkely.

[8] Jewell, W.S. (1973), *Bayesian Regression and Credibility*, RM-75-63, International Institute for Applied Systems Analysis, Laxemburg, Austria.

[9] Kahn P.M. (ed.) (1974), *Credibility: Thery and Applications,* Proceeding of the actuarial Research Conference on Credibility, Berkeley, September 1974, Academic Press, New York

[10] Künsch H.R. (1992), *Robust Methods for Credibility,* ASTIN Bulletin

[11] Kollöffel D. (2003), *Credibility mit spezieller Berücksichtigung von Grossschäden,* diploma thesis, ETH-Zürich.

[12] Schnieper R (2000), *Robust Baysian Inference,* paper presented at the ASTIN colloquium in Sardinia.