ROBUST CREDIBILITY

By Alois Gisler and Peter Reinhard

"Winterthur", Swiss Insurance Company Winterthur [1]

Abstract

Outlier observations caused by big claims or by an event producing a series of claims are a special problem in ratemaking and in tariff calculation. The authors believe that combining credibility and robust statistics is the right answer to this problem The main idea is to robustify the individual claims experience by using a robust estimator T_i instead of the individual mean $\bar{X}_i$ and to look at the credibility estimator based on the robust statistics $\{T_i : i = 1, 2, .. \}$. Choosing a particular influence function leads to data-trimming with an observation-dependent trimming point.

Keywords

Robust statistics; standard credibility models, data trimming.

1. Introduction and Motivation

The data in Figure 1 represent observed loss ratios (claims amount divided by sum insured) of a given risk group in industrial fire over a ten year period. Figure 1 is an example of the following situation often encountered in practice: most of the observations are lying randomly within a band depending on the size of the group and on the line of business, but a few observations are far away and much bigger The smaller the group the more likely are such 'outlier observations'. They are cause by individual big claims or by events producing a series of claims (e.g. storm).

   Assuming you want to estimate the pure risk premium for the given risk group of Figure 1 based on a ten year observation period, the first obvious idea would be to take the mean over the observation period as an estimator, which would give an estimated value of 0.66‰ But if you do the same calculation one year later, then the 'outlier' observation of year 1 is probably replaced by an ordinary observation in year 11, and the ten year average decreases by about 20% Of course such random fluctuations have to be avoided in a professional tariff-calculation. The simple mean is not a suitable estimator and there is a real need for more sophistication The main problem to be solved is how to treat outlier observations in rate making and in tariff calculation.

[1] A first version of the paper was presented at the ASTIN Colloquium 1990 in Switzerland
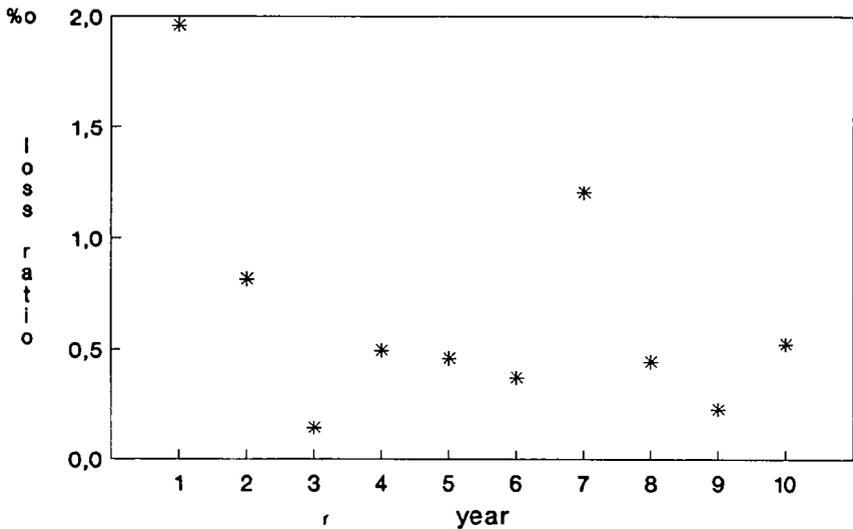
FIGURE 1 Observed loss ratios of a risk group in industrial fire

Before trying to give a theoretical answer it is interesting to see how this problem was tackled in actuarial practice. For this purpose let us have a look at the methods used for the calculation of the pure risk premium in industrial fire in Germany and in Switzerland.

A short description of the German calculation system is given in J. STRAUSS (1984). The annual statistics are built up in a hierarchical way. On the lowest level there are the data (sum insured, loss ratios, etc.) of so-called *risk types*, which are taken together into *risk groups*, which again are combined to *risk categories*. At the top level, we have the data of the total industrial fire business as a whole. The tariff is calculated by some kind of hierarchical procedure from top down. The Bühlmann-Straub credibility model is successively applied at the different levels By doing so the total claim amount is first spread among the risk categories, then within the risk categories among the risk groups and finally within the risk groups among the risk types. Thus the claims load of each claim (for instance a big claim) hitting a particular risk type is successively divided up at the different levels. The higher the level on the hierarchical tree the bigger will be the credibility weight, and the larger will be the portion of the claims load that will remain within the corresponding group or category.

The Swiss calculation is based on a method developed by H. AMMETER (1982). In each position of a hierarchical tree (risk group, risk type etc.) he makes a distinction between ordinary and extraordinary loss ratios. The extraordinary loss ratios are trimmed in an appropriate way. By doing so the corresponding aggregate claim amount is divided up into an 'ordinary part' and an '*xs*-part' But how to distinguish between the two types of loss ratios and where to fix the trimming point? Ammeter's idea was to look at the

influence of an annual observation on the loss ratio taken over the whole observation period. The trimming point is calculated in such a way that the influence of each observation is limited to the influence of a zero-observation. This calculation is carried out at different levels of a hierarchical structure from bottom up. At each level the xs-part of the aggregate claim amount is equally distributed over the neighbouring group at the next hierarchical level.

Both methods—although quite different—were successfully applied in the tariff calculation and led obviously to reasonable results (otherwise they would have been rejected by the practitioners). From both methods we can learn something

The German system is based on credibility theory. Indeed credibility theory in its standard form makes a first step in the right direction, how to charge outlier observations. It explains to us that claims should not be fully charged, but only with their respective credibility weight. However, it turned out in practical applications that outliers might still have distorting effects. On the one hand they cause a substantial reduction of the credibility weights. As a consequence the credibility premiums of 'risks' without large claims are smoothed too much towards the overall mean. On the other hand the credibility premium of a risk might increase tremendously by the occurrence of one single large claim. To overcome this GISLER (1980) combined credibility procedures and data trimming. This method has successfully been used in actuarial applications However, it is not applicable to situations where only claims rates are given and where the corresponding volume measures are different in size The use of hierarchical procedures and the introduction of hierarchical credibility models was certainly a further step on the credibility staircase, which is of great importance for practice. As already mentioned, the German system is based on such an approach

At first sight the Swiss method introduced by Ammeter seems to be an original, rather pragmatic approach. However, looking a little closer, one observes that it is also related to a famous theory. The idea of Ammeter is to limit the influence of single observations. But this is the basic concept behind robust statistics. Indeed Ammeter introduced—perhaps without being aware of it—a robust estimator. We shall come back to this estimator later on

The first to have the idea of combining credibility theory with robust statistics was H. R. KÜNSCH (1992). He already presented some main ideas at a lecture given in February 1990 at the ETH in Zürich The diploma work of REINHARD (1989) had also been written under his guidance. At the 1990 ASTIN colloquium at Montreux the present authors then presented an early version of this paper At the 1991 ASTIN colloquium in Stockholm there was another paper by Kremer on the same subject, which has been published in the meantime in the German actuarial journal (KREMER (1991)). The main objection of the present authors against Kremer's approach is that his estimators are globally (expectation over the whole portfolio) biased.

The present authors believe that combining credibility with robust statistics is the right answer for dealing with outlier-observations and that this idea has a great potential for practical applications In Section 2 some basic definitions

and findings from the theory of robust statistics will be given. To make this paper self-contained, some well-known results of standard credibility are summarized in Section 3. Specific robust credibility estimators as well as the corresponding estimates of the structural parameters are presented in Section 4. A simulation study carried through in Section 5 illustrates the functioning of the robust credibility estimators presented in Section 4.

2. RESULTS FROM THE THEORY OF ROBUST STATISTICS

In this chapter we introduce some basic concepts and summarize some main results out of the theory of robust statistics, which we will need later on. We rely on the presentation in HAMPEL and alii (1986), which we can recommend as an excellent introduction. All results are given without proofs and the interested reader is again referred to HAMPEL and alii (1986).

Robust statistics is an extension of classial parametric statistics, taking into account that parametric models are only an idealized approximation to reality. It studies the behavior of statistical procedures not only under strict parametric models, but also in the neighbourhood of such models. The idea is to construct statistical procedures which still behave fairly well under slight deviations from the assumed model. In a formal sense we might say that *robust statistics is the statistics of approximate parametric models.* The *main aim is to describe what the bulk of the data is telling us.* However in insurance we cannot forget about the deviating observations. A big loss ratio for instance is not simply an 'error noise' in our data, but rather caused by rare events like storms, big fires etc. which make a substantial part of the total claims costs. A *'second' aim of robust statistics is to identify deviating data points for further treatment.* In insurance this is often as important as the description of the bulk of the data.

Suppose we have one-dimensional observations $X_1, \ldots, X_n$, which are assumed to be i.i.d. and distributed according to F_ϑ (density f_ϑ) out of a parametric family $\{F_\vartheta; \vartheta \in \Theta\}$. To be more precise, we know that this is an idealization of reality and we assume that the true distribution lies in the neighbourhood of our model. We want to estimate the expectation of X_t. For simplicity's sake we further assume that the parametrization is chosen in such a way that $\vartheta = E_\vartheta[X_t]$. We denote by G_n the empirical cdf (cumulative distribution function) of a sample with n observations. As estimators of ϑ we consider real-valued statistics $T_n = T(G_n)$, where T are functionals. The simplest idea to look at the influence of a single observation is the so called *empirical influence function.* Given a sample $(x_1, \ldots, x_{n-1})$ it is the plot of $T_n(x_1, \ldots, x_{n-1}, x)$ as a function of x.

By translating and rescaling one obtains *the sensitivity curve*

(1) $SC_n(x) = n[T_n(x_1, \ldots, x_{n-1}, x) - T_{n-1}(x_1, \ldots, x_{n-1})]$

Letting $n \to \infty$ yields the influence function invented by HAMPEL (1968, 1974).

Definition 1

The *influence function (IF)* of T at F is given by

(2) $$IF(x; T, F) = \lim_{t \downarrow 0} \frac{T((1-t)F + t\Delta_x) - T(F)}{t}$$

where Δ_x is the probability measure putting mass 1 at the point x.

The IF is mainly a heuristic tool with an easy heuristic interpretation: it describes the effect of an infinitesimal contamination at the point x on the estimate, standardized by the mass of contamination

Definition 2

The *gross-error sensitivity* of T at F is defined by

(3) $$\gamma^* = \sup_x |IF(x; T, F)|.$$

The gross-error sensitivity measures the *worst influence* which a small amount of contamination of fixed size can have on the value of the estimator. It is desirable that $\gamma^*(T, F)$ be finite. Robustifying an estimator is typically putting a bound on $\gamma^*(T, F)$

If the r.v. $X_i (i = 1, 2, ..)$ are i i d. according to G, then G_n will tend to G by the Glivenko-Cantelli theorem. As a consequence

$$T_n(X_1, ..., X_n) \xrightarrow[n \to \infty]{} T(G)$$

whenever T is continuous with respect to the sup-norm, which will normally be the case

In most cases

$$\sqrt{n} (T_n - T(G)) \xrightarrow[n \to \infty]{} \mathcal{N}(0, V(T, G))$$

i e. T_n is asymptotically normal distributed with expectation $T(G)$ and variance $V(T, G)/n$. $V(T, G)$ is called the asymptotic variance.

In regular cases, the following important relations hold true.

(4) $$\int IF(x; T, G) \, dG(x) = 0$$

(5) $$V(T, G) = \int IF(x; T, G)^2 \, dG(x).$$

Our aim is to find a functional T resp. an estimator T_n with bounded gross-error sensitivity. For this purpose let us first have a look at the maximum likelihood estimator (MLE), which is defined as the value

$$\hat{\vartheta} = T_n(X_1, \ldots, X_n) \text{ which maximizes } \prod_{i=1}^{n} f_\vartheta(X_i),$$

or, what is equivalent, by

$$(6) \qquad \sum_{i=1}^{n} [-\ln f_{T_n}(X_i)] = \min_{T_n} !$$

HUBER (1964) proposed to generalize this to

$$(7) \qquad \sum_{i=1}^{n} \rho(X_i, T_n) = \min_{T_n} !$$

where ρ is some function on $\mathbb{R} \times \Theta$

Suppose that ρ has a derivative $\psi(x, \theta) = (\partial/\partial\theta)\, \rho(x, \theta)$, then the estimate T_n satisfies the implicit equation

$$(8) \qquad \sum_{i=1}^{n} \psi(X_i, T_n) = 0.$$

Definition 3

An estimator defined by (7) or (8) is called an *M-estimator*

If G_n is the empirical cdf generated by the sample, then the solution T_n of (8) can be written as $T(G_n)$, where T is the functional given by

$$(9) \qquad \int \psi(x, T(G))\, dG(x) = 0.$$

As already said, T_n will normally tend to $T(G)$ Or looked at the other way round, $T(G)$ is the asymptotic expectation of T_n, which can be calculated by formula (9).

Theorem 1

Let T_n be an *M*-estimator defined by (8), and IF$(x; \psi, F)$ the influence function of T at F, then

$$(10) \qquad \text{IF}(x; \psi, F) = \frac{\psi(x, T(F))}{-\int (\partial/\partial\theta)\, [\psi(y, \theta)]_{T(F)}\, dF(y)}.$$

Remark

Theorem 1 states that the influence function is proportional to the ψ-function. *Hence using (8) with a bounded ψ-function automatically results in an estimator T_n with bounded gross-error sensitivity*

From (5) we obtain for the asymptotic variance of the M-estimator

(11) $$V(T, G) = \frac{\int \psi^2(x, T(G)) \, dG(x)}{[\int (\partial/\partial\theta) \, [\psi(y, \theta)]_{T(G)} \, dG(y)]^2}$$

3. STANDARD CREDIBILITY

To make this paper self-contained we summarize in this subsection the assumptions and estimators in the Bühlmann and Straub model (1970), which is well known and presumably the most frequently applied credibility model in insurance practice.

Consider a portfolio of risk ('risk' = synonym for things like individual policies, risk classes, risk types etc) numbered $i = 1, 2, .., N$ Assume that each risk i is characterized by a hidden risk parameter Θ_i. $X_i = (X_{i1}, \ldots, X_{in})'$ is the observation vector of risk i (e.g. X_{ij} = loss ratio of risk i in year j).

Assumptions

BS1 · The random variables X_{ij} ($j = 1, 2, .., n$) are conditionally, given Θ_i, independent with

$E[X_{ij}|\Theta_i] = \mu(\Theta_i)$

$\mathrm{Var}\,[X_{ij}|\Theta_i] = \sigma^2(\Theta_i)/V_{ij}$

where V_{ij} are known volume measures.

BS2 The pairs (Θ_1, X_1), (Θ_2, X_2), . are independent, and $\Theta_1, \Theta_2, $. are i.i d.

The aim is to estimate for each risk i the risk premium $\mu(\Theta_i)$.

An estimator $\hat\mu(\Theta_i)$ is said to be better than $\tilde\mu(\Theta_i)$ if

$$E[(\hat\mu(\Theta_i) - \mu(\Theta_i))^2] < E[(\tilde\mu(\Theta_i) - \mu(\Theta_i))^2]$$

that is we use quadratic loss.

The best possible estimator based on $\mathcal{D} = \{X_{ij} \cdot i = 1, .. , N, j = 1, \ldots, n\}$ is $\tilde\mu(\Theta_i) = E[\mu(\Theta_i)|\mathcal{D}]$ which is called the *exact Bayesian estimator* The *credibility estimator*, i.e. the best estimator of the form $\hat\mu(\Theta_i) = a_0 + \sum_{i, j} a_{ij} X_{ij}$, is

(12) $\hat\mu(\Theta_i) = \mu + \alpha_i(\overline{X}_i - \mu)$

where $\mu = E[X_{ij}] = E[\mu(\Theta_i)]$

$$\overline{X}_i = V_i^{-1} \sum_{j=1}^{n} V_{ij} X_{ij} \qquad V_i = \sum_{j=1}^{n} V_{ij}$$

$$\alpha_i = \frac{V_i v}{V_i v + u} \qquad \text{with } v = \text{Var}[\mu(\Theta_i)]$$
$$u = E[\sigma^2(\Theta_i)].$$

In insurance practice the structural parameters μ, u, v are mostly unknown But they can be estimated from the data. By replacing the unknown parameters in (12) by the corresponding estimators one arrives at the *empirical credibility estimator*.

4. ROBUST CREDIBILITY

The individual claims experience enters into the credibility estimator (12) by means of $\overline{X}_i$. It is well known that the credibility estimator is exact Bayesian for various specific models. But it is also known that the credibility estimator might behave rather poorly if such a specific model is disturbed by a process producing only a few outliers (see for instance BÜHLMANN, GISLER, JEWELL (1982)). The idea of Künsch was to robustify the individual claims experience by using a robust estimator $T_i = T_i(X_{i1}, \ldots, X_{in})$ instead of the $\overline{X}_i$. The hope is to get estimators which also perform reasonably well in the neighbourhood of models, where the credibility estimator is exact Bayesian. REINHARD (1989) and KÜNSCH (1992) considered semilinear credibility estimators (see for instance GISLER (1990)) based on the statistics $\{T_i : i = 1, \ldots, N\}$ We suggest a slightly different approach. We propose to divide the pure risk premium itself into an 'ordinary part' and an 'xs-part', and to estimate each component separately.

Formally we write

(13) $\mu_X(\Theta_i) = \mu_o(\Theta_i) + \mu_{xs}(\Theta_i)$

where $\mu_X(\Theta_i) = E[X_{ij}|\Theta_i]$.

The 'ordinary part' $\mu_o(\Theta_i)$ should be interpreted as the expected loss ratio generated by the claims load of 'ordinary losses', whereas the 'xs-part' $\mu_{xs}(\Theta_i)$ is the additional expected claims load generated mainly by extraordinary events (e g. big fires), whose occurence usually lead to an outlier observation of the affected loss ratio.

To estimate *the 'ordinary part'* $\mu_o(\Theta_i)$ we combine credibility and robust statistics, that is we estimate $\mu_o(\Theta_i)$ by a credibility estimator based on a robust statistics $\{T_i \quad i = 1, \ldots, N\}$. Since it is the very nature of a robust statistics to describe what the bulk of the data is telling us, we put by definition·

(14) $\mu_o(\Theta_i) = E[T_i|\Theta_i]$.

As to *the 'xs-part'* $\mu_{xs}(\Theta_i)$ to be interpreted as the additional expected claims load mainly generated by extraordinary events, the present authors believe that the actuary should reflect upon how much the risks in the portfolio are exposed to such 'outlier-events' and to use this knowledge for ratemaking. In mathematical terms this means to put more a priori structure into the model. An often encountered situation will be that all risks in the portfolio can be considered as equally exposed to outlier events. Then the a priori structure is identical with the

(15)   *assumption*: $\mu_{xs}(\Theta_i) = \mu_{xs}$.

It might happen that for instance risk 1 is considered to be twice exposed to 'outlier-events' than risk 2, or generally, that the a priori structure is given by $\mu_{xs}(\Theta_i) = A\,\mu_{xs}$ where A is a known $N \times 1$ matrix. If no such a priori knowledge is available then there still remains the possibility to estimate $\mu_{xs}(\Theta_i)$ by a credibility estimator. But on what statistics should this credibility estimator be based on? One possibility would be to base the estimation of $\mu_{xs}(\Theta_i)$ also on the robust statistics T_i. Then the resulting estimator (13) of $\mu_X(\Theta_i)$ would be identical to the one considered by REINHARD (1989) and KUNSCH (1992) in the case of identical volumes. But this makes only sense if it is natural to believe that the bulk of the data reflected by the robust statistics T_i does also tell us something with respect to the 'xs-part' However, in most practical situations this is hardly the case. The usual situation will be that the bulk of the data contains very little information with respect to 'outlier-events'. Hence if using a credibility estimator for $\mu_{xs}(\Theta_i)$, it will be more natural and more appropriate in most cases to base it on the statistics of the observed xs-loss ratios $XS_{ij} = X_{ij} - T_i$ Very often the resulting credibility-weights will be near to zero, such that the resulting estimators will be similar to the ones obtained on the basis of assumption (15).

In the following we shall work on the hypothesis of assumption (15). However it should be no difficulty for the skilled reader to adapt the results to the other situations mentioned above. With assumption (15) the robust credibility estimator of $\mu_X(\Theta_i)$ equals

(16) $\hat{\mu}_X(\Theta_i) = \mu_{xs} + \hat{\mu}_o(\Theta_i)$

where $\hat{\mu}_o(\Theta_i)$ is a credibility estimator of $\mu_o(\Theta_i)$ based on a robust statistics $\{T_i \cdot i = 1, \ldots, N\}$.

By standard techniques we find that

(17) $\hat{\mu}_o(\Theta_i) = \mu_{T_i} + \alpha_i(T_i - \mu_{T_i})$

where $\alpha_i = \dfrac{\operatorname{Var}\left[\mu_{T_i}(\Theta_i)\right]}{E\left[\operatorname{Var}\left[T_i | \Theta\right]\right] + \operatorname{Var}\left[\mu_{T_i}(\Theta_i)\right]}$

$\mu_{T_i}(\Theta_i) = E[T_i | \Theta_i]$

$\mu_{T_i} = E[T_i]$

Remark on the notation

The careful reader will have noted that $\mu_{T_i}(\Theta_i)$ is identical to $\mu_o(\Theta_i)$ by definition (14). However, we have here and in the following deliberately chosen this notation to indicate explicitly the dependence on the choice of the robust statistics T_i.

(17) is the general formula of a whole class of robust credibility estimators, since we have by now not specified the robust estimators T_i. Indeed, there are a lot of robust estimators proposed in the literature on robust statistics Instead of going through this palette we prefer to present a specific estimator, which is feasible in practice and which performed well in the simulation study carried through in Section 5. By feasible we mean that there is a simple algorithm to calculate T_i and that there are explicit formulae to estimate the structural parameters.

4.1. Robust credibility in the Bühlmann and Straub model with identical volumes

In this subsection we assume that the volume measures in the Bühlmann and Straub model are identical, i.e. $V_{ij} = V$ ($i = 1, 2, . . ., N, j = 1, 2, .. , n$) In this case an M-estimator T_i is implicitly defined by

$$(18) \qquad \sum_{j=1}^{n} \psi(X_{ij}, T_i) = 0.$$

If we assume a scale model, i.e. $F_\vartheta(x) = P(X_{ij} \le x | \Theta_i = \vartheta) = F(x/\vartheta)$, then it is natural to put

$$(19) \qquad \psi(x, \vartheta) = \psi(x/\vartheta).$$

A typical example of a scale model is model I of Section 5, where the X_{ij} are supposed to be Γ-distributed with shape parameter γ and scale parameter Θ_i. The standard credibility estimator is exact Bayesian in this model. However, as soon as the true underlying model deviates only slightly from the assumed model, then the standard credibility estimator might be rather poor. For instance, in model II of Section 5, it is assumed that the bulk of the data is well described by model I, but that some few observations are taken from another 'xs-urn'. In this model II one can show that the influence function of the exact Bayesian estimator is limited and has the following shape: it is first linearly growing, takes somewhere its maximum and tends then to zero for $x \to \infty$. This and the fact that robustifying an estimator means to use a limited influence function, motivates us to choose

$$(20) \qquad \psi(x) = \min(x - 1, 1).$$

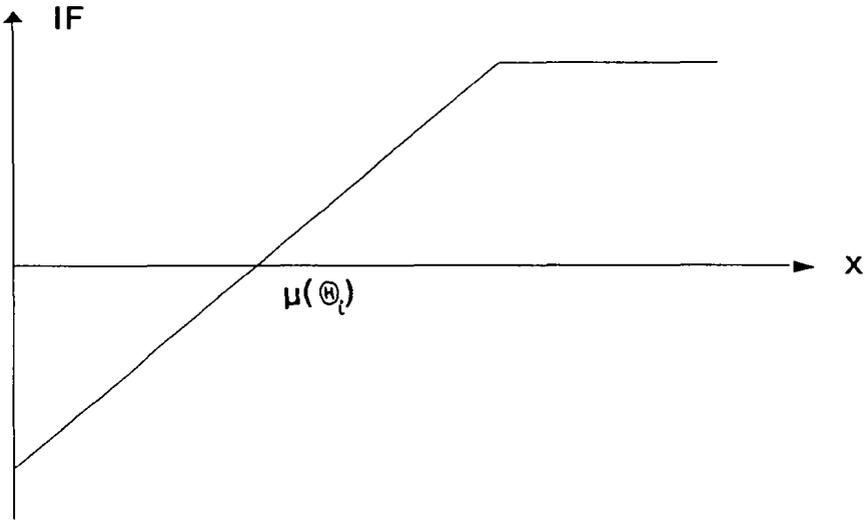This is equivalent to the influence function drawn in Figure 2.

FIGURE 2 Influence function of the estimator T_i given by (18) and (20)

Inserting (19) and (20) into (18) yields

$$(21) \qquad \sum_{j=1}^{n} \min\left(\frac{X_{ij}}{T_i} - 1, 1\right) = 0.$$

Since $\min\left(\dfrac{x}{t} - 1, 1\right) = \dfrac{\min(x, 2t)}{t} - 1$, we obtain

$$(22) \qquad T_i = \frac{1}{n} \sum_{j=1}^{n} \min(X_{ij}, 2\,T_i).$$

Remarks

— Note that T_i is a weighted mean of trimmed data with a data-dependent trimming point.
— The robust estimator T_i is given by an implicit equation.
— By the choice of the ψ-function and by (10) it becomes obvious that the influence of a yearly observation is limited to the influence of a zero-observation. Hence it is not surprising that the estimator (22) is identical to the one used by AMMETER (1982) in the case of identical volume measures.

For finding an algorithm to solve this implicit equation, we consider the function

$$(23) \qquad f(t) = \frac{1}{n} \sum_{j=1}^{n} \min(X_{ij}, 2\,t).$$

Denoting by $\{X_{i(j)} : j = 1, \ldots, n\}$ the order statistics of $\{X_{ij} : j = 1, \ldots, n\}$ and by l_i the number of zero-observations of risk i we obtain

$$
f'(t) = \begin{cases}
\dfrac{2}{n}(n - l_i) & \text{for} \qquad t \le \tfrac{1}{2} X_{i(l_i + 1)} \\[2ex]
2\,\dfrac{n - j}{n} & \text{for} \quad \tfrac{1}{2} X_{i(j)} < t \le \tfrac{1}{2} X_{i(j+1)} \quad j = l_i + 1, \ldots, n - 1 \\[2ex]
0 & \text{for} \quad \tfrac{1}{2} X_{i(n)} < t
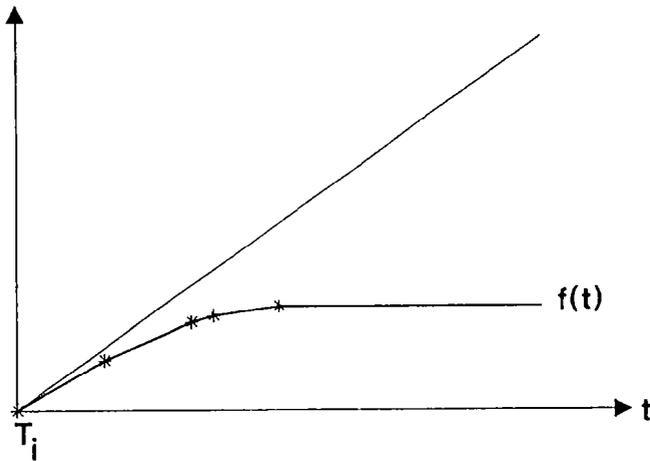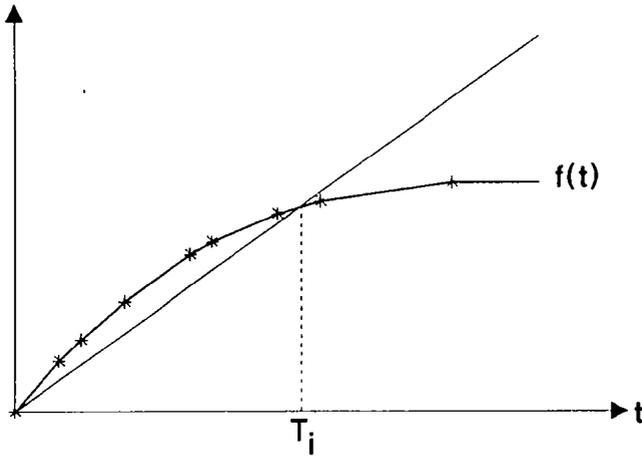\end{cases}
$$

FIGURE 3   Two possible graphes of function $f(t)$ given by (23)

Thus T_i can be calculated by the following procedure:

Calculate $T_i^{(j)} = f(X_{i(j)}/2)$ for $j = n, n-1, \ldots$ until $T_i^{(j)} > X_{i(j)}/2$ and let k_i be the first index, for which this inequality is fulfilled. If such a $k_i \geq 1$ exists then

(24)
$$T_i = \frac{\sum_{j=1}^{k_i} X_{i(j)}}{2k_i - n}, \quad \text{otherwise} \quad T_i = 0.$$

Remarks

— If $X_{i(n)} \leq 2\overline{X}_i$ then $T_i = \overline{X}_i$.
— If half or more of the observations X_{ij} are zero then $T_i = 0$. To be more precise, if exactly half of the observations are zero, then every point in the interval $[0, 0.5 \cdot X_{i, n/2}]$ would be a solution of (21). The algorithm (24) takes the zero-solution in this case.

Figure 4 illustrates the effect of applying the estimator T_i on the data of Figure 1. The observations in year 1 and in year 7 are trimmed down to 1.16‰ with the effect that the arithmetic mean $\overline{X}_i = 0.66‰$ is reduced by 12% to $T_i = 0.58‰$.

To find *the empirical credibility estimator* we have to estimate the structural parameters occurring in (16) and (17). As there exists in general no explicit
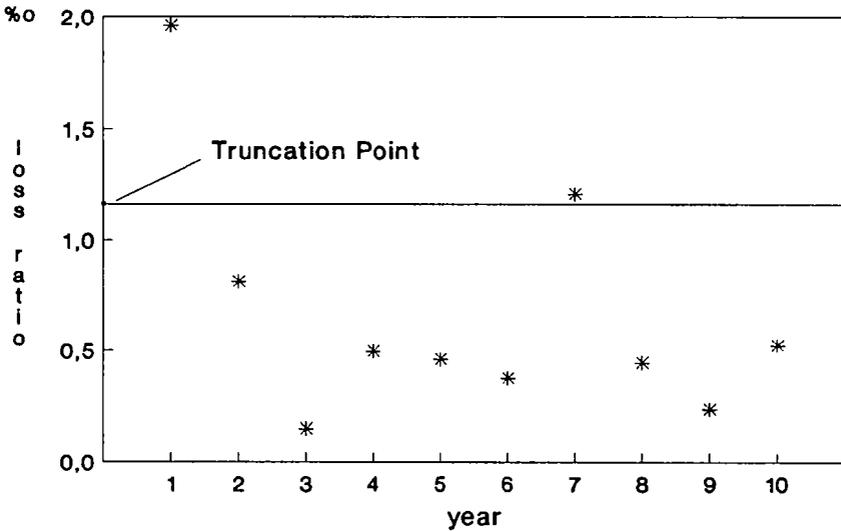

FIGURE 4 Loss ratios as in Figure 1

$\overline{X}_i = 0.66$ $T_i = 0.58$ Truncation Point = 1.16

formula for $\mu_{T_i}(\Theta_i)$ and $\text{Var}[T_i|\Theta_i]$, we replace $\mu_{T_i}(\Theta_i)$ by the asymptotic expectation $T_i(F_{\Theta_i})$ given by (9) and $\text{Var}[T_i|\Theta_i]$ by n^{-1} times the asymptotic variance $V(T_i, F_{\Theta_i})$ given by (11). Then we get

(25) $$\hat{\mu}_o(\Theta_i) \cong \mu_T + \frac{nv_T}{nv_T + u_T}(T_i - \mu_T)$$

$$\text{where} \quad v_T = \text{Var}[T_i(F_{\Theta_i})]$$
$$u_T = E[V(T_i, F_{\Theta_i})]$$
$$\mu_T = E[\mu_{T_i}(\Theta_i)].$$

To estimate the unknown structural parameters μ_{xs}, μ_T, u_T, v_T out of the data, we write (22) as

(26) $$T_i = \frac{1}{n}\sum_{j=1}^{n} T_{ij} \quad \text{with} \quad T_{ij} = \min(X_{ij}, 2\,T_i).$$

For convenience we call

$$T_{ij} \qquad \text{the ordinary claim amount and}$$
$$XS_{ij} = X_{ij} - T_{ij} \qquad \text{the } xs\text{-claim amount.}$$

The situation for estimating μ_T and v_T is identical to the one in the Bühlmann and Straub model. It is slightly different for u_T, since the r v. T_{ij} $(j = 1, \ldots, n)$ are not conditionally independent, given Θ_i. Therefore we estimate $\text{Var}[T_i|\Theta_i]$ by $n^{-1}\hat{u}_i$, where $\hat{u}_i$ is an estimate of the asymptotic variance $V(T_i, F_{\Theta_i})$. By replacing F_{Θ_i} in (11) by the empirical distribution of the $\{X_{ij}; j = 1, \ldots, n\}$ we obtain after some straightforward calculations and after changing the norming constant from n^{-1} to $(n-1)^{-1}$

(27) $$\hat{u}_i = \frac{\dfrac{1}{n-1}\sum_{j=1}^{n}(T_{ij} - T_i)^2}{\left(1 - \dfrac{2}{n}\sum_{j=1}^{n} 1_{[X_{ij} > 2T_i]}\right)^2}$$

where $1_{[\;]}$ is the indicator function

Remark

Note that the denominator of (27) is equal to 1 in the case where all $X_{ij} \leq 2\,T_i$, i e in the case where $T_i = \bar{X}_i$

Thus we finally arrive at the following estimators, which are the analogue of the estimators in the Buhlmann and Straub model:

(28.1) $$\hat{\mu}_T = N^{-1}\sum_{i=1}^{N} T_i$$

$$(28.2) \qquad \hat{\mu}_{xs} = N^{-1} \sum_{i=1}^{N} XS_i \qquad \text{where} \quad XS_i = \frac{1}{n} \sum_{j=1}^{n} XS_{ij}$$

$$(28.3) \qquad \hat{u}_T = N^{-1} \sum_{i=1}^{N} \hat{u}_i \qquad \text{where } \hat{u}_i \text{ is given by (27)}$$

$$(28.4) \qquad \hat{v}_T = (N-1)^{-1} \sum_{i=1}^{N} (T_i - \hat{\mu}_T)^2 - n^{-1} \hat{u}_T.$$

By inserting (28) into (16) and into (25) we get *the empirical robust credibility estimator*

$$(29) \qquad \hat{\hat{\mu}}_X(\Theta\iota) = \hat{\mu}_{xs} + \hat{\mu}_T + \hat{\alpha}(T_i - \hat{\mu}_T)$$

$$\text{where} \quad \hat{\alpha} = \frac{n\hat{v}_T}{n\hat{v}_T + \hat{u}_T}.$$

4.2. Robust credibility in the general Bühlmann and Straub model

Contrary to subsection 4.1 we shall now allow the volume measures V_{ij} to be different. Then we have to generalize (18)-(20).

Assume for the moment that the volume measures are natural numbers. Then we arrive at the general Buhlmann and Straub model by looking at the X_{ij} as being averages of V_{ij} independent (unobservable) random variables $Y_{ij}^{(v)}$, i e. $X_{ij} = V_{ij}^{-1} \sum_{v=1}^{V_{ij}} Y_{ij}^{(v)}$, where the $Y_{ij}^{(v)}$ fulfill the conditions of the Bühlmann and Straub model with identical volumes $V_{ij} \equiv 1$. By replacing the unobservable $Y_{ij}^{(v)}$ ($v = 1, 2, \ldots, V_{ij}$) by the 'observed' average X_{ij} and inserting them into (18) we get

$$(30) \qquad \sum_{j=1}^{n} V_{ij} \, \psi(X_{ij}, T_i) = 0.$$

It is an obvious and natural idea to give more weight to an observation belonging to a cell (i, j) with a big volume measure and to use the volume measures as weights. But we also have to modify the ψ-function (20). With the idea of the X_{ij} being observed averages it becomes obvious that the observed loss ratios will be the more smoothed the bigger the corresponding volumes V_{ij} If we simply used (20), then a risk ι with small volumes V_{ij} would be favoured compared to a risk k with big volumes V_{kj}, since the corresponding yearly observations X_{ij} of risk ι would have a bigger 'chance' of being truncated than the yearly observations X_{kj} of risk k. (20) leads to the estimator (22), where observations X_{ij} belonging to the interval $[0, 2T_i]$ are not truncated and are considered as 'ordinary' observations. Hence we might look at the interval

$[0, 2 T_i]$ as something like a confidence interval. The generalization to a model with different volumes is to make this interval dependent on the volume V_{ij}, i.e. to consider intervals of the form $[0, (1 + f(V_{ij})) T_i]$. This is identical to making the Ψ-function dependent on the volume V_{ij} and to replacing (20) by

(31) $\psi(x, V_{ij}) = \min(x - 1, f(V_{ij}))$.

Since $\sqrt{\mathrm{Var}[X_{ij}|\Theta_i, V_{ij}]} = V_{ij}^{-\frac{1}{2}} \sigma(\Theta_i)$, we put

(32) $f(V_{ij}) = c \cdot V_{ij}^{-\frac{1}{2}}$

> where c is a suitably chosen constant

(31)   should be about the same as (20) for a risk i with average volume. Thus natural candidates for the choice of c are

$$c_1 = \sqrt{\bar{V}} \qquad \text{with } \bar{V} = \frac{1}{nN} \sum_{i=1}^{N} \sum_{j=1}^{n} V_{ij} \quad \text{or}$$

$$c_2 = \sqrt{\mathrm{median}(V_{ij})} \qquad (i = 1, 2, \ldots, N; j = 1, 2, \ldots, n)$$

The authors suggest to use normally c_1 and to give preference to c_2 in cases, where the volumes of the different risks in the portfolio have a distribution, which is very skew.

By putting $\psi(x, \vartheta) = \psi(x/\vartheta)$ as in Subsection 4.1 and the inserting (32) and (31) into (30) we obtain

(33) $\sum_{j=1}^{n} V_{ij} \cdot \min\left(\frac{X_{ij}}{T_i} - 1, c V_{ij}^{-\frac{1}{2}}\right) = 0$

> with $c = c_1$ (or $c = c_2$).

Note that (33) is a generalization of (21) to the case of different volume measures Another derivation and justification of (33) and the resulting estimator (34) are given in the appendix.

By the same arguments as used in the derivation of (22) we easily find that (33) leads to

(34) $T_i = \sum_{j=1}^{n} \frac{V_{ij}}{V_i} \min(X_{ij}, c_{ij} T_i)$

> where $c_{ij} = 1 + c V_{ij}^{-\frac{1}{2}}$

$$V_i = \sum_{j=1}^{n} V_{ij}.$$

Note that T_i is the solution of the implicit equation

(35) $T_i = f(T_i)$

$$\text{where } f(t) = \sum_{j=1}^{n} \frac{V_{ij}}{V_i} c_{ij} \min(Z_{ij}, t)$$

$$Z_{ij} = c_{ij}^{-1} X_{ij}.$$

Denoting by $Z_{i(j)}$ the order statistics of Z_{ij} and by $V_{i(j)}$ and $c_{i(j)}$ the corresponding V_{ij} and c_{ij}, we find by using the same arguments as in Subsection 4.1 the following algorithm to solve (35):

Calculate $T_i^{(j)} = f(Z_{i(j)})$ for $j = n, n-1, \dots$ until $T_i^{(j)} > Z_{i(j)}$ and let k_i be the first index, for which this inequality is fulfilled. If such a $k_i \geq 1$ exists then

(36) $$T_i = \frac{\displaystyle\sum_{j=1}^{k_i} V_{i(j)} c_{i(j)} Z_{i(j)}}{\displaystyle V_i - \sum_{j=k_i+1}^{n} c_{i(j)} V_{i(j)}}, \quad \text{otherwise} \quad T_i = 0.$$

Remarks

— If $Z_{i(n)} \leq \bar{X}_i$ then $T_i = \bar{X}_i = V_i^{-1} \sum_{j=1}^{n} V_{ij} X_{ij}$.

— If l_i denotes the number of zero-observations and if $\sum_{j=l_i+1}^{n} V_{i(j)} c_{i(j)} \leq V_i$,

then $T_i = 0$. To be more precise, if we have strict equality in the above equation, then every point in the interval $[0, Z_{i(l_i+1)}]$ would be a solution of (33). The algorithm (36) takes the zero-solution in this case.

To find *the empirical credibility estimator* we have again to estimate the structural parameters occurring in (16) and (17) from the data. Because $F_{\Theta_i}(x) = P(X_{ij} \leq x | \Theta_i)$ as well as the ψ-function itself depend on the volume measure V_{ij}, a strict mathematical treatment becomes unfeasible. With the modification made in the ψ-function (see (31) and (32)) we can assume that $E[T_i | \Theta_i]$ is approximately independent of the underlying volumes. We approximate it by $T(F_{\Theta_i})$, where $T(F_{\Theta})$ is the asymptotic expectation for a risk with volumes $V_{ij} \equiv 1$. The variance $\text{Var}[T_i | \Theta_i]$ clearly depends on the underlying volumes. The variation of the volume measures over time within a risk, i.e. the variation of V_{ij} ($j = 1, 2, \dots, n$), is in most practical situations rather small. But there might be substantial differences of volumes between risks. Assume for the moment, that the volumes V_{ij} ($j = 1, 2, \dots, n$) within a risk are fixed and identical to $\tilde{V}_i$. Since $\text{Var}[X_{ij} | \Theta_i, \tilde{V}_i] = \sigma^2(\Theta_i)/\tilde{V}_i$, we might assume that $\text{Var}[T_i | \Theta_i, \tilde{V}_i] \cong V(T, F_{\Theta_i})/n\tilde{V}_i$, where $V(T, F_{\Theta_i})$ is the asymptotic variance of

a risk with volume 1. Hence a natural straightforward generalization of (25) to the case of different volume measures is

$$(37) \quad \hat{\mu}_o(\Theta_i) \approx \mu_T + \frac{V_i v_T}{V_i v_T + u_T} (T_i - \mu_T)$$

where $\quad V_i = \sum_{j=1}^{n} V_{ij}$

$v_T = \mathrm{Var}\,[T(F_{\Theta_i})]$

$T(F_{\Theta_i}) =$ asymptotic expectation for a risk with volumes $V_{ij} \equiv 1$

$u_T = E[V(T, F_{\Theta_i})]$

$V(T, F_{\Theta_i}) =$ asymptotic variance for a risk with volumes $V_{ij} \equiv 1$.

To estimate the structural parameters μ_{xs}, μ_T, u_T, v_T write (34) as

$$(38) \qquad T_i = \sum_{j=1}^{n} \frac{V_{ij}}{V_i} T_{ij} \quad \text{with} \quad T_{ij} = \min\,(X_{ij}, c_{ij} T_i).$$

For convenience we call

$\qquad T_{ij}$ $\qquad\qquad\qquad$ the observed *ordinary loss ratio*

$\qquad XS_{ij} = X_{ij} - T_{ij}$ $\qquad$ the observed *xs-loss ratio*

and

$\qquad V_{ij} T_{ij}$ $\qquad\qquad\qquad$ the ordinary claim amount

$\qquad V_{ij} XS_{ij}$ $\qquad\qquad$ the *xs*-claim amount.

Inserting the empirical distribution function G_i of the X_{ij} ($j = 1, 2, \ldots, n$) into (10), we find after some straightforward calculations

$$(39) \qquad \mathrm{IF}\,(x_{ij}; \psi, G_i) = \frac{T_{ij} - T_i}{1 - \sum_{j=1}^{n} \frac{V_{ij}}{V_i} c_{ij}\, 1_{[T_{ij} \neq X_{ij}]}}.$$

Since $\mathrm{Var}\,[X_{ij}|\Theta_i, V_{ij}] = V_{ij}^{-1} \cdot \sigma^2(\Theta_i)$ we can assume that $\mathrm{Var}\,[T_{ij}|\Theta_i, V_{ij}] \cong V_{ij}^{-1} \cdot V(T, F_{\Theta_i})$. Hence we suggest to estimate $V(T, F_{\Theta_i})$ by

$$(40) \qquad \hat{u}_i = \frac{\dfrac{1}{n-1} \sum_{j=1}^{n} V_{ij}(T_{ij} - T_i)^2}{\left(1 - \sum_{j=1}^{n} \dfrac{V_{ij}}{V_i} c_{ij}\, 1_{[T_{ij} \neq X_{ij}]}\right)^2}$$

Of course (40) should be considered as the generalization of (27) to the case of different volume measures. Then by using the analogue estimators as usually used in the Bühlmann and Straub model (for a discussion see DUBEY and GISLER (1981)) we obtain

(41.1) $$\hat{u}_T = N^{-1} \sum_{i=1}^{N} \hat{u}_i$$

(41 2) $$\hat{v}_T = c^{-1} \left\{ \sum_{i=1}^{N} \frac{V_i}{V} (T_i - T)^2 - (N-1) \frac{\hat{u}_T}{V} \right\}$$

where $$V = \sum_{i=1}^{N} V_i$$

$$T = V^{-1} \sum_{i=1}^{N} V_i T_i$$

$$c = \sum_{i=1}^{N} \frac{V_i}{V} \left(1 - \frac{V_i}{V} \right)$$

(41.3) $$\hat{\mu}_T = \left(\sum_i \hat{\alpha}_i \right)^{-1} \sum_i \hat{\alpha}_i T_i$$

where $$\hat{\alpha}_i = \frac{V_i \hat{v}_T}{V_i \hat{v}_T + \hat{u}_T}$$

(41.4) $$\hat{\mu}_{xs} = V^{-1} \sum_{i=1}^{N} V_i XS_i \qquad \left(XS_i = \sum_{j=1}^{n} \frac{V_{ij}}{V_i} XS_{ij} \right).$$

Thus the *empirical robust credibility formula in the case of different volume measures* is given by

(42) $$\hat{\mu}_X(\Theta_i) = \hat{\mu}_{xs} + \hat{\mu}_T + \hat{\alpha}_i (T_i - \hat{\mu}_T)$$

where $$\hat{\alpha}_i = \frac{V_i \hat{v}_T}{V_i \hat{v}_T + \hat{u}_T}$$ and

where the estimators of the structural parameters are given by (41).

Remarks

1) $$\sum_{i=1}^{N} V_i \hat{\mu}_X(\Theta_i) = \sum_{i=1}^{N} V_i X_i.$$

By replacing for each risk i the observed claim amount $V_i X_i$ by the corresponding pure risk premium $V_i \hat{\mu}_X(\Theta_i)$, we get a mathematical

allocation of the total claims amount. Thus our formula provides us with a tariff, which is fair for the portfolio as a whole.

ii)   The authors would like to emphasize once more that a strict mathematical treatment in the case of different volume measures becomes unfeasible. But they believe that the proposed estimators are reasonable and useful for practical purposes. This is also confirmed by the simulation study presented in Section 5. It should also be noted that a suboptimal estimation of the credibility weight α_i is not very sensitive to the quality of the credibility estimator with respect to quadratic loss.

5. A SIMULATION STUDY

In order to test the proposed method and to illustrate its functioning we have carried through a simulation study. We have simulated data of two different models. In both models we consider portfolios of N risks numbered $i = 1, 2, \ldots, N$. We assume that each risk i is characterized by a hidden risk parameter Θ_i. To each risk i belongs an observation vector $X_i = (X_{i1}, \ldots, X_{in})'$, where X_{ij} might be interpreted as the loss ratio of risk i in year j. To each cell (i, j) is given a volume measure V_{ij} (a natural number) and it is assumed that X_{ij} is the mean of (unobservable) r.v $Y_{ij}^{(v)}$, i.e.

$$X_{ij} = (V_{ij})^{-1} \sum_{v=1}^{V_{ij}} Y_{ij}^{(v)}.$$

Model 1

Assumptions

MI1:   Given Θ_i, the random variables $Y_{ij}^{(v)}$ ($j = 1, 2, \ldots, n$; $v = 1, \ldots, V_{ij}$) are independent and Gamma (γ, Θ_i^{-1}) distributed.

MI2:   The pairs $(\Theta_1, X_1), \ldots, (\Theta_N, X_N)$ are independent and $\Theta_1^{-1}, \ldots, \Theta_N^{-1}$ are Gamma (α, β) distributed.

Remarks

—   Note that given Θ_i, the X_{ij} are Gamma $(V_{ij}\gamma, V_{ij}\Theta_i^{-1})$ distributed.

—   The unconditional distribution of the $Y_{ij}^{(v)}$ is, in the terminology of HOGG and KLUGMAN (1984), a generalized Pareto distribution with density function

$$f_0(y) = \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha)\,\Gamma(\gamma)} \left(\frac{\beta}{\beta+y}\right)^\alpha \left(\frac{y}{\beta+y}\right)^\gamma \frac{1}{y}.$$

—   It is well known that in model I the credibility estimator is exact Bayesian and that

(43) $\tilde{\mu}(\Theta_i) = \hat{\mu}(\Theta_i) = \mu + \alpha_i(\overline{X}_i - \mu)$

$$\text{with} \quad \mu = E[X_{ij}] = \gamma \, \frac{\beta}{\alpha - 1}$$

$$\alpha_i = \frac{\gamma V_i}{\gamma V_i + (\alpha - 1)} \, .$$

Now assume that model I is a good description for the bulk of the data. But some of the data (the outliers) are generated by another law. This is formalized by assuming that occasionally, i.e. with probability π, $Y_{ij}^{(v)}$ is drawn from the 'xs-urn' with density $f_e(x)$ This leads to

Model II

Assumptions

MII1 · Given Θ_i, the random variables $Y_{ij}^{(v)}$ ($j = 1, 2, \ldots, n$; $v = 1, \ldots, V_{ij}$) are independent with density function

$$f_{\Theta_i}(y) = (1 - \pi) f(y|\Theta_i) + \pi f_e(y)$$

where $f(y|\Theta_i)$ is the density of the Gamma (γ, Θ_i^{-1}) distribution and

where $f_e(y) = \dfrac{\Gamma(a+c)}{\Gamma(a)\,\Gamma(c)} \left(\dfrac{b}{b+y} \right)^a \left(\dfrac{y}{b+y} \right)^c \dfrac{1}{y}$

is the density of a generalized Pareto distribution with a, b, c fixed constants.

MII2: The pairs $(\Theta_1, X_1), \ldots, (\Theta_N, X_N)$ are independent and $\Theta_1^{-1}, \ldots, \Theta_N^{-1}$ are Gamma (α, β) distributed.

MII3: $a < \alpha$.

Remark

The assumption MII3 means that large $Y_{ij}^{(v)}$ are more likely to come from the 'xs-urn' than from the 'ordinary' urn.

The observation period in our simulation was 6 years. The portfolio contained 100 risks with volume $V_{ij} \equiv 1$ (small volumes), 100 risks with volume $V_{ij} \equiv 3$ (medium volumes) and 100 risks with volume $V_{ij} \equiv 5$ (great volumes)

In model I we have chosen $\gamma = 2$, $\alpha = 5$, $\beta = 2$. This means, that

$$\mu_X = E[X_{ij}] = 1$$
$$u_X = E[\text{Var}\,[X_{ij}|\Theta_i; V_{ij} = 1]] = 0.667$$
$$v_X = \text{Var}\,[E[X_{ij}|\Theta_i]] = 0.333 \, .$$

The simulation gave the following results:

— standard credibility estimator
   structural parameters:
      $\hat{\mu}_X = 1.017$, $\hat{u}_X = 0.618$, $\hat{v}_X = 0.356$

credibility weights (for volumes 1, 3, 5).

$\hat{\alpha}_1 = 0.775$, $\hat{\alpha}_{101} = 0.912$, $\hat{\alpha}_{201} = 0.945$

mean quadratic loss:

$MQL = 0.0352$ $\sqrt{MQL} = 0.188$.

— *robust credibility estimator* ·

structural parameters:

$\hat{\mu}_{xs} = 0.003$, $\hat{\mu}_T = 1\,013$, $\hat{u}_T = 0.666$, $\hat{v}_T = 0.353$

credibility weights (for volumes 1, 3, 5):

$\hat{\alpha}_1 = 0.761$, $\hat{\alpha}_{101} = 0.905$, $\hat{\alpha}_{201} = 0.941$

mean quadratic loss ·

$MQL = 0.0358$ $\sqrt{MQL} = 0.189$.

Remember that the standard credibility estimator is exact Bayesian in this model. It is therefore not surprising that the mean quadratic loss of the robust estimator is greater than the one of the standard credibility estimator. However the difference is only 2%. Hence the loss of efficiency by using the robust credibility estimator instead of the 'optimal estimator' is very small This was confirmed by three other simulations. We refrain from listing the simulated loss ratios of the risk in the portfolio. The differences between the robust and the standard empirical credibility estimators were very small.

In model II we have chosen a probability π of 5% for $Y_{ij}^{(v)}$ being an 'outlier'. The parameters of the 'outlier density' $f_e(y)$ were $a = 3$, $b = 10$, $c = 1$, which gives an expectation of 5 and a variance of 50. Of course the parameters of the 'ordinary density' were the same as in the simulation of model I. Thus the structural parameter of the standard credibility estimator were

$$\mu_X = 1.20$$
$$u_X = 3.893$$
$$v_X = 0.301.$$

The simulation gave the following results:

— *standard credibility estimator*.

structural parameters ·

$\hat{\mu}_X = 1.246$, $\hat{u}_X = 4.308$, $\hat{v}_X = 0.218$

credibility weights (for volumes 1, 3, 5):

$\hat{\alpha}_1 = 0.233$, $\hat{\alpha}_{101} = 0.476$, $\hat{\alpha}_{201} = 0.602$

mean quadratic loss ·

$MQL = 0.1390$ $\sqrt{MQL} = 0.373$.

— *robust credibility estimator:*

structural parameters:

$\hat{\mu}_{xs} = 0.108$, $\hat{\mu}_T = 1.113$, $\hat{u}_T = 1.759$, $\hat{v}_T = 0.243$

credibility weights (for volumes 1, 3, 5):

$\hat{\alpha}_1 = 0.453$, $\hat{\alpha}_{101} = 0.713$, $\hat{\alpha}_{201} = 0.806$

mean quadratic loss:

$$\text{MQL} = 0.0843 \quad \sqrt{\text{MQL}} = 0.290.$$

Using the robust credibility estimator instead of the standard credibility estimator reduces the mean quadratic loss by 40%. This order of magnitude was obtained in several simulation runs. Thus the robust credibility estimator performs substantially better than the standard credibility estimator. Note that the credibility weights are much bigger for the robust estimator. Table 1 shows the simulated figures of the first fifteen risks of each volume group. Risk Nr. 1 has a big outlier. The standard credibility estimator increases drastically whereas the robust estimator reacts reasonably. But there are also differences in cases where no outliers were observed (see for instance risk Nr. 5) The robust estimator is usually nearer to the true value also in this case, due to the greater credibility weight.

Further remark

It would also be interesting to compare the results obtained by the robust credibility estimator with the ones which would be obtained by the method of optimum trimming in credibility proposed by GISLER (1980). The authors have renounced it in order not to overload this paper. The main difference between optimum trimming in credibility and the robust credibility presented in this paper is that in the first case the individual claim amounts $Y_{ij}^{(v)}$ are trimmed at a trimming point which is the same for all risks in the portfolio, whereas in the robust credibility approach the observed loss ratios X_{ij} are trimmed with a trimming point depending on the claims experience of the particular contract. In the first case the individual claim amounts have to be known, whereas in the second case only the knowledge of the loss ratios is necessary

ALOIS GISLER AND PETER REINHARD

TABLE 1
FIGURES OF A SIMULATION OF MODEL II

nsk number i	order statistics $X_{(ij)}$						individual experience X_i	T_i	credibility estimator standard	robust	true value $\mu_X(\Theta_i)$
1	72 3	89 4	96 7	235 7	286 0	4 798 0	929 7	238 7	311 8	238 7	159 0
2	77 7	115 6	133 8	201 4	201 6	667 3	232 9	223 4	149 8	223 4	232 4
3	37 8	74 1	175 2	223 7	332 3	335 6	196 5	196 5	141 3	196 5	261 1
4	11 4	28 3	36 0	45 2	118 5	653 0	148 7	73 3	130 2	73 3	65 4
5	34 9	37 2	118 9	296 1	312 6	313 2	185 5	185 5	138 7	185 5	210 4
6	3 4	41 8	48 2	71 1	160 8	179 4	84 1	84 1	115 2	84 1	87 8
7	3 3	18 3	21 8	30 8	84 2	183 2	56 9	48 5	108 8	48 5	83 5
8	12 2	33 0	38 2	36 7	59 5	71 5	41 5	41 5	105 3	41 5	93 9
9	10 9	24 1	79 7	131 4	133 8	150 3	88 4	88 4	116 2	88 4	119 5
10	58 3	58 9	82 9	111 5	112 2	141 0	94 1	94 1	117 5	94 1	87 9
11	75 7	120 7	146 9	183 4	228 7	523 8	213 2	213 2	145 2	213 2	276 6
12	57 7	156 3	176 3	178 0	254 6	384 2	201 2	201 2	142 4	201 2	167 2
13	39 7	79 0	89 8	158 0	164 3	354 6	147 6	147 6	129 9	147 6	99 8
14	20 0	65 3	93 8	95 2	112 5	121 2	84 7	84 7	115 3	84 7	89 1
15	40 0	64 8	75 9	147 8	179 7	298 9	134 5	134 5	126 9	134 5	134 0
101	41 7	41 8	44 6	65 5	81 7	85 6	60 2	60 2	93 9	86 0	97 8
102	16 8	33 0	33 6	51 2	109 9	310 5	92 5	61 1	109 3	86 7	72 2
103	103 8	116 7	117 3	127 3	136 6	732 0	222 3	150 4	171 1	150 4	121 0
104	26 7	42 3	43 3	50 9	51 7	88 6	50 6	50 6	89 3	79 2	74 1
105	70 3	90 4	100 0	125 1	131 4	160 1	112 9	112 9	119 0	123 6	112 6
106	42 9	83 5	85 3	111 8	120 4	124 2	94 7	94 7	110 3	110 7	115 6
107	35 4	39 3	41 9	44 0	81 9	125 0	61 3	60 6	94 4	86 4	88 6
108	18 7	20 3	34 8	40 8	87 2	433 8	102 6	45 4	114 1	75 6	69 0
109	49 7	70 7	77 9	86 1	104 5	105 4	82 4	82 4	104 5	101 9	106 5
110	132 3	188 2	191 5	211 5	256 3	744 1	287 3	244 9	202 1	217 8	188 2
111	26 5	31 8	35 9	44 9	63 1	233 6	72 6	50 6	99 8	79 2	69 5
112	43 2	83 4	114 1	117 7	128 5	532 0	169 8	121 7	146 1	129 9	142 0
113	32 4	45 4	52 3	67 5	97 1	110 2	67 5	67 5	97 4	91 3	87 4
114	71 7	82 8	86 7	91 0	113 8	303 1	124 9	111 5	124 7	122 7	98 0
115	43 2	45 1	64 6	66 0	118 7	528 2	144 3	84 4	134 0	103 3	114 6
201	108 4	112 6	134 8	134 9	160 8	192 8	140 7	140 7	134 3	146 1	154 1
202	112 8	120 8	177 5	213 6	229 4	281 4	189 2	189 2	163 5	185 2	178 2
203	40 4	40 5	45 0	46 1	80 0	148 2	66 7	59 6	89 7	80 8	78 0
204	86 7	121 1	150 4	163 7	227 6	508 0	209 6	177 4	175 8	175 6	117 0
205	39 8	41 0	48 8	58 8	70 5	234 8	82 3	61 3	99 1	82 1	66 9
206	91 7	95 8	98 8	137 1	164 0	179 1	127 8	127 8	126 5	135 6	146 0
207	101 5	101 6	117 5	119 2	133 4	144 8	119 7	119 7	121 6	129 1	140 2
208	24 7	26 5	28 5	39 1	66 7	122 7	51 4	43 9	80 5	68 1	60 3
209	137 1	208 8	212 0	216 2	245 1	323 4	223 8	223 8	184 3	213 0	235 5
210	29 6	40 4	43 4	52 4	57 6	77 1	50 1	50 1	79 7	73 1	76 6
211	58 6	82 4	97 0	117 4	118 7	141 9	102 7	102 7	111 4	115 4	138 2
212	40 2	50 8	50 9	52 5	55 3	816 9	177 8	59 1	156 6	80 3	80 8
213	37 0	48 5	51 8	55 8	56 0	117 1	61 0	59 0	86 3	80 2	92 3
214	71 6	80 1	93 6	110 6	116 1	135 9	101 3	101 3	110 6	114 3	118 0
215	97 1	111 8	142 4	143 5	159 0	198 2	142 0	142 0	135 1	147 1	153 0

APPENDIX: DERIVATION OF THE ESTIMATOR (33)

a) Derivation based on Huber's estimator

HUBER (1964) (see also HAMPEL and alii (1986), p. 172) studied robust estimation of location by determining M-estimators that are optimal in a minimax sense. The model framework was the following:

The distribution of X_j ($j = 1, 2, \ldots, n$) is F, a symmetric distribution with location parameter ϑ. Let $P_\varepsilon = \{G \ . \ G = (1 - \varepsilon) F + \varepsilon H, \ H \text{ symmetric}\}$ be a set of distributions in the neighbourhood of F.

Huber was looking for the M-estimator $\tilde{\psi}$ minimizing the maximal asymptotic variance over P_ε, that is

$$(44) \qquad V(T(\tilde{\psi}), G) = \min_{\psi} \ \sup_{G \in P_\varepsilon} \ V(T(\psi), G)$$

where $T(\psi)$ is the M-estimator defined by (8).

The solutions of (44) for $F = \Phi$ (normal distribution) are

$$(45) \quad \tilde{\psi}(x, \vartheta) = \psi_c(x, \vartheta) = \text{sign}(x - \vartheta) \min(|x - \vartheta|, c) \qquad \text{(Huber estimator)}$$

In the case of different $\sigma_j^2 = \text{Var}[X_j]$, the Huber estimator T is obtained by applying (45) on the normalized data. Using (7) yields, that T is given by

$$(46) \qquad \sum_{j=1}^{n} \rho_c \left(\frac{X_j - T}{\sigma_j} \right) = \min_{\vartheta} \ \sum_{j=1}^{n} \rho_c \left(\frac{X_j - \vartheta}{\sigma_j} \right)$$

$$\text{where} \quad \rho_c(x) = \begin{cases} \dfrac{1}{2} x^2 & |x| \le c \\[2ex] c \left(|x| - \dfrac{c}{2} \right) & |x| \ge c \end{cases}$$

(note, that $\rho_c'(x) = \psi(x)$)

or equivalently

$$(47) \qquad \sum_{j=1}^{n} \sigma_j^{-1} \psi_c \left(\frac{X_j - T}{\sigma_j} \right) = 0.$$

Since $\psi_c(ax) = a \psi_{c/a}(x)$, (47) is also equivalent to

$$(48) \qquad \sum_{j=1}^{n} \sigma_j^{-2} \psi_{c\sigma_j}(X_j - T) = 0.$$

In the case of the Bühlmann and Straub model we have

$$\sigma_{ij}^2 = \text{Var}[X_{ij}|\Theta_i] = \sigma^2(\Theta_i)/V_{ij}.$$

In a scale model $\sigma^2(\Theta_i)$ is proportional to Θ_i^2.

Hence applying (48) to the data of risk i in the Bühlmann and Straub case yields

$$(49) \qquad \sum_{j=1}^{n} V_{ij} \operatorname{sign} (X_{ij} - T_i) \min \left(\left| \frac{X_{ij}}{T_i} - 1 \right|, \frac{c}{\sqrt{V_{ij}}} \right) = 0$$

or equivalently

$$(50) \qquad T_i = \sum_{j=1}^{n} \frac{V_{ij}}{V_i} \max \left((1 - r_{ij}) T_i, \min (X_{ij}, (1 + r_{ij}) T_i) \right)$$

$$\text{where} \quad V_i = \sum_{j=1}^{n} V_{ij}$$

$$r_{ij} = c V_{ij}^{-1/2}.$$

Since $X_{ij} \geq 0$ we obtain in the case where $r_{ij} \geq 1$ $(j = 1, 2, . , n)$

$$(51) \qquad T_i = \sum_{j=1}^{n} \frac{V_{ij}}{V_i} \min (X_{ij}, (1 + r_{ij}) T_i)$$

which is exactly the same as formula (34).

In the robust credibility approach we use (51) also in the case where some of the r_{ij} are smaller than 1. One reason is that (50) implies also truncation from below, which would be less accepted in practice than truncation from above Another reason is that (51) is easy to calculate and has, except in the very special case mentioned in Section 4.1, a unique solution which is not the case for (50). Moreover in most cases the differences between (50) and (51) will be pretty small. However, from a pure mathematical point of view, there would be no reason to prefer (51) to (50) On the contrary, when considering the limiting case $c V_{ij}^{-1} \to 0$ for $j = 1, 2, . ., n$, then T_i defined by (51) converges to $\min (X_{ij}; j = 1, 2, \ldots, n)$, whereas T_i defined by (50) converges to median $(X_{ij}; j = 1, 2, . ., n)$.

b) Different argument

In a) we started with the symmetric location case and then suddenly switched to the asymmetric scale case This is somewhat questionable, as pointed out by one of the referees, who mentioned the following different argument for the case where $V_{ij} \equiv V_i$ for $j = 1, 2, . . , n$.

Let $X_1, X_2, \ldots$ be i.i.d. $\sim dG (x/\vartheta)$ and consider the M-estimator T_n defined by

$$\sum_{i=1}^{n} \operatorname{sign} (X_i - T_n) \min \left(b, \left| \frac{X_i}{T_n} - 1 \right| \right) = 0$$

Then b determines the gross-error sensitivity and the problem is how this should depend on G. One reasonable requirement is that the ratio gross-error sensitivity to asymptotic standard deviation of T_n should be independent of G, i.e. we compare the maximal influence of outliers to the precision of the estimator. This is the self-standardized sensitivity of HAMPEL and alii (1986). It means that

$$b^2/E\left[\min\left(b^2, \left(\frac{X_{ij}}{T(G)} - 1\right)^2\right)\right] = \text{const.}$$

If G is the Gamma (γ, γ)-distribution, then this implies at least for $\gamma \to \infty$

$$b(\gamma) \sim \text{const.} \cdot \gamma^{-1/2}$$

because Gamma $(\gamma, \gamma) \approx \mathscr{A}(1, \gamma^{-1})$ as $\gamma \to \infty$. For $V_{ij} \equiv V_i$ this gives (49). For the Gamma-distribution it follows from Section 2.4 of HAMPEL and alii (1986) that the above T_n is optimal in the sense defined there.

REFERENCES

AMMETER, H (1982) Mathematisches Modell zur Kalkulation der Tarife in der schweizerischen Feuer- und Elementarschadenversicherung *Internal Paper, available from the author*

BÜHLMANN, H and STRAUB, E (1970) Glaubwürdigkeit für Schadensätze *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* **70**, 11–113

BÜHLMANN, H, GISLER, A and JEWELL, W (1982) Excess Claims and Data Trimming in the Context of Credibility Rating Procedures *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* **82**, 1, 117–147

DUBEY, A and GISLER, A (1981) On Parameter Estimators in Credibility *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* **81**, 2, 187–212

GISLER, A (1980) Optimales Stutzen von Beobachtungen im Credibility-Modell, *ETH-Thesis Nr 6556* See also GISLER, A (1980) Optimum trimming of data in the Credibility Model *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* **80**, 3, 313–326

GISLER, A (1990) Credibility Theory Made Easy *Mitteilungen der Schweizerischen Vereinigung der Versicherungsmathematiker* **90**, 1, 75–100

HAMPEL, F R (1968) Contributions to the theory of robust estimation *Ph D thesis University of California, Berkeley*

HAMPEL, F R (1974) The influence curve and its role in robust estimation *J Am Statist Assoc* **69**, 383–393

HAMPEL, F R, RONCHETTI, E, ROUSSEEUW, P and STAHEL, W (1986) *Robust Statistics* John Wiley & Sons, New York

HOGG, R V and KLUGMAN, S A (1984) *Loss distributions* John Wiley & Sons, New York

HUBER, P (1964) Robust estimation of a location parameter *Annals of Mathematical Statistics* **35**, 73–101

KREMER, E (1991) Large Claims in Credibility *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* **XX**, 123–150

KÜNSCH, H R (1992) Robust Methods for Credibility *ASTIN Bulletin* **22**, 33–49

REINHARD P (1989) Robuste Schätzungen im Credibility-Modell *Diplomarbeit, ETH Zürich*

STRAUSS, J (1984) Calculation of Premium Rates according to the new German Industrial Fire Tariff 82 *Proceedings of the 4 Countries ASTIN-Symposium, Akersloot*, 3–21

ALOIS GISLER AND PETER REINHARD
'Winterthur', Swiss Insurance Company, Box 357,
CH-8401 Winterthur