ROBUST METHODS FOR CREDIBILITY

by

Hans R. Künsch

Seminar für Statistik

ETH-Zentrum

CH-8092 Zürich

Abstract: Excess claims lead to an unsatisfactory behavior of standard linear credibility estimators. We suggest in this paper to use robust methods in order to obtain better estimators. Our first proposal is the linear credibility estimator with the claims replaced by a robust M-estimator of scale calculated from the claims. This corresponds to a truncation of the claims with a truncation point depending on the data and different for each contract. We discuss the properties of the robust M-estimator and present several examples. In order to improve the performance for a very small number of years, we propose a second estimator, which incorporates information from other claims into the M-estimator.

1 Introduction

This paper attempts to introduce robust methods into the area of credibility. In many actuarial applications excess claims (outliers) which are much bigger than ordinary claims do occur. With linear credibility these excess claims lead to a small credibility factor and thus the individual experience is used only to a small extent. Still those contracts which incurred by bad luck such an excess claim have to pay a high premium. This is not satisfactory. Gisler [1980a, 1980b] has proposed the following solution to this problem: He truncates all claims above a certain level which is determined from the whole portfolio so as to minimize mean square error. It seems however that this method is difficult to use in more complex situations, e.g. for hierarchical credibility or when only claim rates with different volume measures are available. As a theoretician I also wondered if a single truncation level for all contracts is always appropriate. For a further discussion of excess claims in credibility we refer the reader also to the paper by Gisler and Reinhard (1990).

Robust statistics has had an enormous development during the past 25 years. According to Hampel et al. (1986) "In a broad informal sense, robust statistics is a body of knowledge, partly formalized into 'theories of robustness', relating to deviations from idealized assumptions in statistics". In particular this body of knowledge contains some clever estimation methods which are much less affected by outliers than their classical counterparts. They were developed when studying heavy tailed deviations from distributional assumptions. These robust estimators are thus good candidates for dealing with the excess claims in credibility. In several respects the situation in credibility theory is, however, different from the one usually considered in robustness: For instance in credibility nonparametric methods prevail whereas robustness studies a neighborhood of a specific parametric model; credibility is interested in all the data, not only in the majority, because premiums have to cover all claims; credibility estimators are estimating random variables, not parameters. Maybe for these reasons the two fields have been largely separated until now. Because of these differences we were not able to derive our estimators from a general principle or an optimality criterion. We just propose some simple estimators which are based on heuristic considerations and seem to work reasonably well. In our first proposal we calculate a robust estimator from the claims of each contract and use then linear credibility based on these estimators. In the second proposal we incorporate an a priori premium into a robust estimator based on the claim sizes in a nonlinear way. Here we are inspired by some Bayesian estimator. The resulting estimator is, however, free from any distributional assumptions. With these proposals we hope to convince actuarians that robustness can make a contribution to the problem of excess claims and that further research is worthwhile.

2 Models and Estimators

We consider the basic credibility model with J contracts and n years of experience. It contains unobservable risk parameters θ_j and claim sizes X_{ij} $(1 \leq i \leq n,\ 1 \leq j \leq J)$. We make the following distributional assumptions:

$$(\theta_j, X_{1j}, \ldots, X_{nj}) \text{ are i.i.d. } (1 \leq j \leq J), \tag{2.1}$$

$$\theta_j \text{ is distributed according to } U(d\theta), \tag{2.2}$$

$$\text{Given } \theta_j, X_{1j}, \ldots, X_{nj} \text{ are i.i.d. with distribution } F_{\theta_j}(dx). \tag{2.3}$$

It will be convenient to distinguish between the following two situations:

`Case I`: U and F_θ are known. In this case a single contract is sufficient, and we drop the index j.

`Case II`: U and F_θ are unknown.

Although Case I is not realistic, it is useful for explaining our ideas.

Our goal is to estimate $\mu_j = \mu(\theta_j) = E_{\theta_j}[X_{ij}]$, the pure risk premium. In Case I we propose

$$\widehat{\mu(\theta)} = \mu + \alpha(T(X_1, \ldots, X_n) - E[T]) \tag{2.4}$$

where $\mu = E[\mu(\theta)] = \int \mu(\theta)U(d\theta)$ is the overall mean and α is the credibility factor. With an optimal α, this is the linear credibility estimator based on T instead of

$X_1, \ldots, X_n$. As our pure experience based estimator T we take a robust estimator defined implicitly as the solution of

$$\sum_{i=1}^{n} \chi(X_i/T) = 0 \tag{2.5}$$

with $\chi(z) = \max\ (-c_1,\ \min\ (z-1, c_2))$. In Case II we replace means by averages:

$$\widehat{\mu_j} = \bar{X}_{..} + \alpha(T_j - \bar{T}_.) \tag{2.6}$$

where $\bar{X}_{..} = (nJ)^{-1} \sum_j \sum_i X_{ij}, \bar{T}_. = J^{-1} \sum_j T_j$ and T_j is defined as the solution of

$$\sum_{i=1}^{n} \chi(X_{ij}/T_j) = 0 \tag{2.7}$$

with χ as above.

We list here some simple properties of our proposal which serve as a first justification:

a) It is scale equivariant: If all $X_i(X_{ij})$ are multiplied by a constant c, then $\widehat{\mu(\theta)}(\widehat{\mu_j})$ is multiplied by the same constant.

b) It includes the linear credibility estimator: If $c_1 = 1, c_2 = \infty$, then $T = \bar{X}_.$ and $T_j = \bar{X}_{.j}$.

c) It is unbiased: $E[\hat{\mu}(\theta)] = \mu$ and $J^{-1} \sum \hat{\mu}_j = \bar{X}_{..}$.

d) It is related to the truncation estimator by Gisler [1980a,b]: (2.5) can be written as

$$T = n^{-1} \sum_{i=1}^{n} \max\ ((1-c_1)T,\ \min\ (X_i, (1+c_2)T)) \tag{2.8}$$

i.e. claims on both ends are truncated. The main difference to Gisler [180a,b] is that the truncation point depends on the contracts experience and is given implicitly.

The estimator (2.5) belongs to the class of so-called M-estimators of scale. These are standard estimators in robust statistics, see Hampel et al. (1986, Chap. 2). Their most important property is that the change due to an additional observation at x is approximately proportional to $\chi(x/T)$. Hence our form of χ means that we bound the influence of each observation from above and below. Further properties will be discussed in Section 3. From this discussion the role of the constants c_1, c_2 will become clearer. It will be seen that their choice is not very crucial and can be done beforehand without having much information about the claims. The credibility factor α on the other hand must depend on the distributions U and F_θ in Case I and on all the data in Case II. How this can be done is the content of Section 4.

3 Discussion of the Robust Estimator

We consider only a single contract.

3.1 Existence, uniqueness and calculation of the estimator.

Denote by $x_{(1)} \leq x_{(2)} \leq \ldots \quad \leq x_{(n)}$ the ordered sample and by k_o the number of zero values in the sample. Furthermore we introduce the set of solutions for the equation defining T :

$$L = L(x_1, \ldots, x_n) = \{t > 0, \sum_{i=1}^{n} \chi(x_i/t) = 0\}.$$

Existence and uniqueness of T is covered in the following:

Lemma 3.1 *If $k_o < nc_2/(c_1 + c_2)$, then L is a closed, finite, non-empty interval. If $k_o = nc_2/(c_1 + c_2)$ then $L = (0, x_{(k_o+1)}/(1 + c_2)]$. If $k_o > nc_2/(c_1 + c_2)$, then L is empty.*

`Proof:` For any $x \geq 0, t \longrightarrow \chi(x/t)$ is continuous and monotone decreasing. For $t > x_{(n)}, \sum \chi(x_i/t) < 0$. For $t \leq x_{(k_0+1)}/(1 + c_2)$ we have

$$\sum \chi(x_i/t) = -k_0 c_1 + (n - k_0) c_2.$$

From this the lemma follows easily. q.e.d.

The case where L contains more than one point does occur, e.g. $n = 4, c_1 = c_2 = 0.5, x_1 = x_2 = 0.4, x_3 = x_4 = 1.8$ gives $L = [0.8, 1.2]$. If a unique definition of T is needed, we will take the midpoint of L. In case where L is empty, we take $T = 0$.

The defining equation for T can also be written in the following equivalent form with $x_{(0)} \equiv 0, x_{(n+1)} \equiv \infty$, compare (2.8):

$$T = \sum_{i=\ell_1+1}^{n-\ell_2} x_{(i)}/(n - \ell_2(1 + c_2) - \ell_1(1 - c_1)), \tag{3.1}$$

$$x_{(\ell_1)} \leq (1 - c_1)T \leq x_{(\ell_1+1)}, \tag{3.2}$$

$$x_{(n-\ell_2)} \leq (1 + c_2)T \leq x_{(n-\ell_2+1)}, \tag{3.3}$$

For given ℓ_1 and ℓ_2, T can be computed from (3.1a) and then (3.1b) and (3.1c) can be checked. Because ℓ_1, ℓ_2 can take only a finite number of values, T can be found by trial and error, at least for n not too big. In a more systematic iterative procedure one determines new values ℓ_1 and ℓ_2 such that (3.1b) and (3.1c) are satisfied and then computes a new T from (3.1a) etc. In our experience this worked very well, but we didn't try to prove the convergence of the algorithm.

3.2 The breakdown point

In Lemma 3.1 we have already seen that with less than $nc_2/(c_1+c_2)$ zero claims keep the estimator away from zero. Here we consider the opposite case: What is the maximal number of claims tending to infinity for which the estimator remains bounded? In robustness this is called the finite sample breakdown point, c.f. Hampel et al. (1986, Sec. 2.2a). The breakdown point plus one is then the minimal number of outlying claims needed to take the estimator to infinity.

Lemma 3.2 *Let $k = min\,\{i \in \mathbb{N}; \mathrm{i} \geq \mathrm{n}\ \mathrm{c}_1/(\mathrm{c}_1+\mathrm{c}_2)\}$. Then T remains bounded if less than k observations tend to infinity, but it tends to infinity if k or more observations tend to infinity.*

Proof: By definition of k

$$k(c_1+c_2) \geq n\ c_1 \tag{3.4}$$

and

$$(k-1)\ (c_1+c_2) < n\ c_1. \tag{3.5}$$

First we assume that $x_{(n-k+1)}$ is fixed and derive an upper bound for T : If $c_1 < 1$ we put $t = x_{(n-k+1)}/(1-c_1)$. Then by (3.3)

$$\sum_{i=1}^{n} \chi(x_i/t) \leq -(n-k+1)c_1 + (k-1)c_2 < 0.$$

Hence by monotonicity $T < t$.

If $c_1 = 1$, we put

$$t = \sum_{i=1}^{n-k+1} x_{(i)}/\left(n-(k-1)\ (1+c_2)\right) \geq x_{(n-k+1)}/(n-(k-1)(1+c_2)).$$

By (3.2) we obtain $(1+c_2)t \geq x_{(n-k+1)}$, hence

$$\sum_{i=1}^{n} \chi(x_i/t) \leq \sum_{i=1}^{n-k+1} x_{(i)}/t - (n-k+1) + (k-1)c_2 = 0.$$

Moreover for $t' > t$ $\chi(x_{(n-k+1)}/t') < \chi(x_{(n-k+1)}/t)$. Therefore $\sum_i \chi(x_i/t') < 0$. This implies $T \leq t$ and thus completes the proof of the first part.

For the second part we put $t = x_{(n-k+1)}k/\left(nc_1 + k(1-c_1)\right)$. By (3.2) $(1+c_2)t \geq x_{(n-k+1)}$. Hence

$$\begin{aligned}\sum_{i=1}^{n} \chi(x_i/t) &\geq -(n-k)c_1 + k\chi(x_{(n-k+1)}/t)\\ &= -(n-k)c_1 + (nc_1 + k(1-c_1) - k) = 0.\end{aligned}$$

Hence the right endpoint of $L = \{t'; \sum_i \chi(x_i/t') = 0\}$ is $\geq t$, and $t \longrightarrow \infty$ if $x_{(n-k+1)} \longrightarrow \infty$. q.e.d.

Lemmas 3.1 and 3.2 show that the tolerance towards zero values and outliers are in conflict. This can be shown to be true in general for scale equivariant continuous estimators. A reasonable compromise might be to take $c_1 = c_2$, but if a priori knowledge about the claims is available other choices are possible.

3.3 Linearization of the estimator.

For the credibility factor α in (2.4) and (2.5) we need the variance of T and T_j. Because of the implicit definition, this seems hopeless. There is, however, a simple asymptotic approximation. With the help of the so-called influence function we can linearize T (see Hampel et al., 1986, Chap. 2):

$$T(X_1,\ldots,X_n) = T(\theta) + n^{-1}\sum_{i=1}^{n} IF(X_i,\theta) + o_p(n^{-1/2}). \tag{3.6}$$

Here $T(\theta)$ is defined implicitly by

$$\int \chi(x/T(\theta))F_\theta(dx) = 0 \tag{3.7}$$

and the influence function IF is given by

$$IF(x,\theta) = \chi(x/T(\theta))\; T(\theta)^2 M(\theta)^{-1} \tag{3.8}$$

where

$$M(\theta) = \int \chi'(x/T(\theta))xF_\theta(dx) = \int_{(1-c_1)T(\theta)}^{(1+c_2)T(\theta)} xF_\theta(dx). \tag{3.9}$$

In particular we have $E_\theta[IF(X_i,\theta)] = 0$, hence $E_\theta[T(X_1,\ldots,X_n)] \approx T(\theta)$, and

$$\mathrm{Var}_\theta[T(X_1,\ldots,X_n)] \cong n^{-1}E_\theta[IF(X_i,\theta)^2] = n^{-1}E_\theta\left[\chi^2(X_i/T(\theta))\right]\cdot T(\theta)^4 M(\theta)^{-2}. \tag{3.10}$$

One can define optimal constants c_1, c_2 as those values which minimize the asymptotic variance of the bias corrected estimator $T\mu(\theta)/T(\theta)$, i.e. $E\left[IF^2(X_i/T(\theta))\right]\mu(\theta)^2$ $T(\theta)^{-2}$. This depends on F_θ, but fortunately a choice of $c_1 = 1$ c_2 between 1 and 2 is often not much worse than the optimum but often much better than $c_1 = 1, c_2 = \infty$ (which gives $T = \bar{X}$), cf. the example in Section 5.1 and the results for the closely related robust location estimator. In view of this and the tolerance to zero's and outliers investigated in the previous sections, we recommend as a standard choice $c_1 = c_2 = 1$ for small samples and $c_1 = 1, c_2 = 1.5$ or 2 for moderate samples.

One might also object that for samples sizes $n \leq 10$ typical in insurance, the approximation (3.4) might be rather crude. However, we use (3.4) and its consequences only to determine α. We conjecture that a suboptimal choice of α does not have a great effect.

4 The Credibility Factor

In Case I we obtain by a straightforward calculation for the estimator (2.4)

$$E[(\widehat{\mu(\theta)} - \mu(\theta))^2] = \mathrm{Var}[\mu(\theta)] + \alpha^2\{\mathrm{E}[\mathrm{Var}_\theta(\mathrm{T})] + \mathrm{Var}[\mathrm{E}_\theta(\mathrm{T})]\} - 2\alpha\mathrm{Cov}[\mathrm{E}_\theta(\mathrm{T}), \mu(\theta)]. \tag{4.1}$$

Hence the α minimizing mean square error is given by

$$\alpha_0 = \frac{\mathrm{Cov}[\mathrm{E}_\theta(\mathrm{T}), \mu(\theta)]}{\mathrm{E}[\mathrm{Var}_\theta(\mathrm{T})] + \mathrm{Var}[\mathrm{E}_\theta(\mathrm{T})]}. \tag{4.2}$$

An unbiased estimator of the denominator in Case II is

$$(J-1)^{-1}\sum_{j=1}^{J}(T_j-\bar{T}.)^2 \tag{4.3}$$

Because

$$E[\sum_j (T_j-\bar{T}.)(\bar{X}_{.j}-\bar{X}_{..})] = (J-1)\{\mathrm{Cov}[\mathrm{E}_\theta(\mathrm{T}),\mu(\theta)] + \mathrm{E}[\mathrm{Cov}_{\theta_j}(\mathrm{T_j},\mathrm{X_{ij}})]\},$$

we need an estimator of $E[\mathrm{Cov}_{\theta_j}(\mathrm{T_j},\mathrm{X_{ij}})]$. The linearization (3.4) suggests to use

$$n^{-1}J^{-1}(n-1)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n} X_{ij}\widehat{IF}(X_{ij},T_j) \tag{4.4}$$

where

$$\widehat{IF}(x,T_j) = \chi(x/T_j)T_j^2 n/\sum_{i=1}^{n} X_{ij}1_{[(1-c_1)T_j\le X_{ij}\le(1+c_2)T_j]} \tag{4.5}$$

We thus take the estimator (2.6) with $\alpha = \widehat{\alpha}_0$ where

$$\widehat{\alpha}_0 = \frac{(J-1)^{-1}\sum_j(T_j-\bar{T}.)(\bar{X}_{.j}-\bar{X}_{..}) - n^{-1}J^{-1}(n-1)^{-1}\sum_j\sum_i X_{ij}\widehat{IF}(X_{ij},T_j)}{(J-1)^{-1}\sum_j(T_j-\bar{T}.)^2} \tag{4.6}$$

In the case $T_j = \bar{X}_{.j}$ (i.e. $c_1 = 1,\ c_2 = \infty$), one usually modifies the unbiased estimators for the numerator and the denominator of (4.2) so that all estimated variances are ≥ 0 and all estimated correlations are between -1 and $+1$. If we want to do this here too, we have to estimate $E[\mathrm{Var}_\theta(\mathrm{T})]$ e.g. by

$$(n-1)^{-1}\sum_{i=1}^{n}\widehat{IF}(X_{ij},T_j)^2.$$

Since this creates additional complextiy and we have no evidence that one really gains by doing so, we use the version with $\widehat{\alpha}_0$ given by (4.6).

5 Examples

The examples in this section are chosen for computational simplicity and for their ability to illustrate the advantages and disadvantages of our robust credibility estimator. We restrict ourselves to the Case I.

5.1 Scale families.

Here we assume that

$$F_\theta(x) = F(x/\theta) \tag{5.1}$$

for a fixed cumulative distribution function F. This implies

$$\begin{aligned} \mu_\theta &= m\theta \text{ with } m = \int xF(dx), \\ \mathrm{Var}_\theta[\mathrm{X}] &= v^2\theta^2 \text{ with } v^2 = \int (x-m)^2 F(dx), \\ E_\theta[T] &= d_n \cdot \theta \text{ for some } d_n \in \mathbb{R}, \\ \mathrm{Var}_\theta[\mathrm{T}] &= \sigma_n^2 \cdot \theta^2 \text{ for some } \sigma_n^2 \in \mathbb{R}. \end{aligned}$$

Moreover the results from Section 3.3 show that as $n \longrightarrow \infty \quad d_n \sim d,\ \sigma_n^2 \sim n^{-1}\sigma^2$ where

$$\int \chi(x/d)F(dx) = 0$$

$$\sigma^2 = \int \chi(x/d)^2 F(dx) d^4 \left(\int x\chi'(x/d)F(dx)\right)^{-2}$$

Inserting these results into (4.2) gives the optitmal credibility factor

$$\alpha_0 = \frac{md_n \ \mathrm{Var}(\theta)}{\mathrm{d_n^2 Var}(\theta) + \sigma_\mathrm{n}^2 \mathrm{E}[\theta^2]}$$

As a side remark we note that as $n \longrightarrow \infty$, $\alpha_0 \longrightarrow \frac{m}{d}$. Because $T \longrightarrow d\theta$ for $n \longrightarrow \infty$, we see that $\widehat{\mu(\theta)} \longrightarrow m\theta$, i.e. $\widehat{\mu(\theta)}$ is consistent. The bias due to the truncation is compensated by the credibility factor. With the credibility factor α_0 we can also calculate the mean square error (MSE) $E[(\widehat{\mu(\theta)} - \mu(\theta))^2]$. We obtain

$$MSE\ (robust) = \frac{m^2\sigma_n^2 E[\theta^2]\mathrm{Var}(\theta)}{\mathrm{d_n^2 Var}(\theta) + \sigma_\mathrm{n}^2 \mathrm{E}[\theta^2]}.$$

On the other hand the mean square error of the linear credibility estimator is known to be

$$MSE\ (linear) = \frac{m^2 v^2 E[\theta^2]\mathrm{Var}(\theta)}{\mathrm{n\ m^2 Var}(\theta) + \mathrm{v^2 E}[\theta^2]}.$$

Defining the relative efficiency [RE] as the ratio of the mean square errors we thus have shown that

$$RE(robust : linear) = \frac{v^2}{\sigma_n^2} \cdot \frac{d_n^2 \mathrm{Var}(\theta) + \sigma_\mathrm{n}^2 \mathrm{E}[\theta^2]}{\mathrm{nm^2 Var}(\theta) + \mathrm{v^2 E}[\theta^2]}.$$

Letting n tend to infinity we obtain the asymptotic relative efficiency (ARE)

$$ARE\ (robust : linear) = \frac{v^2 d^2}{\sigma^2 m^2}.$$

This is nothing else than the usual asymptotic relative efficiency of $m\ d^{-1}T$ versus the arithmetic mean. We illustrate this result numerically in the following situation

$$F(x) = (1 - e^{-x})(1-\varepsilon) + 1_{[x \geq a]} \cdot \varepsilon \tag{5.2}$$

This means that given θ with probability $1-\varepsilon$ a claim is exponential with mean θ (an ordinary claim) and with probability ε it is equal to $a \cdot \theta$ (an outlier claim). Asymptotic relative efficiencies are given in Table 1 for selected values of ε and a. We see that the loss of efficiency in the case $\varepsilon = 0$ is more than compensated by cases where large outlier claims are possible. Note that the truncation method of Gisler [1980,a,b] cannot handle this situation if $\mathrm{Var}(\theta)$ is not close to zero because the size of the outlying claims is also proportional to θ.

5.2 An example with two radically different claim size distributions

In this example the risk parameter and the claim size take only two values:

$$\begin{aligned} P[\theta = 1] &= P[\theta = 2] &&= 0.5, \\ P[X = 1|\theta = 1] &= 0.9,\ P[X &&= 10|\theta = 1] = 0.1, \\ P[X = 1|\theta = 2] &= 0,\ P[X &&= 10|\theta = 2] = 1. \end{aligned}$$

This means that one group of contracts produces only large claims whereas the first group produces usually small claims with occasional outliers. In this case calculations can be made in closed form without any approximations. The results for $n = 10$, and $c_1 = c_2 = 1$ are given in Table 2. We see that the robust credibility estimator is quite close to the posterior mean which is optimal for square loss – at least in those cases which do occur in practice. The linear credibility estimator is obviously bad. It can be shown easily that the truncation estimator of Gisler [1980a,b] coincides with the linear credibility estimator. It is also instructive to see who pays for the outliers which occur in the first group. It is the lucky person in the same group who has not yet incurred a large claim.

5.3 An example by Gisler [1980a]:

This is again a case where both θ and X are discrete: $\theta \in \{1, 2, 3, 4\}, X \in \{0, 2, 4, 6, 40\}$. We have $P[\theta = i] = 0.25$ for all i. The conditional probabilities for X given θ are given in Table 3. We take $n = 3$ so that exact calculations can be made without too much work. We compare here four estimators. The first one is the classical linear credibility estimator

$$\hat{\mu}_j^{lin} = 3.912 + 0.158(\bar{X}_{.j} - 3.912)$$

The second one is our robust credibility estimator with $c_1 = c_2 = 1$ and $\alpha = \alpha_0$ given by (4.2):

$$\hat{\mu}_j^{rob} = 3.912 + 0.351(T_j - 3.089).$$

The third one is the semilinear credibility estimator with optimal truncation point of Gisler [1980a,b].

$$\hat{\mu}_j^{Gisler} = 3.912 + 0.794(\frac{1}{3}\sum_{i=1}^{3} G(X_{ij}) - 2.767)$$

$$\text{where} \quad G(x) = \min(x, 4.89).$$

Finally we consider the optimal estimator

$$\hat{\mu}_j^{Bayes} = E[X_{4j}|X_{1j}, X_{2j}, X_{3j}].$$

The mean square errors for these estimators are $1.90, 1.47, 1.12$ and 1.09. The values of estimators for some realizations of (X_{1j}, X_{2j}, X_{3j}) are given in Table 4. We see that typically the robust estimator is between the linear and Gisler's estimator. The most striking exception occurs for contracts with two or three claims of 40. They are heavily charged by our robust estimator. This difference will become irrelevant for somewhat larger n's because then the probability for a contract to produce a majority of outlier claims is practically zero. Moreover, if we change the model slightly and introduce an additional risk class with $P_{\theta_j}[X_{ij} = 40] >> 0$, then the above bad performance of our estimator turns into an advantage: Whenever there is a majority of outlier claims, we should charge the corresponding contract heavily.

5.4 Discussion

Our estimator performs well if there are outlying claims and the number of years available is not very small. In other situations it seems to be at least acceptable in its performance. It can deal also well with situations where the outlying claims vary considerably with the risk parameter. The reason for this is that our estimator determines a truncation point separately for each contract, based only on the experience of the contract under consideration. If the number of years is small, one might want to use also the experience from other contracts to some extent. How this can be done in a robust way is the topic of the next section.

6 A More Sophisticated Approach

We consider first Case I with known distributions and fix a measure τ for the average claim size. It could be $E[X]$, but it is preferable to choose τ robust so that it is not affected by single outliers and atypical contracts. A concrete proposal is given below. We then suggest the following estimator

$$\widehat{\mu(\theta)} = \mu + (T(X_1, \dots, X_n; \tau) - E[T(X_1, \dots, X_n; \tau)]) \tag{6.1}$$

where $\mu = E[X_i]$ and T is defined implicitly as the solution of

$$\sum_{i=1}^{n} \chi(X_i/T) = \gamma(1 - \tau/T). \tag{6.2}$$

Here $\chi(x) = \max(-c_1, \min(x-1, c_2))$ as before. In order to explain the character of this estimator, consider first the case $c_1 = 1,\ c_2 = \infty$. Then by a simple calculation $T = \gamma(n+\gamma)^{-1}\tau + n(n+\gamma)^{-1}\bar{X}_.$ and $\widehat{\mu(\theta)} = \gamma(n+\gamma)^{-1}\mu + n(n+\gamma)^{-1}\bar{X}_.$. We thus recover the usual linear credibility estimator with credibility factor $n(n+\gamma)^{-1}$. In the general case we introduce the truncated claims

$$X_i^* = \max\left((1-c_1)T,\ \min(X_i, (1+c_2)T)\right).$$

One then finds by the same argument that

$$T = n(n+\gamma)^{-1} n^{-1} \sum_{i=1}^{n} X_i^* + \gamma(n+\gamma)^{-1}\tau, \tag{6.3}$$

i.e. T is a convex combination of the a priori value τ and the mean of the truncated claims. The truncation point depends however on T (and thus on τ and $X_1, \dots, X_n$)

so that (6.3) is not an explicit solution. The estimator T incorporates already the a priori value τ with the weight $\gamma(n+\gamma)^{-1}$. The passage from T to $\widehat{\mu(\theta)}$ serves only to achieve unbiasedness; there is no need to introduce an additional credibility factor there.

The main advantage of this proposal is that the a priori value τ is used to find the truncation points $(1-c_1)T$ and $(1+c_2)T$. This improves the ability of the estimator to detect and truncate outlying claims. It is most visible when we study the breakdown points of T. With similar arguments as in Section 3.1-2 we can prove the following result.

Lemma 6.1 *i) (6.2) has always a unique solution in the interval* $[\gamma(nc_1+\gamma)^{-1}\tau, \infty)$. *ii) The breakdown point, i.e. the maximal number of claims tending to infinity the estimator can tolerate without going to infinity, is given by* $\min\{i \in \mathbb{N}, \mathrm{i} \geq (\mathrm{nc}_1+\gamma)/(\mathrm{c}_1+\mathrm{c}_2)\} - 1$.

Again a choice $c_1 = 1,\ c_2$ between 1 and 2 is expected to work well in most cases.

A different justification of our estimator T can be obtained from the Bayesian viewpoint. It is easily checked that (6.2) is the normal equation for the estimator maximizing the a posteriori density if we choose $F_\theta(dx) = \theta^{-1} f(x\theta^{-1})dx$ with

$$f(x) = \text{const.} \begin{cases} e^{-x}, & x \leq 1+c_2 \\ x^{-c_2}(1+c_2)^{c_2} e^{-1-c_2}, & x \geq 1+c_2, \end{cases}$$

and

$$U(d\theta) = \text{const.}\tau^{\gamma-1}\theta^{-\gamma}\exp(-\gamma^{\tau}/\theta)d\theta$$

provided $c_1 = 0,\ c_2 > 1,\ \gamma > 1$. We thus see that our proposal corresponds to heavy tails in both the distribution of claims and the distribution of the risk parameter. The assumption of scale families for both F and U leads to an estimator which is scale equivariant: If $X_1, \ldots, X_n$ and τ are multiplied by a constant, then T is multiplied by the same constant.

Next we discuss the choice of the a priori value τ. We propose to use the solution of the following equation

$$\int \chi(\tau(\theta)/\tau)U(d\theta) = 0$$

where $\tau(\theta)$ is determined by

$$\int \chi(x/\tau(\theta))F_\theta(dx) = 0$$

and χ is the by now well known truncated linear function. From what has been said before, it is clear that τ is a robust measure of the average claim size. The example of the linear credibility estimator at the beginning showed that the choice of τ is irrelevant if $c_1 = 1$ and $c_2 = \infty$. Presumably in other cases too the value of τ will not be crucial.

In Case II which is relevant for applications our proposal is as follows

$$\widehat{\mu_j} \quad = \quad \bar{X}_{..} + (T_j - \bar{T}_.) \qquad (j = 1, \ldots, J) \tag{6.4}$$

$$\sum_{i=1}^{n} \chi(X_{ij}/T_j) = \gamma(1-\tau/T_j)\ (j=1,\ldots,J) \tag{6.5}$$

$$\sum_{j=1}^{J} \chi(\tau_j/\tau) = 0 \tag{6.6}$$

$$\sum_{i=1}^{n} \chi(X_{ij}/\tau j) = 0 \qquad (j=1,\ldots,J). \tag{6.7}$$

This is a straightforward modification of the previous definitions replacing expectations by averages. Note that τ_j is nothing else than our proposal from Section 2.

In order to complete our discussion we have to determine γ by an objective procedure from the data. We do this by minimizing an estimate of mean square error. In Case I it follows from (6.1) that

$$E[(\widehat{\mu(\theta)} - \mu(\theta))^2] = E\left[(T - E_\theta[T])^2\right] + \tag{6.8}$$

$$\mathrm{Var}(\mathrm{E}_\theta[\mathrm{T}]) + \mathrm{Var}(\mu(\theta)) - 2\mathrm{Cov}(\mathrm{E}_\theta[\mathrm{T}], \mu(\theta)).$$

Because T is nonlinear, $E_\theta[T]$ and $\mathrm{Var}_\theta[\mathrm{T}]$ are difficult to evaluate. An approximation can be obtained like in Section 3.3 by linearizing. If γ is fixed and n tends to infinity, the influence of τ disappears as is easily seen in the example of linear credibility. Hence in order to obtain a nontrivial asymptotics, we take $\gamma = \gamma(n) = n\gamma_\infty$. Then we obtain by a Taylor expansion (cf. Hampel et al., 1986, Chap. 2):

$$T(X_1,\ldots,X_n;\tau) = T(\theta,\tau) + n^{-1}\sum_{i=1}^{n} IF(X_i,\theta,\tau) + o_p(n^{-1/2}) \tag{6.9}$$

where $T(\theta,\tau)$ is defined implicitly by

$$\int \chi(x/T(\theta,\tau))F_\theta(dx) = \gamma_\infty(1-\tau/T(\theta,\tau)) \tag{6.10}$$

and the influence function IF is now

$$IF(x,\theta,\tau) = \{\chi(x/T(\theta,\tau)) - \gamma_\infty(1-\tau/T(\theta,\tau))\}T(\theta,\tau)^2 M(\theta,\tau)^{-1} \tag{6.11}$$

$$\text{with } M(\theta,\tau) = \int \chi'(x/T(\theta,\tau))xF_\theta(dx) + \gamma_\infty\tau. \tag{6.12}$$

Note that $E_\theta[IF(X_i,\theta,\tau)] = 0$. Hence $E_\theta[T] \approx T(\theta,\tau)$ and $\mathrm{Var}_\theta[\mathrm{T}] \approx \mathrm{n}^{-1}\int \mathrm{IF}(\mathrm{x},\theta,\tau)^2\mathrm{F}_\theta(\mathrm{dx})$. This can be plugged into the formula for the mean square error which then can be minimized with respect to γ. Details are left to the reader.

In Case II we need estimates for the different terms on the right hand side of (6.8): An unbiased estimator for $E\,[(T-E_\theta[T])^2]$ is $(J-1)^{-1}\sum_{j=1}^{J}(T_j - \bar{T}_.)^2$. Because $E\,[(T-E_\theta[T])^2] = E[\mathrm{Var}_\theta[\mathrm{T}]] + \mathrm{Var}(\mathrm{E}_\theta[\mathrm{T}])$, we must estimate also $E[\mathrm{Var}_\theta[\mathrm{T}]]$. The linearization (6.9)-(6.12) suggests to do this by

$$n^{-1}J^{-1}(n-1)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\widehat{IF}(X_{ij},T_j,\tau)^2$$

where

$$\widehat{IF}(x,T_j,\tau)=\{\chi(x/T_j)-\gamma n^{-1}(1-\tau/T_j)\}T_j^2\widehat{M}(T_j,\tau)^{-1} \tag{6.13}$$

and

$$\widehat{M}(T_j,\tau)=n^{-1}\sum_{i=1}^{n}X_{ij}1_{[(1-c_1)T_j\le X_{ij}\le(1+c_2)T_j]}+\gamma n^{-1}\tau. \tag{6.14}$$

Similarly we estimate $E[\mathrm{Cov}_\theta(\mathrm{T},\mathrm{X_i})]$ by

$$n^{-1}J^{-1}(n-1)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\widehat{IF}(X_{ij},T_j,\tau)(X_{ij}-\bar{X}_{.j}).$$

Taking all this together we minimize the following expression with respect to γ

$$2(J-1)^{-1}\sum_{j=1}^{J}\{(T_j-\bar{T}.)^2-(T_j-\bar{T}.)(\bar{X}_{.j}-\bar{X}_{..})\} \tag{6.15}$$

$$-n^{-1}J^{-1}(n-1)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}\{\widehat{IF}(X_{ij},T_j,\tau)^2-2\widehat{IF}(X_{ij},T_j,\tau)(X_{ij}-\bar{X}.)\}.$$

A lengthy but straightforward calculation shows that in the case $c_1=1,\ c_2=\infty$ (6.15) becomes

$$(J-1)^{-1}\sum_{j=1}^{J}(\bar{X}_{.j}-\bar{X}_{..})^2(2\alpha^2-2\alpha)$$

$$-n^{-1}J^{-1}(n-1)^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n}(X_{ij}-\bar{X}_{.j})^2(\alpha^2-\alpha)$$

where $\alpha=n(n+\gamma)^{-1}$. So in this case the minimization can be done in closed form and gives the classical result. In general we will have to find the optimal γ numerically by evaluation (6.15) for some values of γ. Since the exact value of γ will not matter too much, a coarse search ought to be sufficient. This completes the presentation of our second proposal.

REFERENCES:

Gisler, A. (1980a): Optimales Stutzen von Beobachtungen im Credibility Modell. Ph.D.thesis, ETH Zürich, Nr. 6556.

Gisler, A. (1980b): Optimum trimming of data in the credibility model. Mitteilungen Ver. Schweiz. Vers. math. 80, 313-326.

Gisler, A., Reinhard, P. (1990): Robust Credibility. Paper presented at the XXIIth Astin Colloquium, Montreux.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986): *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

Table 1: Asymptotic relative efficiency of the robust versus the linear credibility estimator in the mixture model (5.1)-(5.2).

	$\varepsilon = 0$	$\varepsilon = 0.1$	
		a=10	a=100
$c_1 = c_2 = 1$	0.72	1.40	4.60
$c_1 = 1, c_2 = 2$	0.90	1.01	3.31

Table 2: Credibility estimators in the example 5.2 for $n = 10$.

Number of claims	Probability given		Linear	Robust	Posterior
=10	θ=1	θ=2		$c_1 = c_2 = 1$	mean
0	0.3487	0	1.12	1.74	1.9
1	0.3874	0	2.00	1.85	1.9
2	0.1937	0	2.88	2.04	1.9
3	0.0574	0	3.76	2.43	1.9
4	0.0112	0	4.64	3.57	1.9
5	0.0015	0	5.52	5.86	1.9
10	0	1	9.92	9.99	10.0

Table 3: Conditional probabilities in the Example 5.3

θ	$P_\theta[X_i = x]$				
	$x = 0$	$x = 2$	$x = 4$	$x = 6$	$x = 40$
1	0.5445	0.2475	0.0990	0.0990	0.0100
2	0.2940	0.2940	0.2450	0.1470	0.0200
3	0.0970	0.2910	0.3395	0.2425	0.0300
4	0.0480	0.1440	0.2880	0.4800	0.0400

Table 4: Credibility estimators in the Example 5.3 with $n = 3$ and selected claims.

(X_1, X_2, X_3)	Linear	Robust $c_1 = c_2 = 1$	Gisler	Posterior mean
(0,0,0)	3.29	2.83	1.72	2.09
(0,0,6)	3.61	2.83	3.01	2.50
(0,2,2)	3.50	3.30	2.78	2.78
(0,2,6)	3.72	3.53	3.54	3.25
(0,6,6)	3.93	4.23	4.31	4.30
(2,4,6)	3.93	4.23	4.60	4.69
(6,6,6)	4.24	4.93	5.60	5.68
(0,0,40)	5.4l	2.83	3.01	2.58
(0,6,40)	5.72	4.93	4.31	4.13
(2,4,40)	5.72	4.93	4.60	4.53
(6,6,40)	6.04	7.03	5.60	5.46
(6,40,40)	7.84	12.86	5.60	5.58