On parameter estimators in credibility

A. Dubey, Zurich, and A. Gisler, Winterthur

On Parameter Estimators in Credibility

1. Introduction

Bühlmann and Straub (1970) have introduced the credibility model which is nowadays probably most used in practice. Formally this model can be described as follows:

$\theta_j (j = 1, 2, \ldots, N)$ are (possibly vector valued) random variables [RV] and $\underline{X}_j = (X_{1j}, \ldots, X_{nj})$ are vectors of observable RV fulfilling the following conditions:

1) The vectors $(\theta_j, X_{1j}, \ldots, X_{nj}), j = 1, \ldots, N$, are independent,

2) $\theta_1, \theta_2, \ldots, \theta_N$ are independent and identically distributed and

3) given θ_j, the RV $X_{1j}, \ldots, X_{nj}$ are independent and

$$E[X_{ij} \mid \theta_j] = \mu(\theta_j)$$

$$\mathrm{Var}[X_{ij} \mid \theta_j] = \sigma^2(\theta_j)/P_{ij},$$

where P_{ij} are known positive numbers.

In actuarial science the X_{ij} may be interpreted as loss ratios of N contracts in the years $i = 1, \ldots, n$. Each contract, $j = 1, \ldots, N$, is characterised by an unknown risk parameter θ_j. The P_{ij} are known volumes of the contract j in the different years i (the number of risk years, the total amount of wages, the turnover, etc. according to the different lines of insurance). The number of years n is assumed to be the same for all contracts. It is however possible to generalize without difficulty all the formulae for the case where n differs from contract to contract.

Let P_j and X_j (P and X) denote the total volume and the total loss ratio of a contract (of the whole portfolio):

$$P_j = \sum_i P_{ij} \qquad P = \sum_j P_j$$

$$X_j = \sum_i \frac{P_{ij}}{P_j} X_{ij} \qquad X = \sum_j \frac{P_j}{P} X_j$$

In this model we have:

$$E[X_j \mid \theta_j] = \mu(\theta_j) \quad , \operatorname{Var}[X_j \mid \theta_j] = \sigma^2(\theta_j)/P_j .$$

Let

$$\mu = E[\mu(\theta)] \quad , \quad v = E[\sigma^2(\theta)] \quad , \quad w = \operatorname{Var}[\mu(\theta)].$$

Then:

$$\operatorname{Var}[X_{ij}] = E[\operatorname{Var}[X_{ij} \mid \theta_j]] + \operatorname{Var}[E[X_{ij} \mid \theta_j]]$$

$$= \frac{v}{P_{ij}} + w$$

and

$$\operatorname{Var}[X_j] = \frac{v}{P_j} + w .$$

The credibility estimator of $\mu(\theta_j)$, i.e. the best (with respect to quadratic loss) estimator linear in the observations, is:

$$\hat{\mu}(\theta_j) = \alpha_j X_j + (1 - \alpha_j) \mu \tag{1}$$

where

$$\alpha_j = \frac{P_j w}{P_j w + v} = \frac{\operatorname{Var}[\mu(\theta)]}{\operatorname{Var}[X_j]} . \tag{2}$$

The estimator (1) depends on the three parameters μ, v and w, which are usually unknown in practice and have to be estimated. Estimators of these parameters were already proposed by Bühlmann and Straub (1970). In Switzerland other estimators have been used subsequently in practical applications. In the present paper these different estimators are compared and discussed. Furthermore, we have found *classes* of estimators, which are also dealt with in this paper. Our studies overlap in part with investigations of De Vylder (1977, 1980a, 1980b).

2. Estimation of the parameter μ

The best linear unbiased estimator of μ is:

$$\hat{\mu} = \sum_j \frac{\alpha_j}{\alpha} X_j, \text{ where } \alpha = \sum_j \alpha_j \tag{3}$$

(if $w = 0$ we define α_j/α by $\lim_{w \to 0} \alpha_j/\alpha = P_j/P$).

This result is well known and can be proved for instance by the following considerations:

$$\text{Let } \hat{\mu} = \sum_{i,j} a_{ij} X_{ij} = \sum_j a_j \sum_i \frac{a_{ij}}{a_j} X_{ij} = \sum_j a_j Y_j$$

$$\text{with } \sum_{i,j} a_{ij} = 1,\ a_j = \sum_i a_{ij}.$$

Then: $E[Y_j | \theta_j] = \mu(\theta_j)$

$\mathrm{Var}[Y_j] = E[\mathrm{Var}[Y_j | \theta_j]] + w.$

It is known that with independent RV with the same expectation $\tilde{\mu}$ we obtain the best linear unbiased estimator of $\tilde{\mu}$ by taking the weights inversely proportional to the variances. Given θ_j the RV $X_{1j}, \ldots, X_{nj}$ are independent with the same expectation. Because Var $[X_{ij} | \theta_j] = \sigma^2(\theta_j)/P_{ij}$, $\mathrm{Var}[Y_j | \theta_j]$ and thus also $\mathrm{Var}[Y_j]$ are minimal, if a_{ij}/a_j are proportional to P_{ij}, i.e. $Y_j = X_j$. For the same reason the optimal weights a_j are inversely proportional to $\mathrm{Var}[X_j] = w + v/P_j$, i.e. proportional to α_j.

Remarks:

- Bühlmann and Straub (1970) derived the homogeneous credibility estimator (i.e. the best estimator of the form $\mu(\theta_j) = \sum a_{ij} X_{ij}$). Replacing μ by (3) in the (inhomogeneous) estimator (1) yields the homogeneous estimator.
- Observe that α_j depends on the parameters v and w. (3) is an estimator only if v and w are known. But in reality they are not known. Such RV depending on unknown parameters are called "pseudo-estimators" by De Vylder (1980b). In practice v and w in (3) are simply replaced by corresponding estimations and so the final estimator $\hat{\mu}$ is not necessarily unbiased. But this is hardly a serious argument against such an estimator $\hat{\mu}$.

3. Estimation of the parameter v

Bühlmann and Straub (1970) give the following estimator:

$$\hat{v} = \frac{1}{N}\sum_j \frac{1}{n-1}\sum_i P_{ij}(X_{ij}-X_j)^2. \tag{4}$$

In a later application another estimator was proposed:

$$\hat{v} = \frac{1}{N}\sum_j \frac{1}{n}\sum_i \frac{P_{ij}}{Q_{ij}}(X_{ij}-X_j)^2, \tag{5}$$

$$\text{where } Q_{ij} = 1 - \frac{P_{ij}}{P_j}.$$

We shall briefly analyse these two estimators.

Let $S_j = \frac{1}{n-1}\sum_i P_{ij}(X_{ij}-X_j)^2$

$$T_{ij} = \frac{P_{ij}}{Q_{ij}}(X_{ij}-X_j)^2,\ T_j = \frac{1}{n}\sum_i T_{ij}.$$

It is easy to show that

$$E[S_j \mid \theta_j] = E[T_{ij} \mid \theta_j] = E[T_j \mid \theta_j] = \sigma^2(\theta_j).$$

In addition let the RV X_{ij} be conditionally normally distributed. Then $\mathrm{Var}[T_{ij} \mid \theta_j] = 2\sigma^4(\theta_j)$. If $T_{1j}, \ldots, T_{nj}$ were independent given θ_j, we would obtain $\mathrm{Var}[T_j \mid \theta_j] \leq \mathrm{Var}[S_j \mid \theta_j]$ by the same arguments used in section 2. This was the motivation for (5). However, even for given θ_j the RV $T_{1j}, \ldots, T_{nj}$ are dependent. Indeed $\mathrm{Var}[S_j \mid \theta_j] \leq \mathrm{Var}[\tilde{S}_j \mid \theta_j]$ for all estimators $\tilde{S}_j$ with $E[\tilde{S}_j \mid \theta_j] = \sigma^2(\theta_j)$, because (S_j, X_j) is a complete sufficient statistic given θ_j. Furthermore, all S_j have the same variance and thus it is optimal to give each of them the same weight.

Therefore, supposing that the X_{ij} are conditionally normally distributed, (4) is better than (5), while we do not know of any case where (5) has a smaller variance than (4). In our opinion estimator (4) should therefore be preferred to (5).

In the following $\hat{v}$ always denotes the estimator (4). Furthermore, it is supposed that $\hat{v} > 0$. If $\hat{v} = 0$, then every contract has full credibility and we need not estimate w at all.

4. Estimation of the parameter w

The estimation of the parameter w is the most difficult and is the very motive of this investigation. The estimator of Bühlmann and Straub (1970) can be written in the following manner:

$$\hat{w} = \frac{1}{c}\left\{\sum_j \frac{P_j}{P}(X_j - X)^2 - (N-1)\frac{\hat{v}}{P}\right\}, \tag{6}$$

where

$$c = \sum_j \frac{P_j}{P}\left(1 - \frac{P_j}{P}\right).$$

As possibly $\hat{w} < 0$, $\max(\hat{w}, 0)$ is used as estimator.

Bichsel and Straub (1976) proposed the following estimator:

$\hat{w}$ is the positive solution (if such a solution exists) of the equation:

$$\hat{w} = f(\underline{X}, \hat{v}, \hat{w}), \tag{7}$$

where

$$f(\underline{X}, v, w) = \frac{1}{N-1}\sum_j \alpha_j (X_j - \hat{\mu})^2.$$

The RV $f(\underline{X}, v, w)$ depends on the parameters v and w and $E[f(\underline{X}, v, w)] = w$. Given the parameters v and w, $f(\underline{X}, v, w)$ would be an unbiased estimator of w. Thus $f(\underline{X}, v, w)$ is a "pseudo-estimator" in De Vylder's notation. Contrary to (3) this "pseudo-estimator" includes also the parameter w which we want to estimate. If we merely formally replace v and w by $\hat{v}$ and $\hat{w}$, we have equation (7) and $\hat{w}$ appears on both sides of the equation. Thus the estimator (7) is given by an implicit equation.

De Vylder (1976, 1980a) and Norberg (1981) have proposed estimators of the structural parameters in the Hachemeister regression model (1975) based on such "pseudo-estimators". The basic idea consists in considering "pseudo-estimators" which *(given the parameters) are unbiased* and in looking within a certain class of "pseudo-estimators" for the one *with minimal variance (given the parameters)*. De Vylder has investigated certain classes of "pseudo-estimators", whereas after an appropriate parametrization Norberg has applied the Gauss-Markov theorem and has determined the optimal "pseudo-estimator" within all "pseudo-estimators" which are unbiased, linear and based on a chosen statistic. In another context Ammeter (1980) estimates his tariff parameter α by use of a "pseudo-estimator".

4.1 *Two classes of random variables and corresponding estimators*

In (6) the basic RV is

$$T=\sum_j \frac{P_j}{P}(X_j-X)^2$$

and

$$E[T]=\sum_j \frac{P_j}{P}\left(1-\frac{P_j}{P}\right)\left(\frac{v}{P_j}+w\right).$$

If we use instead of P_j other positive numbers a_j, we get instead of T other corresponding RV.
Hence let $\underline{a}>\underline{0}$ (i.e. $a_j>0$ for $j=1, \ldots, N$), $a=\sum_j a_j>0$ and

$$X_a=\sum_j \frac{a_j}{a} X_j.$$

Then

$$E\left[\sum_j \frac{a_j}{a}(X_j-X_a)^2\right]=\sum_j \frac{a_j}{a}\left(1-\frac{a_j}{a}\right)\left(\frac{v}{P_j}+w\right). \tag{8}$$

By this we get the following class of estimators as a direct generalization of the Bühlmann-Straub estimator.

Class I

$$\{\hat{w} \mid \hat{w} \text{ is solution of } \hat{w}=f(\underline{X}, \hat{\underline{a}}, \hat{v}, \hat{w}), f(\underline{X}, \underline{a}, v, w) \in \text{class (10)}\} \tag{9}$$

$$\left\{ f(\underline{X}, \underline{a}, v, w)=\frac{\sum_j \frac{a_j}{a}(X_j-X_a)^2-\sum_j \frac{v}{P_j}\frac{a_j}{a}\left(1-\frac{a_j}{a}\right)}{\sum_j \frac{a_j}{a}\left(1-\frac{a_j}{a}\right)} \;\middle|\; \underline{a} \geqq \underline{0}, a>0 \right\} \tag{10}$$

Remarks:

- The weights a_j may depend on the parameters v and w and on the volumes P_j.
- In the formula only the relative weights a_j/a occur. Without loss of generality we could assume that $a=1$. We will make use of that in the sections 4.4 and 4.5.

Notation: – if $a_j = a_j(v, w)$ then $\hat{a}_j := a_j(\hat{v}, \hat{w})$ and $\underline{\hat{a}} := (\hat{a}_1, \dots, \hat{a}_N)$,

– if the weights a_j do not depend on v and w, then $\hat{a}_j := a_j$ and $\underline{\hat{a}} := \underline{a}$. In this case (9) is an explicit unbiased estimator, where the right side $f(\underline{X}, \underline{\hat{a}}, \hat{v}, \hat{w})$ does not depend on $\hat{w}$.

If $E[X_{ij}] = \mu$ is known we use instead of (9) and (10):

$$\{\hat{w} \mid \hat{w} \text{ is solution of } \hat{w} = \tilde{f}(\underline{X}, \underline{\hat{a}}, \hat{v}, \hat{w}), \tilde{f}(\underline{X}, \underline{a}, v, w) \in \text{class (12)}\} \tag{11}$$

$$\left\{\tilde{f}(\underline{X}, \underline{a}, v, w) = \sum_j \frac{a_j}{a}(X_j - \mu)^2 - \sum_j \frac{a_j}{a}\frac{v}{P_j} \,\middle|\, \underline{a} \geqq \underline{0}, a > 0\right\} \tag{12}$$

Bichsel and Straub use in (7) as basic RV $T = \Sigma \alpha_j (X_j - \hat{\mu})^2$ and we have:

$E[T] = (N-1)w$. If we use any other positive weights a_j instead of the weights α_j, then (8) can be written as

$$E\left[\sum_j a_j (X_j - X_a)^2\right] = w\left(\sum_j a_j \left(1 - \frac{a_j}{a}\right)\alpha_j^{-1}\right).$$

Thus the following class is a direct generalization of the Bichsel-Straub estimator:

Class II

$$\{\hat{w} \mid \hat{w} \text{ is solution of } \hat{w} = g(\underline{X}, \underline{\hat{a}}, \hat{v}, \hat{w}), g(\underline{X}, \underline{a}, v, w) \in \text{class (14)}\} \tag{13}$$

$$\left\{ g(\underline{X}, \underline{a}, v, w) = \frac{\sum_j \frac{a_j}{a}(X_j - X_a)^2}{\sum_j \alpha_j^{-1} \frac{a_j}{a}\left(1 - \frac{a_j}{a}\right)} \,\middle|\, \underline{a} \geqq \underline{0}, a > 0 \right\} \tag{14}$$

Remark:

If $a_j = \alpha_j$ then (13) is the Bichsel-Straub estimator (7).

If $E[X_{ij}] = \mu$ is known, we use instead of (13) and (14):

$$\{\hat{w} \mid \hat{w} \text{ is solution of } \hat{w} = \tilde{g}(\underline{X}, \underline{\hat{a}}, \hat{v}, \hat{w}), \tilde{g}(\underline{X}, \underline{a}, v, w) \in \text{class (16)}\} \tag{15}$$

$$\left\{ \tilde{g}(\underline{X}, \underline{a}, v, w) = \frac{\sum_j a_j (X_j - \mu)^2}{\sum_j a_j \alpha_j^{-1}} \,\middle|\, \underline{a} \geqq \underline{0}, a > 0 \right\} \tag{16}$$

We have obtained these classes I and II by direct generalization of the Bühlmann-Straub and Bichsel-Straub estimators. However, although (10) and (14) are different classes of RV, the corresponding estimator classes (9) and (13) are fully identical. By purely algebraic operations it can be seen that for all weights $\underline{a}$ each solution of $\hat{w} = f(\underline{X}, \hat{\underline{a}}, \hat{v}, \hat{w})$ is at the same time, with the same weights $\hat{\underline{a}}$, also a solution of $\hat{w} = g(\underline{X}, \hat{\underline{a}}, \hat{v}, \hat{w})$ and vice versa. Equally (12) and (16) are different classes of RV, (11) and (15) however fully identical classes of estimators.

4.2 *A basic problem*

Let $\hat{w}_1$ and $\hat{w}_2$ be two implicit estimators given as solutions of:

$\hat{w}_1 = f_1(\underline{X}, \hat{w}_1), \quad \hat{w}_2 = f_2(\underline{X}, \hat{w}_2)$

Furthermore let:

$E[f_1(\underline{X}, w)] = E[f_2(\underline{X}, w)] = w$

$\mathrm{Var}[f_1(\underline{X}, w)] < \mathrm{Var}[f_2(\underline{X}, w)]$

The usual procedure based on "pseudo-estimators" is that in such a case $\hat{w}_1$ is preferred two $\hat{w}_2$, which also seems adequate intuitively. But can such a conclusion be justified? What properties of the RV $f_1(\underline{X}, w)$ and $f_2(\underline{X}, w)$ can be expected to recur in the corresponding estimators? Does it hold for instance that $E[\hat{w}_1] = E[\hat{w}_2] = w$ and $\mathrm{Var}[\hat{w}_1] < \mathrm{Var}[\hat{w}_2]$?

First it has to be noted, that $\hat{w}_1$ and $\hat{w}_2$ are *not necessarily unbiased*. This effect usually occurs if in an unbiased estimator unknown parameters are replaced by corresponding estimates. But with a sufficiently large number of contracts such estimators are very often approximately unbiased.

The second question concerns the variance. Can we draw conclusions under certain circumstances from $\mathrm{Var}[f(\underline{X}, w)]$ about $\mathrm{Var}[\hat{w}]$? With regard to this question let us consider the classes (12) and (16) in an example.

Example:

Let X_j be normally distributed

Then for (12) and (16):

$$\mathrm{Var}[\tilde{f}(\underline{X}, \underline{a}, v, w)] = 2w^2 \sum_j \left(\frac{a_j}{a}\right)^2 \alpha_j^{-2} \quad \text{and} \tag{17}$$

$$\mathrm{Var}\left[\tilde{f}(\underline{X}, \underline{a}, v, w)\right] \text{ is minimal, if } a_j = \alpha_j^2 \tag{18}$$

$$\mathrm{Var}\left[\tilde{g}(\underline{X}, \underline{a}, v, w)\right] = 2w^2 \frac{\sum\limits_j a_j^2 \alpha_j^{-2}}{(\sum\limits_j a_j \alpha_j^{-1})^2} \quad \text{and} \tag{19}$$

$$\mathrm{Var}\left[\tilde{g}(\underline{X}, \underline{a}, v, w)\right] \text{ is minimal, if } a_j = \alpha_j. \tag{20}$$

If we take the variance of the "pseudo-estimator" as criterion for the variance of the estimators, we thus obtain as "optimal" estimators within (11) and (15) respectively:

$$\hat{w}_1 \text{ is solution of } \hat{w}_1 = \tilde{f}(\underline{X}, \hat{\underline{a}}, \hat{v}, \hat{w}_1), \text{ where } a_j = \alpha_j^2 \tag{21}$$

$$\hat{w}_2 \text{ is solution of } \hat{w}_2 = \tilde{g}(\underline{X}, \hat{\underline{a}}, \hat{v}, \hat{w}_2), \text{ where } a_j = \alpha_j. \tag{22}$$

Note that (22) is the Bichsel-Straub estimator if μ is known. As exposed in section 4.1, (11) and (15) are identical and hence $\hat{w}_1$ belongs also to the class (15) and $\hat{w}_2$ belongs also to the class (11). Thus we have the paradoxical situation that with respect to class (12) $a_j = \alpha_j^2$ are the optimal weights and with respect to class (16) $a_j = \alpha_j$ are the optimal weights. This is obviously a contradiction demonstrating that we have to be careful about drawing conclusions from the variance of the "pseudo-estimators" about the variance of the estimators.

But the question remains: which of the following estimators should be used: Bühlmann-Straub, Bichsel-Straub or an estimator with quadratic credibility-weights $a_j = \alpha_j^2$?

4.3 *An asymptotic solution*

In statistics there are estimators having a certain formal similarity with these two classes; for instance the maximum likelihood estimator of a parameter p is often calculated in the following way:

Let $X_1, \ldots, X_N$ be the observations:

then $\hat{p}$ is the solution of an equation $h(\hat{p}, X_1, \ldots, X_N) = 0$.

Each generally valid property of these estimators (consistency, efficiency, etc.) is an asymptotic property. In our problem we cannot expect any generally valid statement with a finite number of observations.

Let us assume that v is known. In the classes I and II the estimated value $\hat{w}$ is the solution of an equation

$$\hat{w} = F(\underline{X}, \hat{w}),$$

where

$$E[F(\underline{X}, w)] = w.$$

It would be more precise to write $\hat{w}_N = F_N(\underline{X}, \hat{w}_N)$, because all these quantities depend on the number N of contracts. Let the following assumptions be fulfilled:

a) For all N and for $y \geq 0$ the first and the second derivatives of $F(\underline{X}, y)$ with respect to y both exist and are continuous.

b) The estimator is consistent, i.e. $\hat{w} \xrightarrow{P} w$ as $N \longrightarrow \infty$.

c) There exists a real number z (possibly depending on w) such that $F'(\underline{X}, w) \xrightarrow{P} z$ as $N \longrightarrow \infty$.

d) There is a number C and an open interval I containing w such that $P[|F''(\underline{X}, y)| > C] \longrightarrow 0$ as $N \longrightarrow \infty$ for $y \in I$.

Here $\xrightarrow{P}$ denotes the convergence in probability. Under these assumptions we have

$$\begin{aligned} \hat{w} &= \hat{w} - w + w = F(\underline{X}, \hat{w}) \\ &= F(\underline{X}, w) + (\hat{w} - w)\, F'(\underline{X}, w) + \frac{1}{2}(\hat{w} - w)^2\, F''(\underline{X}, w^*), \end{aligned}$$

where $w^* \in [\hat{w}, w]$ (or $[w, \hat{w}]$).

Thus

$$\sqrt{N}(F(\underline{X}, w) - w) = \sqrt{N}(\hat{w} - w) \cdot \{1 - \underset{\substack{\downarrow P \\ z}}{F'(\underline{X}, w)} - \frac{1}{2}\underset{\substack{\downarrow P \\ 0}}{(\hat{w} - w)}\, F''(\underline{X}, w^*)\}.$$

If $\hat{w} \xrightarrow{P} w$, then $P[w^* \in I] \longrightarrow 1$. Hence the bracket $\{\ldots\}$ converges in probability to $1 - z$.

We thus have

$$L\{(1-z)\sqrt{N}(\hat{w}-w)\} = L\{\sqrt{N}(F(\underline{X}, w)-w)\}, \tag{23}$$

where $L\{Y_N\}$ denotes the asyptotic distribution of a random sequence. In particular for the asymptotic variance we have:

$$(1-z)^2 \,\text{As.Var}\left[\sqrt{N}\,\hat{w}\right] = \text{As.Var}\left[\sqrt{N}\,F(\underline{X}, w)\right]. \tag{24}$$

Here As.Var[.] denotes the variance of the asymptotic distribution. This asymptotic variance is not necessarily identical with the limit of the variance. But in our two classes of RV we have the following situation: the stochastic part of $F(\underline{X}, w)$ is:

$$\sum_j \frac{a_j(w)}{a(w)}(X_j - X_a)^2 = \sum_j \frac{a_j(w)}{a(w)}(X_j-\mu)^2 - (X_a-\mu)^2,$$

i.e. essentially a sum of independent random variables plus a random variable converging to 0 in probability. By standardizing $F(\underline{X}, w)$ we obtain (with some necessary conditions about $\underline{X}$)

$$\frac{F(\underline{X}, w)-w}{\sqrt{\text{Var}[F(\underline{X}, w)]}} \longrightarrow N(0,1),\ N \longrightarrow \infty,$$

i.e.

$$L\left\{\sqrt{N}(F(\underline{X}, w)-w)\right\} = N\{0, \lim(N\cdot\text{Var}[F(\underline{X}, w)])\}.$$

Thus the equation (24) may be written as follows:

$$(1-z)^2 \,\text{As.Var}[\sqrt{N}\,\hat{w}] = \lim_{N\to\infty}\{N\ \text{Var}[F(\underline{X}, w)]\}. \tag{24'}$$

To get an "optimal" estimator from the class I or II we minimize the variance of $F(\underline{X}, w)$. Equation (24)' shows that minimizing $\text{Var}[F(\underline{X}, w)]$ does not necessarily imply the minimization of $\text{Var}[\hat{w}]$, because $z = \lim F'(\underline{X}, w)$ may depend on w. So we can explain the paradox in the example of section 4.2.

In the class (10), differentiating with respect to w yields:

$$E[f'(\underline{X}, \underline{a}, v, w)] = 0 \text{ for all } \underline{a},$$

$$\text{i.e. } z = \lim_{N\to\infty} f'(\underline{X}, \underline{a}, v, w) = 0.$$

However in the class (14) we obtain:

$$0 < z = \lim_{N\to\infty} \frac{\sum_j \frac{v}{P_j} a_j \left(1-\frac{a_j}{a}\right)}{\sum_j \left(\frac{v}{P_j}+w\right) a_j \left(1-\frac{a_j}{a}\right)} < 1. \tag{26}$$

The procedure of section 4.2 is therefore admissible for the class I, but it is incorrect for the class II. Hence in the example of the section 4.2 the estimator (21) is optimal by taking the weights $a_j = \alpha_j^2$.

Remarks:

- With reasonable assumptions about the portfolio it can be proved that the Bichsel-Straub estimator fulfills conditions a)–d). For a whole class of estimators we will examine the question of consistency in section 4.5.
- In the above considerations we have assumed v to be known. We were in a similar situation in section 2. If v is estimated by (4), we get:

$$L\left\{(1-z)\sqrt{N}\,(\hat{w}-w)\right\} = L\left\{\sqrt{N}(F(\underline{X}, v, w)-w)+y\,\sqrt{N}(\hat{v}-v)\right\},$$

where

$$y = \lim_{N\to\infty} \frac{d}{dv}\; F(\underline{X}, v, w) \quad \text{(in probability)}.$$

In class I we have

$$y = -\lim_{N\to\infty} \frac{\sum_j \frac{1}{P_j} a_j\,(1-a_j)}{\sum_j a_j\,(1-a_j)}.$$

Hence y depends on $\underline{a}$. For instance in the normal-normal case (i.e. $X_{ij} \sim N(\mu(\theta_j), v/P_{ij})$ distributed conditionally, $\mu(\theta_j) \sim N(\mu, w)$), $\hat{v}$ and $F(\underline{X}, v, w)$ are independent so that

$$\text{As.Var}\left[\sqrt{N}\cdot\hat{w}\right] = \text{As.Var}\left[\sqrt{N}\,F(\underline{X}, v, w)\right] + y(\underline{a})^2\,\text{As.Var}\left[\sqrt{N}\,\hat{v}\right].$$

As a matter of fact we would have to consider also the variance of v in the determination of the optimal weights.

4.4 *Optimal estimation for normally distributed observations*

Let us now assume that the RV X_j are normally distributed with expected value μ and variance $v/P_j+w=w\alpha_j^{-1}$. If μ is known, then, as we have seen, the estimator (11) is (asymptotically) optimal if $a_j=\alpha_j^2/\Sigma\alpha_k^2$. If μ is unknown, we have to determine those a_j for which the variance of $f(\underline{X}, \underline{a}, v, w)$ in (10) becomes minimum. By direct calculation we get:

$$y := \mathrm{Var}\,[f(\underline{X}, \underline{a}, v, w)]$$

$$= 2w^2 \frac{\Sigma a_j^2\,\alpha_j^{-2}+(\Sigma a_j^2\alpha_j^{-1})^2-2\Sigma a_j^3\,\alpha_j^{-2}}{(1-\Sigma a_j^2)^2} \tag{27}$$

where

$$a=\Sigma a_j=1.$$

Let $P_0>0$ be the smallest volume within the portfolio and α_0 the corresponding credibility factor. It is easy to verify that for all j

$$\frac{\partial y}{\partial\alpha_j^{-1}}\geqq 0.$$

As $1\leqq\alpha_j^{-1}\leqq\alpha_0^{-1}$ we have

$$2w^2\frac{\Sigma a_j^2+(\Sigma a_j^2)^2-2\Sigma a_j^3}{(1-\Sigma a_j^2)^2}\leqq y\leqq 2w^2\,\alpha_0^{-2}\,\frac{\Sigma a_j^2+(\Sigma a_j^2)^2}{(1-\Sigma a_j^2)^2}. \tag{28}$$

As $\Sigma a_j^3\leqq\sqrt{\Sigma a_j^2}\;\Sigma a_j^2$, the lefthand side of the inequality is larger than

$$\frac{\Sigma a_j^2\,(1-\sqrt{\Sigma a_j^2})^2}{(1-\Sigma a_j^2)^2}>0.$$

Let us assume that $P_0=\mathrm{Min}\,P_j\geqq c>0$ as $N\longrightarrow\infty$. Because of (28), we have $y\longrightarrow 0$ if and only if $\Sigma a_j^2\longrightarrow 0$ (i.e. $\mathrm{Var}\,[f(\underline{X}, \underline{a}, v, w)]\longrightarrow 0\Longleftrightarrow \mathrm{Var}\,[X_a]\longrightarrow 0$).

Thus we can confine our investigations to weights $\underline{a}$ fulfilling

$$\Sigma a_j^2\longrightarrow 0,\; N\longrightarrow\infty \tag{29}$$

In (27) we have (inequality of Schwarz)

$$(\Sigma a_j^2\ \alpha_j^{-1})^2 = (\Sigma a_j\ a_j\ \alpha_j^{-1})^2 \leqq \Sigma a_j^2\ \Sigma a_j^2\ \alpha_j^{-2}$$

and

$$\Sigma a_j^3\ \alpha_j^{-2} \leqq \operatorname*{Max}_k\ a_k \sum_j a_j^2\ \alpha_j^{-2} \leqq \sqrt{\Sigma a_j^2}\ \Sigma a_j^2\ \alpha_j^{-2}.$$

Thus

$$y = 2w^2\ \Sigma a_j^2\ \alpha_j^{-2}\{1 + O_N\}, \tag{30}$$

where $O_N \longrightarrow 0$ as $\Sigma a_j^2 \longrightarrow 0$.

Hence asymptotically the optimal weights in class (10) are those that minimize $\Sigma a_j^2 \alpha_j^{-2}$ and thus are identical with the optimal weights in class (12). The estimation of μ has no influence on the choice of the optimal weights which are:

$$a_j = \frac{\alpha_j^2}{\Sigma \alpha_k^2}. \tag{31}$$

4.5 *Existence, uniqueness and consistency*

In section 4.1 we have defined the estimator $\hat{w}$ as the solution of an equation $\hat{w} = F(\underline{X}, \underline{\hat{a}}, \hat{v}, \hat{w})$. Thereby we have not examined at all the question whether such a solution exists and is unique. The following example shows that a more precise definition is required. Consider two contracts with $P_1 = 10$, $P_2 = 1$, $v = 10$ and assume μ to be known. We will estimate w by (15) with the quadratic weights $a_j = \alpha_j^2 / \Sigma \alpha_k^2$.

Let the observations be such that

$(X_1 - \mu)^2 = 0.807018$
$(X_2 - \mu)^2 = 47.087719.$

Then the equation (16) has four different non-negative solutions: $\hat{w}_0 = 0$, $\hat{w}_1 = 1$, $\hat{w}_2 = 2$, $\hat{w}_3 = 4.4474$.
In this section we give an exact definition of the estimators. As we have seen, the classes I and II defined in section 4.1 lead to the same estimated values. For practical reasons we will work in the class II.

Suppose the chosen weights are functions $a_j(v, w, \underline{P}, N)$, which fulfill the following conditions for any given $v, \underline{P}$ and N.

i) $0 \leqq a_j(v, c, \underline{P}, N) < 1, \; \sum_{j=1}^{N} a_j(v, c, \underline{P}, N) = 1, \quad 0 \leqq c < \infty$

ii) $a_j(v, c, \underline{P}, N)$ is continuous in c, $0 \leqq c < \infty$

iii) $\lim_{c \to \infty} a_j(v, c, \underline{P}, N)$ exists and is < 1.

As we have remarked in section 4.1, the normalization $\Sigma a_j = 1$ does not entail a loss of generality.
For the sake of simplicity we will write $a_j(c)$ instead of $a_j(v, c, \underline{P}, N)$.

Within class II we have to solve the equation (13), i.e.

$$g(c) = \frac{\Sigma\, a_j(c)\,(X_j - X_{a(c)})^2}{\Sigma\left(c + \frac{\hat{v}}{P_j}\right) a_j(c)\,(1 - a_j(c))}\, c = c. \tag{32}$$

For a given realisation $\underline{X}$ it is evident that

1) $c = 0$ is always a solution of (32).

2) All other solutions are also solutions of $g(c)/c = 1$.

3) $$\lim_{c \downarrow 0} \frac{g(c)}{c} = \frac{\Sigma a_j(0)\,(X_j - X_{a(0)})^2}{\hat{v}\, \Sigma \frac{1}{P_j} a_j(0)\,(1 - a_j(0))}. \tag{33}$$

4) $\lim_{c \uparrow \infty} \dfrac{g(c)}{c} = 0.$

Let

$$h(c) = \begin{cases} \dfrac{g(c)}{c} & , 0 < c < \infty \\[2ex] \lim_{c \downarrow 0} \dfrac{g(c)}{c}, & c = 0 \end{cases}$$

Thus $h(c)$ is continuous with $\lim_{c \to \infty} h(c) = 0$. If $h(0) > 1$, then the equation $h(c) = 1$ has at least one strictly positive solution which is also a solution of (32). Hence the following definition is valid:

Definition:

a) If $h(0) = \dfrac{\Sigma a_j(0)\,(X_j - X_{a(0)})^2}{\hat{v}\,\Sigma \dfrac{1}{P_j} a_j(0)\,(1 - a_j(0))} > 1,$ (34)

$\hat{w}$ is the smallest positive solution of equation (32).

b) Otherwise: $\hat{w} = 0$.

Remarks:

- With this definition in the above example (with $a_j(0) = P_j^2/\Sigma P_k^2$), the estimate of w is $\hat{w} = 0$.
- In the case $a_j(c) = P_j/\Sigma P_k$, our definition is the same as the Bühlmann-Straub estimator.
- The real reason for proposing this definition is the following. For known v let (with v instead of $\hat{v}$ in (32)):

$$G(c) = E\left[g(c)\right] = \frac{\sum\left(w + \frac{v}{P_j}\right) a_j(c)\,(1 - a_j(c))}{\sum\left(c + \frac{v}{P_j}\right) a_j(c)\,(1 - a_j(c))}\, c$$

$$H(c) = E\left[h(c)\right] = \frac{G(c)}{c}$$

The equation $G(c) = c$ has the following solutions:

a) 0 and w if $H(0) > 1$ (i.e. $w > 0$)

b) 0 if $H(0) = 1$ (i.e. $w = 0$)

Having a large portfolio we can expect these properties to be valid for the estimators too.

Theorem 1:
If:

$v > 0$ is known

$\mathrm{Var}\left[(X_j - \mu)^2\right] \leqq V < \infty$, for all j.

$0 < P_0 \leqq P_j \leqq Q < \infty$, for all j.

The weights $a_j(c)$ fulfill the above stated conditions (*i*), (*ii*), (*iii*) and in addition the following one

(*iv*) $\quad \Sigma a_j^2(c) \longrightarrow 0,\ N \longrightarrow \infty$

uniformly in c, $0 \leqq c < \infty$.

Then:

The above defined estimator w is consistent, i.e. $\hat{w} \xrightarrow{P} w,\ N \longrightarrow \infty$.

Remark:

The Bichsel-Straub estimator $(a_j(c) = \alpha_j(c)/\Sigma\, \alpha_k(c))$ as well as the estimator with the quadratic weights $(a_j(c) = \alpha_j^2(c)/\Sigma \alpha_k^2(c))$ fulfills assumptions (*i*)–(*iv*).

In the Bichsel-Straub estimator (with $\alpha(c) = \Sigma \alpha_j(c)$) we have

$$\frac{d}{dc}\,\Sigma a_j(c)^2 = \frac{d}{dc}\,\Sigma\left(\frac{\alpha_j(c)}{\alpha(c)}\right)^2$$

$$= \frac{2}{c\alpha(c)^3}\sum_{j,\,k} \alpha_k(c)\, \alpha_j^2(c)\, (\alpha_k(c) - \alpha_j(c))$$

$$= -\frac{1}{c\alpha(c)^3}\sum_{j,\,k} \alpha_j(c)\, \alpha_k(c)\, (\alpha_j(c) - \alpha_k(c))^2 \leq 0.$$

With the quadratic weights $(\beta(c) = \Sigma \alpha_j^2(c))$ we have

$$\frac{d}{dc}\,\Sigma a_j(c)^2 = \frac{d}{dc}\,\Sigma\left(\frac{\alpha_j^2(c)}{\beta(c)}\right)^2$$

$$= \frac{4}{c\beta(c)^3}\sum_{j,\,k} \alpha_j^2(c)\, \alpha_k^2(c)\, (\alpha_j^2(c)\, \alpha_k(c) - \alpha_j^3(c))$$

$$= -\frac{2}{c\beta(c)^3}\sum_{j,\,k} \alpha_j^2(c)\, \alpha_k^2(c)\, (\alpha_j(c) + \alpha_k(c))\, (\alpha_j(c) - \alpha_k(c))^2 \leq 0,$$

i.e. for both estimators we have

$$\Sigma a_j^2(c) \leqq \Sigma a_j^2(0) \longrightarrow 0,\ N \longrightarrow \infty$$

uniformly in c.

Thus condition (*iv*) is fulfilled. Conditions (*i*)–(*iii*) are clearly fulfilled. In particular we have for both estimators:

$$a_j(\infty) = 1/N.$$

For the Bühlmann-Straub estimator ($a_j(c) = P_j/P$) all the conditions are clearly fulfilled.

Proof of Theorem 1:

For each N we have as before

$$h_N(c) = \frac{\Sigma a_j(c)\,(X_j - X_{a(c)})^2}{\Sigma\left(c + \dfrac{v}{P_j}\right) a_j(c)\,(1 - a_j(c))}, \quad 0 \leqq c < \infty, \tag{35}$$

$$H_N(c) = \frac{\Sigma\left(w + \dfrac{v}{P_j}\right) a_j(c)\,(1 - a_j(c))}{\Sigma\left(c + \dfrac{v}{P_j}\right) a_j(c)\,(1 - a_j(c))}, \quad 0 \leqq c < \infty. \tag{36}$$

Due to (*iv*) there is a number $A < 1$ such that

$$\Sigma a_j^2(c) \leqq A \quad \text{for } N \geqq N_0,\ 0 \leqq c < \infty.$$

Let $B_N(c)$ be the common denominator of (35) and (36).

$$B_N(c) \geqq \frac{v(1 - A)}{Q} \quad \text{for } N \geqq N_0,$$

i.e. $1/B_N(c)$ is uniformly bounded in c for $N \geqq N_0$.
The numerator of $h_N(c) - H_N(c)$ can be written as

$$\underbrace{\Sigma a_j(c)\left[(X_j - \mu)^2 - \left(w + \frac{v}{P_j}\right)\right]}_{A_1(c)} \underbrace{-(X_{a(c)} - \mu)^2}_{A_2(c)} + \underbrace{\Sigma\left(w + \frac{v}{P_j}\right) a_j^2(c)}_{A_3(c)}.$$

We have

1) $E[A_1(c)] = 0$

   $\text{Var}[A_1(c)] \leqq \Sigma a_j^2(c).\ V \longrightarrow 0,\ N \longrightarrow \infty$, uniformly in c.

   Hence we have (due to the inequality of Chebyshev)

   $A_1(c) \xrightarrow{P} 0,\ N \longrightarrow \infty$, uniformly in c.

2) $E[X_{a(c)}] = \mu$, $\mathrm{Var}[X_{a(c)}] \leqq \Sigma a_j^2(c)\left(w+\frac{v}{P_0}\right) \longrightarrow 0,$

uniformly in c.
Thus

$X_{a(c)} \xrightarrow{P} \mu$, $N \longrightarrow \infty$, uniformly in c
and also

$A_2(c) \xrightarrow{P} 0$, $N \longrightarrow \infty$, uniformly in c.

3) $A_3(c) \leqq \left(w+\frac{v}{P_0}\right) \Sigma a_j^2(c) \longrightarrow 0$, $N \longrightarrow \infty$, uniformly in c.

Together we have

$$\sup_{0 \leqq c < \infty} |h_N(c) - H_N(c)| \xrightarrow{P} 0, \quad N \longrightarrow \infty.$$

Let $\hat{W}_N$ be the estimated value of w. In respect to the above definition we have

$h_N(\hat{W}_N) = 1$ if $h_N(0) > 1$ and

$\hat{W}_N = 0$ if $h_N(0) \leqq 1$.

In the case $w > 0$ we have for $c = 0$

$$H_N(0) \geqq 1 + \frac{w\,P_0}{v} > 1, \text{ for all } N.$$

Hence with a probability increasing to 1 we have $h_N(0) > 1$ and the equation $h_N(c) = 1$ has at least one strictly positive solution. Thus we have in this case:

$$P\left[h_N(\hat{W}_N) = 1\right] \longrightarrow 1, \quad N \longrightarrow \infty.$$

In the case $w = 0$ we have $H_N(0) = 1$ and

$$P[|h_N(\hat{W}_N) - 1| \geqq \varepsilon] = P[(|h_N(\hat{W}_N) - 1| \geqq \varepsilon) \cap (\hat{W}_N = 0)]$$
$$\leqq P[|h_N(0) - 1| \geqq \varepsilon] = P[|h_{N(0)} - H_N(0)| \geqq \varepsilon] \longrightarrow 0.$$

Thus we have in both cases

$$P[|h_N(\hat{W}_N) - 1| \geqq \varepsilon] \longrightarrow 0, \quad N \longrightarrow \infty.$$

Now

$$|H_N(\hat{W}_N)-1|\leqq|H_N(\hat{W}_N)-h_N(\hat{W}_N)|+|h_N(\hat{W}_N)-1|$$
$$\leqq\sup_c|H_N(c)-h_N(c)|+|h_N(\hat{W}_N)-1|$$
$$\xrightarrow{P} 0,\ N\longrightarrow\infty,$$
$$\text{i.e. } H_N(\hat{W}_N)\xrightarrow{P} 1,\ N\longrightarrow\infty.$$

Due to $H_N(c)\leqq\dfrac{w+\dfrac{v}{P_0}}{c+\dfrac{v}{Q}}\longrightarrow 0,\ c\longrightarrow\infty$, uniformly in N, there is a $C_1<\infty$ such that

$$c>C_1 \Rightarrow H_N(c)<\varepsilon \text{ for all } N.$$

Thus $$P[\hat{W}_N\geqq C_1]\leqq P[H_N(\hat{W}_N)<\varepsilon]\longrightarrow 0,N\longrightarrow\infty.$$

From (36) it follows that

$$|H_N(c)-1|\geqq\frac{1}{c+\dfrac{v}{P_0}}|w-c|, \text{ for all } N.$$

Thus

$$P[|\hat{W}_N-w|\geqq\delta]\leqq P[\hat{W}_N>C_1]+P[(|\hat{W}_N-w|\geqq\delta)\cap(\hat{W}_N\leqq C_1)]$$
$$\leqq P[\hat{W}_N>C_1]+P\left[|H_N(\hat{W}_N)-1|\geqq\frac{\delta}{C_1+\dfrac{v}{P_0}}\right]\longrightarrow 0,\ N\longrightarrow\infty.$$

And finally

$$\hat{W}_N\xrightarrow{P} w,\ N\longrightarrow\infty.$$

q.e.d.

4.6 *Numerical aspects of the estimators*

If the weights a_j are fixed numbers not depending on v and w then the corresponding estimators (9) and (11) are explicit unbiased estimators and there are no numerical difficulties. For instance the Bühlmann-Straub estimator

(6) is of this type. However, if the weights a_j depend on w, then the estimators are defined as the solution of an implicit equation. Such a solution can always be determined approximatively by calculating $f(\underline{X}, \hat{v}, c)$ (respectively $\tilde{f}(\underline{X}, \hat{v}\ c)$) for different values $c = c_1, c_2, \ldots$ and then by choosing $\hat{w} = c_k$ such that $f(\underline{X}, \hat{v}, c_k) \approx c_k$ (or $\tilde{f}(\underline{X}, \hat{v}, c_k) \approx c_k$). Of course it would be desirable to have a generally valid, simple algorithm for solving such an equation. As the following theorem shows, there is such an algorithm for the Bichsel-Straub estimator (7). However, for general weights (and also for the quadratic weights $a_j = \alpha_j^2$) we have not found such a generally valid method.

Theorem 2:

a) The Bichsel-Straub estimator (7) has one and only one strictly positive solution if and only if the Bühlmann-Straub estimator (6) gives a strictly positive value.

b) Let $w_0 > 0$, $w_{n+1} = f(\underline{X}, \hat{v}, w_n)$ for $n = 0, 1, 2, \ldots$

where $f(\underline{X}, v, w) = \dfrac{1}{N-1} \sum_j \alpha_j(c) \cdot (X_j - X_{\alpha(c)})^2.$

Then:
For every $w_0 > 0$ the sequence $\{w_n | n = 0, 1, 2, \ldots\}$ converges. The limit value is the strictly positive solution of (7) if such a solution exists. Otherwise it is 0.

Remarks:

- The condition in a) is equivalent to (34).
- The Bichsel-Straub estimation can be calculated by iteration. Bichsel and Straub (1976) used this method in their practical application.
- It is useful to calculate first the Bühlmann-Straub estimation and to use this value as starting point for the iteration. If the estimated value is negative no iteration is needed.
- In addition to the sequence $\{w_n\}$ we can also calculate the following two sequences of estimated values: for each step in the iteration we choose the weights $a_j = \alpha_j(w_n)$ and $b_j = \alpha_j^2(w_n)$ and then determine the corresponding (explicit) estimations $\hat{w}_{n+1}$ and $\hat{\hat{w}}_{n+1}$ according to (9). As w_n is a function of $\underline{X}$, $\hat{w}_{n+1}$ and $\hat{\hat{w}}_{n+1}$ are not unbiased. Nevertheless, as for any fixed non-stochastic weights a_j the estimator (9) is unbiased, we may hope that all

the values $\hat{w}_n$ and $\hat{\hat{w}}_n$ are of about the same size. Thus the two sequences $\{\hat{w}_n\}$ and $\{\hat{\hat{w}}_n\}$ (but not $\{w_n\}$) give us an empirical measure for the stability of the estimation.

- De Vylder (1980b) has proved the same theorem.

Proof of Theorem 2:

i) Let

$$h(\underline{X}, \hat{v}, c) = \frac{f(\underline{X}, \hat{v}, c)}{c} = \frac{1}{N-1} \sum_j \frac{P_j}{P_j c + \hat{v}} (X_j - X_{\alpha(c)})^2.$$

Then

$$h'(\underline{X}, \hat{v}, c) = -\frac{1}{N-1} \sum_j \left(\frac{P_j}{P_j c + \hat{v}}\right)^2 (X_j - X_{\alpha(c)})^2$$

$$-\frac{d}{dc} X_{\alpha(c)} \cdot \left\{\frac{2}{N-1} \sum_j \frac{P_j}{P_j c + \hat{v}} (X_j - X_{\alpha(c)})\right\}.$$

Looking at the definition of $X_{\alpha(c)}$, we see that

$$\sum_j \frac{P_j}{P_j c + \hat{v}} (X_j - X_{\alpha(c)}) = 0$$

and thus $h'(\underline{X}, \hat{v}, c) < 0$ for $c \geqq 0$.
Hence there is one and only one strictly positive solution of (7) if and only if $h(\underline{X}, \hat{v}, 0) > 1$.

$$\text{As } h(\underline{X}, \hat{v}, 0) > 1 \Longleftrightarrow \frac{1}{N-1} \sum_j \frac{P_j}{\hat{v}} (X_j - X)^2 > 1$$

$$\Longleftrightarrow \sum_j P_j (X_j - X)^2 - (N-1)\hat{v} > 0 \Longleftrightarrow (6) > 0,$$

(a) is proved.

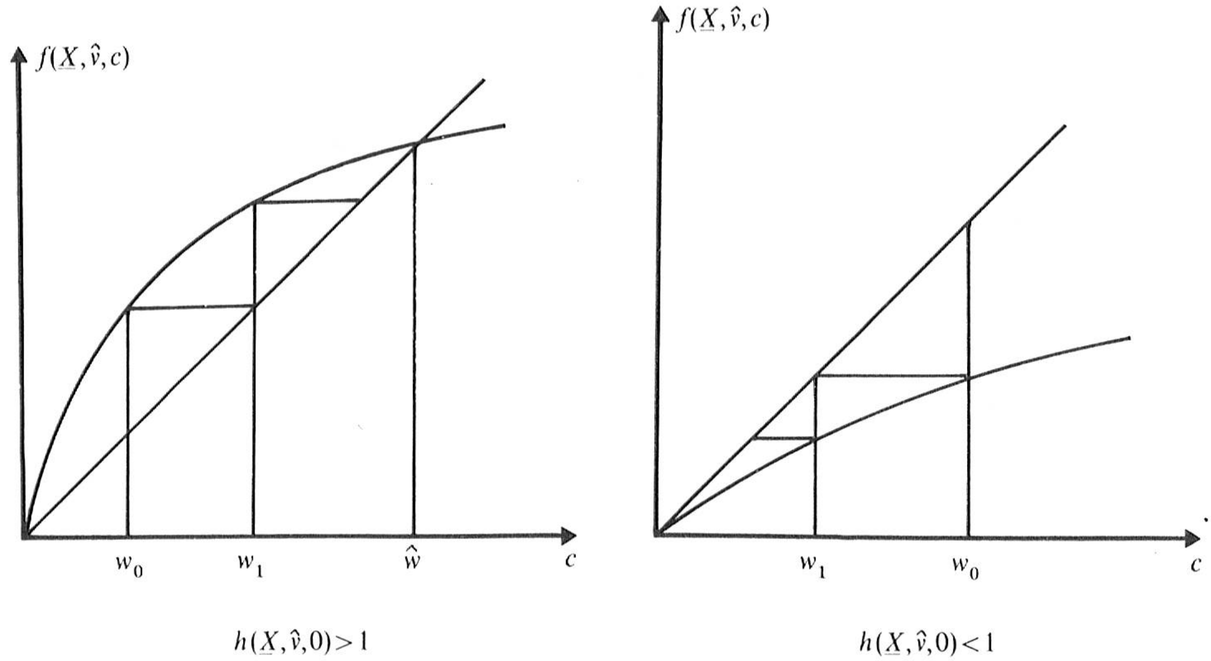
ii) Analogously as in the derivation of $h'(\underline{X}, \hat{v}, c)$ we obtain

$$f'(\underline{X}, \hat{v}, c) = \frac{1}{N-1} \sum_j \frac{P_j \hat{v}}{(P_j c + \hat{v})^2} (X_j - X_{\alpha(c)})^2 > 0 \text{ for } c \geqq 0.$$

Note that $f'(\underline{X}, \hat{v}, 0) = h(\underline{X}, \hat{v}, 0)$.

Let be $f'(\underline{X}, \hat{v}, 0) = h(\underline{X}, \hat{v}, 0) > 1$ and x_0 be the strictly positive solution of (7). Then $f(\underline{X}, \hat{v}, c) > c$ for $0 < c < x_0$ and $f(\underline{X}, \hat{v}, c) < c$ for $c > x_0$. Furthermore $f(\underline{X}, \hat{v}, c)$ is increasing with respect to c. Hence for $0 < w_0 < x_0$ we get $w_0 < w_1 = f(\underline{X}, \hat{v}, w_0) < f(\underline{X}, \hat{v}, x_0) = x_0$. By induction it follows that the sequence $\{w_n | n = 0, 1, \dots\}$ is increasing and bounded and therefore convergent. Obviously the limiting value is x_0. Similarly, if $w_0 > x_0$, then $\{w_n | n = 0, 1 \dots\}$ is monotonically decreasing with limit x_0.
If $h(\underline{X}, \hat{v}, 0) \leq 1$ then $f(\underline{X}, \hat{v}, c) < c$ for $c > 0$. The same arguments then show that $\{w_n\}$ decreases to 0.

The idea of the proof can be illustrated by the following figures:



5. Conclusions

The estimator of μ is undisputed. In order to estimate v there are good arguments for preferring (4) to (5). For the estimation of w the situation is more complicated. In particular the following estimators all belonging to the class (9) are to be discussed:

the Bühlmann-Straub estimator with the weights $a_j = P_j$;

the Bichsel-Straub estimator with the weights $a_j = \alpha_j$;

the estimator with the quadratic weights $a_j = \alpha_j^2$.

The following table shows some properties of the three estimators:

Estimator of w	Unbiased	Consistent	Asymptotically optimal*	Computational work demanded
Bühlmann-Straub	yes	yes	no	few
Bichsel-Straub	no	yes	no	medium
Quadratic weights	no	yes	yes	much

*Asymptotically optimal if the RV X_j are normally distributed.

In the "normal case" (i.e. X_j normally distributed) the estimator with the quadratic weights is asymptotically optimal, but unfortunately we are confronted with numerical difficulties. In Switzerland up to now only the Bühlmann-Straub and the Bichsel-Straub estimators have been used in practical applications. Even in the "normal case" neither of the two estimators is universally better than the other. The following example may serve as an illustration:

Example:

$6N$ contracts; $P_1 = P_2 = \ldots = P_{5N} = 1$, $P_{5N+1} = P_{5N+2} = \ldots = P_{6N} = 8$; $v = 5$, μ known

For all $N \geqq 1$ we get:	$w = 1$	$w = 5$
$N \cdot$(variance of the Bühlmann-Straub estimator):	4.13	29.88
$N \cdot$(variance of the Bichsel-Straub estimator):	5.72	26.12

Hence for $w = 1$ the Bühlmann-Straub estimator is better, whereas for $w = 5$ the Bichsel-Straub estimator is better.

The results found so far suggest as a reasonable approach the following procedure: First we determine the value $\hat{w}_1$ of the Bühlmann-Straub estimator. If $\hat{w}_1 > 0$ then we calculate the weights

$$a_j = \left(\frac{P_j \hat{w}_1}{P_j \hat{w}_1 + \hat{v}} \right)^2$$

and afterwards we make out the corresponding estimator $\hat{w}_2$ within the class (9). Note that the weights a_j are now fixed and $\hat{w}_2$ is given by an explicit formula. It can, however, not be proved that $\hat{w}_2$ is better than $\hat{w}_1$.
Finally we would like to mention that the theoretically optimal weights a_j in the class (11) are proportional to $1/\mathrm{Var}[X_j]$. If the RV X_j are not normally

distributed, these optimal weights depend also on the fourth moments of X_j (see Norberg (1981)). If we took these fourth moments into account too, the estimators would become even more complicated, and it is questionable whether improvements would be achieved by doing so.

Although we are not in the position to state definitely which estimator of w should be used in practice in every case, we do hope that this paper has brought about some clarification and suggestions in the discussion of this estimation problem.

Acknowledgement

This paper has been strongly motivated by the ideas and results of F. Bichsel and E. Straub and was written while the first-named author was working at the Forschungsinstitut für Mathematik, ETH Zürich.

A. Dubey
"Zürich" Insurances
8022 Zürich

A. Gisler
Winterthur Insurances
General Guisan-Strasse 40
8401 Winterthur

References

Ammeter, H. (1980). "Potenzmittel-Credibility." MVSVM, vol. 80/3, p. 327–340.
See also: "Potenzmittel-Credibility." Thesis, University of Bern.

Bühlmann, H. and *Straub, E.* (1970): "Glaubwürdigkeit für Schadensätze." MVSVM, vol. 70/1, p. 111–133.

Bichsel, F. and *Straub, E.* (1976): "Erfahrungstarifierung in der Kollektiv-Krankenversicherung." Unpublished.

De Vylder, F. (1977): "Optimal Parameter Estimation in semidistributionfree Credibility." Presented to the XIII ASTIN Colloquium in Washington, D.C.

– (1980a): "Practical Credibility Theory with Emphasis on optimal Parameter Estimation." Forthcoming in ASTIN-Bulletin.

– (1980b): "Hierarchical Credibility Theory." Not yet published.

Hachemeister, C. (1975): "Credibility for regression models with application to trend." In *Credibility: Theory and applications* (ed. P.M. Kahn), pp. 129–163. Academic Press, New York.

Norberg, R. (1981): "On optimal Parameter Estimation in credibility." Presented to the XV ASTIN Colloquium in Loen.

Summary

In Switzerland different estimators of the structural parameters in the credibility model of Bühlmann and Straub (1970) have been used in practical applications. These estimators are compared and discussed. Furthermore, by generalization, classes of estimators are obtained. In these classes an estimator of the parameter w is often defined as the solution of an implicit equation $\hat{w} = f(\underline{X}, \hat{w})$. It is investigated how far conclusions can be drawn from the random variable $f(\underline{X}, w)$ ("pseudo-estimator") about the estimator.

Zusammenfassung

Bei praktischen Anwendungen des Kredibilitätsmodells von Bühlmann und Straub (1970) in der Schweiz wurden verschiedene Schätzer der Strukturparameter benützt. Diese Schätzer werden verglichen und diskutiert. Überdies wurden durch Verallgemeinerung Klassen von Schätzern erhalten. In diesen Klassen ist ein Schätzer des Parameters w oft als Lösung einer impliziten Gleichung $\hat{w} = f(\underline{X}, \hat{w})$ definiert. Es wird untersucht, inwieweit Rückschlüsse von den Zufallsvariablen $f(\underline{X}, w)$ («Pseudoschätzer») auf den Schätzer gezogen werden können.

Résumé

Lors d'applications pratiques du modèle de credibility de Bühlmann et Straub (1970) en Suisse, différents estimateurs des paramètres de structure ont été utilisés. Ces estimateurs sont comparés et discutés. Par généralisation, on obtient des classes d'estimateurs. Dans ces classes un estimateur du paramètre w est souvent défini comme la solution d'une équation implicite $\hat{w} = f(\underline{X}, \hat{w})$. On étudie dans quelle mesure des conclusions peuvent être transportées de la variable aléatoire $f(\underline{X}, w)$ («pseudo-estimateur») à l'estimateur.