Robust Theory Applied to Jewell's Hierarchical Credibility Model

CAROLINE WARD

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of Requirements

for the Degree of Master of Science at

Concordia University

Montreal, Quebec, Canada

November 1997

0-612-39960-5

ABSTRACT

Robust Theory Applied to
Jewell's Hierarchical Credibility Model

CAROLINE WARD

Introduction of robust theory is considered in Jewell's hierarchical credibility model. A comparison is made with the classical linear credibility models, Künsch's model and Gisler & Reinhard's model.

Acknowledgement

I would like to thank Dr. Jose Garrido for his help and support throughout the completion of this thesis. Without his patience and encouragement, this thesis would not have been achieved.

To my family and friends.

Contents

Chapter 1

Credibility Theory

1.1 Introduction

What is credibility theory ? In statistical inference, credibility theory achieves the compromise between the notions of stability, precision and responsiveness to the most recent events. In order to obtain a good data analysis that allows for extrapolating into the future, a given population has to be divided into homogeneous classes or cells. Having homogeneous groups enables one to gain precision and to eliminate the bias caused by the evolution of the distribution of the population. It also allows to anticipate the future tendencies. Credibility theory studies how to incorporate the information that is cell-specific with the information gathered for the whole portfolio premium calculation. For the insurance industry, equity between different classes of risk can be achieved through credibility, allowing the company to react more swiftly to competition by recognizing individual and group characteristics.

Credibility theory is mostly used in the insurance field. "Credibility theory is the study of a weighting process, including development of formulas for assigning the credibility weights and estimation of the parameters or values that appear in these formulas"[Venter et al.(1990), p.380). The general precepts of credibility can also be applied outside of insurance. As an illustration, take a hockey example: at the beginning of the season, one player can compare himself to his entire team based on past seasons statistics. How does one achieve this goal? One way is through the means of credibility.

According to Goovaerts et al.(1987), p.7, "Credibility theory provides us with techniques to determine insurance premiums for contracts that belong to a more or less heterogeneous portfolio, in case there is limited or irregular claim experience for each contract but ample experience for the portfolio. It is both the art and science to adjust the insurance premiums and to improve their accuracy".

In the insurance industry, it is agreed that not all risks deserve the same premium adjustment based on past experience and on their respective volumes. A correction factor is introduced through the credibility factor. The theory of credibility is concerned with the value that should be given to this credibility factor, say Z, i.e. how much can one rely on the current data based on one's own experience and the industry's experience?

When referring to the data, it is understood that individual premiums could also be contract premiums and that the collective premiums could as well be the portfolio premiums. The general credibility formula originated in the United States and was firstly used in the field of Worker's Compensation insurance:

$$C = (1 - Z)B + ZA$$

where

- A is the mean of the current data at hand (or premium of an individual contract)
- B is the prior data or past experience (mean premium of all contracts)
- C is the compromise or the update of the data
- Z is the weight given to the current data at hand, such that $0 \leq Z \leq 1$

What happens if the volume of an individual class is sufficient? The credibility factor, Z, is then equal to 1 and there is no need to use the entire portfolio experience because the individual class is credible enough. When Z is equal to zero, the comprise premium is entirely based on the portfolio premium and the data at hand is not taken into account.

In credibility models, the heterogeneity is characterized by a random parameter value Θ. For an individual, then consider the conditional distribution of the claim size variable X given the value of the risk parameter Θ. The variance of the risk is composed of the variance of the conditional mean $\mathrm{E}(X|\Theta)$ and the expected value of the conditional variance $\mathrm{Var}(X|\Theta)$.

1.2 Classical Credibility

The remainder of this chapter relies mostly on Goovaerts et Hoogstad(1987) and Goovaerts et al.(1990).

1.2.1 Notations and definitions

Consider a portfolio of k contracts observed over t periods. The parameter Θ_{js} describes the risk characteristics of contract j ($j = 1, ..., k$) in period s ($s = 1, ..., t$). This means that for each contract, at each period, there exists a parameter Θ_{js} that describes the unique characteristics of this specific contract. The problem in practice is that the risk parameters Θ_{js}'s are unobservable and our goal is to estimate premiums that depend on these parameters.

In simple credibility models, it is assumed that contract risk characteristics are homogeneneous through time, i.e. they do not evolve through time. Therefore the subscript s in Θ_{js} is dropped. In spite of the risk time homogeneity, the claim experiences, X_{js}, of the random variables Θ_j are not necessarily homogeneous through time. The X_{js}'s are also considered as random variables, with observable realizations x_{js}. Note that while here X_{js}'s will be referred to as claims severities, they could as well represent claim ratios or claim frequencies.

If time homogeneity is assumed then Θ_j is a structural parameter and its distribution $U(\theta_j)$ is called the structure function.

Consider X for a fixed contract and a given period then

$$f(x) = \int_{\Omega} f(x|\theta)dU(\theta).$$

The portfolio premium is given by

$$\mathrm{E}(X) = \int_{\mathcal{X}} xf(x)dx,$$

and the contract premium is given by

$$\mathrm{E}(X|\Theta = \theta) = \int_{\mathcal{X}} xf(x|\theta)dx.$$

Hence $\mathrm{E}(X|\Theta)$ is a random variable. Assuming that the claim amounts are positive, the integrals can be interchanged

$$\mathrm{E}[\mathrm{E}(X|\Theta)] = \int_\Omega (\int_\mathcal{X} x f(x|\theta)dx)dU(\theta).$$

Then the following general result is given:

Lemma 1.1

$$\mathrm{E}[\mathrm{E}(X|\Theta)] = \mathrm{E}(X).$$

Consider $X_1, X_2, \ldots, X_t$ the claims for a fixed contract over t time periods. Then $f(x_1, x_2, \ldots, x_t|\theta)$ is the joint conditional distribution and

$$f(x_1, x_2, \ldots, x_t) = \int_\mathcal{X} f(x_1, x_2, \ldots, x_t|\theta)dU(\theta)$$

is the corresponding marginal distribution. From this formula one obtains the second general result:

Lemma 1.2 $\forall r, s = 1, 2, \ldots, t,$

$$\mathrm{Cov}(X_r, X_s) = \mathrm{E}[\mathrm{Cov}(X_r, X_s)|\Theta] + \mathrm{Cov}[\mathrm{E}(X_r|\Theta), \mathrm{E}(X_s|\Theta)]$$

Proof: Consider
$\mathrm{Cov}(X_r, X_s|\Theta = \theta)$

$$\begin{aligned} &= \mathrm{E}(X_r X_s|\Theta = \theta) - \mathrm{E}(X_r|\Theta = \theta)\mathrm{E}(X_s|\Theta = \theta) \\ &= \int_\mathcal{X}\int_\mathcal{X} [x_r - \mathrm{E}(X_r|\Theta = \theta)][x_s - \mathrm{E}(X_s|\Theta = \theta)] f(x_r, x_s|\theta)dx_s dx_r \end{aligned}$$

$\Rightarrow \mathrm{Cov}(X_r, X_s|\Theta)$ is also a random variable and hence

$$\mathrm{Cov}(X_r, X_s|\Theta) = \mathrm{E}(X_r X_s|\Theta) - \mathrm{E}(X_r|\Theta)\mathrm{E}(X_s|\Theta)$$

Now by Lemma 1.1

$$\begin{aligned} &\mathrm{E}[\mathrm{Cov}(X_r, X_s|\Theta)] \\ &\quad= \mathrm{E}[\mathrm{E}(X_r X_s|\Theta) - \mathrm{E}(X_r|\Theta)\mathrm{E}(X_s|\Theta)] \\ &\quad= \mathrm{E}(X_r X_s) - \mathrm{E}[\mathrm{E}(X_r|\Theta)\mathrm{E}(X_s|\Theta)] \end{aligned}$$

And also by Lemma 1.1

$$\begin{aligned} &\mathrm{Cov}[\mathrm{E}(X_r|\Theta), \mathrm{E}(X_s|\Theta)] \\ &= \mathrm{E}[\mathrm{E}(X_r|\Theta)\mathrm{E}(X_s|\Theta)] - \mathrm{E}[\mathrm{E}(X_r|\Theta)]\mathrm{E}[\mathrm{E}(X_s|\Theta)] \\ &= \mathrm{E}[\mathrm{E}(X_r|\Theta)\mathrm{E}(X_s|\Theta)] - \mathrm{E}(X_r)\mathrm{E}(X_s) \end{aligned}$$

Then $\mathrm{E}[\mathrm{Cov}(X_r, X_s|\Theta)] + \mathrm{Cov}[\mathrm{E}(X_r|\Theta), \mathrm{E}(X_s|\Theta)]$

$$\begin{aligned} &= \mathrm{E}(X_r X_s) - \mathrm{E}[\mathrm{E}(X_r|\Theta)\mathrm{E}(X_s|\Theta)] + \mathrm{E}[\mathrm{E}(X_r|\Theta)\mathrm{E}(X_s|\Theta)] - \mathrm{E}(X_r)\mathrm{E}(X_s) \\ &= \mathrm{Cov}(X_r, X_s) \end{aligned}$$

□

From the previous Lemma, we have:

Corollary 1.1

$$\mathrm{Var}(X) = \mathrm{E}[\mathrm{Var}(X|\Theta)] + \mathrm{Var}[\mathrm{E}(X|\Theta)]$$

1.2.2 Bühlmann's model for a single contract

In order to get a better grasp of Bühlmann(1967) credibility model, first define it for a single contract: i.e. $j = 1$ and $r = 1, ..., t$.

Assumptions:

(i) Conditionally on Θ, the random variables $X_1, X_2, \ldots, X_t$ are independent and identically distributed, with known common distribution $F(x|\theta)$;

(ii) $\mathrm{E}(X^2)$ is finite;

(iii) The structure function $U(\theta)$ is known.

Structural parameters:

(1) $\mu(\Theta) = \mathrm{E}(X_r|\Theta)$ is the policy or individual risk premium;

(2) $m = \mathrm{E}[\mu(\Theta)]$ is the portfolio or expected average claim amount over the entire portfolio;

(3) $\sigma^2(\Theta) = \mathrm{Var}(X_r|\Theta)$ is the measure of total claim dispersion for the portfolio;

(4) $s^2 = \mathrm{E}[\sigma^2(\Theta)]$ is the weighted portfolio claims variance;

(5) $a = \mathrm{Var}[\mu(\Theta)]$ shows the heterogeneity in the portfolio; the within variance or the variance of the individual risk premium.

Bühlmann suggests to find the "best" function $g(x_1, x_2, \ldots, x_t)$ and to estimate $\mu(\Theta)$. He defines "best" as the function $g(x_1, x_2, \ldots, x_t)$ that minimizes

$$\mathrm{E}\{[\mu(\Theta) - g(X_1, X_2, \ldots, X_t)]^2\} \tag{1.1}$$

It is well known [Goovaerts and Hoogstad(1987) pp.25-26] that

$$g^*(X_1, X_2, \ldots, X_t) = \mathrm{E}[\mu(\Theta)|X_1, X_2, \ldots, X_t] \tag{1.2}$$

is the exact value that minimizes (1.1). To approximate $g^*(x_1, x_2, \ldots, x_t)$ with a simpler estimator, restrict the class of g functions to linear ones. A linear non-homogeneous credibility estimator would be given by:

$$g(x_1, x_2, \ldots, x_t) = c_0 + \sum_{r=1}^{t} c_r x_r$$

Hence the following objective function has to be minimized

$$\mathrm{E}\{[\mu(\Theta) - c_0 - \sum_{r=1}^{t} c_r X_r]^2\}$$

with respect to $c_0, c_1, \ldots, c_t$. Differentiating with respect to c_0 and setting equal to zero gives

$$c_0 = m(1 - \sum_{r=1}^{t} c_r)$$

Now differentiate with respect to $c_{r'}$, $r' = 1, 2, \ldots, t$, and set equal to 0 to obtain

$$\mathrm{E}\{[\mu(\Theta) - m - \sum_{r=1}^{t} c_r(X_r - m)]X_{r'}\} = 0$$

Which can be rewritten as

$$\mathrm{Cov}[\mu(\Theta), X_{r'}] = \sum_{r=1}^{t} c_r \mathrm{Cov}(X_r, X_{r'})$$

Which in turn, with the help of Lemmas 1.1 and 1.2, can be rewritten as

$$a = \sum_{r=1}^{t} c_r (a + \delta_{rr'} s^2) \quad \text{where } \delta_{rr'} \text{ is Kroneckers' symbol.}$$

Solving recursively each equation for $r' = 1, 2, \ldots, t$, one finds that

$$c_1 = c_2 = \ldots = c_t = c = \frac{a}{s^2 + at} = \frac{Z}{t}$$

$$\text{where} \quad Z = \frac{at}{s^2 + at}$$

Therefore

$$c_0 = (1 - Z)m$$

and hence

$$g^*(X_1, X_2, \ldots, X_t) = (1 - Z)m + Z\bar{X} \quad \text{with } \bar{X} = \frac{1}{t} \sum_{r=1}^{t} X_i \tag{1.3}$$

is Bühlmann's credibility estimator for $\mu(\Theta)$.

Properties of Bühlmann's credibility estimator:

Bühlmann's credibility estimator is not a statistical estimator as it depends on parameter values m, s^2 and a. Nevertheless the following properties hold:

(i) Asymptotic properties:

(a) if $t \to \infty$ then $Z \to 1$
i.e. the contract number of claims is infinite, there is no need for portfolio claims data;

(b) if $s^2 \to \infty$ then $Z \to 0$
i.e. the claim experience variable for fixed Θ shows a high degree of randomness and we cannot rely on the information contained in the contract;

(c) if $a \to \infty$ then $Z \to 1$.
i.e. the portfolio is extremely heterogeneous and it does not contain credible information on the specific risk;

(d) if $a = 0$ then $Z = 0$
i.e. the portfolio is perfectly homogeneous and there is no need for the credibility factor.

(ii) Unbiased: since

$$E[g^*(X_1, X_2, \ldots, X_t)] \doteq E[(1-Z)m + Z\bar{X}] = m$$

(iii) Forecasting: for future claims, the linear approximation to $E(X_{t+1}|X_1, \ldots, X_t)$ equals the linear (approximate) premium $E[\mu(\Theta)|X_1, \ldots, X_t]$.

1.3 Exact Credibility

In Bayesian theory, if the prior $U(\theta)$ and the posterior $f(\theta|x_1, x_2, \ldots, x_t)$ belong to the same distribution family then $f(x_1, x_2, \ldots, x_t|\theta)$ and $U(\theta)$ are said to be conjuguate distributions. Jewell (1974) studied natural conjuguates. He particularly considered conjuguates of the exponential family.

For the uni-dimensional exponential family, which he defined by

$$f(x|\theta) = \frac{p(x)e^{-\theta x}}{q(\theta)}, \quad x \in \mathcal{X} \text{ and } \theta \in \Omega \tag{1.4}$$

he derives the following relations:

$$E(X|\Theta = \theta) = \frac{-q'(\theta)}{q(\theta)}$$

and hence

$$\mu(\Theta) = \frac{-q'(\Theta)}{q(\Theta)} \quad \text{and} \quad \mathrm{Var}(X|\Theta = \theta) = \frac{\partial}{\partial\theta}[\frac{q'(\theta)}{q(\theta)}] = -\mu'(\theta)$$

For natural conjuguates of the uni-dimensional family, Jewell(1974) proved the following theorem:

Theorem 1.1 (Jewell) Let $f(x|\theta)$ be a member of the one-dimensional exponential family in (1.4), then

(i) The natural conjuguate of $f(x|\theta)$, $u(\theta)$, is given by:

$$u(\theta) = \frac{q(\theta)^{-t_0}e^{-\theta x_0}}{c(t_0, x_0)}$$

where

$$t_0 = \frac{s^2}{a} \text{ and } x_0 = \frac{m}{t_0}$$

(ii) $\bar{X} = \frac{1}{t}\sum_{i=1}^{t} X_i$ is sufficient for θ.

(iii) If $u(\theta)$ is the natural conjuguate of $f(x|\theta)$ in (i), exact credibility occurs i.e., $g^*(X_1, \ldots, X_t)$ is equal to the optimal in equation (1.3) and

$$g^*(X_1, \ldots, X_t) = \mathrm{E}[\mu(\Theta)|X_1, \ldots, X_t] = (1 - Z)m + Z\bar{X}$$

Proof: the proof of this theorem can be found in the original paper by Jewell(1974).

1.4 The Classical Model of Bühlmann

Bühlmann(1967) adapted his model described in the previous section to portfolio data. The jth-contract is described by a vector $(\Theta_j, X_{j1}, ..., X_{jt}) = (\Theta_j, \underline{X}_j)$. Hence for each contract $j = 1, ..., k$ we have:
Assumptions:

(B1) The contracts $j = 1, ..., k$ i.e. the pairs $(\Theta_j, \underline{X}_j)$ are independent and the random variables Θ_j are identically distributed;

(B2) For every contract $j = 1, ..., k$ conditionally on Θ_j, the random variables $X_{j1}, ..., X_{jt}$ are independent and identically distributed.

(B3) $\forall r, s = 1, ..., t$ and $\forall j = 1, ..., k$

$$\mathrm{E}(X_{jr}|\Theta_j) = \mu(\Theta_j)$$

$$\mathrm{Cov}(X_{jr}, X_{js}|\Theta_j) = \delta_{rs}\sigma^2(\Theta_j)$$

Structural parameters:

(1) $m = \mathrm{E}(X_{jr}) = \mathrm{E}[\mu(\Theta_j)]$

(2) $a = \mathrm{Var}[\mu(\Theta_j)]$

(3) $s^2 = \mathrm{E}[\sigma^2(\Theta_j)]$

Notation:

$$Z = \frac{at}{s^2 + at}$$

Theorem 1.2 Under assumptions (B1) to (B3), the non-homogeneous linear credibility estimate for $\mu(\Theta_j)$ is given by

$$\hat{\mu}(\Theta_j) = (1 - Z)m + Z\bar{X}_{j.}$$

Proof: The proof is as in section 1.2.2 and it can be found in Goovaerts et al.(1990), pp.143-144. One has to minimize

$$\mathrm{E}\{[\mu(\Theta_j) - c_0 - \sum_1^t c_r X_{jr}]^2\}$$

Again $\hat{\mu}(\Theta_j)$ is a credibility estimator but not a statistical one.

Lemma 1.3 The following estimators are unbiased for the structural parameters m, s^2 and a respectively:

$$\hat{m} = \frac{1}{kt}\sum_{r=1}^{t}\sum_{j=1}^{k} X_{jr}$$

$$\hat{s}^2 = \frac{1}{k(t-1)}\sum_{r=1}^{t}\sum_{j=1}^{k}(X_{jr} - \bar{X}_j)^2$$

$$\hat{a} = \frac{1}{(k-1)}\sum_{j=1}^{k}(\bar{X}_j - \bar{X})^2 - \frac{\hat{s}^2}{t}$$

Proof: see Goovaerts et al.(1990), p.145.

Remark: Although $\hat{\mu}(\Theta_j) = (1 - Z)m + Z\bar{X}_{j.}$ is unbiased, in the sense that $\mathrm{E}[\hat{\mu}(\Theta_j)] = m = \mathrm{E}[\mu(\Theta_j)]$, note that $\hat{\hat{\mu}}(\Theta_j) = (1-\hat{Z})\hat{m} + \hat{Z}\bar{X}_{j}.$ does not have an expectation of m.

1.5 Bühlmann and Straub's Model

The Bühlmann and Straub (1970) model allows for the introduction of weights w_{js} in the variance expressions of X_{js}. This is a first generalization of the i.i.d. assumption associated with contracts, enlarging the field of applications.

Assumptions:

Here the observations are still assumed independent and with constant mean, but the variance is allowed to vary with the values of known weights:

(BS1) The contracts $j = 1, ..., k$ [i.e. the pair vectors $(\Theta_j, \underline{X}_j)$] are independent and the variables Θ_j's are identically distributed;

(BS2) $\forall r, s = 1, ..., t$ and $\forall j = 1, ..., k$

$$E(X_{jr}|\Theta_j) = \mu(\Theta_j)$$

$$\text{Cov}(X_{jr}, X_{js}|\Theta_j) = \frac{\delta_{rs}}{w_{jr}}\sigma^2(\Theta_j)$$

where the w_{jr} are known weights and $\mu(\Theta_j)$ and $\sigma^2(\Theta_j)$ are unknown functions.

Structural parameters:

(1) $m = E(X_{jr}) = E[\mu(\Theta_j)]$

(2) $a = \text{Var}[\mu(\Theta_j)]$

(3) $s^2 = E[\sigma^2(\Theta_j)]$

Notation:

$$w_{..} = \sum_{j=1}^{k} w_{j.} = \sum_{j=1}^{k}\sum_{r=1}^{t} w_{jr}$$

$$Z_j = \frac{aw_{j.}}{aw_{j.} + s^2} \quad \text{and} \quad Z_. = \sum_{j=1}^{k} Z_j$$

$$X_{jw} = \sum_{r=1}^{t} \frac{w_{jr}}{w_{j.}} X_{jr}$$

$$X_{ww} = \sum_{j=1}^{k} \frac{w_{j.}}{w_{..}} X_{jw}$$

$$X_{zw} = \sum_{j=1}^{k} \frac{Z_j}{Z_.} X_{jw}$$

From the previous hypotheses, one obtains the following Lemmas:

Lemma 1.4 $\forall i, j = 1, \ldots, k$ and $r = 1, \ldots, t$

$$\text{Cov}[\mu(\Theta_j), X_{ir}] = \delta_{ij} a$$

Proof: Follows from Lemma 1.2:

$$\begin{aligned}
\mathrm{Cov}[\mu(\Theta_j), X_{ir}] &= \mathrm{Cov}\{\mathrm{E}[\mu(\Theta_j)|\Theta_j], \mathrm{E}(X_{ir}|\Theta_j)\} + \mathrm{E}\{\mathrm{Cov}[\mu(\Theta_j), X_{ir}|\Theta_j]\} \\
&= \mathrm{Cov}[\mu(\Theta_j), \mathrm{E}(X_{ir}|\Theta_j)] + \mathrm{E}\{\mathrm{Cov}[\mu(\Theta_j), X_{ir}|\Theta_j]\} \\
&= \delta_{ij}a + 0 \\
&= \delta_{ij}a
\end{aligned}$$

since $\mu(\Theta_j)$, conditionally on Θ_j, is a constant and the covariance of a constant with any random variable is 0. □

Lemma 1.5 $\forall r, u = 1, \ldots, t$

$$\mathrm{Cov}(X_{jr}, X_{iu}) = 0, \qquad \text{if } j \neq i$$

Proof: by Lemma 1.2

$$\begin{aligned}
\mathrm{Cov}(X_{jr}, X_{iu}) &= \mathrm{E}[\mathrm{Cov}(X_{jr}, X_{iu}|\Theta_j)] + \mathrm{Cov}[\mathrm{E}(X_{jr}|\Theta_j), \mathrm{E}(X_{iu}|\Theta_j)] \\
&= 0 + \mathrm{Cov}[\mathrm{E}(X_{jr}|\Theta_j), \mathrm{E}(X_{iu})] \\
&= \mathrm{Cov}[\mathrm{E}(X_{jr}|\Theta_j), m] \\
&= 0
\end{aligned}$$

□

Lemma 1.6 $\forall j = 1, \ldots, k$ and $\forall r, u = 1, \ldots, t$

$$\mathrm{Cov}(X_{jr}, X_{ju}) = a + \frac{\delta_{ru}s^2}{w_{jr}}$$

Proof: by Lemma 1.2 and by hypothesis (BS1)

$$\begin{aligned}
\mathrm{Cov}(X_{jr}, X_{ju}) &= \mathrm{E}[\mathrm{Cov}(X_{jr}, X_{ju}|\Theta_j)] + \mathrm{Cov}[\mathrm{E}(X_{jr}|\Theta_j), \mathrm{E}(X_{ju}|\Theta_j)] \\
&= \mathrm{E}[\frac{\delta_{ru}\sigma^2(\Theta_j)}{w_{jr}}] + \mathrm{Cov}[\mu(\Theta_j), \mu(\Theta_j)] \\
&= \frac{\delta_{ru}s^2}{w_{jr}} + a
\end{aligned}$$

□

With the previous results, the following theorem can be shown:

Theorem 1.3 Under assumptions (BS1) and (BS2) the non-homogeneous linear credibility estimator is given by

$$\hat{\mu}(\Theta_j) = (1 - Z_j)m + Z_j X_{jw} \tag{1.5}$$

Proof: For fixed $j = 1, \ldots, k$, $\mu(\Theta_j)$ is estimated by a function

$$g_j(X_{11}, X_{12}, \ldots, X_{kt}) = c_0^j + \sum_{i=1}^{k} \sum_{r=1}^{t} c_{ir}^j X_{ir}$$

Again a square-loss function

$$\mathrm{E}\{[\mu(\Theta_j) - c_0^j - \sum_{i=1}^{k} \sum_{r=1}^{t} c_{ir}^j X_{ir}]^2\}$$

has to be minimized with respect to $c_0^j, c_{11}^j, \ldots, c_{kt}^j$, for $j = 1, \ldots, k$. For a given $j = 1, \ldots, k$ take the derivative with respect to c_0^j and set equal to 0 to obtain:

$$\mathrm{E}[\mu(\Theta_j)] - c_0^j - m \sum_{i=1}^{k} \sum_{r=1}^{t} c_{ir}^j = 0$$

Hence

$$c_0^j = [1 - \sum_{i=1}^{t} \sum_{r=1}^{t} c_{ir}^j] m \tag{1.6}$$

Take the partial derivatives with respect to the $c_{i'r'}^j$, replace c_0^j by (1.6) and set equal to 0. This gives the following system of equations, $\forall i' = 1, \ldots, k$ and $\forall r' = 1, \ldots, t$:

$$\begin{aligned} \mathrm{Cov}[\mu(\Theta_j), X_{i'r'}] &= \sum_{i=1}^{k} \sum_{r=1}^{t} c_{ir}^j \, \mathrm{Cov}(X_{ir}, X_{i'r'}) \\ &= \sum_{r=1}^{t} c_{i'r}^j [a + \frac{\delta_{rr'} s^2}{w_{i'r'}}] \end{aligned} \tag{1.7}$$

The last line is obtained from Lemma 1.6 and Lemma 1.5; we can say it is true only if $i = i'$, otherwise it is equal to 0. There are 2 cases:

1. $i' \neq j$
Then by Lemma 1.4 equation (1.7) is equivalent to

$$\begin{aligned} 0 &= \sum_{r=1}^{t} c^j_{i'r}[a + \frac{\delta_{rr'} s^2}{w_{i'r}}] \\ &= a c^j_{i'.} + \frac{s^2 c^j_{i'r'}}{w_{i'r'}} \end{aligned}$$

which implies

$$\begin{aligned} 0 &= a c^j_{i'.} w_{i'.} + s^2 c^j_{i'.} \\ &= c^j_{i'.}(a w_{i'.} + s^2) \end{aligned}$$

$\forall i' = 1, \ldots, k$. Then since by assumption s^2 and a are non-zero values

$$c^j_{i'.} = 0 \quad \forall i' = 1, \ldots, k$$

And hence

$$c^j_{i'r'} = 0 \quad \forall i' = 1, \ldots, k \text{ and } r' = 1, \ldots, t$$

2. $i' = j$
Then by Lemma 1.4 equation (1.7) is equivalent to

$$a = a c^j_{j.} + \frac{c^j_{jr'} s^2}{w_{jr'}} \quad \text{for} \quad r' = 1, \ldots, t \tag{1.8}$$

where

$$c^j_{j.} = \sum_{r=1}^{t} c^j_{jr}$$

Again multiplying each side of the last equation by $w_{jr'}$ and summing each side over r' gives

$$\sum_{r'=1}^{t} a w_{jr'} = a c^j_{j.} \sum_{r'=1}^{t} w_{jr'} + s^2 c^j_{j.}$$

Since $w_{j.} = \sum_{r'=1}^{t} w_{jr'}$, isolate $c^j_{j.}$ to obtain

$$c^j_{j.} = \frac{a w_{j.}}{a w_{j.} + s^2} \tag{1.9}$$

Therefore replace equation (1.9) into equation (1.8) to yield

$$
\begin{aligned}
c^j_{jr} &= \frac{aw_{jr}}{aw_{j.} + s^2} \\
&= [\frac{aw_{j.}}{aw_{j.} + s^2}][\frac{w_{jr}}{w_{j.}}] \\
&= Z_j[\frac{w_{jr}}{w_{j.}}]
\end{aligned}
$$

Now, since for $i \neq j$, $c^j_{ir} = 0$, then

$$
\begin{aligned}
c^j_0 &= m(1 - \sum_{r=1}^{t} c^j_{jr}) \\
&= m(1 - Z_j \sum_{r=1}^{t} \frac{w_{jr}}{w_{j.}}) \\
&= m(1 - Z_j)
\end{aligned}
$$

And

$$
\begin{aligned}
g_j(X_{11}, X_{12}, \ldots, X_{kt}) = \hat{\mu}(\Theta_j) &= c^j_0 + \sum_{r=1}^{t} c^j_{jr} X_{jr} \\
&= m(1 - Z_j) + Z_j \sum_{r=1}^{t} \frac{w_{jr}}{w_{j.}} X_{jr} \\
&= m(1 - Z_j) + Z_j X_{jw}
\end{aligned}
$$

□

Lemma 1.7 The following are unbiased estimators for the structural parameters m, a and s^2 respectively:

$$\hat{m} = X_{zw}$$

$$\hat{a} = w_{..} \frac{\sum_{j=1}^{k} w_{j.}(X_{jw} - X_{ww})^2 - (k-1)\hat{s}^2}{w_{..}^2 - \sum_{j=1}^{k} w_{j.}^2}$$

$$\hat{s}^2 = \frac{1}{k} \frac{1}{(t-1)} \sum_{j=1}^{k} \sum_{r=1}^{t} w_{jr}(X_{jr} - X_{jw})^2$$

Proof: see Goovaerts et al.(1990), p.152.

Remark: $\hat{m} = X_{zw}$ is unbiased with Z_j weights but not with

$$\hat{Z}_j = \frac{\hat{a}w_{j.}}{\hat{a}w_{j.} + \hat{s}^2}.$$

Also remark that

$$\mathrm{E}[\hat{\mu}(\Theta_j)] = \mathrm{E}[(1 - \hat{Z}_j)\hat{m} + \hat{Z}_j X_{jw}] \neq m.$$

1.6 Jewell's Hierarchical Model

In certain cases, portfolios can be subdivided into subportfolios (or sectors) which show more homogeneity within subportfolios that between them. Jewell (1975) proposed a direct extension of Bühlmann & Straub's model to this problem. The model is constructed in such a way that

(1) Each subportfolio is given a structural variable Θ_p where $p = 1, 2, \ldots, P$;

(2) Each contract within subportfolio p is given structural variables (Θ_p, Θ_{pj}) for $j = 1, \ldots, k_p$;

(3) Each year, within a given contract j, we observe a claim X_{pjr}, $r = 1, \ldots, t_{pj}$, possibly paired to a known weight w_{pjr} given in advance.

The data for subportfolio p is the set of variables $(\Theta_p, \Theta_{pj}, X_{pjr})$ while for the contract pj it is defined by the set (Θ_{pj}, X_{pjr}).

Assumptions

(J1) The subportfolios $p = 1, \ldots, P$ [i.e. the pairs $(\Theta_p, \Theta_{pj}, X_{pjr})$] are independent $\forall p \neq p'$;

(J2) For each $p = 1, \ldots, P$, the contracts $pj = p1, \ldots, pk_p$ [i.e. the pairs (Θ_{pj}, X_{pjr})] are conditionally independent given Θ_p;

(J3) $\forall p = 1, \ldots, P$ and $\forall j = 1, \ldots, k_p$, the claims $X_{pj1}, \ldots, X_{pjt_{pj}}$ are conditionally independent given (Θ_p, Θ_{pj});

(J4) All pairs of variables (Θ_p, Θ_{pj}), for $p = 1, \ldots, P$ and $j = 1, \ldots, k_p$, are identically distributed;

(J5) $\forall p, j$ and r

$$\mathrm{E}(X_{pjr}|\Theta_p, \Theta_{pj}) = \mu(\Theta_p, \Theta_{pj}) \quad \forall r = 1, \ldots, t_{pj}$$

and

$$\mathrm{Var}(X_{pjr}|\Theta_p, \Theta_{pj}) = \frac{1}{w_{pjr}} \sigma^2(\Theta_p, \Theta_{pj}) \quad \forall r = 1, \ldots, t_{pj}$$

where μ and σ^2 do not depend on the subscripts p, j and r and w_{pjr} are known weights.

Define

$$\nu(\Theta_p) = \mathrm{E}[\mu(\Theta_p, \Theta_{pj})|\Theta_p] = \mathrm{E}(X_{pjr}|\Theta_p)$$

Structural parameters:

(1) $m = \mathrm{E}[\nu(\Theta_p)] = \mathrm{E}[\mu(\Theta_p, \Theta_{pj})] = \mathrm{E}(X_{pjr})$ represents the combined expectation for the entire portfolio;

(2) $s^2 = \mathrm{E}[\sigma^2(\Theta_p, \Theta_{pj})]$ measures the degree of fluctuation of the individual contract i.e. the heterogeneity in time of the data;

(3) $a = \mathrm{E}\{\mathrm{Var}[\mu(\Theta_p, \Theta_{pj})|\Theta_p]\}$ is now the quantity measuring the degree of variability (or heterogeneity) in a subportfolio;

(4) $b = \mathrm{Var}[\nu(\Theta_p)]$ measures the heterogeneity between the different subportfolios.

Notation

$$w_{p..} = \sum_{j=1}^{k_p} w_{pj.} = \sum_{j=1}^{k_p} \sum_{r=1}^{t_{pj}} w_{pjr}$$

$$X_{pjw} = \sum_{r=1}^{t_{pj}} \frac{w_{pjr}}{w_{pj.}} X_{pjr}$$

$$X_{pzw} = \sum_{j=1}^{k_p} \frac{Z_{pj}}{Z_{p.}} X_{pjw}$$

$$X_{zzw} = \sum_{p=1}^{P} \frac{Z_p}{Z_{..}} X_{pzw}$$

The credibility factor at the contract level is

$$Z_{pj} = \frac{aw_{pj.}}{s^2 + aw_{pj.}}$$

The credibility factor at the subportfolio level is

$$Z_p = \frac{bZ_{p.}}{a + bZ_{p.}}$$

In order to avoid confusion, one has to be careful in differentiating between

$$Z_p \text{ and } Z_{p.} = \sum_{j=1}^{k_p} Z_{pj}$$

Lemma 1.8 For any $p, q = 1, 2, \ldots, P$, $i = 1, 2, \ldots, k_q$, $j = 1, 2, \ldots, k_p$ and $r = 1, 2, \ldots, t_{qi}$

$$\text{Cov}[\mu(\Theta_p, \Theta_{pj}), X_{qir}] = \delta_{pq}(\delta_{ij}a + b)$$

Proof:
$\text{Cov}[\mu(\Theta_p, \Theta_{pj}), X_{qir}]$

$$\begin{aligned} &= \text{Cov}\{\text{E}[\mu(\Theta_p, \Theta_{pj})|\Theta_p], \text{E}(X_{qir}|\Theta_p)\} + \text{E}\{\text{Cov}[\mu(\Theta_p, \Theta_{pj}), X_{qir}|\Theta_p]\} \\ &= \text{Cov}\{\nu(\Theta_p), \text{E}(X_{qir}|\Theta_p)\} + \text{E}\{\text{Cov}[\mu(\Theta_p, \Theta_{pj}), X_{qir}|\Theta_p]\} \end{aligned}$$

But $\text{Cov}[\mu(\Theta_p, \Theta_{pj}), X_{qir}|\Theta_p]$

$$\begin{aligned} &= \text{E}[\text{Cov}\{\mu(\Theta_p, \Theta_{pj}), X_{qir}|\Theta_{pj}\}|\Theta_p] + \text{Cov}[\text{E}\{\mu(\Theta_p, \Theta_{pj})|\Theta_{pj}\}, \text{E}\{X_{qir}|\Theta_{pj}\}|\Theta_p] \\ &= \delta_{pq}\delta_{ij}\text{Cov}[\mu(\Theta_p, \Theta_{pj}), \text{E}(X_{pjr}|\Theta_p, \Theta_{pj})] \\ &= \delta_{pq}\delta_{ij}\text{Var}[\mu(\Theta_p, \Theta_{pj})] \end{aligned}$$

And hence $\text{Cov}[\mu(\Theta_p, \Theta_{pj}), X_{qir}]$

$$\begin{aligned} &= \delta_{pq}\text{Cov}[\nu(\Theta_p), \text{E}(X_{pir}|\Theta_p)] + \delta_{pq}\delta_{ij}\text{E}\{\text{Var}[\mu(\Theta_p, \Theta_{pj})]\} \\ &= \delta_{pq}\text{Var}[\nu(\Theta_p)] + \delta_{pq}\delta_{ij}a \\ &= \delta_{pq}b + \delta_{pq}\delta_{ij}a \\ &= \delta_{pq}(\delta_{ij}a + b) \end{aligned}$$

□

Lemma 1.9 For any $p, q = 1, 2, , \ldots, P, i = 1, 2, , \ldots, k_q$ and $r = 1, 2, \ldots, t_{q_i}$

$$\text{Cov}[\nu(\Theta_p), X_{qir}] = \delta_{pq} b$$

Proof:

$$\begin{aligned}
\text{Cov}[\nu(\Theta_p), X_{qir}] &= \text{Cov}\{\text{E}[\nu(\Theta_p)|\Theta_p], \text{E}(X_{qir}|\Theta_p)\} + \text{E}\{\text{Cov}[\nu(\Theta_p), X_{qir}|\Theta_p]\} \\
&= \text{Cov}[\nu(\Theta_p), \delta_{pq}\text{E}(X_{pir}|\Theta_p) + (1 - \delta_{pq})m] + 0 \\
&= \delta_{pq}\text{Var}[\nu(\Theta_p)] \\
&= \delta_{pq} b
\end{aligned}$$

□

Lemma 1.10 By (J1), for any $p \neq q, = 1, 2, \ldots, P, j = 1, 2, \ldots, k_p, j' = 1, 2, , \ldots, k_q, r = 1, 2, \ldots, t_{pj}$ and $r' = 1, 2, \ldots, t_{qj'}$

$$\text{Cov}[X_{pjr}, X_{qj'r'}] = 0$$

Lemma 1.11 For any $p, q = 1, 2, \ldots, P$ and $j = 1, 2, \ldots, k_q$

$$\text{Cov}[\nu(\Theta_p), X_{qjw}] = \delta_{pq} b$$

Proof:

$$\begin{aligned}
\text{Cov}[\nu(\Theta_p), X_{qjw}] &= \text{Cov}[\nu(\Theta_p), \sum_{r=1}^{t_{qj}} \frac{w_{qjr}}{w_{qj\cdot}} X_{qjr}] \\
&= \sum_{r=1}^{t_{qj}} \frac{w_{qjr}}{w_{qj\cdot}} \text{Cov}[\nu(\Theta_p), X_{qjr}] \\
&= \sum_{r=1}^{t_{qj}} \frac{w_{qjr}}{w_{qj\cdot}} \delta_{pq} b, \quad \text{from above,} \\
&= \delta_{pq} b
\end{aligned}$$

□

Lemma 1.12 For any $p = 1, 2, \ldots, P, j = 1, 2, \ldots, k_p$ and $r, r' = 1, 2, \ldots, t_{pj}$

$$\text{Cov}[X_{pjr}, X_{pjr'}] = a + b + \frac{\delta_{rr'} s^2}{w_{pjr}}$$

Proof:

$$\begin{aligned}
\mathrm{Cov}[X_{pjr}, X_{pjr'}] &= \mathrm{Cov}[\mathrm{E}(X_{pjr}|\Theta_p, \Theta_{pj}), \mathrm{E}(X_{pjr'}|\Theta_p, \Theta_{pj})] + \mathrm{E}[\mathrm{Cov}(X_{pjr}, X_{pjr'}|\Theta_p, \Theta_{pj})] \\
&= \mathrm{Var}[\mu(\Theta_p, \Theta_{pj})] + \delta_{rr'}\mathrm{E}[\mathrm{Var}(X_{pjr}|\Theta_p, \Theta_{pj})] \\
&= \mathrm{Var}\{\mathrm{E}[\mu(\Theta_p, \Theta_{pj})|\Theta_p]\} + \mathrm{E}\{\mathrm{Var}[\mu(\Theta_p, \Theta_{pj})|\Theta_p]\} + \frac{\delta_{rr'}s^2}{w_{pjr}} \\
&= \mathrm{Var}\{\mathrm{E}[\nu(\Theta_p)]\} + a + \frac{\delta_{rr'}s^2}{w_{pjr}} \\
&= a + b + \frac{\delta_{rr'}s^2}{w_{pjr}}
\end{aligned}$$

□

Lemma 1.13 For any $p = 1, 2, \ldots, P, j, j' = 1, 2, \ldots, k_p, r = 1, 2, \ldots, t_{pj}$ and $r' = 1, 2, \ldots, t_{pj'}$

$$\mathrm{Cov}[X_{pjr}, X_{pj'r'}] = \delta_{jj'}[a + \frac{\delta_{rr'}s^2}{w_{pjr}}] + b$$

Proof:

$$\begin{aligned}
\mathrm{Cov}[X_{pjr}, X_{pj'r'}] &= \mathrm{E}[\mathrm{Cov}(X_{pjr}, X_{pj'r'}|\Theta_p)] + \mathrm{Cov}[\mathrm{E}(X_{pjr}|\Theta_p), \mathrm{E}(X_{pj'r'}|\Theta_p)] \\
&= \delta_{jj'}\mathrm{E}[\mathrm{Cov}(X_{pjr}, X_{pjr'}|\Theta_p)] + \mathrm{Var}[\nu(\Theta_p)] \\
&= \delta_{jj'}\{\mathrm{Cov}(X_{pjr}, X_{pj'r'}) - \mathrm{Cov}[\mathrm{E}(X_{pjr}|\Theta_p), \mathrm{E}(X_{pj'r'}|\Theta_p)]\} + b \\
&= \delta_{jj'}\{\mathrm{Cov}(X_{pjr}, X_{pj'r'}) - \mathrm{Var}[\nu(\Theta_p)]\} + b \\
&= \delta_{jj'}[a + b + \frac{\delta_{rr'}s^2}{w_{pjr}} - b] + b \\
&= \delta_{jj'}[a + \frac{\delta_{rr'}s^2}{w_{pjr}}] + b
\end{aligned}$$

□

Lemma 1.14 For any $p = 1, 2, \ldots, P, j, j' = 1, 2, \ldots, k_p, r = 1, 2, \ldots, t_{pj}$ and $r' = 1, 2, \ldots, t_{pj'}$

$$\mathrm{Cov}[X_{pjr}, X_{pj'w}] = \delta_{jj'}\frac{a}{Z_{pj}} + b$$

Proof:

$$\begin{aligned}
\mathrm{Cov}[X_{pjr}, X_{pj'w}] &= \mathrm{Cov}(X_{pjr}, \sum_{r'=1}^{t_{pj'}} \frac{w_{pj'r'}}{w_{pj'.}} X_{pj'r'}) \\
&= \sum_{r'=1}^{t_{pj'}} \frac{w_{pj'r'}}{w_{pj'.}} \mathrm{Cov}(X_{pjr}, X_{pj'r'}) \\
&= \delta_{jj'}[\sum_{r'=1}^{t_{pj'}} \frac{w_{pj'r'}}{w_{pj'.}} (\frac{\delta_{rr'} s^2}{w_{pj'r'}} + a)] + b \\
&= \delta_{jj'}(\frac{s^2}{w_{pj'.}} + a) + b \\
&= \delta_{jj'} \frac{a}{Z_{pj}} + b
\end{aligned}$$

□

Lemma 1.15 For any $p = 1, 2, \ldots, P, j, j' = 1, 2, \ldots, k_p, r = 1, 2, \ldots, t_{pj}$ and $r' = 1, 2, \ldots, t_{pj'}$

$$\mathrm{Cov}[X_{pjw}, X_{pj'w}] = \delta_{jj'} \frac{a}{Z_{pj}} + b$$

Proof:

$$\begin{aligned}
\mathrm{Cov}[X_{pjw}, X_{pj'w}] &= \mathrm{Cov}(\sum_{r=1}^{t_{pj}} \frac{w_{pjr}}{w_{pj.}} X_{pjr}, X_{pj'w}) \\
&= \sum_{r=1}^{t_{pj}} \frac{w_{pjr}}{w_{pj.}} \mathrm{Cov}(X_{pjr}, X_{pj'w}) \\
&= \sum_{r=1}^{t_{pj}} \frac{w_{pjr}}{w_{pj.}} (\delta_{jj'} \frac{a}{Z_{pj}} + b) \\
&= \delta_{jj'} \frac{a}{Z_{pj}} + b
\end{aligned}$$

□

The credibility premiums at the portfolio and contract levels, for a two-level model, i.e. $p = 2$, are given by the following theorem.

Theorem 1.4 Under assumptions (J1) to (J5), the non-homogeneous linear credibility estimate at the portfolio level is given by:

$$\hat{\nu}(\Theta_p) = (1 - Z_p)m + Z_p X_{pzw}$$

Proof:
Again minimize a mean-square expression which takes the form of

$$\mathrm{E}\{[\nu(\Theta_p) - c_0 - \sum_{q=1}^{P}\sum_{j=1}^{k_q}\sum_{r=1}^{t_{qj}} c_{qjr}X_{qjr}]^2\}$$

By symmetry arguments, it can be shown that c_{qjr} will be proportional to w_{qjr}. Therefore the following simplified expression has to be minimized

$$\mathrm{E}\{[\nu(\Theta_p) - c_0 - \sum_{q=1}^{P}\sum_{j=1}^{k_q} c_{qj}X_{qjw}]^2\} \tag{1.10}$$

Differentiate the last equation with respect to c_0 and set equal to 0 to obtain

$$c_o = \mathrm{E}[\nu(\Theta_p)] - \sum_{q=1}^{P}\sum_{j=1}^{k_q} c_{qj}\mathrm{E}(X_{qjw})$$

Replace c_0 in equation (1.10) by the value previously obtained to minimize the following objective function:

$$\mathrm{E}[\{\nu(\Theta_p) - \mathrm{E}[\nu(\Theta_p)] - \sum_{q=1}^{P}\sum_{j=1}^{k_q} c_{qj}X_{qjw} - \mathrm{E}(X_{qjw})\}^2]$$

Differentiate with respect to $c_{q'j'}$ ($q' = 1, \ldots, P$ and $j' = 1, \ldots, k_q$) and set equal to 0 to obtain:

$$\mathrm{Cov}[\nu(\Theta_p), X_{q'j'w}] = \sum_{q=1}^{P}\sum_{j=1}^{k_q} c_{qj}\mathrm{Cov}(X_{qjw}, X_{q'j'w})$$

Replace the respective covariances by the results obtained in the previous Lemmas to get

$$\delta_{pq'}b = \sum_{q=1}^{P}\sum_{j=1}^{k_q} c_{qj}\delta_{qq'}(\frac{\delta_{jj'}a}{Z_{qj}} + b)$$

As with the Bühlmann and Straub model, distinguish between two cases:

1. $p \neq q'$

$$\begin{aligned} 0 &= \sum_{q=1}^{P} \sum_{j=1}^{k_q} \delta_{qq'} c_{qj} (\frac{\delta_{jj'} a}{Z_{qj}} + b) \\ &= \sum_{j=1}^{k_q} c_{qj} (\frac{\delta_{jj'} a}{Z_{q'j}} + b) \\ &= c_{q.} b + \frac{c_{qj'} a}{Z_{q'j'}} \\ &= Z_{q'j'} c_{q.} b + c_{q'j'} a \end{aligned}$$

Now sum on each side to obtain

$$\begin{aligned} 0 &= \sum_{j'=1}^{k_q} c_{q.} b Z_{q'j'} + \sum_{j'=1}^{k_q} a c_{q'j'} \\ &= c_{q'.} Z_{q'.} b + a c_{q'.} \\ &= c_{q'.} (b Z_{q'.} + a) \end{aligned}$$

Hence $c_{q'.} = 0 \ \ \forall q' = 1, \ldots, P$ and $c_{q'j'} = 0 \ \ \forall j' = 1, \ldots, k_{q'}$.

2. $p = q'$

$$\begin{aligned} b &= \sum_{q=1}^{P} \sum_{j=1}^{k_q} \delta_{qq'} c_{qj} (\frac{\delta_{jj'} a}{Z_{qj}} + b) \\ &= \sum_{j=1}^{k_q} c_{qj} (\frac{\delta_{jj'} a}{Z_{qj}} + b) \\ &= \frac{c_{pj'} a}{Z_{pj'}} + b c_{p.} \end{aligned}$$

Let

$$\alpha_p = \frac{c_{pj'}}{Z_{pj'}} = \frac{c_{p.}}{Z_{p.}} \quad \forall j' = 1, \ldots, k$$

Then

$$\begin{aligned} b &= \alpha_p a + \alpha_p Z_{p.} b \\ &= \alpha_p (a + Z_{p.} b) \end{aligned}$$

Isolate α_p to obtain

$$\alpha_p = \frac{b}{a + Z_{p.}b} = \frac{Z_p}{Z_{p.}}$$

And hence

$$\begin{aligned} c_{pj'} &= \alpha_p Z_{pj'} \\ &= \frac{Z_{pj'}}{Z_{p.}} Z_p \end{aligned}$$

Hence

$$\begin{aligned} c_0 &= \mathrm{E}[\nu(\Theta_p)] - \sum_{j=1}^{k_p} \frac{Z_{pj}}{Z_{p.}} Z_p \mathrm{E}(X_{qjw}) \\ &= m(1 - Z_p) \end{aligned}$$

Finally

$$\begin{aligned} \nu(\hat{\Theta}_p) &= m(1 - Z_p) + \sum_{j=1}^{k_p} \frac{Z_{pj'}}{Z_{p.}} Z_p X_{qjw} \\ &= m(1 - Z_p) + Z_p X_{pjz} \end{aligned}$$

□

Theorem 1.5 Under assumptions (J1) to (J5), the non-homogeneous linear credibility estimate on the contract level is given by

$$\hat{\mu}(\Theta_p, \Theta_{pj}) = (1 - Z_{pj}) m_p + Z_{pj} X_{pjw}$$

Proof:
The following least-square equation has to be minimized

$$\mathrm{E}(\mathrm{E}\{[\mu(\Theta_p, \Theta_{pj}) - c_0 - \sum_{q=1}^{P} \sum_{i=1}^{k_q} \sum_{r=1}^{t_{qi}} c_{qir} X_{qir}]^2 | \Theta_p\}) \tag{1.11}$$

Differentiate with respect to c_0 and set the derivative equal to 0 gives

$$\begin{aligned} c_0 &= \mathrm{E}\{\mathrm{E}[\mu(\Theta_p, \Theta_{pj})] | \Theta_p\} - \sum_{q=1}^{P} \sum_{i=1}^{k_q} \sum_{r=1}^{t_{qi}} c_{qir} \mathrm{E}[\mathrm{E}(X_{qir} | \Theta_p)] && (1.12) \\ &= \mathrm{E}[\mu(\Theta_p, \Theta_{pj}) | \Theta_p] - \sum_{q=1}^{P} \sum_{i=1}^{k_q} \sum_{r=1}^{t_{qi}} c_{qir} \mathrm{E}(X_{qir}) && (1.13) \end{aligned}$$

Replacing equation (1.13) into equation (1.11) yields

$$E(E\{[\mu(\Theta_p, \Theta_{pj}) - E[\mu(\Theta_p, \Theta_{pj})] - \sum_{q=1}^{P}\sum_{i=1}^{k_q}\sum_{r=1}^{t_{qi}} c_{qir}[X_{qir} - E(X_{qir})]]^2|\Theta_p\})$$

Differentiate the last equation with respect to $c_{q'i'r'}$ and set equal to 0 now gives

$$E\{\text{Cov}[\mu(\Theta_p, \Theta_{pj}), X_{q'i'r'}|\Theta_p]\} = \sum_{q=1}^{P}\sum_{i=1}^{k_q}\sum_{r=1}^{t_{qi}} c_{qir} E[\text{Cov}(X_{qir}, X_{q'i'r'}|\Theta_p)] \quad (1.14)$$

But by Lemma 1.8, $\text{Cov}[\mu(\Theta_p, \Theta_{pj}), X_{q'i'r'}]$ is

$$\begin{aligned} &= E\{\text{Cov}[\mu(\Theta_p, \Theta_{pj}), X_{q'i'r'}|\Theta_p] + \text{Cov}\{E[\mu(\Theta_p, \Theta_{pj})|\Theta_p], E(X_{q'i'r'}|\Theta_p)\} \\ &= E\{\text{Cov}[\mu(\Theta_p, \Theta_{pj}), X_{q'i'r'}|\Theta_p] + \text{Cov}[\nu(\Theta_p), \delta_{pq'}\nu(\Theta_p)]\} \\ &= E\{\text{Cov}[\mu(\Theta_p, \Theta_{pj}), X_{q'i'r'}|\Theta_p] + \delta_{pq}b \end{aligned}$$

And hence

$$\begin{aligned} E\{\text{Cov}[\mu(\Theta_p, \Theta_{pj}), X_{q'i'r'}|\Theta_p]\} &= \delta_{pq'}(\delta_{i'j}a + b) - \delta_{pq'}b \\ &= \delta_{pq'}\delta_{i'j}a \end{aligned}$$

And also by Lemma 1.13, $\text{Cov}(X_{qir}, X_{q'i'r'})$

$$\begin{aligned} &= E[\text{Cov}(X_{qir}, X_{q'i'r'}|\Theta_p)] + \text{Cov}[E(X_{qir}|\Theta_p), E(X_{q'i'r'}|\Theta_p)] \\ &= E[\text{Cov}(X_{qir}, X_{q'i'r'}|\Theta_p)] + \delta_{pq}\delta_{pq'}\text{Var}[\nu(\Theta_p)] \end{aligned}$$

And hence

$$E[\text{Cov}(X_{qir}, X_{q'i'r'}|\Theta_p)] = \delta_{qq'}[\delta_{ii'}(\frac{\delta_{rr'}}{w_{qir}}s^2 + a) + b] - \delta_{pq'}\delta_{qq'}b$$

Equation (1.14) is then equal to

$$\delta_{pq'}\delta_{i'j}a = \sum_{q=1}^{P}\sum_{i=1}^{k_q}\sum_{r=1}^{t_{qi}} c_{qir}\{\delta_{qq'}[\delta_{ii'}(\frac{\delta_{rr'}}{w_{qir}}s^2 + a) + b] - \delta_{pq'}\delta_{qq'}b\} \quad (1.15)$$

It can be shown that for all other values than $p = q'$ and $i' = j$, $c_{qir} = 0$. So now let $p = q'$ and $i' = j$ to obtain

$$\begin{aligned} a &= \sum_{q=1}^{P}\sum_{i=1}^{k_q}\sum_{r=1}^{t_{qi}} c_{qir}\{\delta_{pq}[\delta_{ij}(\frac{\delta_{rr'}}{w_{qir}}s^2 + a)]\} \\ &= \sum_{i=1}^{k_p}\sum_{r=1}^{t_{pi}} c_{pir}\delta_{ij}(\frac{\delta_{rr'}}{w_{pir}}s^2 + a) \\ &= \sum_{r=1}^{t_{pj}} c_{pjr}(\frac{\delta_{rr'}}{w_{pjr}}s^2 + a) \\ &= c_{pjr'}\frac{s^2}{w_{pjr'}} + ac_{pj.} \end{aligned}$$

Now let

$$\alpha_{pj} = \frac{c_{pj1}}{w_{pj1}} = \ldots = \frac{c_{pjt}}{w_{pjt}} = \frac{c_{pj.}}{w_{pj.}}$$

Then obtain

$$\begin{aligned} a &= \alpha_{pj}s^2 + a\alpha_{pj}w_{pj.} \\ &= \alpha_{pj}(s^2 + aw_{pj.}) \end{aligned}$$

and hence

$$\alpha_{pj} = \frac{a}{s^2 + aw_{pj.}} = \frac{Z_{pj}}{w_{pj.}}$$

Then

$$c_{pjr} = \alpha_{pjr}w_{pjr} = \frac{w_{pjr}}{w_{pj.}}Z_{pj}$$

Now replace c_0

$$\begin{aligned} c_0 &= m_p - \sum_{r=1}^{t_{pj}} \frac{w_{pjr}}{w_{pj.}} Z_{pj}\mathrm{E}(X_{pjr}) \\ &= m_p(1 - Z_{pj}) \end{aligned}$$

and hence

$$\begin{aligned} \hat{\mu}(\Theta_p, \Theta_{pj}) &= (1 - Z_{pj})m_p + \sum_{r=1}^{t_{pj}} \frac{w_{pjr}}{w_{pj.}} Z_{pj}X_{pjr} \\ &= (1 - Z_{pj})m_p + Z_{pj}X_{pjw} \end{aligned}$$

□

Lemma 1.16 The following are unbiased estimators for the structural parameters m_p, m, s^2, a *and* b respectively:

$$\hat{m_p} = X_{pzw}$$

$$\hat{m} = X_{zzw}$$

$$\hat{s^2} = \frac{\sum_{p=1}^{P}\sum_{j=1}^{k_p}\sum_{r=1}^{t_{pj}} w_{pjr}(X_{pjr} - X_{pjw})^2}{\sum_{p=1}^{P}\sum_{j=1}^{k_p}(t_{pj} - 1)_+}$$

$$\hat{a} = \frac{\sum_{p=1}^{P}\sum_{j=1}^{k_p} Z_{pj}(X_{pjw} - X_{pzw})^2}{\sum_{p=1}^{P}(k_p - 1)_+}$$

$$\hat{b} = \frac{\sum_{p=1}^{P} Z_p(X_{pzw} - X_{zzw})^2}{(P-1)}$$

Proof: see Goovaerts et al.(1990), pp.168-169.

1.7 Hachemeister's Regression Model

The Hachemeister(1975) model, like Jewell's model, is a further extension of Bühlmann and Straub's model. It introduces regressors. Hachemeister relaxes the identical conditional expectation of X_{jr}'s given Θ_j's for $r = 1, ..., t$ to

$$\mathrm{E}(X_{jr}|\Theta_j) = \mu_r(\Theta_j)$$

where $\mu_r(\Theta_j) = \underline{Y}'_{jr}\underline{\beta}(\Theta_j)$ is the product of a vector $\underline{Y}_{jr}$ of regressors and a vector $\underline{\beta}(\Theta_j)$ of unknown parameters.

Assumptions:

This model departs from the i.i.d.case by allowing for different means, explained through regressors, and for different variances, which are again function of known weights as in Bühlmann & Straub's case. In addition it allows for possible covariances between observations and contracts.

(H1) For $j = 1, ..., k$ the contracts $(\Theta_j, \underline{X}_j)$ are pair-wise independent and the Θ_j's are indenpendent and identically distributed;

(H2) $\underline{\mu}(\Theta_j) = [\mu_1(\Theta_j), ..., \mu_t(\Theta_j)]'$ where

$$\begin{aligned} E(\underline{X}_j|\Theta_j) = \underline{\mu}(\Theta_j) &= \mathbf{Y}\underline{\beta}(\Theta_j) \\ &= \mathbf{Y}[\beta_1(\Theta_j), ..., \beta_n(\Theta_j)]'; \end{aligned}$$

where $\mathbf{Y}$ is a $t_{\times}n$ design matrix (of rank $n < t$) of known coefficients and $\underline{\beta}(\Theta_j)$ is a $n_{\times}1$ vector of unknown parameters.

(H3) $\text{Cov}(\underline{X}_j|\Theta_j) = \sigma^2(\Theta_j)\mathbf{V}_j$ where $\mathbf{V}_j$ is a $t \times t$ positive semi-definite matrix of known inverse weights and $\sigma^2(\Theta_j)$ a scalar function.

Structural Parameters:

(1) $s^2 = E[\sigma^2(\Theta_j)]$ is a scalar;

(2) $\mathbf{A} = \text{Cov}[\underline{\beta}(\Theta_j)]$ is a $n \times n$ matrix;

(3) $\underline{b} = E[\underline{\beta}(\Theta_j)]$ is a $n \times 1$ vector.

Notation:

$$\mathbf{U}_j = (\mathbf{Y}'\mathbf{V}_J^{-1}\mathbf{Y})^{-1} \quad \text{is a } n \times n \text{ matrix}$$

$$\mathbf{C}_j = s^2\mathbf{V}_J + \mathbf{YAY}' \quad \text{is a } n \times n \text{ matrix}$$

$$\mathbf{Z}_j = \mathbf{A}(s^2\mathbf{U}_J + \mathbf{A})^{-1} \quad \text{is a } n \times n \text{ matrix}$$

Lemma 1.17 The weighted least squares estimators are

$$\underline{\hat{\beta}}(\Theta_j) = (\mathbf{Y}'\mathbf{V}_J^{-1}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{V}_J^{-1}\underline{X}_j \quad \text{and} \quad \underline{\hat{\mu}}(\Theta_j) = \mathbf{Y}\underline{\hat{\beta}}(\Theta_j)$$

Proof: The following sum of squares has to be minimized:

$$Q_j = [\underline{X}_j - \mathbf{Y}\underline{\beta}(\Theta_j)]'\mathbf{V}_J^{-1}[\underline{X}_j - \mathbf{Y}\underline{\beta}(\Theta_j)]$$

Differentiating with respect to $\underline{\beta}(\Theta_j)$ and setting equal to 0, one obtains:

$$-2\mathbf{Y}'\mathbf{V}_J^{-1}[\underline{X}_j - \mathbf{Y}\underline{\hat{\beta}}(\Theta_j)] = 0$$

Which is equivalent to

$$\mathbf{Y}'\mathbf{V}_J^{-1}\underline{X}_j = \mathbf{Y}'\mathbf{V}_J^{-1}\mathbf{Y}\underline{\hat{\beta}}(\Theta_j)$$

And hence

$$\underline{\hat{\beta}}(\Theta_j) = (\mathbf{Y}'\mathbf{V}_\mathrm{J}^{-1}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{V}_\mathrm{J}^{-1}\underline{X}_j$$

□

A simple algebric manipulation shows that $\underline{\hat{\beta}}(\Theta_j)$ can also be rewritten as

$$\underline{\hat{\beta}}(\Theta_j) = (\mathbf{Y}'\mathbf{C}_\mathrm{J}^{-1}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{C}_\mathrm{J}^{-1}\underline{X}_j$$

Theorem 1.6 Under assumptions (H1) to (H3), the non-homogeneous linear Bayes estimator for $\underline{\beta}(\Theta_j)$ is given by

$$B_j^a = \mathbf{Z}_j\underline{\hat{\beta}}(\Theta_j) + (\mathbf{I} - \mathbf{Z}_j)\underline{b}$$

Proof: Consider

$$d(\epsilon) = \mathrm{E}\{[\underline{\beta}(\Theta_j) - B_j^a - \epsilon\underline{D}]'\mathbf{P}[\underline{\beta}(\Theta_j) - B_j^a - \epsilon\underline{D}]\}$$

where $\underline{D}$ is a $n_\times 1$ linear combination of the X_{jr}'s and $\mathbf{P}$ is a positive definite matrix. The theorem holds for $d'(0) = 0 \;\; \forall \underline{D}$. Now differentiate to obtain

$$d'(\epsilon) = -2\mathrm{E}\{\underline{D}'\mathbf{P}[\underline{\beta}(\Theta_j) - B_j^a - \epsilon\underline{D}]\}$$

Define the following centered variables

$$\underline{\beta}^0(\Theta_j) = \underline{\beta}(\Theta_j) - \underline{b}$$

$$\underline{\hat{\beta}}^0(\Theta_j) = \underline{\hat{\beta}}(\Theta_j) - \underline{b}$$

$$\underline{X}_j^0 = \underline{X}_j - \mathbf{Y}\underline{b}$$

Then substituting B_j^a by the estimator defined in the theorem, and the above centered variables in $d'(\epsilon)$ for $\epsilon = 0$, for every $\underline{D}$ we have to prove that

$$\mathrm{E}\{\underline{D}'\mathbf{P}[\underline{\beta}^0(\Theta_j) - \mathbf{Z}_j\underline{\hat{\beta}}^0(\Theta_j)]\} = 0$$

But since $\underline{D}$ is a linear combination of X_{jr}'s, it may be written as

$$\underline{D} = \underline{h}_0 + \mathbf{h}_1\underline{X}_j^0$$

where $\underline{h}_0$ is a $n_\times 1$ vector and $\mathbf{h}_1$ is an $n_\times t$ matrix. Therefore one has to prove that

$$\mathrm{E}\{[\underline{h}'_0 + (\underline{X}_j^0)'\mathbf{h}_1']\mathbf{P}[\underline{\beta}^0(\Theta_j) - \mathbf{Z}_j\underline{\hat{\beta}}^0(\Theta_j)]\} = 0$$

Because of the centered variables, the linear term disappears, so what remains of the last equation is the following expression for the left hand side:

$$\begin{aligned}
\mathrm{E}\{(\underline{X}_j^0)'\mathbf{h}'_1\mathbf{P}[\underline{\beta}^0(\Theta_j) - \mathbf{Z}_j\underline{\hat{\beta}}^0(\Theta_j)]\} &= \mathrm{E}[\mathrm{Tr}\{(\underline{X}_j^0)'\mathbf{h}'_1\mathbf{P}[\underline{\beta}^0(\Theta_j) - \mathbf{Z}_j\underline{\hat{\beta}}^0(\Theta_j)]\}] \\
&= \mathrm{E}[\mathrm{Tr}\{\mathbf{h}'_1\mathbf{P}[\underline{\beta}^0(\Theta_j) - \mathbf{Z}_j\underline{\hat{\beta}}^0(\Theta_j)](\underline{X}_j^0)'\}] \\
&= \mathrm{Tr}[\mathbf{h}'_1\mathbf{P}\mathrm{E}\{[\underline{\beta}^0(\Theta_j) - \mathbf{Z}_j\underline{\hat{\beta}}^0(\Theta_j)](\underline{X}_j^0)'\}]
\end{aligned}$$

where the fact that a scalar random variable trivially equals its trace is used and also the fact that $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$. But the last equation is proven to be equal to 0 as can be seen by:

$$\begin{aligned}
\mathrm{E}\{[\underline{\beta}^0(\Theta_j) - \mathbf{Z}_j\underline{\hat{\beta}}^0(\Theta_j)](\underline{X}_j^0)'\} &= \mathrm{E}[\underline{\beta}^0(\Theta_j)(\underline{X}_j^0)'] - \mathbf{Z}_j\mathrm{E}[\underline{\hat{\beta}}^0(\Theta_j)(\underline{X}_j^0)'] \\
&= \mathrm{Cov}[\underline{\beta}^0(\Theta_j), \underline{X}_j^0] - \mathbf{Z}_j\mathrm{Cov}[\underline{\hat{\beta}}^0(\Theta_j), \underline{X}_j^0] \\
&= \mathrm{Cov}[\underline{\beta}(\Theta_j), \underline{X}_j] - \mathbf{Z}_j\mathrm{Cov}[\underline{\hat{\beta}}^0(\Theta_j), \underline{X}_j] \\
&= \mathbf{AY}' - \mathbf{Z}_j(\mathbf{A} + s^2\mathbf{U}_j)\mathbf{Y}' \\
&= 0
\end{aligned}$$

□

Lemma 1.18 The following estimators are unbiased for the structural parameters s^2, $\mathbf{A}$ and $\underline{b}$ respectively:

$$\hat{s}^2 = \frac{1}{k}\frac{1}{(t-n)}\sum_{j=1}^{k}[\underline{X}_j - \mathbf{Y}\underline{\hat{\beta}}(\Theta_j)]'\mathbf{V}_j^{-1}[\underline{X}_j - \mathbf{Y}\underline{\hat{\beta}}(\Theta_j)]$$

$$\hat{\mathbf{A}} = \frac{1}{(k-1)}\sum_{j=1}^{k}\mathbf{Z}_j[\underline{\hat{\beta}}(\Theta_j) - \underline{\hat{b}}][\underline{\hat{\beta}}(\Theta_j) - \underline{\hat{b}}]'$$

where

$$\underline{\hat{b}} = (\sum_{j=1}^{k}\mathbf{Z}_j)^{-1}\sum_{j=1}^{k}\mathbf{Z}_j\underline{\hat{\beta}}(\Theta_j)$$

Proof: see Goovaerts et al.(1990), pp.183-185.

Chapter 2

Robust Inference

2.1 Introduction

"Robust statistics is often defined by being the statistics of approximate parametric models" [Gisler and Reinhard (1993)]. A robust estimator is an estimator that behaves well for a wide variety of underlying distributions but it is not necessarily the best for any one distribution.

Apart from sensitivity to model specification errors, estimation in parametric distributions is affected by outliers. For instance location estimators, such as the arithmetic mean, are highly sensitive to outliers. Outliers are observations which fall outside the bulk of the data. They often occur in observations found on the tail of distribution functions.

In insurance, data outliers might be caused by rare events such as hurricanes, earth quakes, etc. Outliers can also come from observation or data entry errors or noise in the model. When rating a risk portfolio, one has to look at the information given by the bulk of the data and identify outliers to treat them accordingly. In more simple words, robust statistics is a way to limit the influence of model specification error and of outlying observations on the parameter estimation.

One must remark that nonparametric statistics is different from robust statistics. Nonparametric statistics considers distributions subject to restrictions, such as a certain quantile p of a distribution, distribution symmetry or linearity of estimators. But often nonparametric methods are robust (e.g. the median estimator, Wilcoxon test).

We give here a brief introduction to robust statistical theory using Hampel's approach. Hampel et al.(1986) states three questions that one has to be able to answer in order to understand robustness:

(1) What is the effect of an outlier on a statistic (e.g. the arithmetic mean)?

(2) How much contaminated data can an estimator stand before it becomes useless?

(3) How does one assert robustness of an estimator?

2.2 M-Estimators

Consider a sample of observations $X_1, X_2, \ldots, X_n$ that are independent and identically distributed and whose distribution $G_\theta(x)$ belongs to a parametric family $\{G_\theta;\ \theta \in \Theta\}$. In classical statistics, one would assume that the observations follow a distribution of the above parametric family and estimate θ based on these sample values.

More realistically, robust theory assumes that the model $\{G_\theta;\ \theta \in \Theta\}$ is an approximation of reality, the exact distribution being in the neighborhood of the parametric family.

Let G_n be the empirical cumulative distribution function of the sample $X_1, X_2, \ldots, X_n$. By definition G_n is:

$$G_n(x) = \frac{1}{n}\sum_{i=1}^{n} \Delta_{X_i}(x)$$

where $\Delta_{X_i}(x)$ is the probability point mass 1 in x.

To estimate θ, consider the statistic $T_n = T_n(X_1, X_2, \ldots, X_n) = T(G_n)$ defined by some functional T. This restricts the estimators considered to functionals of G_n or estimators that can asymptotically be replaced by functionals, i.e. in probability

$$\lim_{n\to\infty} T_n(X_1, X_2, \ldots, X_n) \to T(G_\theta)$$

The estimators T_n considered here are assumed asymptotically normally distributed with expectation $T(G_\theta)$ and variance

$$\frac{\text{Var}(T, G_\theta)}{n}$$

i.e. in distribution

$$\sqrt{n}[T_n - T(G_\theta)] \longrightarrow N[0, \mathrm{Var}(T, G_\theta)] \text{ when } n \rightarrow \infty.$$

They are also assumed Fisher consistent [see Kallianpur and Rao(1955)], i.e.:

$$T(G_\theta) = \theta \ \forall \, \theta \in \Theta$$

Consider the statistic $T_{n-1} = T_{n-1}(X_1, X_2, \ldots, X_{n-1})$. How much does T_{n-1} change if an observation x is added to the sample $X_1, X_2, \ldots, X_{n-1}$?

Definition 2.1 The sensitivity curve of the estimator T_n, defined for $n \geq 2$ at the sample values $x_1, x_2, \ldots, x_{n-1}$ is given by

$$SC_n(x; x_1, \ldots, x_{n-1}, T_n) = n[T_n(x_1, x_2, \ldots, x_{n-1}, x) - T_{n-1}(x_1, x_2, \ldots, x_{n-1})]$$

Note that the sensitivity curve could also be defined if, say x_{n-1}, is replaced by x instead of merely adding the extra observation x.

Example: The sensitivity curve of the statistic $T_n(x_1, x_2, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$ (i.e the sample mean) is given by:

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

In general, the sensitivity curve can be complicated to compute and depends on the particular sample values on which it is calculated. A more global measure is needed.

Definition 2.2 The Influence Function (IF) of T at G_θ is given by:

$$IF(x; T, G_\theta) = \lim_{\epsilon \rightarrow 0} \frac{T[(1-\epsilon)G_\theta + \epsilon\Delta_x] - T(G_\theta)}{\epsilon}$$

$\forall x \in \mathcal{X}$ where the limit exists.

If the influence function is identically zero over some set, then the contaminated data points with values in this set have no influence on the functional. The link between the sensitivity curve and the influence function can be seen if G_θ is replaced by G_{n-1} and ϵ by $\frac{1}{n}$. The influence function then measures approximately n times the change in T caused by the additional observation in x when T is applied to a sample of size $n-1$. The influence function also measures the asymptotic bias caused by the outlier or contamination in the observations. The sensitivity curve can be rewritten as:

$$SC_n(x) = \frac{T[(1-\frac{1}{n})G_{n-1} + \frac{1}{n}\Delta_x] - T(G_{n-1})}{\frac{1}{n}}$$

In many cases $SC_n(x)$ will converge to $IF(x;T,G_\theta)$ as $n \to \infty$.

Under appropriate regularity conditions on $IF(x;T,G_\theta)$ [Hoaglin et al.(1983)], the following relations hold:

$$\int_{-\infty}^{\infty} IF(x;T,G_\theta)dG_\theta(x) = 0,$$

and

$$\mathrm{Var}(T,G_\theta) = \int_{-\infty}^{\infty} IF(x;T,G_\theta)^2 dG_\theta(x). \tag{2.1}$$

The asymptotic relative efficiency of two estimators say $\{T_n; n \geq 1\}$ and $\{S_n; n \geq 1\}$ can now be compared with the help of (2.1). One way to measure the worst effect that a contamination of fixed size can have on the value of an estimator is through the gross-error sensitivity.

Definition 2.3 The gross-error sensitivity of T at G_θ is defined by

$$\gamma^*(T,G_\theta) = \sup_x |IF(x;T,G_\theta)|$$

It is desirable that $\gamma^*(T,G_\theta)$ be finite. The bound obtained by the gross-error sensitivity can be looked at as the upper limit for the bias of an estimator. The largest influence an observation, say x_i, can have on the statistic T_n is approximatively equal to

$$\frac{\gamma^*(T,G_\theta)}{n}$$

In order to understand how large the gross-error sensitivity can be, one can compare it to the asymptotic standard deviation, i.e.:

$$\gamma^{**}(T,G_\theta) = \frac{\sup_x |IF(x;T,G_\theta)|}{\sqrt{\int_{-\infty}^{\infty} IF(x;T,G_\theta)^2 dG_\theta(x)}}$$

One can show that $\gamma^{**}(T, G_\theta)$ is always greater or equal to 1, in particular for T=median(G_θ), $\gamma^{**}(T, G_\theta) = 1$. Hence, in a certain sense, the median is the most robust statistic.

Putting a bound on $\gamma^*(T, G_\theta)$ will often conflict with the aim of asymptotic efficiency. Usually, robustifying an estimator is to try to put a bound on $\gamma^*(T, G_\theta)$.

Another function measuring the effect of changing an observation for another in the same neighborhood is the local-shift sensitivity.

Definition 2.4 The local-shift sensitivity is defined by

$$\lambda^* = \sup_{x \neq y} \frac{|IF(y; T, G_\theta) - IF(x; T, G_\theta)|}{|y - x|}$$

where $x, y \in \mathcal{X}$.

If the underlying distribution F_θ is symmetric and its center of symmetry is zero, one can define the rejection point.

Definition 2.5 The rejection point is defined by

$$\rho^* = \inf\{r > 0 \ ; \ IF(x; T, G_\theta) = 0 \ \text{ when } |x| > r\}$$

All observations beyond the rejection point ρ^* are not taken into account. Hence it is desirable that ρ^* be finite.

An estimator is resistant if contaminated data affects it in a limited way. If it is not resistant, the estimator breaks down as the proportion of contaminated data becomes too large. This remark leads to the following definition by Hampel et al.(1986):

Definition 2.6 The breakdown point of an estimator is the largest possible fraction of the observations for which there is a bound on the change in the estimate when that fraction of the sample is altered without restriction.

An estimator is resistant only if its breakdown point is greater than zero. For instance let T be the functional for the mean. If one observation is increased in value then $T_n = T(F_n)$ increases without bound. Therefore the breakdown point of the mean is 0. It can be shown that there exists no estimator that treats observations equivariantly with a breakdown bound greater than $\frac{1}{2}$.

Our goal is to find a functional T that could define an estimator T_n with bounded gross-error sensitivity. Consider the maximum likelihood estimator. Given a sample $X_1, X_2, \ldots, X_n$, it is defined to maximize

$$\prod_{i=1}^{n} f_\theta(X_i)$$

with respect to θ. This is equivalent to minimizing

$$\sum_{i=1}^{n}[-\ln f_\theta(X_i)].$$

Huber (1964) proposes to reduce this to

$$\sum_{i=1}^{n}\rho(X_i,\theta)$$

where ρ is some function on $\mathcal{R} \times \Theta$. Now suppose that the derivative of $\rho(x,\theta)$ exists and is given by

$$\psi(x,\theta) = \frac{\partial}{\partial\theta}\rho(x,\theta)$$

Hence, Huber generalized the idea to define "Maximum likelihood type" estimators:

Definition 2.7 A functional T defined implicitely by

$$\int_{-\infty}^{\infty}\psi[x,T(G)]dG(x) = 0$$

where $\psi : \mathcal{R}^2 \rightarrow \mathcal{R}$ is called an M-functional, and $T_n = T(G_n)$, an M-estimator defined implicitely by

$$\sum_{i=1}^{n}\psi(X_i,T_n) = 0 \qquad (2.2)$$

M-estimators minimize more general objective functions than the sum of squared residuals associated with the sample mean. For instance

(i) For the median, one wants to solve for t the following equation:

$$\sum_{i=1}^{n} \psi(X_i, t) = \sum_{i=1}^{n} \text{sign}(X_i - t) = 0$$

where sign() is the sign function.

(ii) For the mean, one wants to solve for t the following equation:

$$\sum_{i=1}^{n} \psi(X_i, t) = \sum_{i=1}^{n} (X_i - t) = 0$$

Theorem 2.1 Let $T_n = T(G_n)$ be an M-estimator as defined above and $IF(x; T, G_\theta)$ the influence function of T at G_θ then

$$\begin{aligned} IF(x; T, G_\theta) &= \frac{\psi[x, T(G_\theta)]}{-\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta}[\psi(x, \theta)]_{T(G_\theta)} dG_\theta(x)} \\ &= -\frac{\psi[x, T(G_\theta)]}{\frac{\partial}{\partial\theta} \int_{-\infty}^{\infty} [\psi(x, \theta)]_{T(G_\theta)} dG_\theta(x)} \end{aligned}$$

Hence it is easy to see that $IF(x; T, G_\theta)$ is directly proportional to $\psi[x, T(G_\theta)]$. For instance:

(i) if $\psi[x, t] = \text{sign}(x - t)$ then the influence function (of the median) is given by:

$$IF(x; T, G_\theta) = \frac{\text{sign}[x - T(G_\theta)]}{2 f_\theta[T(G_\theta)]}$$

which is bounded in x.

(ii) if $\psi[x, t] = x - t$, then the influence function (of the mean) is given by:

$$IF(x; T, G_\theta) = x - T(G_\theta)$$

This means that the desired shape of the influence function is achieved by an appropriate selection of the function ψ. By using equation (2.1) and Theorem 2.1, it is easy to rewrite the asymptotic variance of the M-estimator:

$$\text{Var}(T, G_\theta) = \frac{\int \psi^2[x, T(G_\theta)] dG_\theta(x)}{\{\int \frac{\partial}{\partial\theta}[\psi(y, \theta)]_{T(G_\theta)} dG_\theta(y)\}^2} \tag{2.3}$$

Corollary 2.1 Consider the cumulative distribution function G_θ and the density f_θ where $\theta \in \Theta$. Assume that $T(G_\theta)$ is Fisher consistent, i.e.

$$\int_{-\infty}^{\infty} \psi[x, T(G_\theta)] dG_\theta(x) = 0,$$

or equivalently $T(G_\theta) = \theta$. Then integration by part gives

$$IF(x; T, G_\theta) = \frac{\psi(x, \theta)}{-\int_{-\infty}^{\infty} \psi(x, \theta) s(x, \theta) dG_\theta(x)}$$

where

$$s(x, \theta) = \frac{\partial}{\partial \theta} \ln[f_\theta(x)]$$

no derivatives of $\psi(x, \theta)$ are required.

With the above result we can conclude that $\psi(x, \theta) = IF(x; T, G_\theta)$ for some Fisher consistent M-functional if and only if

$$\int_{-\infty}^{\infty} \psi[x, T(G_\theta)] dG_\theta(x) = 0$$

and

$$\int_{-\infty}^{\infty} \psi(x, \theta) s(x, \theta) dG_\theta(x) = 1$$

2.3 Optimal Robust Estimators

From the above discussion it seems reasonable to seek Fisher-consistent statistics T which minimize the asymptotic variance under the model, subject to a limit L on the gross-error sensitivity. Restated this means that one needs to find a $\psi(x, \theta)$ such that

(i) $\int_{-\infty}^{\infty} \psi(x, \theta) dG_\theta(x) = 0$

(ii) $\int_{-\infty}^{\infty} \psi(x, \theta) s(x, \theta) dG_\theta(x) = 1$

(iii) $\frac{|\psi(x,\theta)|}{\sqrt{\int_{-\infty}^{\infty} IF(x;T,G_\theta)^2 dG_\theta(x)}} \leq L$

that will minimize

$$\int_{-\infty}^{\infty} [\psi(x,\theta)]^2 dG_\theta(x)$$

Now define the following function:

$$[y]_b^c = \begin{cases} c & \text{if } y \geq c \\ y & \text{if } b \leq y \leq c \\ b & \text{if } y \leq b \end{cases}$$

Then it can be shown [see Hampel et al.(1986)] that the solution is given by

$$\psi(x,\theta) = \frac{[s(x,\theta) - a(\theta)]_{-b(\theta)}^{b(\theta)}}{M(\theta)}$$

where

$$M(\theta) = \int_{-\infty}^{\infty} [s(x,\theta) - a(\theta)]_{-b(\theta)}^{b(\theta)} s(x,\theta) dG_\theta(x)$$

and $a(\theta)$ and $b(\theta)$ are such that the other constraints are satisfied (which can be done if $L \geq 1$).

Chapter 3

Robust Credibility Models

3.1 Künsch's Model

Künsch(1992) proposes to replace the claim averages $\bar{X}_{j.}$ by a robust M-estimator of location. He truncates claims at a truncation point which depends on the data and is different for each contract. This robustifies the classical model of Bühlmann by replacing the contract sample averages by location M-estimators of Huber's type.

Consider the basic credibility model with k contracts and t time periods. As usual, the risk parameters Θ_j are unobservable and all claims $X_{jr} \geq 0, \forall r = 1, \ldots, t$.

<u>Assumptions</u>:

(K1) The contracts $(\Theta_j, \underline{X_j}) = (\Theta_j, X_{j1}, \ldots, X_{jt})$ are independent and identically distributed for $j = 1, \ldots, k$;

(K2) Θ_j is distributed according to the same U, for $j = 1, \ldots, k$;

(K3) Given Θ_j, the claims $X_{j1}, X_{j2}, \ldots, X_{jt}$ are conditionally independent and identically distributed with distribution $F_{X|\Theta}$.

Künsch distinguishes between two cases:

Case I: The distributions U and $F_{X|\Theta}$ are known;

Case II: The distributions U and $F_{X|\Theta}$ are unknown and a semi-parametric approach is used.

We will examine each case separately.

3.1.1 Case I: The Distributions U and F_{X_Θ} are known

In order to estimate $\mu(\Theta_j) = \mathrm{E}(X_{jr}|\Theta_j)$ for some $j = 1, \ldots, k$, Künsch proposes to use Bühlmann's non-homogeneous linear credibility estimator and to replace the sample averages $\bar{X}_{j.}$ by an M-estimator T_j :

$$M_j^I = m + Z[T_j - \mathrm{E}(T_j)] \tag{3.1}$$

where

$$m = \mathrm{E}[\mu(\Theta_j)] = \int_\Omega \mu(\theta) dU(\theta)$$

is the known portfolio mean and $T_j = T_j(X_{j1}, X_{j2}, \ldots, X_{jt})$ an M-estimator defined implicitly as the solution of

$$\sum_{r=1}^{t} \chi(\frac{X_{jr}}{T_j}) = 0 \tag{3.2}$$

where

$$\chi(k) = \max\{-c_1, \min(k-1, c_2)\}$$

and $0 < c_1 \leq 1$ and $c_2 > 0$. It can be shown that T_j exists and is unique [Künsch(1992), p.36].

With some algebraic manipulations equation (3.2) can be rewritten as

$$\frac{1}{t}\sum_{r=1}^{t} \tilde{\chi}(\frac{X_{jr}}{T_j}) = 1 \tag{3.3}$$

where

$$\tilde{\chi}(k) = \max\{1 - c_1, \min(k, c_2 + 1)\}$$

i.e. claims on both ends are truncated if $c_1 < 1$.

Properties of Künsch's estimator M_j^I:

(i) M_j^I is scale-equivariant, i.e. if all X_{jr} are multiplied by a constant c, then so will M_j^I;

(ii) If $c_1 = 1$ and $c_2 = \infty$, then Bühlmann's estimator is reproduced and $T_j = \frac{1}{t}\sum_{r=1}^{t} X_{jr} = \bar{X}_{j.}$;

(iii) It is unbiased, since by definition, an M-estimator is Fisher-consistent and $m = \mathrm{E}[\mu(\Theta_j)]$;

(iv) Under this pure Bayesian approach, the exact credibility premium $E[\mu(\Theta_j)|X_{j1}, X_{j2}, \ldots, X_{jt}]$ can be computed and is optimal. In most cases it will be non-linear but possibly robust. To ensure linearity, Künsch suggests instead to use a non-robust optimal linear approximation to the estimator and to robustify it. However, after robustification, this destroys the optimality property.

To obtain an estimate of T_j, (3.3) suggests the following iterative algorithmic formulation

$$T_j^{(n+1)} = [\frac{1}{t}\sum_{r=1}^{t}\tilde{\chi}(\frac{X_{jr}}{T_j^{(n)}})]^{\frac{1}{2}} T_j^{(n)} \quad \text{for } n \geq 0, \tag{3.4}$$

with a robust starting value of $T_j^{(0)} = \text{median}\{X_{j1}, X_{j2}, \ldots, X_{jt}\}$. The convergence of this algorithm can be found in Huber(1981), section 8.6.

To obtain the credibility factor Z in M_j^I, the variance of T_j is needed. With the help of Hampel et al.(1986), Chapter 2 and the influence function, as defined in Theorem 2.1, consider the linearization of T_j:

$$T_j(X_{j1}, X_{j2}, \ldots, X_{jt}) = T_j(F_{X|\Theta_j}) + \frac{1}{t}\sum_{r=1}^{t} IF(X_{jr}; T_j, F_{X|\Theta_j}) + \vartheta_p(\frac{1}{\sqrt{t}}) \tag{3.5}$$

where $\vartheta_p(\frac{1}{\sqrt{t}})$ is a remainder term vanishing to 0 at the stated rate. $T_j(F_{X|\Theta_j})$ is defined implicitely as the solution of

$$\int \chi[\frac{x}{T_j(F_{X|\Theta_j})}] dF_{X|\Theta_j}(x) = 0 \tag{3.6}$$

and by Theorem 2.1, the influence function (IF) is given by

$$IF(x; T_j, F_{X|\Theta_j}) = \chi[\frac{x}{T_j(F_{X|\Theta_j})}] T_j^2(F_{X|\Theta_j}) M(\Theta_j)^{-1}$$

where

$$M(\Theta_j) = \int \chi'[\frac{x}{T_j(F_{X|\Theta_j})}] x dF_{X|\Theta_j}(x) = \int_{(1-c_1)T_j(F_{X|\Theta_j})}^{(1+c_2)T_j(F_{X|\Theta_j})} x dF_{X|\Theta_j}(x)$$

is a normalization constant. If we calculate the expectation of (3.5), since $E[IF(X_{jr}; T_j, F_{X|\Theta_j})|\Theta_j] = 0$, by (3.6), we get that $E[T_j(X_{j1}, X_{j2}, \ldots, X_{jt})|\Theta_j] \approx$

$T_j(F_{X|\Theta_j})$ and we can now calculate the variance

$$\begin{aligned}\text{Var}[T_j(X_{j1}, X_{j2}, \ldots, X_{jt})|\Theta_j] &\cong \frac{1}{t}\text{E}[IF(X_{jr}; T_j, F_{X|\Theta_j})^2|\Theta_j] \\ &= \frac{1}{t}\text{E}\{\chi^2[\frac{X_{jr}}{T_j(F_{X|\Theta_j})}]\}T_j^4(F_{X|\Theta_j})M^{-2}(\Theta_j)\end{aligned}$$

The credibility factor Z is straightforward and is given in the following theorem:

Theorem 3.1 Under assumptions (K1) to (K3), the non-homogeneous linear credibility estimator is given by:

$$M_j^I = \hat{\mu}(\Theta_j) = m + Z[T_j - \text{E}(T_j)]$$

where

$$Z = \frac{\text{Cov}[\text{E}(T_j|\Theta_j), \mu(\Theta_j)]}{\text{E}[\text{Var}(T_j|\Theta_j)] + \text{Var}[\text{E}(T_j|\Theta_j)]} \tag{3.7}$$

Proof: we want to minimize the sum of squares

$$\begin{aligned}\text{E}\{[M_j^I - \mu(\Theta_j)]^2\} &= \text{E}\{[m + Z[T_j - \text{E}(T_j)] - \mu(\Theta_j)]^2\} \\ &= \text{Var}[\mu(\Theta_j)] + Z^2\{\text{E}[T_j - \text{E}(T_j)]^2\} \\ &\quad -2Z\text{E}\{[\mu(\Theta_j) - m][T_j - \text{E}(T_j)]\}\end{aligned}$$

with respect to Z. Differentiate with respect to Z and set equal to 0 to obtain:

$$Z\text{E}\{[T_j - \text{E}(T_j)]^2\} = \text{Cov}[T_j, \mu(\Theta_j)]$$

Hence

$$Z = \frac{\text{Cov}[T_j, \mu(\Theta_j)]}{\text{Var}(T_j)}$$

which is equivalent to

$$Z = \frac{\text{Cov}[T_j, \mu(\Theta_j)]}{\text{Var}[\text{E}(T_j|\Theta_j)] + \text{E}[\text{Var}(T_j|\Theta_j)]}$$

Now

$$\begin{aligned}\text{Cov}[T_j, \mu(\Theta_j)] &= \text{Cov}\{\text{E}(T_j|\Theta_j), \text{E}[\mu(\Theta_j)|\Theta_j]\} + \text{E}\{\text{Cov}[T_j, \mu(\Theta_j)|\Theta_j]\} \\ &= \text{Cov}[\text{E}(T_j|\Theta_j), \mu(\Theta_j)]\end{aligned}$$

since $E\{Cov[\mu(\Theta_j), T_j|\Theta_j]\}$ is equal to 0. We hence obtain a robust version of Bühlmann's credibility factor:

$$Z = \frac{Cov[E(T_j|\Theta_j), \mu(\Theta_j)]}{E[Var(T_j|\Theta_j)] + Var[E(T_j|\Theta_j)]} \tag{3.8}$$

The exact value of Z can be calculated if U and $F_{X|\Theta}$ are known.

3.1.2 Case II: The Distributions U and F_{X_Θ} are unknown

In Case II, since the portfolio mean m and $E(T_j)$ are unknown, we replace the sample means by averages

$$M_j^{II} = \bar{X}_{..} + Z[T_j - \bar{T}_.] \tag{3.9}$$

where

$$\bar{X}_{..} = \frac{1}{kt}\sum_{j=1}^{k}\sum_{r=1}^{t} X_{jr} \text{ and } \bar{T}_. = \frac{1}{k}\sum_{j=1}^{k} T_j \quad .$$

Again $T_j = T_j(X_{j1}, X_{j2}, \ldots, X_{jt})$ is an M-estimator defined implicitly as the solution of

$$\sum_{r=1}^{t} \chi(\frac{X_{jr}}{T_j}) = 0 \tag{3.10}$$

where

$$\chi(k) = \max\{-c_1, \min(k-1, c_2)\}$$

and $0 < c_1 \leq 1$ and $0 < c_2$. It can be shown that T_j exists and is unique [Künsch(1992), p.36]. Rewrite (3.10) to obtain

$$\frac{1}{t}\sum_{r=1}^{t} \tilde{\chi}(\frac{X_{jr}}{T_j}) = 1 \tag{3.11}$$

where

$$\tilde{\chi}(k) = \max\{1 - c_1, \min(k, c_2 + 1)\}$$

i.e. claims on both ends are truncated if $c_1 < 1$.

Properties of Künsch's estimator M_j^{II}:

(i) M_j^{II} is scale-equivariant, i.e. if all X_{jr} are multiplied by a constant c, then so will M_j^{II};

(ii) If $c_1 = 1$ and $c_2 = \infty$, then Bühlmann's estimator is reproduced and $T_j = \bar{X}_{j.} = \frac{1}{t}\sum_{r=1}^{t} X_{jr}$;

(iii) To achieve unbiasedness, the non-robust estimator $\bar{X}_{..}$ is used for the portfolio mean. From a robust point of view, the credibility premium $(1-Z)\bar{T}_{.} + ZT_j$ is preferable, but the latter is biased.

As in Theorem 3.1, an optimal credibility factor Z is given by Künsch:

Theorem 3.2 Under assumptions (K1) to (K3), the non-homogeneous linear credibility estimator is given by:

$$M_j^{II} = \hat{\mu}(\Theta_j) = \bar{X}_{..} + Z[T_j - \bar{T}_{.}]$$

where

$$Z = \frac{\text{Cov}[\text{E}(T_j|\Theta_j), \mu(\Theta_j)]}{\text{E}[\text{Var}(T_j|\Theta_j)] + \text{Var}[\text{E}(T_j|\Theta_j)]} \tag{3.12}$$

Proof: To minimize the sum of squares

$$\begin{aligned}\text{E}\{[M_j^{II} - \mu(\Theta_j)]^2\} &= \text{E}\{[\bar{X}_{..} + Z(T_j - \bar{T}_{.}) - \mu(\Theta_j)]^2\} \\ &= \text{E}\{\bar{X}_{..}^2 + 2Z\bar{X}_{..}(T_j - \bar{T}_{.}) - 2\bar{X}_{..}\mu(\Theta_j) \\ &\quad + Z^2(T_j - \bar{T}_{.})^2 + \mu^2(\Theta_j) - 2Z(T_j - \bar{T}_{.})\mu(\Theta_j)\}\end{aligned}$$

with respect to Z, differentiate with respect to Z and set equal to 0 to obtain:

$$Z\text{E}[(T_j - \bar{T}_{.})^2] = \text{E}\{(T_j - \bar{T}_{.})[\mu(\Theta_j) - \bar{X}_{..}]\}$$

and hence

$$\begin{aligned}Z &= \frac{\text{E}\{(T_j - \bar{T}_{.})[\mu(\Theta_j) - \bar{X}_{..}]\}}{\text{E}[(T_j - \bar{T}_{.})^2]} \\ &= \frac{\text{Cov}[T_j, \mu(\Theta_j)]}{\text{Var}(T_j)} \\ &= \frac{\text{Cov}[\text{E}(T_j|\Theta_j), \mu(\Theta_j)]}{\text{E}[\text{Var}(T_j|\Theta_j)] + \text{Var}[\text{E}(T_j|\Theta_j)]}\end{aligned} \tag{3.13}$$

Since the distributions U and $F_{X|\Theta}$ are unknown, the numerator and denominator of (3.13) need to be estimated. The denominator $\mathrm{E}[(T_j - \bar{T}_.)^2]$ can be estimated by (see Künsch(1992), p.40):

$$\frac{1}{(k-1)}\sum_{j=1}^{k}(T_j - \bar{T}_.)^2 \tag{3.14}$$

Now to estimate the numerator of (3.13), by definition we have

$$\mathrm{Cov}[\mathrm{E}(T_j|\Theta_j), \mu(\Theta_j)] = \mathrm{Cov}[\mathrm{E}(T_j|\Theta_j), \mathrm{E}(\bar{X}_{j.}|\Theta_j)]$$

and also

$$\mathrm{Cov}(T_j, \bar{X}_{j.}) = \mathrm{E}[\mathrm{Cov}(T_j, \bar{X}_{j.}|\Theta_j)] + \mathrm{Cov}[\mathrm{E}(T_j|\Theta_j), \mathrm{E}(\bar{X}_{j.}|\Theta_j)]$$

It is then easy to see that

$$\mathrm{Cov}[\mathrm{E}(T_j|\Theta_j), \mathrm{E}(\bar{X}_{j.}|\Theta_j)] = \mathrm{Cov}(T_j, \bar{X}_{j.}) - \mathrm{E}[\mathrm{Cov}(T_j, \bar{X}_{j.}|\Theta_j)]$$

where an estimator for $\mathrm{Cov}(T_j, \bar{X}_{j.})$ is given by

$$\frac{\sum_{j=1}^{k}(T_j - \bar{T}_.)(\bar{X}_{j.} - \bar{X}_{..})}{(k-1)}$$

An estimator for $\mathrm{E}[\mathrm{Cov}(T_j, \bar{X}_{j.}|\Theta_j)]$ can also be given using (3.5):

$$\frac{1}{kt(t-1)}\sum_{j=1}^{k}\sum_{r=1}^{t}\hat{IF}(X_{jr}; T_j)(X_{jr} - \bar{X}_{j.})$$

where the empirical influence function is given by

$$\hat{IF}(X_{jr}; T_j) = \frac{t\chi(\frac{X_{jr}}{T_j})T_j^2}{\sum_{r=1}^{t} X_{jr} I_{[(1-c_1)T_j \le X_{jr} \le (1+c_2)T_j]}}$$

The empirical credibility estimator for $M_j^{II} = \hat{\mu}(\Theta_j)$ is then obtained by replacing the parameter Z in (3.12) by a robust sample version:

$$\hat{Z} = \frac{\frac{1}{k-1}\sum_{j=1}^{k}(T_j - \bar{T}_.)(\bar{X}_{j.} - \bar{X}_{..}) - \frac{1}{kt(t-1)}\sum_{j=1}^{k}\sum_{r=1}^{t}\hat{IF}(X_{jr}; T_j)(X_{jr} - \bar{X}_{j.})}{\frac{1}{k-1}\sum_{j=1}^{k}(T_j - \bar{T}_.)^2}$$

Remark: $\hat{Z}$ is not optimal in the sense of Theorem 3.2 since the parameters in the numerator and denominator are replaced by estimated values. Also note that instead of the estimator of $\mathrm{Var}(T_j)$ in (3.14), a more consistent robust estimation would use [see Künsch(1992), p.40]:

$$\frac{1}{(t-1)}\sum_{r=1}^{t} \hat{IF}(X_{jr};T_j)^2.$$

3.2 Gisler & Reinhard's Model

3.2.1 Definitions

Künsch's model uses robust location estimators $T_j = T_j(X_{j1},\ldots,X_{jt})$ instead of the usual contract averages $\bar{X}_{j.}$. These estimators perform reasonably well in the neighborhood of the true model when Bayesian credibility is exact. Gisler and Reinhard(1993) propose to divide the pure risk premium into two components: an ordinary part for average claims, and an excess part for outlying claims, which can be estimated separately. We will also see later that it is possible to include weights in their model. Formally:

$$\mu_x(\Theta_j) = \mu_0(\Theta_j) + \mu_{xs}(\Theta_j)$$

where $\mu_x(\Theta_j) = \mathrm{E}(X_{jr}|\Theta_j)$, $\mu_0(\Theta_j)$ is the ordinary part and $\mu_{xs}(\Theta_j)$ is the excess part. The ordinary part $\mu_0(\Theta_j)$ is the expected loss-ratio generated by the claim load of ordinary losses, whereas the excess-part $\mu_{xs}(\Theta_j)$ is the additional expected claims load generated mainly by extraordinary events such as big fires, hurricanes, etc. The excess-part is the part that usually generates outlier observations and hence affects the outlier ratio in a dramatic way.

To estimate the ordinary part $\mu_0(\Theta_j)$, credibility and robust statistics are combined, i.e. a credibility estimator based on a robust stastistic $T_j = T_j(X_{j1},\ldots,X_{jt})$, $j = 1,\ldots,k$ is used. By definition

$$\mu_0(\Theta_j) = \mathrm{E}[T_j|\Theta_j].$$

All risks in the portfolio are assumed equally exposed to outliers events. This can be expressed in the following manner:

$$\mu_{xs}(\Theta_j) = \mu_{xs} \quad \forall j = 1,\ldots,k.$$

Note that if an a priori assumption can be made to establish how certain risks are more exposed to outlier events than others, then one can always define a known matrix $\mathbf{A}_{k\times 1}$ such that $\mu_{xs}(\Theta_j) = \mathbf{A}\mu_{xs}$.

The robust credibility estimator of $\mu_x(\Theta_j)$ is thus given by

$$\hat{\mu}_x(\Theta_j) = \hat{\mu}_0(\Theta_j) + \mu_{xs} \tag{3.15}$$

where $\hat{\mu}_0(\Theta_j)$ is estimated by standard robust techniques, without regard to bias:

$$\begin{aligned} \hat{\mu}_0(\Theta_j) &= \mathrm{E}(T_j) + Z_j[T_j - \mathrm{E}(T_j)] && (3.16) \\ &= \mu_{T_j} + Z_j[T_j - \mu_{T_j}] && (3.17) \end{aligned}$$

where

$$Z_j = \frac{\mathrm{Var}[\mathrm{E}(T_j|\Theta_j)]}{\mathrm{E}[\mathrm{Var}(T_j|\Theta_j)] + \mathrm{Var}[\mathrm{E}(T_j|\Theta_j)]} = \frac{\mathrm{Var}[\mu_{T_j}(\Theta_j)]}{\mathrm{E}[\mathrm{Var}(T_j|\Theta_j)] + \mathrm{Var}[\mu_{T_j}(\Theta_j)]}$$

Note that $\mu_{T_j} = \mathrm{E}(T_j)$ and $\mu_{T_j}(\Theta_j) = \mathrm{E}(T_j|\Theta_j)$. The subscript T_j is used to emphasize the fact that the estimations are based on T_j and not on the X_{jr}'s. To summarize, the main differences between Künsch's model and Gisler and Reinhard's model are:

(i) The pure risk premium $\mu_x(\Theta_j) = \mathrm{E}(X_{jr}|\Theta_j)$ is divided into two components: an ordinary part $\mu_0(\Theta_j)$ and an excess part μ_{xs}, which is assumed constant $\forall j = 1, \ldots, k$;

(ii) No use of $\bar{X}_{..}$ is made in (3.16), disregarding the bias. Only T_j and $\mathrm{E}(T_j)$ are used;

(iii) An introduction of different weights is allowed, giving a more representative description of reality.

3.2.2 Weighted model with identical volumes

To simplify the mathematical derivations, consider Bühlmann & Straub's model, as defined in Chapter 1, but for identical volumes, i.e. k contracts and t time periods with equal weights $\forall j = 1, \ldots, k$ and $\forall r = 1, \ldots, t$. As usual the risk parameters Θ_j are unobservable, all claims $X_{jr} \geq 0$ and

(BS1) The contracts $j = 1, ..., k$ [i.e. the pair vectors $(\Theta_j, \underline{X}_j)$] are independent and the variables Θ_j's are identically distributed;

(BS2) $\forall r, s = 1, ..., t$ and $\forall j = 1, ..., k$,

$$E(X_{jr}|\Theta_j) = \mu(\Theta_j)$$

$$\mathrm{Cov}(X_{jr}, X_{js}|\Theta_j) = \frac{\delta_{rs}}{w}\sigma^2(\Theta_j)$$

where the w are known weights (in this case all identical), δ_{rs} is Kroneckers' symbol and where $\mu(\Theta_j)$ and $\sigma^2(\Theta_j)$ are unknown functions.

Now consider the M-estimator defined implicitly by

$$\sum_{r=1}^{t} \psi(X_{jr}, T_j) = 0 \tag{3.18}$$

where

$$\psi(x, \theta) = \phi(\frac{x}{\theta}) \tag{3.19}$$

as in a scale model, with

$$\phi(x) = \min(x - 1, 1) \tag{3.20}$$

The M-estimator is thus defined implicitely by

$$\sum_{r=1}^{t} \min(\frac{X_{jr}}{T_j} - 1, 1) = 0 \tag{3.21}$$

which can be rewritten as

$$T_j = \frac{1}{t}\sum_{r=1}^{t} \min(X_{jr}, 2T_j) \tag{3.22}$$

An algorithmic solution to (3.22) can be given to calculate T_j.

Let d be the estimator of T. We use the letter d instead of t, in order not to confuse it with t, the number of observations in contract j. We hence rewrite (3.22) as

$$f(d) = \frac{1}{t}\sum_{r=1}^{t} \min(X_{jr}, 2d)$$

Consider the order statisitics $X_{j(r)}, r = 1, \ldots, t$ of $X_{j1}, \ldots, X_{jt}$. Let l_j be the number of observations equal to 0 in contract j. Therefore there are $t - l_j$ observations that are greater than 0 in this contract j. Now let $X_{j(l_j+1)}$ be the first observation not equal to 0. We are looking for $f'(d)$, the derivative of $f(d)$.

Distinguish between three cases:

1. $X_{j(l_j+1)} \geq 2d$, then $f'(d) = \frac{2(t-l_j)}{t}$
2. $X_{j(t)} < 2d$, then all observations are smaller than $2d$ and therefore $f'(d) = 0$.
3. $X_{j(m)} < 2d \leq X_{j(m+1)}$, for some $m = l_j + 1, \ldots, t-1$. Then we get $f'(d) = \frac{2(t-m)}{t}$

Hence

$$f'(d) = \begin{cases} \frac{2(t-l_j)}{t} & \text{if } X_{j(l_j+1)} \geq 2d \\ \frac{2(t-m)}{t} & \text{if } X_{j(m)} < 2d \leq X_{j(m+1)} \\ 0 & \text{if } X_{j(t)} < 2d \end{cases}$$

We obtain T_j with the following procedure: Calculate $T_j^{(r)} = f[\frac{X_{j(r)}}{2}]$ for $r = t, t-1, \ldots$ until $T_j^{(r)} > \frac{X_{j(r)}}{2}$ and let m_j be the first index for which this inequality is fulfilled. If $m_j \geq 1$ exists then

$$T_j = \frac{\sum_{j=1}^{m_j} X_{j(r)}}{2m_j - t}, \quad \text{otherwise } T_j = 0$$

Note that if $X_{j(t)} \leq 2\bar{X}_{j.}$, then $T_j = \bar{X}_{j.}$ and if half or more of the observations X_{jr} are zero, then $T_j = 0$.

Now to find the empirical credibility estimator, the structural parameters of Z_j in (3.16) need to be estimated. In general there is no explicit formula for $\mu_{T_j}(\Theta_j)$ and for $\mathrm{Var}(T_j|\Theta_j)$; $\mu_{T_j}(\Theta_j)$ is replaced by the asymptotic expectation of $T_j(F_{X|\Theta_j})$ as in Definition 2.7, and $\mathrm{Var}(T_j|\Theta_j)$ by t^{-1} times the asymptotic variance $\mathrm{Var}(T_j, F_{X|\Theta_j})$ as in (2.3). Therefore we obtain the following asymptotic non-homogeneous linear credibility estimator

$$\hat{\mu}_0(\Theta_j) \cong \mu_T + Z_j[T_j - \mu_T]$$

where

$$Z_j = \frac{ta_T}{ta_T + s_T^2}$$

and where

$$a_T = \text{Var}[T_j(F_{X|\Theta_j})]$$
$$s_T^2 = \text{E}[\text{Var}(T_j, F_{X|\Theta_j})]$$
$$\mu_T = \text{E}[\mu_{T_j}(\Theta_j)] = \text{E}(T_j)$$

To complete the estimation of $\mu_x(\Theta_j)$, estimators of the unknown structural parameters μ_{xs}, μ_T, a_T and s_T^2 are needed. The M-estimator T_j can be rewritten as

$$T_j = \frac{1}{t}\sum_{r=1}^{t} T_{jr} \quad \text{with} \quad T_{jr} = \min(X_{jr}, 2T_j).$$

We see from this expression that all losses included in an interval $[0, 2T_j]$ will not be truncated and can be considered as ordinary losses. For convenience, denote by

$$\begin{cases} T_{jr} & \text{the ordinary portion of a claim amount} \\ XS_{jr} = X_{jr} - T_{jr} & \text{the excess portion of a claim amount} \end{cases}$$

Note that the random variables T_{jr}, $j = 1, \ldots, k$ and $r = 1, \ldots, t$, are not conditionally independent given Θ_j. Therefore an estimator of the asymptotic variance $\text{Var}(T_j, F_{X|\Theta_j})$ is needed. Replace $F_{X|\Theta_j}$ by the empirical distribution of the $X_{jr}, r = 1, \ldots, t, j = 1, \ldots, k$ in (2.3). After some straightforward calculations and a change of normalizing constant from t^{-1} to $(t-1)^{-1}$ we get

$$\hat{s}_j^2 = \frac{\frac{1}{t-1}\sum_{r=1}^{t}(T_{jr} - T_j)^2}{(1 - \frac{2}{t}\sum_{r=1}^{t} I_{[X_{jr} > 2T_j]})^2} \tag{3.23}$$

where $I_{[X_{jr} > 2T_j]}$ is an indicator function. Note that the denomimator in (3.23) is equal to 1 in the case where all $X_{jr} \le 2T_j$, i.e. in the case where $T_j = \bar{X}_j$. In summary, the various estimators for the identical volumes case of the Bühlmann & Straub model can be written as:

$$\hat{\mu}_T = \frac{1}{k}\sum_{j=1}^{k} T_j$$

$$\hat{\mu}_{xs} = \frac{1}{k}\sum_{j=1}^{k} \bar{X}S_{j.} = \frac{1}{k}\frac{1}{t}\sum_{j=1}^{k}\sum_{r=1}^{t} XS_{jr}$$

$$\hat{s}_T^2 = \frac{1}{k}\sum_{j=1}^{k} \hat{s}_j^2$$

$$\hat{a}_T = \sum_{j=1}^{k}[\frac{(T_j - \hat{\mu}_T)^2}{(k-1)} - \frac{1}{kt}\hat{s}_T^2]$$

Hence the empirical robust credibility estimator is given by

$$\hat{\mu}_x(\Theta_j) = \hat{\mu}_{xs} + \hat{\mu}_T + \hat{Z}_j[T_j - \hat{\mu}_T] \quad (3.24)$$

where

$$\hat{Z}_j = \frac{t\hat{a}_T}{t\hat{a}_T + \hat{s}_T^2}$$

3.2.3 Weighted model with different volumes

The estimators derived in the previous section can be extended to allow for different weights w_{jr}. Define new random variables

$$X_{jr} = \frac{1}{w_{jr}}\sum_{\nu=1}^{w_{jr}} Y_{jr}^{(\nu)}$$

where the $Y_{jr}^{(\nu)}$ fulfill the conditions of section 3.2.2 with identical volumes $w_{jr} = 1$. The random variables X_{jr} are averages of w_{jr} independent (although unobservable) random variables $Y_{jr}^{(\nu)}$. Now replace the unobservable $Y_{jr}^{(\nu)}$, $(\nu = 1, \ldots, w_{jr})$ by the observed averages X_{jr} and use the fact that our M-estimator is scale invariant. Insert X_{jr} into (3.18), the M-estimator definition, to give

$$\sum_{r=1}^{t} w_{jr}\psi(X_{jr}, T_j) = 0.$$

In section 3.2.2 we had noted that, by the definition of the M-estimator, all observations belonging to the interval $[0, 2T_j]$ could be considered ordinary losses. In the generalization of the model with different volumes, this interval should now depend on some function w_{jr}, for instance $[0,(1+f(w_{jr}))T_j]$. This yields a new definition corresponding to (3.21):

$$\sum_{r=1}^{t} \frac{w_{jr}}{t\bar{w}_{j.}} \min[\frac{X_{jr}}{T_j} - 1, f(w_{jr})] = 0; \quad \bar{w}_{j.} = \frac{1}{t}\sum_{r=1}^{t} w_{jr}$$

Since $\sqrt{\text{Var}(X_{jr}|\Theta_j)} = \frac{1}{\sqrt{w_{jr}}}\sigma(\Theta_j)$, an optimal choice of function f is

$$f(w_{jr}) = \frac{1}{\sqrt{w_{jr}}}c$$

where c is a suitably chosen constant. Gisler and Reinhard suggest the following natural choices for the constant c:

$$c_1 = \sqrt{\bar{w}_{..}} \quad \text{with } \bar{w}_{..} = \frac{1}{k}\sum_{j=1}^{k} \bar{w}_{j.} = \frac{1}{kt}\sum_{j=1}^{k}\sum_{r=1}^{t} w_{jr}$$

or

$$c_2 = \sqrt{\text{median}(w_{jr})} \quad (j = 1,\ldots,k; r = 1,\ldots,t)$$

They also recommend to use constant c_1 except when the w_{jr}'s have a very sweked distribution, when constant c_2 would be prefered. The M-estimator can thus be rewritten as

$$\sum_{r=1}^{t} w_{jr}\min(\frac{X_{jr}}{T_j} - 1, cw_{jr}^{-1/2}) = 0$$

After algebraic manipulation, we find that this implicit definition is equivalent to

$$T_j = \sum_{r=1}^{t} \frac{w_{jr}}{t\bar{w}_{j.}} \min(X_{jr}, c_{jr}T_j)$$

where

$$c_{jr} = 1 + cw_{jr}^{-1/2}$$

Again, an algorithmic formulation can be used to calculate T_j. Let d be the estimator of T. The letter d is used not to confuse it with t, the number of claims. Consider the following function

$$f(d) = \sum_{r=1}^{t} \frac{w_{jr}}{t\bar{w}_{j.}} c_{jr}\min(Z_{jr}, d)$$

and $Z_{j(r)}, r = 1, \ldots, t$, the order statisitics of $Z_{jr} = \frac{X_{jr}}{c_{jr}}$, with $w_{j(r)}$ and $c_{j(r)}$, the corresponding order statistics of w_{jr} and c_{jr}. Let l_j be the number of observations equal to 0 in contract j and $Z_{j(l_j+1)}$ the first observation that is greater than zero. We want to find $f'(d)$. Consider the following cases:

1. $Z_{j(l_j+1)} \geq d$. Then there are $(t - l_j)$ observations fulfilling this inequality. Hence we obtain $f'(d) = \sum_{r=l_j+1}^{t} \frac{w_{jr}}{t\bar{w}_{j.}} c_{j(r)}$

2. $Z_{j(t)} < d$. All observations are smaller than d. Then $f'(d) = 0$

3. $Z_{j(m)} < d \leq Z_{j(m+1)}$, for some $m = l_{j+1}, \ldots, t-1$. Then $f'(d) = \sum_{r=m+1}^{t} \frac{w_{j(r)}}{t\bar{w}_{j.}} c_{j(r)}$

Hence

$$f'(d) = \begin{cases} \sum_{r=l_j+1}^{t} \frac{w_{jr}}{t\bar{w}_{j.}} c_{j(r)} & \text{if } Z_{j(l_j+1)} \geq d \\ \sum_{r=m+1}^{t} \frac{w_{j(r)}}{t\bar{w}_{j.}} c_{j(r)} & \text{if } Z_{j(m)} < d \leq Z_{j(m+1)} \\ 0 & \text{if } Z_{j(t)} < d \end{cases}$$

We obtain T_j with the following procedure: Calculate $T_j^{(r)} = f(Z_{j(r)})$ for $r = t, t-1, \ldots$ until $T_j^{(r)} > Z_{j(r)}$ and let m_j be the first index for which this inequality is fulfilled. If $m_j \geq 1$ exists then

$$T_j = \frac{\sum_{r=1}^{m_j} w_{j(r)} c_{j(r)} Z_{j(r)}}{t\bar{w}_{j.} - \sum_{r=m_j+1}^{t} c_{j(r)} w_{j(r)}}, \quad \text{otherwise } T_j = 0$$

Note that $Z_{j(k)} \leq \bar{X}_{j.}$ then $T_j = \bar{X}_{j.} = \frac{1}{t\bar{w}_{j.}} \sum_{r=1}^{t} w_{jr} X_{jr}$. If $\sum_{r=l_j+1}^{t} w_{j(r)} c_{j(r)} \leq t\bar{w}_{j.}$ then $T_j = 0$.

To find the empirical credibility estimator $\hat{\mu}_0(\Theta_j)$, we have again to estimate the structural parameters in (3.16). Because the distribution $F_{X|\Theta_j}(x)$ as well as the ψ-function depend on w_{jr}, Gisler and Reinhard argue that a strict mathematical treatment becomes unfeasible. They believe that the proposed estimators are reasonable and useful for practical purposes.

With the modification of the ψ-function, they assume that $\mathrm{E}(T_j|\Theta_j)$ is approximately independent of the underlying volumes w_{jr}. $\mathrm{E}(T_j|\Theta_j)$ is approximated by $T_j(F_{X|\Theta_j})$, the asymptotic expectation for a risk with identical volumes ($w_{jr} \equiv 1$) as defined in section 3.2.2. As for the within variance

$\mathrm{Var}(T_j|\Theta_j)$, it is assumed small in comparison to the between variance and hence

$$w_{jr} = \tilde{w}_j, \text{ for } j = 1, \ldots, k \text{ and } r = 1, \ldots, t.$$

is used in

$$\mathrm{Var}(X_{jr}|\Theta_j) \cong \frac{\sigma^2(\Theta_j)}{\tilde{w}_j} \tag{3.25}$$

Furthermore, assume that

$$\mathrm{Var}(T_j|\Theta_j) \cong \frac{\mathrm{Var}(T_j, F_{X|\Theta_j})}{t\tilde{w}_j} \tag{3.26}$$

where $\mathrm{Var}(T_j, F_{X|\Theta_j})$ is the asymptotic contract variance with $w_{jr} \equiv 1$. Hence we obtain for $\hat{\mu}_0(\Theta_j)$

$$\hat{\mu}_0(\Theta_j) \approx \mu_T + Z_j[T_j - \mu_T]$$

where

$$Z_j = \frac{t\bar{w}_{j.}a_T}{t\bar{w}_{j.}a_T + s_T^2}$$

and

$$\bar{w}_{j.} = \frac{1}{t}\sum_{r=1}^{t} w_{jr}$$

where again

$$\begin{aligned} T_j(F_{X|\Theta_j}) &= \text{asymptotic expectation for risks with volumes } w_{jr} \equiv 1 \\ s_T^2 &= \mathrm{E}[\mathrm{Var}(T_j, F_{X|\Theta_j})] \\ \mathrm{Var}(T_j, F_{X|\Theta_j}) &= \text{asymptotic variance for risks with volumes } w_{jr} \equiv 1 \\ a_T &= \mathrm{Var}[T_j(F_{X|\Theta_j})] \end{aligned}$$

Now to estimate the structural parameters μ_{xs}, μ_T, a_T and s_T^2, consider

$$T_j = \sum_{r=1}^{t} \frac{w_{jr}}{t\bar{w}_{j.}} T_{jr} \quad \text{with } T_{jr} = \min(X_{jr}, c_{jr}T_j)$$

For convenience, denote by

$$\begin{cases} T_{jr} & \text{the observed ordinary claim} \\ XS_{jr} = X_{jr} - T_{jr} & \text{the observed excess claim} \end{cases}$$

and

$$\begin{cases} w_{jr}T_{jr} & \text{the ordinary claim total} \\ w_{jr}XS_{jr} = w_{jr}(X_{jr} - T_{jr}) & \text{the excess claim total} \end{cases}$$

Now by use of the empirical distribution function of the X_{jr}, $(r = 1, \ldots, t)$, we find that:

$$\hat{I}F(X_{jr}; T_j) = \frac{T_{jr} - T_j}{1 - \sum_{r=1}^{t} \frac{w_{jr}}{t\bar{w}_{j.}} c_{jr} I_{[T_{jr} \neq X_{jr}]}}$$

By (3.25) and (3.26) suggested variance estimators are

$$\hat{s}_j^2 = \frac{\frac{1}{t-1} \sum_{r=1}^{t} w_{jr}(T_{jr} - T_j)^2}{(1 - \sum_{r=1}^{t} \frac{w_{jr}}{t\bar{w}_{j.}} c_{jr} I_{[T_{jr} \neq X_{jr}]})^2}$$

This implies

$$\hat{s}_T^2 = \frac{1}{k} \sum_{j=1}^{k} \hat{s}_j^2$$

and

$$\hat{a}_T = \frac{1}{h} [\sum_{j=1}^{k} \frac{\bar{w}_{j.}}{k\bar{w}_{..}} (T_j - \bar{T}_.)^2 - (k-1) \frac{\hat{s}_T^2}{kt\bar{w}_{..}}]$$

where

$$\bar{T}_. = \sum_{j=1}^{k} \frac{\bar{w}_{j.}}{k\bar{w}_{..}} T_j$$

$$h = \sum_{j=1}^{k} \frac{\bar{w}_{j.}}{k\bar{w}_{..}} (1 - \frac{\bar{w}_{j.}}{k\bar{w}_{..}})$$

$$\hat{\mu}_T = \frac{\sum_{j=1}^{k} \hat{Z}_j T_j}{\sum_{j=1}^{k} \hat{Z}_j}$$

Hence

$$\hat{Z}_j = \frac{t\bar{w}_{j.}\hat{a}_T}{t\bar{w}_{j.}\hat{a}_T + \hat{s}_T^2}$$

$$\hat{\mu}_{xs} = \frac{1}{k\bar{w}_{..}} \sum_{j=1}^{k} \bar{w}_{j.} \bar{X}S_{j.}$$

where

$$\bar{X}S_{j.} = \sum_{r=1}^{t} \frac{w_{jr}}{t\bar{w}_{j.}} XS_{jr}$$

Thus the empirical robust credibility formula in the case of different volumes is given by

$$\hat{\hat{\mu}}_X(\Theta_j) = \hat{\mu}_{xs} + \hat{\mu}_T + \hat{Z}_j[T_j - \hat{\mu}_T]. \qquad (3.27)$$

Chapter 4

Robustification of Jewell's Hierarchical Model

4.1 Introduction

The new proposed model is a direct extension of Gisler and Reinhard's model as described in Chapter 3. Gisler and Reinhard introduce robust inference in Bühlmann and Straub's credibility model. We take it a step further when we allow robust estimation in Jewell's hierarchical credibility model. The first two subsections will give the robust credibility estimator at the contract level with and without weights. The last subsection hints at the robustification of the estimator at the subportfolio level.

As with Gisler and Reinhard's model, we propose to divide the pure risk premium into an ordinary part and an excess part, which will be estimated separately. Formally we write:

$$\mu_x(\Theta_p, \Theta_{pj}) = \mu_0(\Theta_p, \Theta_{pj}) + \mu_{xs}(\Theta_p, \Theta_{pj})$$

where $\mu_x(\Theta_p, \Theta_{pj}) = \mathrm{E}(X_{pjr}|\Theta_p, \Theta_{pj})$, $\mu_0(\Theta_p, \Theta_{pj})$ is the ordinary part and $\mu_{xs}(\Theta_p, \Theta_{pj})$ the excess part.

To estimate the ordinary part $\mu_0(\Theta_p, \Theta_{pj})$, we use a robust statistics $T_{pj} = T_{pj}(X_{pj1}, \ldots, X_{pjt})$, $j = 1, \ldots, k_p$ and $t = 1, \ldots, t_{p_j}$.

By definition,

$$\mu_0(\Theta_p, \Theta_{pj}) = \mathrm{E}[T_{pj}|\Theta_p, \Theta_{pj}].$$

All risks are again assumed equally exposed to outlier events, i.e.

$$\mu_{xs}(\Theta_p, \Theta_{pj}) = \mu_{xs}$$

The robust credibility estimator of $\mu_X(\Theta_p, \Theta_{pj})$ is then given by

$$\hat{\mu}_X(\Theta_p, \Theta_{pj}) = \hat{\mu}_0(\Theta_p, \Theta_{pj}) + \hat{\mu}_{xs} \tag{4.1}$$

where $\hat{\mu}_0(\Theta_p, \Theta_{pj})$ will be estimated through standard robust techniques, without regard to bias:

$$\begin{aligned} \hat{\mu}_0(\Theta_p, \Theta_{pj}) &= \mathrm{E}(T_{pj}) + Z_{pj}[T_{pj} - \mathrm{E}(T_{pj})] \\ &= \mu_{T_p} + Z_{pj}(T_{pj} - \mu_{T_p}) \end{aligned} \tag{4.2}$$

where

$$\begin{aligned} Z_{pj} &= \frac{\mathrm{Var}[\mathrm{E}(T_{pj}|\Theta_p, \Theta_{pj})]}{\mathrm{E}[\mathrm{Var}(T_{pj}|\Theta_p, \Theta_{pj})] + \mathrm{Var}[\mathrm{E}(T_{pj}|\Theta_p, \Theta_{pj})]} \\ &= \frac{\mathrm{Var}[\mu_{T_p}(\Theta_p, \Theta_{pj})]}{\mathrm{E}[\mathrm{Var}(T_{pj}|\Theta_p, \Theta_{pj})] + \mathrm{Var}[\mu_{T_p}(\Theta_p, \Theta_{pj})]} \end{aligned} \tag{4.3}$$

and

$$\mu_{T_p} = \mathrm{E}(T_{pj})$$
$$\mu_{T_p}(\Theta_p, \Theta_{pj}) = \mathrm{E}(T_{pj}|\Theta_p, \Theta_{pj})$$

4.2 Robust Estimator at the Contract Level with identical weights

Consider Jewell's hierarchical model such that:

(1) Each subportfolio is given a structural variable Θ_p where $p = 1, 2, \ldots, P$;

(2) Each contract within subportfolio p is given structural variables (Θ_p, Θ_{pj}) for $j = 1, \ldots, k_p$;

(3) Each year, within a given contract j, we observe a claim X_{pjr}, $r = 1, \ldots, t_{pj}$, possibly paired to a known weight w_{pjr} given in advance. Here we assume that $w_{pjr} = w, \forall p = 1, \ldots, P, \forall j = 1, \ldots, k_p$, and $\forall t = 1, \ldots, t_{pj}$.

The data for subportfolio p is the set of variables $(\Theta_p, \Theta_{pj}, X_{pjr})$ while for the contract pj it is defined by the set (Θ_{pj}, X_{pjr}).

Assumptions

(J1) The subportfolios $p = 1, \dots, P$ [i.e. the pairs $(\Theta_p, \Theta_{pj}, X_{pjr})$] are independent $\forall p \neq p'$;

(J2) For each $p = 1, \dots, P$, the contracts $pj = p1, \dots, pk_p$ [i.e. the pairs (Θ_{pj}, X_{pjr})] are conditionally independent given Θ_p;

(J3) $\forall p = 1, \dots, P$ and $\forall j = 1, \dots, k_p$, the claims $X_{pj1}, \dots, X_{pjt_{pj}}$ are conditionally independent given (Θ_p, Θ_{pj});

(J4) All pairs of variables (Θ_p, Θ_{pj}), for $p = 1, \dots, P$ and $j = 1, \dots, k_p$, are identically distributed;

(J5) $\forall p, j$ and r

$$\mathrm{E}(X_{pjr}|\Theta_p, \Theta_{pj}) = \mu(\Theta_p, \Theta_{pj}) \quad \forall r = 1, \dots, t_{pj}$$

and

$$\mathrm{Var}(X_{pjr}|\Theta_p, \Theta_{pj}) = \frac{1}{w}\sigma^2(\Theta_p, \Theta_{pj}) \quad \forall r = 1, \dots, t_{pj}$$

where μ and σ^2 do not depend on the subscripts p, j and r and w are known equal weights.

Also define

$$\nu(\Theta_p) = \mathrm{E}[\mu(\Theta_p, \Theta_{pj})|\Theta_p] = \mathrm{E}(X_{pjr}|\Theta_p)$$

Consider the M-estimator defined implicitly by

$$\sum_{r=1}^{t_{pj}} \chi(\frac{X_{pjr}}{T_{pj}}) = 0 \tag{4.4}$$

where

$$\chi(k) = \max\{-c_1, \min(k-1, c_2)\}$$

and $0 < c_1 \leq 1$ and $c_2 > 0$. It can be shown that T_{pj} exists and is unique [Künsch(1992), p.36].

For simplicity purposes, instead of using the M-estimator proposed by Gisler and Reinhard, we chose to use the one proposed by Künsch with $c_1 = 1$. The choice of the M-estimator defines the truncation point.

Rewrite (4.4)

$$\sum_{r=1}^{t_{pj}} \min(\frac{X_{pjr}}{T_{pj}} - 1, c_2) = 0 \tag{4.5}$$

After algebraic manipulations, equation (4.5) can be rewritten as

$$\frac{1}{t_{pj}} \sum_{r=1}^{t_{pj}} \min(\frac{X_{pjr}}{T_{pj}}, c_2 + 1) = 1 \tag{4.6}$$

The M-estimator is now implicitely defined by

$$T_{pj} = \frac{1}{t_{pj}} \sum_{r=1}^{t_{pj}} \min[X_{pjr}, (c_2 + 1)T_{pj}] \tag{4.7}$$

An algorithmic solution to (4.7) can be given to calculate T_{pj} as in section 3.1.1.

We now need to find estimators for the structural parameters in order to calculate Z_{pj}. In general, we assume that there are no explicit formulas for $\mu_{T_p}(\Theta_p, \Theta_{pj}) = \mathrm{E}[T_{pj}|\Theta_p, \Theta_{pj}]$ and for $\mathrm{Var}(T_{pj}|\Theta_p, \Theta_{pj})$. $\mu_{T_p}(\Theta_p, \Theta_{pj})$ is replaced by the asymptotic expectation of $T_{pj}(F_{X|\Theta_p,\Theta_{pj}})$ and $\mathrm{Var}(T_{pj}|\Theta_p, \Theta_{pj})$ by $(t_{pj})^{-1}$ times the asymptotic variance $\mathrm{Var}(T_{pj}, F_{X|\Theta_p,\Theta_{pj}})$. We therefore obtain the following asymptotic non-homogeneous linear credibility estimator

$$\hat{\mu}_0(\Theta_p, \Theta_{pj}) \approx \mu_{T_p} + Z_{pj}(T_{pj} - \mu_{T_p}) \tag{4.8}$$

where

$$Z_{pj} = \frac{t_{pj} a_T}{t_{pj} a_T + s_T^2}$$

$$a_T = \mathrm{Var}\{\mathrm{E}[T_{pj}(F_{X|\Theta_p,\Theta_{pj}})]\}$$

$$s_T^2 = \mathrm{E}[\mathrm{Var}(T_{pj}, F_{X|\Theta_p,\Theta_{pj}})]$$

$$\mu_{T_p} = \mathrm{E}(T_{pj})$$

We use the subscript T to emphasize the fact that a and s^2 will be estimated from T_{pj} values instead of using the raw X_{pjr}.

We now need estimators for the unknown structural parameters μ_{xs}, μ_{T_p}, a_T and s_T^2. The M-estimator T_{pj} can be rewritten as

$$T_{pj} = \frac{1}{t_{pj}} \sum_{r=1}^{t_{pj}} T_{pjr} \quad \text{with } T_{pjr} = \min[X_{pjr}, (1 + c_2)T_{pj}].$$

From this last definition, we see that all losses included in the intervall $[0, (1+c_2)T_{pj}]$ are considered ordinary losses and will not be truncated. Note that if $c_2 = 1$, then we obtain the same intervall as with Gisler and Reinhard's model. Denote by

$$\begin{cases} T_{pjr} & \text{the ordinary loss ratio} \\ XS_{pjr} = X_{pjr} - T_{pjr} & \text{the excess loss ratio} \end{cases}$$

Note that again the random variables T_{pjr}, $j = 1, \ldots, k_p$, $t = 1, \ldots, t_{pj}$ are not conditionally independent given (Θ_p, Θ_{pj}).

We need an estimator of the asymptotic variance $\text{Var}[(T_{pj}, F_{X|\Theta_p,\Theta_{pj}})]$. We replace $F_{X|\Theta_p,\Theta_{pj}}$ by the empirical distribution. After some calculations and a change of normalizing constant from (t_{pj}) to $(t_{pj} - 1)^{-1}$, we get

$$\hat{s}_{pj}^2 = \frac{\frac{1}{t_{pj}-1} \sum_{r=1}^{t_{pj}} (T_{pjr} - T_{pj})^2}{(1 - \frac{(c_2+1)}{t_{pj}} \sum_{r=1}^{t_{pj}} I_{[X_{pjr} > (1+c_2)T_{pj}]})^2} \tag{4.9}$$

where $I_{[X_{pjr} > (1+c_2)T_{pj}]}$ is an indicator function. Note that the denominator is equal to 1 when all $X_{pjr} \leq (c_2 + 1)T_{pj}$, i.e. when $T_{pj} = \bar{X}_{pj\cdot}$.

$$\hat{\mu}_{T_p} = \frac{1}{k_p} \sum_{j=1}^{k_p} T_{pj}$$

$$\hat{\mu}_{p-xs} = \frac{1}{k_p} \sum_{j=1}^{k_p} \bar{X}S_{pj.} = \frac{1}{k_p} \frac{1}{t_{pj}} \sum_{j=1}^{k_p} \sum_{r=1}^{t_{pj}} XS_{pjr}$$

$$\hat{s}_T^2 = \frac{1}{P} \sum_{p=1}^{P} \sum_{j=1}^{k_p} \frac{1}{k_p} \hat{s}_{pj}^2$$

$$\hat{a}_{T_p} = \sum_{j=1}^{k_p}[\frac{(T_{pj} - \hat{\mu}_{T_p})^2}{(k_p - 1)} - \frac{\hat{s}_T^2}{k_p t_{pj}}]$$

and then

$$\hat{a}_T = \frac{1}{P}\sum_{p=1}^{P}\hat{a}_{T_p} = \frac{1}{P}\sum_{p=1}^{P}\sum_{j=1}^{k_p}[\frac{(T_{pj} - \hat{\mu}_{T_p})^2}{(k_p - 1)} - \frac{\hat{s}_T^2}{k_p t_{pj}}]$$

Hence the empirical robust credibility estimator at the contract level is given by

$$\hat{\hat{\mu}}_x(\Theta_p, \Theta_{pj}) = \hat{\mu}_{p-xs} + \hat{\mu}_{T_p} + \hat{Z}_{pj}(T_{pj} - \hat{\mu}_{T_p}) \qquad (4.10)$$

where

$$\hat{Z}_{pj} = \frac{t_{pj}\hat{a}_T}{t_{pj}\hat{a}_T + \hat{s}_T^2}$$

4.3 Robust Estimator at the Contract Level with different weights

As with Gisler and Reinhard's model, the estimators derived in the previous sections can be extended to allow for different weights w_{pjr}. Define the following new random variables

$$X_{pjr} = \frac{1}{w_{pjr}}\sum_{n=1}^{w_{pjr}} Y_{pjr}^{(n)}$$

where the $Y_{pjr}^{(n)}$ fulfill the conditions of section 4.2 with identical volumes $w_{pjr} = 1$. The random variables X_{pjr} can be interpreted as averages of w_{pjr} independent (although unobservable) random variables $Y_{pjr}^{(n)}$. Now replace the unobservable $Y_{pjr}^{(n)}$, $(n = 1, \ldots, w_{pjr})$ by the observed averages X_{pjr} and use the fact that our M-estimator is scale invariant. Inserting X_{pjr} into (4.4), the M-estimator definition, to give

$$\sum_{r=1}^{t_{pj}} w_{pjr}\chi(\frac{X_{pjr}}{T_{pj}}) = 0 \qquad (4.11)$$

where

$$\chi(k) = \max\{-c_1, \min(k-1, c_2)\}$$

and $0 < c_1 \leq 1$ and $c_2 > 0$.

Again use $c_1 = 1$, the corresponding equation to (4.5) is

$$\sum_{r=1}^{t_{pj}} w_{pjr} \min(\frac{X_{pjr}}{T_{pj}} - 1, c_2) = 0$$

In section 4.2, by the definition of the M-estimator, all observations belonging to the interval $[0, (c_2+1)T_{pj}]$ could be considered ordinary losses. In the generalization of the model with different volumes, this interval should now depend on some function of the different volumes w_{pjr}, for instance $\{0, [1 + c_2 f(w_{pjr})]T_{pj}\}$. This yields a new definition corresponding to (4.6):

$$\sum_{r=1}^{t_{pj}} \frac{w_{pjr}}{t_{pj}\bar{w}_{pj.}} \min[\frac{X_{pjr}}{T_{pj}} - 1, c_2 f(w_{pjr})] = 0; \quad \bar{w}_{pj.} = \frac{1}{t_{pj}} \sum_{r=1}^{t_{pj}} w_{pjr}$$

Since $\sqrt{\text{Var}(X_{pjr}|\Theta_p, \Theta_{pj})} = \frac{1}{\sqrt{w_{pjr}}}\sigma(\Theta_p, \Theta_{pj})$, an optimal choice of function f is

$$f(w_{pjr}) = \frac{1}{\sqrt{w_{pjr}}} c$$

where c is a suitably chosen constant. As with Gisler and Reinhard's model, we suggest the same following natural choices for the constant c:

$$c_1 = \sqrt{\bar{w}_{p..}} \quad \text{with } \bar{w}_{p..} = \sum_{j=1}^{k_p} \sum_{r=1}^{t_{pj}} \frac{w_{pjr}}{k_p t_{pj}}$$

or

$$c_2 = \sqrt{\text{median}(w_{pjr})} \quad (j = 1, \ldots, k_p; r = 1, \ldots, t_{pj})$$

We assume the same position as Gisler and Reinhard: to use constant c_1 except when the w_{pjr}'s have a very sweked distribution, where constant c_2 would then be prefered. The M-estimator can thus be rewritten as

$$\sum_{r=1}^{t_{pj}} w_{pjr} \min(\frac{X_{pjr}}{T_{pj}} - 1, c_2 c w_{pjr}^{-1/2}) = 0$$

After algebraic manipulations, we find that this implicit definition is equivalent to

$$T_{pj} = \sum_{r=1}^{t_{pj}} \frac{w_{pjr}}{t_{pj}\bar{w}_{pj.}} \min(X_{pjr}, c_{pjr}T_{pj})$$

where

$$c_{pjr} = 1 + c_2 c w_{pjr}^{-1/2}$$

and

$$\bar{w}_{pj.} = \frac{1}{t_{pj}} \sum_{r=1}^{t_{pj}} w_{pjr}.$$

Again, an algorithmic formulation can used to calculate T_{pj}. To find the empirical credibility estimator $\hat{\mu}_0(\Theta_p, \Theta_{pj})$, we have again to estimate the structural parameters in (4.8). Because the distribution $F_{X|\Theta_p,\Theta_{pj}}(x)$ as well as the χ-function depend on w_{pjr}, we use the same argumentation as Gisler and Reinhard: that a strict mathematical treatment becomes unfeasible. Let us assume that the proposed estimators are reasonable and useful for practical purposes.

With the modification of the χ-function, assume that $E(T_{pj}|\Theta_p, \Theta_{pj})$ is approximately independent of the underlying volumes w_{pjr}. $E(T_{pj}|\Theta_p, \Theta_{pj})$ is approximated by $T_{pj}(F_{X|\Theta_p,\Theta_{pj}})$, the asymptotic expectation for a risk with identical volumes ($w_{pjr} \equiv 1$) as defined in section 4.2. As for the within variance $\mathrm{Var}(T_{pj}|\Theta_p, \Theta_{pj})$, it is assumed small in comparison to between variance and hence

$$w_{pjr} = \tilde{w}_{pj}, \text{ for } j = 1, \ldots, k_p \text{ and } r = 1, \ldots, t_{pj}.$$

is used in

$$\mathrm{Var}(X_{pjr}|\Theta_p, \Theta_{pj}) = \frac{\sigma^2(\Theta_p, \Theta_{pj})}{\tilde{w}_{pj}} \tag{4.12}$$

Furthermore, assume that

$$\mathrm{Var}(T_{pj}|\Theta_p, \Theta_{pj}) \cong \frac{\mathrm{Var}(T_{pj}, F_{X|\Theta_p,\Theta_{pj}})}{t_{pj}\tilde{w}_{pj}} \tag{4.13}$$

where $\mathrm{Var}(T_{pj}, F_{X|\Theta_p,\Theta_{pj}})$ is the asymptotic contract variance with $w_{pjr} = 1$. Hence we obtain for $\hat{\mu}_0(\Theta_p, \Theta_{pj})$

$$\hat{\mu}_0(\Theta_p, \Theta_{pj}) \approx \mu_{T_p} + Z_{pj}[T_{pj} - \mu_{T_p}] \tag{4.14}$$

where

$$Z_{pj} = \frac{t_{pj}\bar{w}_{pj.}a_T}{t_{pj}\bar{w}_{pj.}a_T + s_T^2}$$

and again

$$\begin{aligned}
T_{pj}(F_{X|\Theta_p,\Theta_{pj}}) &= \text{asymptotic expectation for risks with volumes } w_{pjr} \equiv 1 \\
s_T^2 &= \mathrm{E}[\mathrm{Var}(T_{pj}, F_{X|\Theta_p,\Theta_{pj}})] \\
\mathrm{Var}(T_{pj}, F_{X|\Theta_p,\Theta_{pj}}) &= \text{asymptotic variance for risks with volumes } w_{pjr} \equiv 1 \\
a_T &= \mathrm{Var}[T_{pj}(F_{X|\Theta_p,\Theta_{pj}})]
\end{aligned}$$

Now to estimate the structural parameters $\mu_{p-xs}, \mu_{T_p}, a_T$ and s_T^2, consider

$$T_{pj} = \sum_{r=1}^{t_{pj}} \frac{w_{pjr}}{t_{pj}\bar{w}_{pj.}} T_{pjr} \quad \text{with } T_{pjr} = \min(X_{pjr}, c_{pjr}T_{pj})$$

For convenience, denote by

$$\begin{cases} T_{pjr} & \text{the ordinary claim} \\ XS_{pjr} = X_{pjr} - T_{pjr} & \text{the excess claim} \end{cases}$$

$$\begin{cases} w_{pjr}T_{pjr} & \text{the ordinary claim total} \\ w_{pjr}XS_{pjr} = w_{pjr}(X_{pjr} - T_{pjr}) & \text{the excess claim total} \end{cases}$$

Now by use of the empirical distribution function of the X_{pjr}, $(p = 1, \ldots, P, j = 1, \ldots, k_p, r = 1, \ldots, t_{pj})$, we find that:

$$\hat{I}F(X_{pjr}; T_{pj}) = \frac{T_{pjr} - T_{pj}}{1 - \sum_{r=1}^{t_{pj}} \frac{w_{pjr}}{t_{pj}\bar{w}_{pj.}} c_{pjr} I_{[T_{pjr} \neq X_{pjr}]}}$$

By (4.10) and (4.17) suggested variance estimators are

$$\hat{s}_{pj}^2 = \frac{\frac{1}{t_{pj}-1}\sum_{r=1}^{t_{pj}} w_{pjr}(T_{pjr} - T_{pj})^2}{1 - \sum_{r=1}^{t_{pj}} \frac{w_{pjr}}{t_{pj}\bar{w}_{pj.}} c_{pjr} I_{[T_{pjr} \neq X_{pjr}]}}$$

This implies

$$\hat{s}_T^2 = \frac{1}{P}\sum_{p=1}^{P} \hat{s}_{T_p} = \frac{1}{P}\sum_{p=1}^{P}\sum_{j=1}^{k_p} \frac{\hat{s}_{pj}^2}{k_p}$$

and

$$\hat{a}_{T_p} = \frac{1}{h}\{\sum_{j=1}^{k_p}[\frac{\bar{w}_{pj.}}{k_p\bar{w}_{p..}}(T_{pj} - \bar{T}_{p.})^2 - \frac{(k_p-1)}{k_p^2 t_{pj}\bar{w}_{p..}}\hat{s}_T^2]\}$$

where

$$\bar{w}_{p..} = \frac{1}{k_p}\sum_{j=1}^{k_p}\bar{w}_{pj.}$$

$$h = \sum_{j=1}^{k_p}\frac{\bar{w}_{pj.}}{k_p\bar{w}_{p..}}(1 - \frac{\bar{w}_{pj.}}{k_p\bar{w}_{p..}})$$

$$\bar{T}_{p.} = \sum_{j=1}^{k_p}\frac{\bar{w}_{pj.}}{k_p\bar{w}_{p..}}T_{pj}$$

hence

$$a_T = \frac{1}{P}\sum_{p=1}^{P}a_{T_p} = \frac{1}{Ph}\sum_{p=1}^{P}\{\sum_{j=1}^{k_p}[\frac{\bar{w}_{pj.}}{k_p\bar{w}_{p..}}(T_{pj} - \bar{T}_{p.})^2 - \frac{(k_p-1)}{t_{pj}\bar{w}_{p..}}\hat{s}_T^2]\}$$

$$\hat{\mu}_{T_p} = \frac{\sum_{j=1}^{k_p}\hat{Z}_{pj}T_{pj}}{\sum_{j=1}^{k_p}\hat{Z}_{pj}}$$

$$\hat{\mu}_{p-xs} = \frac{1}{k_p\bar{w}_{p..}}\sum_{j=1}^{k_p}\bar{w}_{pj.}\bar{X}S_{pj.} = \frac{1}{k_p\bar{w}_{p..}}\sum_{j=1}^{k_p}\sum_{r=1}^{t_{pj}}\frac{w_{pjr}.XS_{pjr}}{t_{pj}}$$

Thus the empirical robust credibility formula in the case of different volumes is given by

$$\hat{\mu}_x(\Theta_p, \Theta_{pj}) = \hat{\mu}_{p-xs} + \hat{\mu}_{T_p} + \hat{Z}_{pj}[T_{pj} - \hat{\mu}_{T_p}]. \qquad (4.15)$$

where

$$\hat{Z}_{pj} = \frac{t_{pj}\bar{w}_{pj.}\hat{a}_T}{t_{pj}\bar{w}_{pj.}\hat{a}_T + \hat{s}_T^2}$$

4.4 Robust Estimator at the Subportfolio Level With Identical Weights

In Jewell's hierarchical model, the credibility estimator at the subportfolio level $\nu(\Theta_p)$ is given by

$$\hat{\nu}(\Theta_p) = m + Z_p(X_{pzm} - m)$$

In our proposed model, the robustified estimator at the subportfolio level will be given by

$$\hat{\nu}(\Theta_p) = m + Z_p[T_p - \mathrm{E}(T_p)] \tag{4.16}$$

where

$$\begin{aligned} Z_p &= \frac{\mathrm{Var}[\mathrm{E}(T_p|\Theta_p)]}{\mathrm{Var}[\mathrm{E}(T_p|\Theta_p)] + \mathrm{E}[\mathrm{Var}(T_p|\Theta_p)]} \\ &= \frac{\mathrm{Var}(\mu_{T_p})}{\mathrm{Var}(\mu_{T_p}) + \mathrm{E}[\mathrm{Var}(T_p|\Theta_p)]} \end{aligned} \tag{4.17}$$

and where $\mu_{T_p} = \mathrm{E}(T_p|\Theta_p)$.

At the subportfolio level, we use m instead of $\mathrm{E}(T_p)$. In the insurance industry, one the properties that is highly regarded is the unbiasness of an estimator. At the subportfolio level, there is a lot of data and using m instead of $\mathrm{E}(T_p)$ will help to attain the unbiasness objective.

Now consider for T_p the following M-estimator defined implicitely by

$$\sum_{j=1}^{k_p}\sum_{r=1}^{t_{pj}} \chi(X_{pjr}, T_p) = 0$$

The χ-function is defined in the same manner as in section 4.2.

We rewrite this M-estimator to obtain

$$T_p = \frac{1}{k_p}\sum_{j=1}^{k_p}\sum_{r=1}^{t_{pj}} \frac{1}{t_{pj}} \min(X_{pjr}, 2T_p)$$

We again assume that there are no explicit formula for $\mathrm{E}(T_p|\Theta_p) = \mu_{T_p}$ and $\mathrm{Var}(T_p|\Theta_p)$. $\mathrm{E}(T_p|\Theta_p)$ is replaced by the asymptotic expectation of $T_p(F_{X|\Theta_p})$ and $\mathrm{Var}(T_p|\Theta_p)$ by $(k_p t_{pj})^{-1}$ times the asymptotic variance $\mathrm{Var}(T_p, F_{X|\Theta_p})$. We therefore obtain the following asymptotic non-homogeneous linear credibility estimator

$$\hat{\nu}(\Theta_p) \approx m + \tilde{Z}_p[T_p - \mathrm{E}(T_p)] \tag{4.18}$$

where

$$\tilde{Z}_p = \frac{b_T Z_{p.}}{a_T + b_T Z_{p.}}$$

and

$$Z_{p.} = \sum_{j=1}^{k_p} Z_{pj}$$

$$b_T = \mathrm{Var}\{\mathrm{E}[T_p(F_{X|\Theta_p})]\}$$

$$a_T = \mathrm{E}[\mathrm{Var}(T_p, F_{X|\Theta_p})]$$

Again, one has to be careful to distinguish between $Z_{p.}$ and Z_p.

We stop here as we believe that obtaining robust estimators at the contract level is sufficient when the data set is important enough. A robust estimator at the subportfolio level can be obtained with the robust estimators at the contract level. A suggested way would be using an average of the robust estimators at the contract level.

Chapter 5

Results and Conclusions

This chapter is divided in the following manner: we first start by presenting the data set used to illustrate some of the models presented in the previous chapters. The following section consists of a brief summary of each of the models and the results obtained for each of these models with the illustrative data set. Finally the third and last section consists of a global comparison of the results and some suggestions on how to use the models, the limitations of the models and the data sets they may apply to.

5.1 Presentation of the Data Set

The illustrative data set we use is the well known Hachemeister data set (see Goovaerts et al.(1987), pp.31-32). Hachemeister considered five different states and twelve quarters of claim experience. This experience consists of average claim amounts for total private passenger bodily injury insurance from July 1970 until June 1973. Hence we have $k = 5$ contracts and $t = 12$ periods. We also remark that since the Hachemeister data set consists of claim averages, all averages are greater than zero. To illustrate the limitations of the models and the effects of outliers, we will contaminate the data for the last period of state 5 (contract $j = 5$) and observe the differences in the estimation of the parameters.

Table 5.1: Hachemeister Claim Data Set

Claims	j=1	j=2	j=3	j=4	j=5
$r = 1$	1,738	1,364	1,759	1,223	1,456
2	1,642	1,408	1,685	1,146	1,499
3	1,794	1,597	1,479	1,010	1,609
4	2,051	1,444	1,763	1,257	1,741
5	2,079	1,342	1,674	1,426	1,482
6	2,234	1,675	2,103	1,532	1,572
7	2,032	1,470	1,502	1,953	1,606
8	2,035	1,448	1,622	1,123	1,735
9	2,115	1,464	1,828	1,343	1,607
10	2,262	1,831	2,155	1,243	1,573
11	2,267	1,612	2,233	1,762	1,613
12	2,517	1,471	2,059	1,306	1,690

For each of the claim averages above, there is an associated weight. These weights reflect the number of claims corresponding to these averages.

Table 5.2: Hachemeister Associated Weights

Weights	j=1	j=2	j=3	j=4	j=5
$r = 1$	7,861	1,622	1,147	407	2,902
2	9,251	1,742	1,357	396	3,172
3	8,706	1,523	1,329	348	3,046
4	8,575	1,515	1,204	341	3,068
5	7,917	1,622	998	315	2,693
6	8,263	1,602	1,077	328	2,910
7	9,456	1,964	1,277	352	3,275
8	8,003	1,515	1,218	331	2,697
9	7,365	1,527	896	287	2,663
10	7,832	1,748	1,003	384	3,017
11	7,849	1,654	1,108	321	3,242
12	9,077	1,861	1,121	342	3,425

Observe that the number of claims in state 1 is rather high compare to other states. The same observation applies but to a lesser extent, in state 5.

5.2 Results

5.2.1 Bühlmann's Classical Model

Consider a portfolio consisting of k contracts described by $(\Theta_j, \underline{X}_j)$ with $t = 12$ periods, $\forall j = 1, \ldots, 5$. We recall that Bühlmann's credibility estimation of $\mu(\Theta_j)$ is given in Theorem 1.2:

$$\hat{\mu}(\Theta_j) = m + Z(\bar{X}_{j.} - m)$$

Note that m is estimated by $\bar{X}_{..}$ and Z by the empirical credibility estimator in Lemma 1.3.

The results obtained with Hachemeister's data set are:

Table 5.3: Bühlmann's Premiums

Outlier	1,690	5,000	6,000	7,000
$\bar{X}_{1.}$	2,064	2,064	2,064	2,064
$\bar{X}_{2.}$	1,511	1,511	1,511	1,511
$\bar{X}_{3.}$	1,822	1,822	1,822	1,822
$\bar{X}_{4.}$	1,360	1,360	1,360	1,360
$\bar{X}_{5.}$	1,599	1,874	1,958	2,041
$\hat{Z}$	0.9496	0.7546	0.6533	0.5521
$\hat{\mu}(\Theta_1)$	2,044	1,981	1,953	1,928
$\hat{\mu}(\Theta_2)$	1,519	1,563	1,591	1,622
$\hat{\mu}(\Theta_3)$	1,814	1,798	1,794	1,794
$\hat{\mu}(\Theta_4)$	1,376	1,450	1,493	1,539
$\hat{\mu}(\Theta_5)$	1,602	1,838	1,883	1,915
$\bar{X}_{..}$	1,671	1,726	1,743	1,760

Note that as the outlier value increases from 1,690 to 7,000, the estimated credibility factor $\hat{Z}$ decreases from 0.9496 to 0.5521.

5.2.2 Künsch's Model

Since Künsch's model is a robustified version of Bühlmann's classical model, we now compare the results obtained under his model to those above. Here the estimator of $\mu(\Theta_j)$ is given in Theorem 3.2 by

$$\hat{\mu}(\Theta_j) = \bar{X}_{..} + Z(T_j - \bar{T}_.)$$

We recall that the only differences between Bühlmann's classical model and Künsch's model are that m is unknown and that $\bar{X}_{j.}$ is replaced by the M-estimator T_j. For the derivation of the M-estimator T_j, Künsch recommends to use $c_1 = c_2 = 1$ for small samples (Künsch(1992), p.39). c_2 being the upper truncation point, we illustrate different choices of c_2 to observe its effect. We use $c_1 = 1$ since our portfolio of data consists of claim averages and hence they are all greater than zero.

Table 5.4: Künsch's Premiums

$\mathbf{c_1 = 1, c_2 = 1}$

Outlier	1,690	5,000	6,000	7,000
T_1	2,064	2,064	2,064	2,064
T_2	1,511	1,511	1,511	1,511
T_3	1,822	1,822	1,822	1,822
T_4	1,360	1,360	1,360	1,360
T_5	1,599	1,749	1,749	1,749
$\hat{Z}$	0.9496	0.8247	0.7958	0.7668
$\hat{\mu}(\Theta_1)$	2,044	2,025	2,031	2,038
$\hat{\mu}(\Theta_2)$	1,519	1,569	1,591	1,613
$\hat{\mu}(\Theta_3)$	1,814	1,826	1,839	1,852
$\hat{\mu}(\Theta_4)$	1,376	1,445	1,472	1,498
$\hat{\mu}(\Theta_5)$	1,602	1,766	1,781	1,796
$T_.$	1,671	1,701	1,701	1,701

When the claim data for state 5 is not contaminated (i.e. when outlier value is equal to 1,690), we obtain the same results as with Bühlmann's classical model.

However, as the outlier value increases, we note that with Künsch's truncated means, the estimation of the credibility factor decreases more slowly than with Bühlmann's model. Also note that the M-estimator for state 5 (T_5) becomes constant and hence so does the average of all the M-estimators ($T_.$). The resulting estimators of $\mu(\Theta_j)$ are also more stable, i.e. they slowly increase with the outlier value.

Table 5.5: Künsch's Premiums

$\mathbf{c_1 = 1, c_2 = 0.5}$

Outlier	1,690	5,000	6,000	7,000
T_1	2,064	2,064	2,064	2,064
T_2	1,511	1,511	1,511	1,511
T_3	1,822	1,822	1,822	1,822
T_4	1,360	1,360	1,360	1,360
T_5	1,599	1,666	1,666	1,666
$\hat{Z}$	0.9496	0.8697	0.8419	0.8207
$\hat{\mu}(\Theta_1)$	2,044	2,056	2,063	2,071
$\hat{\mu}(\Theta_2)$	1,519	1,575	1,596	1,617
$\hat{\mu}(\Theta_3)$	1,814	1,846	1,859	1,872
$\hat{\mu}(\Theta_4)$	1,376	1,444	1,469	1,493
$\hat{\mu}(\Theta_5)$	1,602	1,710	1,727	1,744
$T_.$	1,671	1,685	1,685	1,685

With $c_2 = 0.5$, the truncation occurs at a smaller value than with $c_2 = 1$ and we therefore obtain an estimated credibility factor $\hat{Z}$ that decreases more slowly.

Table 5.6: Künsch's Premiums

$\mathbf{c_1 = 1, c_2 = 2}$

Outlier	1,690	5,000	6,000	7,000
T_1	2,064	2,064	2,064	2,064
T_2	1,511	1,511	1,511	1,511
T_3	1,822	1,822	1,822	1,822
T_4	1,360	1,360	1,360	1,360
T_5	1,599	1,874	1,944	1,944
$\hat{Z}$	0.9496	0.7546	0.5681	0.5268
$\hat{\mu}(\Theta_1)$	2,044	1,981	1,927	1,930
$\hat{\mu}(\Theta_2)$	1,519	1,563	1,612	1,639
$\hat{\mu}(\Theta_3)$	1,814	1,798	1,789	1,803
$\hat{\mu}(\Theta_4)$	1,376	1,450	1,527	1,559
$\hat{\mu}(\Theta_5)$	1,602	1,838	1,859	1,867
$T_.$	1,671	1,726	1,740	1,740

With $c_2 = 2$, the truncation point is larger than with $c_2 = 1$ and it takes more time for the estimation of $\hat{Z}$ to stabilize to a smaller value.

5.2.3 Bühlmann and Straub's Model

Here contract j is still defined by $(\Theta_j, X_{j1}, \ldots, X_{j12})$ but to each claim average X_{jr} there is an associated weight w_{jr}. The estimator of $\mu(\Theta_j)$ is given in Theorem 1.3 by

$$\hat{\mu}(\Theta_j) = m + Z_j(\bar{X}_{jw} - m)$$

Note that m is estimated by $\bar{X}_{zw}$.

Table 5.7:

Bühlmann and Straub's Premiums

Outlier	1,690	5,000	6,000	7,000
$\bar{X}_{1w}$	2,061	2,061	2,061	2,061
$\bar{X}_{2w}$	1,511	1,511	1,511	1,511
$\bar{X}_{3w}$	1,806	1,806	1,806	1,806
$\bar{X}_{4w}$	1,353	1,353	1,353	1,353
$\bar{X}_{5w}$	1,600	1,914	2,009	2,103
$\hat{Z}_1$	0.9847	0.8130	0.6077	0.1954
$\hat{Z}_2$	0.9276	0.4634	0.2353	0.0460
$\hat{Z}_3$	0.8985	0.3735	0.1752	0.0322
$\hat{Z}_4$	0.7279	0.1527	0.0603	0.0096
$\hat{Z}_5$	0.9588	0.6105	0.3583	0.0805
$\hat{\mu}(\Theta_1)$	2,055	2,018	1,997	1,979
$\hat{\mu}(\Theta_2)$	1,524	1,684	1,806	1,938
$\hat{\mu}(\Theta_3)$	1,793	1,823	1,881	1,954
$\hat{\mu}(\Theta_4)$	1,443	1,760	1,864	1,953
$\hat{\mu}(\Theta_5)$	1,603	1,883	1,937	1,971
$\bar{X}_{zw}$	1,684	1,834	1,897	1,959

We observe that as the outlier value increases, it affects the estimated credibility factor $\hat{Z}_j$ in a drastic way.

5.2.4 Gisler and Reinhard's Model

Again we illustrate Gisler and Reinhard's model after Bühlmann and Straub's to show the effect of robustification. Recall that here the pure risk premium is divided into two components: an ordinary part and an excess part.

The estimator of $\mu(\Theta_j)$ is given by (3.27):

$$\hat{\mu}_x(\Theta_j) = \hat{\mu}_{xs} + \hat{\mu}_T + \hat{Z}_j(T_j - \hat{\mu}_T)$$

We observe that the $\bar{X}_{jw}$ in Bühlmann and Straub model are replaced by M-estimators T_j and m by an estimator based on the T_j's.

Table 5.8: Gisler and Reinhard's Premiums

$\mathbf{c_1 = 1, c_2 = 1}$

Outlier	1,690	5,000	6,000	7,000
T_1	2,061	2,061	2,061	2,061
T_2	1,511	1,511	1,511	1,511
T_3	1,806	1,806	1,806	1,806
T_4	1,353	1,353	1,353	1,353
T_5	1,600	1,777	1,777	1,777
$\hat{Z}_1$	0.9848	0.9315	0.9315	0.9315
$\hat{Z}_2$	0.9278	0.7300	0.7300	0.7300
$\hat{Z}_3$	0.8987	0.6511	0.6511	0.6511
$\hat{Z}_4$	0.7283	0.3606	0.3606	0.3606
$\hat{Z}_5$	0.9589	0.8307	0.8307	0.8307
$\hat{\mu}(\Theta_1)$	2,055	2,071	2,091	2,111
$\hat{\mu}(\Theta_2)$	1,524	1,608	1,628	1,648
$\hat{\mu}(\Theta_3)$	1,793	1,820	1,839	1,859
$\hat{\mu}(\Theta_4)$	1,443	1,642	1,661	1,681
$\hat{\mu}(\Theta_5)$	1,603	1,793	1,813	1,832
$\hat{\mu}_{xs}$	0	31.26	50.94	70.62
$\hat{\mu}_T$	1,684	1,756	1,756	1,756

Again, when the data from state 5 is not contaminated (i.e. when the outlier value is equal to 1,690), we obtain the same results as with Bühlmann and Straub's model. Also notice that even if the outlier value increases, the estimated credibility factors $\hat{Z}_j$ become constant as the ordinary part is no longer affected by the increase. The estimation for μ_T, the mean ordinary claims, becomes constant as well. The only part affected by the increase in the outlier value is the estimation of the mean excess claims, μ_{xs}.

5.2.5 Jewell's Hierarchical Model

To illustrate Jewell's hierarchical model, consider a split of Hachemeister's data set and form the following subportfolios:

- subportfolio 1: states 1 and 3
- subportfolio 2: states 2, 4 and 5.

Each contract is defined by the set $(\Theta_p, \Theta_{pj}, X_{pj1}, \ldots, X_{pjt})$ with $P = 2$ subportfolios and $k_p = 5$ and $t_{pj} = 12$, for all contracts. The estimator of $\mu(\Theta_p, \Theta_{pj})$ is given in Theorem 1.5 by:

$$\hat{\mu}(\Theta_p, \Theta_{pj}) = m_p + Z_{pj}(\bar{X}_{pjw} - m_p)$$

Note that m_p is estimated by $\bar{X}_{zjw}$. We kept the original contract subscripts 1 and 3 to allow for a comparison between the previous models and the results for subportfolio 1 and 2.

Table 5.9: Jewell's Premiums
$\mathbf{subportfolio1, c_1 = 1, c_2 = 1}$

Outlier	1,690	5,000	6,000	7,000
X_{1w1}	2,061	2,061	2,061	2,061
X_{1w3}	1,806	1,806	1,806	1,806
$\hat{Z}_1$	0.4174	0.5006	0.3971	0.2842
$\hat{Z}_3$	0.0895	0.1209	0.0828	0.0516
$\hat{\mu}(\Theta_1, \Theta_{11})$	2,035	2,036	2,035	2,033
$\hat{\mu}(\Theta_1, \Theta_{13})$	1,997	1,986	1,999	2,011
X_{1zw}	2,016	2,011	2,017	2,022

Table 5.10: Jewell's Premiums
subportfolio2, $c_1 = 1, c_2 = 1$

Outlier	1,690	5,000	6,000	7,000
$\bar{X}_{2w2}$	1,511	1,511	1,511	1,511
$\bar{X}_{2w4}$	1,353	1,353	1,353	1,353
$\bar{X}_{2w5}$	1,600	1,914	2,009	2,103
$\hat{Z}_2$	0.1247	0.1660	0.1157	0.0731
$\hat{Z}_4$	0.0288	0.0400	0.0266	0.0162
$\hat{Z}_5$	0.2053	0.2655	0.1919	0.1252
$\hat{\mu}(\Theta_2, \Theta_{22})$	1,544	1,689	1,753	1,821
$\hat{\mu}(\Theta_2, \Theta_{24})$	1,544	1,710	1,773	1,837
$\hat{\mu}(\Theta_2, \Theta_{25})$	1,560	1,775	1,827	1,877
$\bar{X}_{2zw}$	1,549	1,725	1,784	1,845

The estimation of the credibility factors become unstable for all states, but is particularly bad for state 3 and 4. This is partly explained by the small weight values for both of these states.

5.2.6 Proposed Robustified Jewell's Model

As with Gisler and Reinhard's model, the pure premium risk is divided into an ordinary part and an excess part. By anology, the estimator of $\mu(\Theta_p, \Theta_{pj})$ is given by (4.15):

$$\hat{\mu}_x(\Theta_p, \Theta_{pj}) = \hat{\mu}_{p-xs} + \hat{\mu}_{T_p} + \hat{Z}_{pj}(T_{pj} - \hat{\mu}_{T_p})$$

The $\bar{X}_{pjw}$ are replaced by M-estimators T_{pj} and m_p is replaced by an estimator based on T_{pj}.

Table 5.11: Proposed Model's Premiums subportfolio1, $c_1 = 1, c_2 = 1$

Outlier	1,690	5,000	6,000	7,000
T_{11}	2,061	2,061	2,061	2,061
T_{13}	1,806	1,806	1,806	1,806
$\hat{Z}_{11}$	0.8778	0.8667	0.8667	0.8667
$\hat{Z}_{13}$	0.4962	0.4714	0.4714	0.4714
$\hat{\mu}(\Theta_1, \Theta_{11})$	2,050	2,049	2,049	2,049
$\hat{\mu}(\Theta_1, \Theta_{13})$	1,888	1,893	1,893	1,893
$\hat{\mu}_{1-xs}$	0	0	0	0
$\hat{\mu}_{T_1}$	1,969	1,971	1,971	1,971

Table 5.12: Proposed Model's Premiums subportfolio2, $c_1 = 1, c_2 = 1$

Outlier	1,690	5,000	6,000	7,000
T_{22}	1,511	1,511	1,511	1,511
T_{24}	1,353	1,353	1,353	1,353
T_{25}	1,600	1,716	1,716	1,716
$\hat{Z}_{22}$	0.5879	0.5636	0.5636	0.5636
$\hat{Z}_{24}$	0.2274	0.2123	0.2123	0.2123
$\hat{Z}_{25}$	0.7214	0.7010	0.7010	0.7010
$\hat{\mu}(\Theta_2, \Theta_{22})$	1,519	1,664	1,721	1,778
$\hat{\mu}(\Theta_2, \Theta_{24})$	1,489	1,659	1,716	1,773
$\hat{\mu}(\Theta_2, \Theta_{25})$	1,580	1,804	1,861	1,918
$\hat{\mu}_{2-xs}$	0	118.73	175.67	232.60
$\hat{\mu}_{T_2}$	1,529	1,590	1,590	1,590

When comparing the above results to those obtained from Jewell's classical model, one notes how the estimators for the credibility factor $\hat{Z}_{pj}$ are less affected by an increase in the outlier value. They in fact become constant. Since the claim averages are truncated, the estimation of the ordinary part also becomes constant. The only estimation still affected by an increase in the outlier value is the excess part.

5.3 Conclusions

One cannot rule out the possibility of outlier contamination of data sets nor can one eliminate them completely. Since the classical models (such as Bühlmann, Bühlmann and Straub and Jewell's models) use linear estimators composed of averages, robustification allows to take the outliers into account but limits their effect on the parameter estimation.

It becomes obvious by looking at the comparison of the classical models to the robustified models (Künsch, Gilser and Reinhard and the new proposed robustified Jewell's models) that there is a gain in robustifying. The estimators of the credibility factors are more stable, sometimes become constant and are often greater than with the classical models, especially when the data set is contaminated. The estimators of the pure risk premium are also more stable than when compared to the classical models.

Also note that Hachemeister's data set, used as an illustration, is rather small. When we split this data set into two subportfolios, the effect of the outlier is magnified. We suggest to use the robustified models with a larger data set.

It would be interesting to apply robustification to regression credibility models such as Hachemeister's model.

Bibliography

[1] Bühlmann, H., (1967). "Experience Rating and Credibility", *ASTIN Bulletin*, **4**, 3, pp.199-207.

[2] Bühlmann, H. and Straub, E., (1970). "Glaubwürdigkeit für Schadensätze", *Mitt. Ver. Schweiz. Ver.*, **70**, 1, pp.111-133.

[3] Garrido, J. and Romera, R., (1995). "Robust Credibility with the Kalman Filter", *Actuarial Research Clearing House*, **1**, pp.163-171.

[4] Gisler, A. and Reinhard, P., (1993). "Robust Credibility", *ASTIN Bulletin*, **23**, 1, pp.117-143.

[5] Goovaerts, M.J., Kaas, R., Van Heerwaarden, A.E. and Bauwelinckx, T., (1990). *Effective Actuarial Methods, Insurance Series Volume 3*, Elsevier Science Publishers B.V.

[6] Gooavaerts, M.J. and Hoogstad, W.J., (1987). *Survey of Actuarial Studies, Credibility Theory*, Nationale-Nederlanden N.V., City.

[7] Hachemeister, C.A., (1975). "Credibility for Regression Models with Application to Trend", in *Credibility, Theory and Applications*, Proc. Berkeley Act. Res. Conf. on Cred., Academic Press, New-York.

[8] Hampel, F.R. et al., (1986). *Robust Statistics; the Approach Based on Influence Functions*, Wiley, New-York.

[9] Herzog, T.N., (1985). "An Introduction to Bayesian Credibility and Related Topics", *Study Kit of the Casualty Actuarial Society.*

[10] Hoaglin, D.C., Mosteller F., Tukey, J.W. (1983). *Understanding Robust and Explaratory Data Analysis*, Wiley, New-York.

[11] Huber, P., (1964). "Robust Estimation of a Location Parameter", *Annals of Mathematical Statistics*, **35**, pp.73-101.

[12] Huber, P., (1981). *Robust Statistics*, Wiley, New-York.

[13] Jewell, W.S., (1974). "Credible Means are Exact Bayesian for Exponential Families", *Astin Bulletin*, **8**, 1, pp.77-90.

[14] Jewell, W.S., (1975). *The Use of Collateral Data in Credibility Theory: a Hierarchical Model*, International Institute for Applied Systems Analysis, Schloss Laxenburg, Austria.

[15] Kallianpur, G. and Rao, C.R., (1955). "On Fisher's Lower Bound to Asymptotic Variance of a Consistent Estimate", *Sankhyā A*, **15**, pp.331-342.

[16] Künsch, H.R., (1992). "Robust Methods for Credibility", *ASTIN Bulletin*, **22**, 1, pp.33-49.

[17] Longley-Cook, L.H., (1962). "An Introduction to Credibility Theory", *Study kit of the Casualty Actuarial Society.*

[18] Philbrick, S.W., (1981). "An Examination of Credibility Concepts", *Study kit of the Casualty Actuarial Society*, P.C.A.S. LXVIII, 1981, pp.195-212

[19] Venter, Gary G. et al., (1990). "Credibility Theory", in *Foundations of Casualty Actuarial Science*, R&S Financial Printing, pp. 375-483.