INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI®

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Robust Credibility and Kalman Filtering

Mike Tam

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfilment of the Requirements

for the Degree of Master of Science at

Concordia University

Montreal, Quebec, Canada

October 1998

© Mike Tam, 1998

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-39075-6

Canada

NOTE TO USERS

Page(s) not included in the original manuscript are unavailable from the author or university. The manuscript was microfilmed as received.

ii

UMI

Abstract

Robust Credibility and Kalman Filtering

Mike Tam

Credibility theory is an experience rating technique in insurance used to combine an estimate of the expected claims of a contract with the estimate of the expected claims of a portfolio of similar contracts. However, the credibility estimate remains sensitive to large (outlying) claims. In this thesis, robustification of some classical credibility models are presented via robust Kalman filtering. Credibility theory has been shown to be a special case of the Kalman filter (De Jong and Zehnwirth, 1983), thus existing research on the robustification of the Kalman filter, for example, Cipra and Romera (1991), can be applied to robustifying Kalman filter credibility models (Kremer, 1994). After describing in some detail the classical and robust models of credibility, we present an implementation of a robust Kalman filter credibility model and apply it to Hachemeister's dataset (Hachemeister, 1975).

Acknowledgements

It has been both a privilege and a pleasure to write this thesis under the guidance of Professor José Garrido. His constant encouragement and abundant patience were pivotal in the completion of this thesis. To José goes my sincere thanks. I also would like to express my deepest gratitude and appreciation to Philippe, Claudette, George, and my parents who have given me their unwavering support, without which this thesis would not have been written. Finally, I dedicate this thesis to my wife, Sandy, for her refusal to let me give up.

Contents

List of Tables

List of Figures

Introduction

Credibility theory is an experience rating technique in actuarial science. Experience rating is the process whereby the experience of an individual risk is used to calculate the premium rate for that individual. In Credibility theory, the premium for an individual risk is computed by combining the experience of the individual with the experience of a larger collective. However, estimation of the credibility premium can be adversely affected by larger than normal claims. In that case, robust methods can be used to provide a robust credibility premium. The Kalman filter is also useful in credibility theory. Various credibility models have been shown be special cases of the Kalman filter. In applying the Kalman filter to credibility, we can also look to a robust Kalman filter to provide a robust credibility premium.

In this thesis, I review the application of robust statistical methods and the Kalman filter to credibility theory. Using some results available in the credibility literature, I also discuss an empirical robust Kalman filter credibility model and some of the difficulties that can arise in the implementation. Finally, some numerical results are presented.

In chapter 1, a review of credibility theory is given. The limited fluctuation approach to credibility is discussed first. This is followed by a discussion of exact credibility. The greatest accuracy approach to credibility is presented next. This approach includes the models of Bühlmann, Bühlmann and Straub, and Hachemeister. Finally, some empirical estimators of the parameters of the credibility premium are shown.

In chapter 2, the application of robust statistics to credibility is examined. A brief review of some results from robust statistics which are useful in robust credibility is followed by an elaboration of the robust credibility models of Künsch, Gisler and Reinhard, and Kremer.

In chapter 3, the Kalman filter and its connection with credibility theory is presented. Two methods of obtaining empirical credibility estimators in the Kalman filter framework are then given. Finally, a robust Kalman filter is considered and one of the empirical non-robust Kalman filter credibility models is adapted to provide an empirical robust Kalman filter credibility model.

In chapter 4, numerical results of the the various models and estimators discussed in the previous chapters are presented. In this chapter, an attempt is made to compare the performance of the various models using a dataset of claim amounts both with and without an outlier.

In the conclusion, a summary of the thesis is given as well as what further work may be done in the area of robust credibility and Kalman filtering.

Chapter 1

Credibility Theory

Credibility theory is an experience rating technique in actuarial science. It is used to determine the expected claims experience of an individual risk when those risks are not homogeneous, or the claims history of the individual risks are scarce, but the experience of the collective is extensive. Given that the individual risks are embedded in a heterogeneous collective, the objective of the various credibility formulas is to calculate the weight which should be assigned to the individual risk data to determine a *credible* mean of that risk. Here, we define risk to be either uncertainty arising from the possible occurrence of given events; or individuals or entities covered by financial security systems.

In determining an insurance premium, we would like the premiums collected to cover the expected severity of future claims, and that each individual risk be assigned a premium commensurate with the risk that it contributes to the collective. If the second condition is not met, the preferred risks would be overcharged, while the substandard risks would be undercharged. A premium which is not experience-rated would tend to drive away "safe" risks and attract selection against the insurer.

The following solution was suggested by Whitney (1918),

$$M_j^a = (1 - Z)m + ZM_j\,.$$

Here, M_j^a is the credibility adjusted mean. It is a weighted average of the overall mean m, and M_j, the mean for individual risk j. The credibility factor Z is a number between 0 and 1, which is assigned to the individual risk premium. When the individual data is profuse, Z is close to 1.

Credibility theory is used extensively in setting rates for automobile insurance, but we can find credibility theory in other insurance applications. Credibility is often used in experience rating claims in group insurance or in determining the worker's compensation premium for a particular employer. Use of credibility can also be found in loss reserving. Loss reserves are funds set aside to pay the benefits of existing and future obligations. When claims are incurred but not yet reported to the insurer, credibility theory can be used to estimate this amount.

Credibility theory can also be used in more general statistical problems such as predicting economic factors. In the regression model by Hachemeister (1975), the effects of inflation in each U.S. state are modeled by a simple trend line. To better predict the effects of inflation, the state trend is consolidated with the country-wide trend to form the credibility adjusted trend line. We discuss the Hachemeister model in more detail in section 1.5.

In this chapter, we review some credibility models. We first discuss the limited fluctuation approach of credibility which was introduced by Mowbray (1914). In the ensuing section, the relationship between credibility theory and the exponential family of distributions is developed. Afterwards, the following three sections review the greatest accuracy approach of credibility, in particular, the models of Bühlmann (1969), Bühlmann and Straub (1970), and Hachemeister (1975). In the final section, some empirical estimators are described.

1.1 Limited Fluctuation Credibility

Limited fluctuation credibility is used to provide the exposure that is required to assign full credibility to the individual data. We may say, for example, that the estimator $\hat{\theta}_j$ of θ_j is given full credibility when, for some $k > 0$, the probability that $\hat{\theta}_j$ is within $100k\%$ of θ_j is at least $1 - \epsilon$. Thus for $\epsilon > 0$,

$$\Pr[|\hat{\theta}_j - \theta_j| \leq k\theta_j] \geq 1 - \epsilon\,. \tag{1.1}$$

This is analogous to determining the minimum sample size required to generate a $100(1-\epsilon)\%$ confidence interval of length k. If n is the number of observations associated with the parameter θ_j, and a Normal approximation is used, we can derive a value n_0, such that (1.1) holds for $n \geq n_0$.

Example 1.1 Let $\Phi(\cdot)$ represent the distribution function of a normal random variable with mean 0 and variance 1. If $\hat{\theta}_j$ is a Binomial proportion estimator, we can show that full credibility can be given to $\hat{\theta}_j$ if $n \geq n_0 = y^2(1-\theta_j)/(k^2\theta_j)$, where $\Phi(y) = 1 - \epsilon/2$. Based on n observations, we have $\mathrm{E}[\hat{\theta}_j] = \theta_j$, and $\mathrm{Var}[\hat{\theta}_j] = \theta_j(1-\theta_j)/n$. Then

$$\begin{aligned}
\Pr[|\hat{\theta}_j - \theta_j| \leq k\theta_j] &= \Pr[-k\theta_j \leq \hat{\theta}_j - \theta_j \leq k\theta_j] \\
&= \Pr\left[\frac{-k\theta_j}{\sqrt{\theta_j(1-\theta_j)/n}} \leq \frac{\hat{\theta}_j - \theta_j}{\sqrt{\theta_j(1-\theta_j)/n}} \leq \frac{k\theta_j}{\sqrt{\theta_j(1-\theta_j)/n}}\right] \\
&\geq 1 - \epsilon\,.
\end{aligned}$$

By the Central Limit Theorem, $\frac{\hat{\theta}_j - \theta_j}{\sqrt{\theta_j(1-\theta_j)/n}}$ will be approximately distributed as a standard normal random variable if n is large. We have

$$\Phi\left[\frac{k\theta_j}{\sqrt{\theta_j(1-\theta_j)/n}}\right] - \Phi\left[\frac{-k\theta_j}{\sqrt{\theta_j(1-\theta_j)/n}}\right] = 1 - \epsilon\,,$$

hence,

$$\Phi\left[\frac{k\theta_j}{\sqrt{\theta_j(1-\theta_j)/n}}\right] = 1 - \epsilon/2\,,$$

then

$$y = \frac{k\theta_j}{\sqrt{\theta_j(1-\theta_j)/n}}\,,$$

and

$$n_0 = \frac{y^2(1-\theta_j)}{k^2\theta_j}$$

is the smallest number of observations needed for full credibility.

When $n \leq n_0$, we can also find Z for partial credibility, that is, for $0 \leq Z < 1$. Let $\theta_j^a = (1-Z)\,m + Z\hat{\theta}_j$, where m is the mean of the collective data. The error of using θ_j^a is

$$\theta_j^a - \theta_j = Z(\hat{\theta}_j - \theta_j) + (1-Z)(m - \theta_j)\,.$$

The term $Z(\hat{\theta}_j - \theta_j)$ describes the error due to the estimate of the individual mean. If we require this error to be bounded absolutely by $k\theta_j$ with probability $1-\epsilon$, then we have

$$\Pr[Z|\hat{\theta}_j - \theta_j| \leq k\theta_j] = 1 - \epsilon\,.$$

Using the normal approximation and the result from Example 1.1, we can derive Z in the case of partial credibility. We have $y = \dfrac{k\theta_j}{Z\sqrt{\theta_j(1-\theta_j)/n}}$, so

$$Z^2 = \frac{k^2\theta_j n}{y^2(1-\theta_j)} = \frac{n}{n_0}\,.$$

Since Z cannot be greater than 1, we take $Z = \min(\sqrt{\frac{n}{n_0}}, 1)$. If $Z < 1$, there is partial credibility.

1.2 Exact Credibility

For a fixed time period t, the yearly claim amounts X_r for $r = 1, \ldots, t$, occur depending on an unobservable risk parameter Θ, which has the structure distribution function $U(\theta)$. The claim amounts are not dependent on time, therefore given $\Theta = \theta$, the X_r are conditionally independent and identically distributed. The conditional distribution of the claims experience random variables X_r given the risk parameter $\Theta = \theta$, is described by the distribution function $F(x_1, \ldots, x_t|\theta)$.

To derive the individual risk premium $\mu(\Theta) = \mathrm{E}[X|\Theta]$, based on the observations $X_1, \ldots, X_t$, consider approximating $\mu(\Theta)$ by a function $g(X_1, \ldots, X_t)$ which minimizes the squared-error loss function. That is, we wish to determine g such that given $X_1, \ldots, X_t$,

$$\mathcal{R}(g) = \mathrm{E}[\mu(\Theta) - g(X_1, \ldots, X_t)]^2$$

is minimized. We know that when $\mathcal{R}$ is of this form, it is minimized when

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

which is the posterior Bayes estimator. The conditional expectation in (1.2) is called the exact credibility estimator.

In practice, since we do not always know the distribution functions $F(x|\theta)$ and $U(\theta)$, we may wish to adopt a semi-parametric approach by considering only those functions g which are linear combinations of the observations. Thus, by minimizing

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

over all $c_0, \ldots, c_t$, we find that $c_r = \dfrac{a}{s^2 + at} = \dfrac{Z}{t}$, for $r = 1, \ldots, t$, or equivalently

$$Z = \frac{at}{s^2 + at}\,, \tag{1.3}$$

where $a = \mathrm{Var}\{\mathrm{E}[X|\Theta]\}$ and $s^2 = \mathrm{E}\{\mathrm{Var}[X|\Theta]\}$. Also, $c_0 = (1 - Z)\,\mu$ with $\mu = \mathrm{E}\{[\mu(\Theta)]\}$. Hence,

$$\hat{\mu}(\Theta_j) = c_0 + \sum_{r=1}^{t} c_r X_r = (1 - Z)\,\mu + Z\bar{X}\,, \tag{1.4}$$

where $\hat{\mu}(\Theta_j)$, is the optimal linear credibility premium under the least-squares criterion.

Exact credibility refers to the situation where the exact credibility premium coincides with the optimal linear credibility premium. Exact credibility appears in the original model of Bühlmann (1967), where only one contract is considered.

The parameters μ, s^2, and a are called the *structural* parameters. According to Goovaerts and Hoogstad (1987), the parameters are called structural since these parameters should estimated from the data. Since the structural parameters μ, s^2, and a would have to be known for $\hat{\mu}(\Theta_j)$ to be calculated, we cannot consider the linear credibility estimator a statistical estimator. We will return to this point in section 1.6.

Even when both $F(x|\theta)$ and $U(\theta)$ are known, the resultant credibility formula may not be tractable. Bailey (1950) showed that if the conditional distribution is given by the Binomial distribution and the prior distribution is the Beta distribution, then exact credibility occurs. Bailey, in the same paper, also demonstrated exact credibility in the Poisson-Gamma case. For similar results, see also Mayerson (1964).

Jewell (1974) generalized the results of Bailey and Mayerson. He showed that, for the exponential family of functions and its conjugate priors, the exact credibility premium equals the linear credibility formula. We now prove this result.

Proposition 1.1 (Jewell, 1974) Given $F(x|\theta)$ and its conjugate prior $U(\theta)$, the Bayesian credibility premium in the least-squares sense, based on these distributions, coincides with the linear credibility formula if $F(x|\theta)$ is a function from the single-parameter exponential family.

Proof Consider the single-parameter exponential family with natural parametrization

$$f(x|\theta) = \frac{p(x)e^{-\theta x}}{q(\theta)}, \tag{1.5}$$

where $p(x)$ and $q(\theta)$ are arbitrary functions such that $f(x|\theta)$ is a proper density. Let $\theta \in \vartheta$ and $x \in \Omega$. Consider also, the conjugate prior distribution

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

To see that $u(\theta)$ is the natural conjugate prior of $f(x|\theta)$, note that

$$\begin{aligned} g(x,\theta) &= f(x|\theta)u(\theta) \\ &\propto \frac{p(x)e^{-\theta(x+x_0)}}{q(\theta)^{t_0+1}} . \end{aligned}$$

Let

$$k(x) = \int_{\vartheta} \frac{p(x)e^{-\theta(x+x_0)}}{q(\theta)^{t_0+1}} \, d\theta \, ,$$

then the posterior distribution of θ given x is

$$\pi(\theta|x) = \frac{q(\theta)^{-(t_0+1)}p(x)e^{-\theta(x+x_0)}}{k(x)} .$$

Let $c(x) = p(x)/k(x)$, then

$$\pi(\theta|x) = \frac{c(x)e^{-\theta(x+x_0)}}{q(\theta)^{t_0+1}} .$$

Since the posterior density $\pi(\theta|x)$, with parameters t_0+1 and x_0+x, is also a member of the exponential family, it follows that $u(\theta)$ is conjugate for $f(x|\theta)$.

The likelihood function of θ is given by

$$L(\theta|x_1,\dots,x_t) = \frac{p(x_1)\cdots p(x_t)e^{-\theta\sum_{r=1}^{t} x_r}}{[q(\theta)]^t} ,$$

then

$$\mathrm{E}[\mu(\Theta)|x_1,\dots,x_t] = \frac{\displaystyle\int_{\vartheta} \frac{\mu(\theta)p(x_1)\cdots p(x_t)e^{-\theta\sum_{r=1}^{t} x_r}e^{-\theta x_0}}{[q(\theta)]^t q(\theta)^{t_0} c(t_0,x_0)} \, d\theta}{\displaystyle\int_{\vartheta} \frac{p(x_1)\cdots p(x_t)e^{-\theta\sum_{r=1}^{t} x_r}e^{-\theta x_0}}{[q(\theta)]^t q(\theta)^{t_0} c(t_0,x_0)} \, d\theta} . \tag{1.6}$$

Let

$$\upsilon(\theta) = \frac{q(\theta)^{-(t_0+t)}e^{-\theta(x_0+\sum_{r=1}^{t} x_r)}}{c(t_0+t, x_0+\sum_{r=1}^{t} x_r)} ,$$

which is analogous to $u(\theta)$, so $\int_{\vartheta} \upsilon(\theta)\, d\theta = 1$. After simplifying the equation in (1.6), we can write the expectation as

$$\begin{aligned} \mathrm{E}[\mu(\Theta)|x_1,\dots,x_t] &= \int_{\vartheta} \frac{\mu(\theta)e^{-\theta(x_0+\sum_{r=1}^{t} x_r)}}{q(\theta)^{t_0+t}} d\theta, \\ &= \int_{\vartheta} \mu(\theta)\upsilon(\theta)\, d\theta . \end{aligned}$$

Next, we differentiate $\upsilon(\theta)$ with respect to θ,

$$\frac{d}{d\theta}\upsilon(\theta) = \left[-(t_0+t)\frac{q'(\theta)}{q(\theta)} - \left(x_0 + \sum_{r=1}^{t} x_r\right)\right]\upsilon(\theta)\,.$$

But from equation (1.5), we have

$$\int_\Omega \frac{p(x)e^{-\theta x}}{q(\theta)}\,dx = 1\,.$$

Thus,

$$q(\theta) = \int_\Omega p(x)e^{-\theta x}\,dx\,.$$

Since

$$q'(\theta) = -\int_\Omega xp(x)e^{-\theta x}\,dx\,,$$

it follows that $\mathrm{E}[X|\theta] = \mu(\theta) = -q'(\theta)/q(\theta)$. So now we have

$$\frac{d}{d\theta}\upsilon(\theta) = \left[(t_0+t)\mu(\theta) - \left(x_0 + \sum_{r=1}^{t} x_r\right)\right]\upsilon(\theta)\,. \tag{1.7}$$

If $\upsilon(\theta)$ equals zero at both the upper and lower limits of ϑ, integrating both sides of (1.7), we get

$$\int_\vartheta \frac{d}{d\theta}\upsilon(\theta)\,d\theta = \int_\vartheta \left[(t_0+t)\mu(\theta)\upsilon(\theta) - \left(x_0 + \sum_{r=1}^{t} x_r\right)\upsilon(\theta)\right]d\theta$$

or

$$0 = (t_0+t)\int_\vartheta \mu(\theta)\upsilon(\theta)\,d\theta - \left(x_0 + \sum_{r=1}^{t} x_r\right)\,.$$

Hence,

$$\int_\vartheta \mu(\theta)\upsilon(\theta)\,d\theta = \frac{x_0 + \sum_{r=1}^{t} x_r}{t_0+t}\,,$$

and therefore,

$$\begin{aligned}
\mathrm{E}[\mu(\Theta)|x_1,\dots,x_t] &= \frac{x_0+x_1+\cdots+x_t}{t_0+t} \\
&= \left(\frac{t_0}{t_0+t}\right)\frac{x_0}{t_0} + \left(\frac{t}{t_0+t}\right)\frac{x_1+\cdots+x_t}{t} \\
&= (1-Z)\mu + Z\bar{x},
\end{aligned}$$

where $Z = \dfrac{t}{t_0 + t}$.

So for distributions in the exponential family and their conjugate priors, the exact credibility formula can be written as a linear combination of the prior mean and the individual data.

To show that $t_0 = \dfrac{\mathrm{E}\{\mathrm{Var}[X|\Theta]\}}{\mathrm{Var}[\mu(\Theta)]} = \dfrac{s^2}{a}$, note that, as in equation (1.7),

$$\frac{d}{d\theta}u(\theta) = [t_0\mu(\theta) - x_0]u(\theta),$$

and

$$\frac{d^2}{d\theta^2}u(\theta) = [t_0\mu(\theta) - x_0]u'(\theta) + t_0\mu'(\theta)u(\theta)\,.$$

But with

$$q''(\theta) = \int_\Omega x^2\,p(x)e^{-\theta x}\,dx\,,$$

we have $\mathrm{E}[X^2|\theta] = q''(\theta)/q(\theta)$. It then follows that

$$\frac{d}{d\theta}\mu(\theta) = -\,\mathrm{Var}[X|\theta]\,.$$

Then

$$\frac{d^2}{d\theta^2}u(\theta) = [t_0\mu(\theta) - x_0]^2u(\theta) - t_0\,\mathrm{Var}[X|\theta]u(\theta). \tag{1.8}$$

If we assume that $u(\theta)$ equals zero at both endpoints of ϑ, we have

$$\mu = \int_\vartheta \mu(\theta)u(\theta)\,d\theta = \frac{x_0}{t_0}.$$

Then, integrating both sides of (1.8) with respect to θ, we obtain

$$\begin{aligned}
0 &= \mathrm{E}[t_0\mu(\Theta) - x_0]^2 - t_0\,\mathrm{E}\{\mathrm{Var}[X|\Theta]\}\\
&= t_0^2\,\mathrm{E}\left[\mu(\Theta) - \frac{x_0}{t_0}\right]^2 - t_0\,\mathrm{E}\{\mathrm{Var}[X|\Theta]\}\\
&= t_0^2\,\mathrm{Var}[\mu(\Theta)] - t_0\,\mathrm{E}\{\mathrm{Var}[X|\Theta]\}.
\end{aligned}$$

Thus,

$$t_0 = \frac{\mathrm{E}\{\mathrm{Var}[X|\Theta]\}}{\mathrm{Var}[\mu(\Theta)]} = \frac{s^2}{a}.$$

□

1.3 The Classical Model of Bühlmann

In this section, we discuss the classical credibility model of Bühlmann (1969). In this model, an entire portfolio of contracts is now considered, and a linear credibility estimator is sought. Instead of a single risk parameter Θ, we now have risk parameters Θ_j for $j = 1, 2, \ldots, k$, where k is the number of contracts in the portfolio. The claim amount for the jth contract at time r for $r = 1, \ldots, t$ is given by X_{jr}. The assumptions of the model are:

(B1) The contracts (X_j, Θ_j) for $j = 1, \ldots, k$ are independent and identically distributed.

(B2) For every contract $j = 1, \ldots, k$, and for a given Θ_j, the claim amount random variables $X_{j1}, \ldots, X_{jt}$ are conditionally independent and identically distributed.

The first assumption implies that claim amounts from one contract are independent of claim amounts from another contract. The second assumption asserts that within a contract, the claim amounts at time $r = r'$ are independent of claim amounts occurring at times $r \neq r'$. For $\Theta_j = \Theta$, the classical Bühlmann model coincides with the exact credibility model.

When a semi-parametric approach is used to obtain the credibility estimator, we can relax assumption (B2) to equality of the first two moments of the conditional distribution of X_{jr} given Θ_j. That is, for $j = 1, \ldots, k$,

(B2′) $\mathrm{E}[X_{jr}|\Theta_j] = \mu(\Theta_j)$ and the covariance matrix of the claim amounts at time periods $r = 1, 2, \ldots, t$ equals $\sigma^2(\Theta_j)\,\mathbf{I}$,

where $\sigma^2(\Theta_j) = \mathrm{Var}[X_{jr}|\Theta_j]$ and $\mathbf{I}$ is the $t \times t$ identity matrix.

Before we state the main results of this section, we will need to prove the following covariance relationships.

Lemma 1.1 For any $i, j = 1, \dots, k$ and $r, r' = 1, \dots, t$, let $\mu = \mathrm{E}[\mu(\Theta_j)]$, $a = \mathrm{Var}[\mu(\Theta_j)]$, and $s^2 = \mathrm{E}\{\mathrm{Var}[X_{jr}|\Theta_j]\}$. Then

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

(ii) $\mathrm{Cov}[X_{jr'}, X_{jr}] = a + s^2\delta_{rr'}$, and

(iii) $\mathrm{Cov}[X_{jr'}, X_{ir}] = 0$, for $i \neq j$.

The function δ_{ij} is the Kronecker delta, which is defined such that, for any i and j,

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Proof The following proof can also be found in Goovaerts et al. (1990). We first prove (i). For any random variables X, Y, and Θ the covariance of X and Y can be written as

$$\mathrm{Cov}[X, Y] = \mathrm{Cov}\{\mathrm{E}[X|\Theta], \mathrm{E}[Y|\Theta]\} + \mathrm{E}\{\mathrm{Cov}[X, Y|\Theta]\}. \tag{1.9}$$

Setting $X = X_{ir}$, $Y = \mu(\Theta_j)$, and $\Theta = \Theta_j$, we have

$$\mathrm{Cov}[X_{ir}, \mu(\Theta_j)] = \mathrm{Cov}\{\mathrm{E}[X_{ir}|\Theta_j], \mathrm{E}[\mu(\Theta_j)|\Theta_j]\} + \mathrm{E}\{\mathrm{Cov}[X_{ir}, \mu(\Theta_j)|\Theta_j]\}. \tag{1.10}$$

The conditional expected value of $\mu(\Theta_j)$ given Θ_j is again $\mu(\Theta_j)$. If $i = j$, then by definition, the conditional expectation of X_{jr} given Θ_j is equal to $\mu(\Theta_j)$. Conversely, if $i \neq j$, the conditional expectation of X_{ir} given Θ_j is equal to μ, and so $\mathrm{Cov}[\mu, \mu(\Theta_j)] = 0$. Therefore, the first term on the right of (1.10) reduces to $\mathrm{Cov}[\mu(\Theta_j), \mu(\Theta_j)] = \mathrm{Var}[\mu(\Theta_j)]\,\delta_{ij}$. Given Θ_j, the random variable $\mu(\Theta_j)$ is degenerate, so the second term on the right of (1.10) vanishes. We are then left with

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

To prove (ii), let $X = X_{jr}$, $Y = X_{jr'}$, and $\Theta = \Theta_j$ in (1.9). Then

$$\mathrm{Cov}[X_{jr'}, X_{jr}] = \mathrm{Cov}\{\mathrm{E}[X_{jr'}|\Theta_j], \mathrm{E}[X_{jr}|\Theta_j]\} + \mathrm{E}\{\mathrm{Cov}[X_{jr'}, X_{jr}|\Theta_j]\}.$$

The first term on the right side can be written as

$$\mathrm{Cov}[\mu(\Theta_j), \mu(\Theta_j)] = \mathrm{Var}[\mu(\Theta_j)] = a.$$

For a given Θ_j, the X_{jr} for $r = 1, \dots, t$ are conditionally independent, so for $r \neq r'$, the conditional covariance of $X_{jr'}$ and X_{jr} given Θ_j equals zero. When $r = r'$,

$$\mathrm{E}\{\mathrm{Cov}[X_{jr'}, X_{jr}|\Theta_j]\} = \mathrm{E}\{\mathrm{Var}[X_{jr'}|\Theta_j]\} = s^2.$$

Thus,

$$\mathrm{Cov}[X_{jr'}, X_{jr}] = a + s^2\delta_{rr'}.$$

Finally, to see that (iii) is true, let $X = X_{jr'}$, $Y = X_{ir}$, and $\Theta = \Theta_j$, to get

$$\mathrm{Cov}[X_{jr'}, X_{ir}] = \mathrm{Cov}\{\mathrm{E}[X_{jr'}|\Theta_j], \mathrm{E}[X_{ir}|\Theta_j]\} + \mathrm{E}\{\mathrm{Cov}[X_{jr'}, X_{ir}|\Theta_j]\}.$$

By assumption (B1), the conditional covariance of $X_{jr'}$ and X_{ir}, given Θ_j, equals zero. The conditional expectation, given Θ_j, of $X_{jr'}$ and X_{ir} equals $\mu(\Theta_j)$ and μ, respectively. Thus, the covariance of these two conditional expectations equal zero, again by (B1). It follows that the covariance of $X_{jr'}$ and X_{ir} also equals zero. $\square$

We now prove two theorems relating to the best linear approximation to the conditional expectation $\mathrm{E}[\mu(\Theta)|X_1, \dots, X_t]$. The first theorem gives the optimal inhomogeneous credibility estimator for the individual risk premium.

Theorem 1.1 (Bühlmann, 1969) If the hypotheses (B1) and (B2′) are satisfied, then the optimal inhomogeneous linear estimator $\hat{\mu}(\Theta_j)$ of $\mu(\Theta_j)$, in the least-squares sense is

$$\hat{\mu}(\Theta_j) = (1 - Z)\mu + Z\bar{X}_j\,, \tag{1.11}$$

where $\bar{X}_j = \frac{1}{t}\sum_{r=1}^{t} X_{jr}$, and $\mu = \mathrm{E}[\mu(\Theta_j)]$.

Proof The procedure here follows the presentation in Goovaerts et al. (1990). For a given j, we wish to minimize

$$\mathcal{R}_{nh}(c_0, \ldots, c_{jt}) = \mathrm{E}\left[\mu(\Theta_j) - c_0 - \sum_{r=1}^{t} c_{jr} X_{jr}\right]^2 ,$$

over c_0 and all c_{jr} for $r = 1, \ldots, t$. Differentiating $\mathcal{R}_{nh}$ with respect to c_0 and $c_{jr'}$ for each $r' = 1, \ldots, t$ and setting the results equal to zero, we get

$$\frac{\partial}{\partial c_0}\mathcal{R}_{nh} = -2\,\mathrm{E}\left[\mu(\Theta_j) - c_0 - \sum_{r=1}^{t} c_{jr} X_{jr}\right] = 0$$

and

$$\frac{\partial}{\partial c_{jr'}}\mathcal{R}_{nh} = -2\,\mathrm{E}\left\{X_{jr'}\left[\mu(\Theta_j) - c_0 - \sum_{r=1}^{t} c_{jr} X_{jr}\right]\right\} = 0\,.$$

Simplifying the equations, we obtain

$$\mathrm{E}\left[\mu(\Theta_j) - c_0 - \sum_{r=1}^{t} c_{jr} X_{jr}\right] = 0, \tag{1.12}$$

and for every $r' = 1, 2, \ldots, t$,

$$\mathrm{E}\{X_{jr'}[\mu(\Theta_j) - c_0 - \sum_{r=1}^{t} c_{jr} X_{jr}]\} = 0\,. \tag{1.13}$$

Multiplying (1.12) by $\mathrm{E}[X_{jr'}]$ and subtracting it from (1.13), we obtain

$$\begin{aligned} 0 &= \mathrm{Cov}[X_{jr'}, \mu(\Theta_j) - c_0 - \sum_{r=1}^{t} c_{jr} X_{jr}] \\ &= \mathrm{Cov}[X_{jr'}, \mu(\Theta_j)] - \mathrm{Cov}[X_{jr'}, c_0] - \mathrm{Cov}[X_{jr'}, \sum_{r=1}^{t} c_{jr} X_{jr}]\,. \end{aligned}$$

Since $\mathrm{Cov}[X_{jr'}, c_0] = 0$ and $\mathrm{Cov}[X_{jr'}, \sum_{r=1}^{t} c_{jr} X_{jr}] = \sum_{r=1}^{t} \mathrm{Cov}[X_{jr'}, X_{jr}]$, we have

$$\sum_{r=1}^{t} c_{jr}\,\mathrm{Cov}[X_{jr'}, X_{jr}] = \mathrm{Cov}[\mu(\Theta_j), X_{jr'}]\,. \tag{1.14}$$

From (1.12) and (1.14), we obtain the following system of equations:

$$\begin{aligned} c_0 + \mu \sum_{r=1}^{t} c_{jr} &= \mu\,, \\ s^2 c_{jr'} + a \sum_{r=1}^{t} c_{jr} &= a\,, \end{aligned}$$

for $r' = 1, \ldots, t$. The system of equations is symmetric with respect to the c_{jr}, so we can write $c_{j1} = c_{j2} = \cdots = c_{jt} = c$. The system then reduces to

$$c_0 + \mu t c = \mu \,,$$

$$s^2 c + atc = a \,.$$

We find that

$$c = \frac{a}{s^2 + at} = \frac{Z}{t} \,,$$

with $Z = \dfrac{at}{s^2 + at}$ as postulated in (1.3). It follows that

$$c_0 = (1 - Z)\mu$$

and so

$$\hat{\mu}(\Theta_j) = (1 - Z)\mu + Z\bar{X}_j \,. \qquad \square$$

In the foregoing theorem, we can see that equation (1.12) guarantees the unbiasedness of the linear estimator. In the ensuing theorem, where we investigate an homogeneous estimator, we will need to impose a condition which will provide us with the property of unbiasedness. To understand why unbiasedness is important in an insurance framework, consider the concept of unbiasedness as one of the principles of premium calculation, that the expected financial loss to the insurer is zero. Adherence to this rule ensures that the expected value of future claims be equal to the expected value of future premium payments.

In practice, inhomogeneous premium rules are more logical than homogeneous rules. This can be seen intuitively since no past claims should not imply that there is no risk of future claims. Under an inhomogeneous rule, even if $X_{jr} = 0$, a premium would still be assessed. However, if we do not want to consider linear affine functions of the past observations, we have the following theorem for the homogeneous case.

Theorem 1.2 (Bühlmann, 1969) If the hypotheses (B1) and (B2′) are satisfied, then the optimal unbiased homogeneous linear estimator $\hat{\mu}(\Theta_j)$ of $\mu(\Theta_j)$, in the least-squares sense is

$$\hat{\mu}(\Theta_j) = (1 - Z)\,\bar{X} + Z\bar{X}_j\,, \tag{1.15}$$

where $\bar{X}_j = \frac{1}{t}\sum_{r=1}^{t} X_{jr}$, and $\bar{X} = \frac{1}{k}\sum_{j=1}^{k} \bar{X}_j = \frac{1}{kt}\sum_{j=1}^{k}\sum_{r=1}^{t} X_{jr}$.

Proof In the inhomogeneous case, c_0 was the amount of collateral data to be used. In the homogeneous case, if we merely minimized $\mathrm{E}[\mu(\Theta_j) - \sum_{r=1}^{t} c_{jr}X_{jr}]^2$, subject to the constraint $\mathrm{E}[\mu(\Theta_j)] = \sum_{r=1}^{t} c_{jr}\,\mathrm{E}[X_{jr}]$, over all $c_{j1}, \dots, c_{jt}$, the solution would be $c_{jr} = \frac{1}{t}$ for $r = 1, \dots, t$. The estimator for $\mu(\Theta_j)$ would then be $\bar{X}_j$, which satisfies the unbiasedness constraint. However, this estimator does not incorporate any information from the collective. This can also be seen by noting that if there are no claims for risk j, $\sum_{r=1}^{t} c_{jr}X_{jr}$ would equal zero, so no premium would be charged for this risk. If we want to incorporate the collateral data into our estimator, we must re-formulate the minimizing equation as:

$$\mathcal{R}_h(c_{jir}) = \mathrm{E}[\mu(\Theta_j) - \sum_{i=1}^{k}\sum_{r=1}^{t} c_{jir}X_{jir}]^2,$$

for each $i, j = 1, \dots, k$ and $r = 1, \dots, t$.

Since we require the unbiasedness of the linear estimator, we must have

$$\mathrm{E}[\mu(\Theta_j)] = \sum_{i=1}^{k}\sum_{r=1}^{t} c_{jir}\,\mathrm{E}[X_{jir}].$$

But this can be written as $\mu = \mu\sum_{i=1}^{k}\sum_{r=1}^{t} c_{jir}$, hence $\sum_{i=1}^{k}\sum_{r=1}^{t} c_{jir} = 1$.

Multiplying $\sum_{i=1}^{k}\sum_{r=1}^{t} c_{jr} - 1$ by the Lagrange multiplier λ and subtracting the result from $\mathcal{R}_h$ results in

$$\mathcal{R}_{h\lambda}(c_{jir}) = \mathrm{E}[\mu(\Theta_j) - \sum_{i=1}^{k}\sum_{r=1}^{t} c_{jir}X_{jir}]^2 - \lambda\left[\sum_{i=1}^{k}\sum_{r=1}^{t} c_{jir} - 1\right].$$

Differentiating $\mathcal{R}_{h\lambda}$ with respect to $c_{ji'r'}$ for every i' and r', and setting the result equal to zero gives us

$$\frac{\partial}{\partial c_{ji'r'}}\mathcal{R}_{h\lambda} = -2\,\mathrm{E}\left\{X_{ji'r'}\left[\mu(\Theta_j) - \sum_{i=1}^{k}\sum_{r=1}^{t} c_{jir}X_{jir}\right]\right\} - \lambda = 0\,, \tag{1.16}$$

for $i' = 1, \dots, k$ and $r' = 1, \dots, t$. Thus

$$
\begin{aligned}
-\frac{\lambda}{2} &= \mathrm{E}[X_{ji'r'}\mu(\Theta_j)] - \sum_{i=1}^{k}\sum_{r=1}^{t} c_{jir}\,\mathrm{E}[X_{ji'r'}X_{jir}] \\
&= \mathrm{E}[X_{ji'r'}\mu(\Theta_j)] - m^2 - \sum_{i=1}^{k}\sum_{r=1}^{t} c_{jir}\{\mathrm{E}[X_{ji'r'}X_{jir}] - m^2\} \\
&= \mathrm{Cov}[X_{ji'r'}, \mu(\Theta_j)] - \sum_{i=1}^{k}\sum_{r=1}^{t} c_{jir}\,\mathrm{Cov}[X_{ji'r'}, X_{jir}] .
\end{aligned}
$$

Since $\mathrm{Cov}[X_{ji'r'}, \mu(\Theta_j)] = a\delta_{i'j}$ and $\mathrm{Cov}[X_{ji'r'}, X_{jir}] = a + s\delta_{rr'}$ for $i' = i$ and zero for $i' \neq i$, we have

$$
\begin{aligned}
-\frac{\lambda}{2} &= a\delta_{i'j} - \sum_{i=1}^{k}\sum_{r=1}^{t} c_{jir}(a + s\delta_{rr'}) \\
&= a\delta_{i'j} - a\sum_{r=1}^{t} c_{ji'r} - s^2\, c_{ji'r'} .
\end{aligned}
\tag{1.17}
$$

Since the system of equations is symmetric with respect to the c_{jir}, we write the c_{jir} as $c_{ji'r'}$. Then $\sum_{r=1}^{t} c_{ji'r'} = tc_{ji'r'}$. From (1.17),

$$
a\delta_{i'j} + \frac{\lambda}{2} = atc_{ji'r'} + s^2\, c_{ji'r'} .
$$

So,

$$
c_{ji'r'} = \frac{a\delta_{i'j} + \lambda/2}{at + s^2}. \tag{1.18}
$$

Since $\sum_{i=1}^{k}\sum_{r=1}^{t} c_{jir} = 1$, it follows that

$$
\begin{aligned}
1 &= \sum_{i'=1}^{k}\sum_{r'=1}^{t} \frac{a\delta_{i'j} + \lambda/2}{at + s^2} \\
&= \frac{at}{at + s^2} + \frac{\lambda}{2}\frac{kt}{at + s^2}.
\end{aligned}
$$

Let $Z = \dfrac{at}{at + s^2}$, then

$$
\frac{\lambda}{2} = \frac{(1 - Z)}{kZ/a}.
$$

Inserting this into (1.18), we get

$$
\begin{aligned}
c_{ji'r'} &= \frac{a\delta_{i'j} + (1-Z)a/kZ}{at+s^2} \\
&= \frac{(1-Z+kZ\delta_{i'j})a}{(at+s^2)kZ}.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\sum_{i'=1}^{k}\sum_{r'=1}^{t} c_{ji'r'}X_{ji'r'} &= \frac{(1-Z)}{ktZ}\sum_{i'=1}^{k}\sum_{r'=1}^{t}\frac{at}{at+s^2}X_{ji'r'} + \frac{1}{t}\sum_{i'=1}^{k}\sum_{r'=1}^{t}\frac{at\delta_{i'j}}{at+s^2}X_{ji'r'} \\
&= (1-Z)\bar{X} + Z\bar{X}_j\,. \qquad \square
\end{aligned}
$$

The asymptotic behaviour of Z and the structural parameters a and s^2 are intuitively appealing. As the number of time periods becomes arbitrarily large, Z will approach one. As the individual data increases in size, the collective data will no longer be required. However, in practice, it is difficult to partition the collective into strictly homogeneous sub-classes. Even when t approaches infinity, Z may remain significantly less than one.

When a decreases to zero, Z will tend to zero. This can be seen by noting that a is the "between contracts" variance. When $a = 0$, there is no variation between contracts and the entire portfolio is homogeneous. The best linear estimate under the least-squares criterion then is the mean of the collective. If a approaches infinity, Z approaches one. The collateral data is so heterogeneous that the individual data should not be combined.

When s^2 approaches zero, Z will tend towards one. Since s^2 is the "within contract" variance, if $s^2 = 0$, then the individual data is completely homogeneous and the collective data is not required. If s^2 increases without bound, however, the individual data contains so much heterogeneity that it is not useful in estimating the individual mean.

1.4 The Bühlmann-Straub Model

The Bühlmann-Straub (1970) model is a generalization of the classical Bühlmann model. In the Bühlmann-Straub model, natural weights are assigned to the data and are allowed to vary with time. If a portfolio can be divided into sub-groups, with each contract in the jth sub-group having the same risk parameter Θ_j, and if the number of contracts in the jth sub-group is w_j, then the w_j contracts in the jth sub-group can be replaced by their average. Then allowing the weights to vary with time, we add the index r to w_j to indicate the dependence on time. The natural weights are then written as w_{jr} for $j = 1, \ldots, k$ and $r = 1, \ldots, t$, and are considered as the number of contracts grouped into an average contract. We may also consider cases where the weights are given by other types of exposure such as premium volumes. The special case which coincides with the classical Bühlmann model is just the Bühlmann-Straub model with constant weights.

So now, each contract $j = 1, \ldots, k$ is made up of the average of a group of contracts with the weights $w_{j1}, \ldots, w_{jt}$ varying with time. We would also like all contracts to have the same expectation of claim size as a function of the risk parameter Θ_j. The assumptions of the Bühlmann-Straub model are as follows: For $j = 1, \ldots, k$ and $r, r' = 1, \ldots, t$,

(BS1) $\mathrm{E}[X_{jr}|\Theta_j] = \mu(\Theta_j)$,
$\mathrm{Cov}[X_{jr}, X_{jr'}|\Theta_j] = \delta_{rr'}\sigma^2(\Theta_j)/w_{jr}$.

(BS2) The contracts (X_j, Θ_j) for $j = 1, \ldots, k$ are independent. The variables $\Theta_1, \ldots, \Theta_k$ are identically distributed. The observations X_{jr} have finite variance.

As is evident in assumption (BS2), the independence between the contracts still holds. In (BS1), since $\mathrm{Cov}[X_{jr}, X_{jr'}|\Theta_j] = 0$; for $r, r' = 1, \ldots, t$ and $r \neq r'$, the independence within the contracts remains as well. The equality of the first moment of the observations is still true; however, due to the introduction of the weights which vary with time, the variance of the observations are no longer homogeneous with respect to time.

We introduce the following notation for convenience:

$$
\begin{aligned}
w_j &= \sum_{r=1}^{t} w_{jr} \,, \\
w &= \sum_{j=1}^{k} w_j = \sum_{j=1}^{k}\sum_{r=1}^{t} w_{jr} \,, \\
Z &= \sum_{j=1}^{k} Z_j \,, \\
X_{jw} &= \sum_{j=1}^{k} \frac{w_{jr}}{w_j} X_{jr} \,, \\
X_{ww} &= \sum_{j=1}^{k} \frac{w_j}{w} X_{jw} = \sum_{j=1}^{k}\sum_{r=1}^{t} \frac{w_{jr}}{w} X_{jr} \,, \\
X_{zw} &= \sum_{j=1}^{k} \frac{Z_j}{Z} X_{jw} = \sum_{j=1}^{k}\sum_{r=1}^{t} \frac{Z_j}{Z}\frac{w_{jr}}{w} X_{jr} \,.
\end{aligned}
$$

In the Bühlmann-Straub model, the individual estimator is X_{jw} and the estimator for the collective in the homogeneous case is X_{zw}. The credibility weights Z_j, are such that

$$Z_j = \frac{a w_j}{a w_j + s^2} \,, \tag{1.19}$$

where $a = \mathrm{Var}[\mu(\Theta_j)]$ and $s^2 = \mathrm{E}[\sigma^2(\Theta_j)]$. The credibility estimator for $\mu(\Theta_j)$ then is

$$\hat{\mu}(\Theta_j) = (1 - Z_j)\, X_{zw} + Z_j X_{jw} \,. \tag{1.20}$$

To prove the optimality under the least-squares criterion of these credibility estimators, we need the following covariance relations:

Lemma 1.2 The following covariance relations hold:

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

(ii) $\mathrm{Cov}[X_{jr'}, X_{ir}] = 0$ for $i \neq j$,

(iii) $\mathrm{Cov}[X_{jr'}, X_{jr}] = a + s^2\delta_{rr'}/w_{jr}$,

(iv) $\mathrm{Cov}[X_{jr}, X_{jw}] = \mathrm{Cov}[X_{jw}, X_{jw}] = a + s^2/w_j$,

(v) $\mathrm{Cov}[X_{jw}, X_{zw}] = \mathrm{Cov}[X_{zw}, X_{zw}] = a/Z$,

(vi) $\mathrm{Cov}[X_{jw}, X_{ww}] = s^2/w + aw_j/w$,

(vii) $\mathrm{Cov}[X_{ww}, X_{ww}] = s^2/w + a\sum_{j=1}^{k}(w_j/w)^2$.

The proof of Lemma 1.2 is similar to that of Lemma 1.1. Using computations analogous to those in the proof of Lemma 1.1, and the notation specified in this section, Lemma 1.2 can be easily demonstrated.

Equation (1.20) specifies the homogeneous credibility estimators in the Bühlmann-Straub model. If X_{zw} is replaced by $\mu = \mathrm{E}[\mu(\Theta_j)]$, we obtain the inhomogeneous credibility estimators.

Theorem 1.3 If the Bühlmann-Straub assumptions (BS1) and (BS2) hold, then the optimal linearized inhomogeneous credibility estimator of $\mu(\Theta_j)$ is

$$\hat{\mu}(\Theta_j) = (1 - Z_j)\mu + Z_j X_{jw}\,, \tag{1.21}$$

where $\mu = \mathrm{E}[\mu(\Theta_j)]$ and Z_j is as in (1.19).

The derivation of (1.21) is similar to that of the inhomogeneous optimal credibility estimators in Bühlmann's classical model. There, the solution was provided by minimizing

$$\mathcal{R}_{nh}(c_0, \ldots, c_{jt}) = \mathrm{E}[\mu(\Theta_j) - c_0 - \sum_{r=1}^{t} c_{jr}X_{jr}]^2\,.$$

To prove Theorem 1.3, we would use the same technique. However, the number of claims associated with each claim amount random variable X_{jr}, is now no longer necessarily equal. Therefore, the term $a\sum_{r=1}^{t} c_{jr}$ is not equal to atc, but is instead equal to aw_jc_j. And the term $c_0 = (1 - Z_j)\,\mu$, is not uniform across contracts as in the classical Bühlmann case. For complete details, refer to Goovaerts et al. (1990).

In the homogeneous case, we have the following theorem.

Theorem 1.4 The solution of the following minimization problem,

$$\min_{c_{ji1}, c_{ji2}, \ldots, c_{jit}} \mathrm{E}\{[\mu(\Theta_j) - \sum_{j=1}^{k} \sum_{r=1}^{t} c_{jir} X_{ir}]^2\}\,,$$

for $i = 1, \ldots, k$ and for each $j = 1, \ldots, k$, subject to the constraint

$$\mathrm{E}[\mu(\Theta_j)] = \sum_{j=1}^{k} \sum_{r=1}^{t} c_{jir}\,\mathrm{E}[X_{ir}]\,,$$

is

$$\hat{\mu}(\Theta_j) = (1 - Z_j)\,X_{zw} + Z_j X_{jw}\,, \tag{1.22}$$

where Z_j is as in (1.19).

The proof of Theorem 1.4 can also be found in Goovaerts et al. (1990). As in the classical Bühlmann case, it is necessary that the minimization be accomplished over all linear combinations of claim amounts of the portfolio for each contract. Then, with an application of the Lagrange multiplier method to the unbiasedness restriction as the constraint equation, the result follows.

1.5 The Hachemeister Regression Model

The regression credibility model was originally proposed by Hachemeister (1975). Because the effects of inflation in claim figures had become a major problem in rate-making, Hachemeister developed a model which could be used to evaluate the credibility of state (or contract) trends against country (or portfolio) trends when estimating the expected severity of claims.

Viewing inflation as a factor which varies with time, Hachemeister proposed a simple linear regression model with time as the independent variable. In this credibility model, the net risk premiums are no longer time-independent, and inflation is modeled by a linear trend, which can be seen as an extension of the Bühlmann-Straub model.

The assumptions of the Hachemeister regression model are:

(H1) The risk parameters, $\Theta_1, \ldots, \Theta_j$ are independent and identically distributed. The contracts $j = 1, \ldots, k$ are independent.

(H2) $\mathrm{E}[\underline{X}_j|\Theta_j] = \mathbf{Y}_j\underline{\beta}(\Theta_j)$, for $j = 1, \ldots, k$, where $\underline{\beta}$ is an unknown regression vector of length n and $\mathrm{Cov}[\underline{X}_j|\Theta_j] = \sigma^2(\Theta_j)\,\mathbf{W}_j^{-1}$.

The design matrix $\mathbf{Y}_j$ has full column rank n and dimension $t \times n$, where n is an arbitrary value denoting the number of factors being considered. The matrix is chosen in advance and determines the type of trend that is modeled. For example, if we wish to model a quadratic trend, then the design matrix for each contract would have the same form,

$$\mathbf{Y}_j = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ \vdots & \vdots & \vdots \\ 1 & t & t^2 \end{bmatrix}.$$

In general, for each contract,

$$\mathbf{Y}_j = (\underline{Y}_{j1}, \ldots, \underline{Y}_{jt})',$$

where $\underline{Y}_{jr} = (y_{jr1}, y_{jr2}, \ldots, y_{jrn})$. In the foregoing example, $\underline{Y}_{jr} = (1, r, r^2)$.

The fixed weight matrix $\mathbf{W}_j^{-1}$ has dimension $t \times t$, and assumes the following form

$$\mathbf{W}_j^{-1} = \begin{bmatrix} 1/w_{j1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 1/w_{jt} \end{bmatrix},$$

where the w_{jr} for $r = 1, \ldots, t$, are the average number of claims for X_{jr}.

Since Hachemeister's model no longer requires that the conditional expectation of the claims be the same for all time periods, we write the the conditional expected value at time r as

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

The mean values of the model at different points in time for a given contract are represented by the $t \times 1$ vector

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

for $j = 1, \ldots, k$. The vector of regression coefficients is

$$\underline{\beta}(\Theta_j) = [\beta_0(\Theta_j), \ldots, \beta_{n-1}(\Theta_j)]',$$

so that $\underline{\mu}(\Theta_j) = \mathbf{Y}_j\underline{\beta}(\Theta_j)$, as in (H2).

Example 1.2 If $\mathbf{Y}_j = (1, \ldots, 1)'$ and $\mathbf{W}_j^{-1} = \mathrm{diag}(1/w_{j1}, \ldots, 1/w_{jt})$, Hachemeister's model reduces to Bühlmann and Straub's model, since

$$\mu_r(\Theta_j) = \beta_0(\Theta_j)$$

and

$$\mathrm{Var}[X_{jr}|\Theta_j] = \sigma^2(\Theta_j)/w_{jr}\,,$$

for $r = 1, \ldots, t$.

We now derive the credibility adjusted regression coefficients for a contract. Denote by $\underline{X}_j$, the vector of claim amount random variables $(X_{j1}, \ldots, X_{jt})$ for contract j. We will require the following definitions:

$$\mathrm{E}[\underline{\beta}(\Theta_j)] = \underline{\beta}\,,$$

$$\mathrm{Cov}[\underline{\beta}(\Theta_j)] = \mathbf{\Lambda}\,,$$

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

Lemma 1.3 The following relations hold:

(i) $\mathrm{E}[\underline{\mu}(\Theta_j)] = \mathrm{E}[\underline{X}_j] = \mathbf{Y}_j\underline{\beta}$,

(ii) $\mathrm{Cov}[\underline{\beta}(\Theta_j), \underline{X}_j'] = \mathbf{\Lambda}\,\mathbf{Y}_j'$,

(iii) $\mathrm{Cov}[\underline{\mu}(\Theta_j), \underline{X}_j'] = \mathrm{Cov}\{\mathrm{E}[\underline{X}_j|\Theta_j]\} = \mathbf{Y}_j\mathbf{\Lambda}\mathbf{Y}_j'$,

(iv) $\mathrm{E}\{\mathrm{Cov}[\underline{X}_j|\Theta_j]\} = s^2\,\mathbf{W}_j^{-1}$.

Proof To prove (i), note that $\mathrm{E}[\underline{\mu}(\Theta_j)] = \mathrm{E}[\underline{X}_j]$ is obvious since $\underline{\mu}(\Theta_j) = \mathrm{E}[\underline{X}_j|\Theta_j]$. Then,

$$\begin{aligned}\mathrm{E}[\underline{\mu}(\Theta_j)] &= \mathrm{E}[\mathbf{Y}_j\underline{\beta}(\Theta_j)] \\ &= \mathbf{Y}_j\underline{\beta}\,.\end{aligned}$$

We next prove (ii). Since

$$\begin{aligned}\mathrm{Cov}[\underline{\beta}(\Theta_j), \underline{X}_j'] &= \mathrm{Cov}[\underline{\beta}(\Theta_j), \underline{\beta}(\Theta_j)']\mathbf{Y}_j' \\ &= \mathbf{\Lambda}\,\mathbf{Y}_j'\,.\end{aligned}$$

To prove (iii), we use the proof of (ii),

$$\begin{aligned}\mathrm{Cov}[\underline{\mu}(\Theta_j), \underline{X}_j] &= \mathrm{Cov}[\mathbf{Y}_j\underline{\beta}(\Theta_j), \underline{X}_j'] \\ &= \mathrm{Cov}[\mathbf{Y}_j\underline{\beta}(\Theta_j)] \\ &= \mathbf{Y}_j\mathbf{\Lambda}\mathbf{Y}_j'\,.\end{aligned}$$

Finally, to prove (iv), we have

$$\begin{aligned}\mathrm{E}\{\mathrm{Cov}[\underline{X}_j|\Theta_j]\} &= \mathrm{E}[\sigma^2(\Theta_j)\,\mathbf{W}_j^{-1}] \\ &= s^2\,\mathbf{W}_j^{-1}.\end{aligned}$$
□

Suppose the claim severity random vector $\underline{X}_j$ for contract j, can be written as the sum of the mean claim amounts of the contract and some random error term, that is,

$$\begin{aligned}\underline{X}_j &= \underline{\mu}(\Theta_j) + \underline{\epsilon}_j \\ &= \mathbf{Y}_j\underline{\beta}(\Theta_j) + \underline{\epsilon}_j \qquad (1.23)\end{aligned}$$

where $\mathrm{E}[\underline{\epsilon}_j] = \underline{0}$, and

$$\begin{aligned}\mathrm{Cov}[\underline{\epsilon}_j] &= \mathrm{E}\{\mathrm{Cov}[\underline{X}_j - \underline{\mu}(\Theta_j)|\Theta_j]\} + \mathrm{Cov}\{\mathrm{E}[\underline{X}_j - \underline{\mu}(\Theta_j)|\Theta_j]\} \\ &= \mathrm{E}\{\mathrm{Cov}[\underline{X}_j|\Theta_j]\} \\ &= s^2\,\mathbf{W}_j^{-1}\,.\end{aligned}$$

Let $\mathbf{\Phi}_j = s^2\,\mathbf{W}_j^{-1}$. As a measure of accuracy, we wish to minimize the sum of squares

$$\mathcal{S}[\underline{\beta}(\Theta_j)] = [\underline{X}_j - \mathbf{Y}_j\underline{\beta}(\Theta_j)]'\,\mathbf{\Phi}_j^{-1}\,[\underline{X}_j - \mathbf{Y}_j\underline{\beta}(\Theta_j)]\,. \tag{1.24}$$

The weighted least-squares solution of $\mathcal{S}(\underline{\beta})$ is the individual estimator of $\underline{\beta}(\Theta_j)$:

$$\hat{\underline{\beta}}_j = (\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,\mathbf{Y}_j)^{-1}\,\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,\underline{X}_j \tag{1.25}$$

$$= (\mathbf{Y}_j'\,\mathbf{W}_j\,\mathbf{Y}_j)^{-1}\,\mathbf{Y}_j'\,\mathbf{W}_j\,\underline{X}_j\,, \tag{1.26}$$

with the latter equation being a consequence of assumption (H2).

Our optimal (credibility) estimator for $\underline{\beta}(\Theta_j)$ is to be restricted to estimators of the form

$$\hat{\underline{\mu}}(\Theta_j) = \underline{\gamma} + \mathbf{\Gamma}\,\underline{X}_j\,, \tag{1.27}$$

where $\underline{\gamma}$ is an arbitrary $t \times 1$ vector and $\mathbf{\Gamma}$ is an arbitrary $t \times t$ matrix. We seek to minimize the expected squared error

$$\mathcal{R}(\underline{\gamma}, \mathbf{\Gamma}) = \mathrm{E}\{[\underline{\beta}(\Theta_j) - \underline{\gamma} + \mathbf{\Gamma}\,\underline{X}_j]'\,[\underline{\beta}(\Theta_j) - \underline{\gamma} + \mathbf{\Gamma}\,\underline{X}_j]\}\,, \tag{1.28}$$

over all vectors $\underline{\gamma}$ and matrices $\mathbf{\Gamma}$ of appropriate dimensions. The first order derivatives of $\mathcal{R}(\underline{\gamma}, \mathbf{\Gamma})$ set equal to zero are

$$\frac{\partial}{\partial\underline{\gamma}}\mathcal{R} = \mathrm{E}[\underline{\beta}(\Theta_j) - \underline{\gamma} - \mathbf{\Gamma}\,\underline{X}_j] = \underline{0}\,,$$

and

$$\frac{\partial}{\partial\mathbf{\Gamma}}\mathcal{R} = \mathrm{E}\{[\underline{\beta}(\Theta_j) - \underline{\gamma} - \mathbf{\Gamma}\,\underline{X}_j]\,\underline{X}_j'\} = \underline{0}.$$

It follows that

$$E[\underline{\beta}(\Theta_j)] = \underline{\gamma} + \mathbf{\Gamma}\ E[\underline{X}_j]\,, \tag{1.29}$$

and

$$E[\underline{\beta}(\Theta_j)\,\underline{X}_j'] = \underline{\gamma}\,E[\underline{X}_j'] + \mathbf{\Gamma}\ E[\underline{X}_j\underline{X}_j']. \tag{1.30}$$

Multiplying (1.29) by $E[\underline{X}_j']$ and subtracting the result from (1.30), we obtain

$$\mathrm{Cov}[\underline{\beta}(\Theta_j), \underline{X}_j'] = \mathbf{\Gamma}\ \mathrm{Cov}[\underline{X}_j], \tag{1.31}$$

or

$$\mathbf{\Gamma} = \mathrm{Cov}[\underline{\beta}(\Theta_j), \underline{X}_j']\,\{\mathrm{Cov}[\underline{X}_j]\}^{-1}. \tag{1.32}$$

And from (1.29), we get

$$\underline{\gamma} = E[\underline{\beta}(\Theta_j)] - \mathrm{Cov}[\underline{\mu}(\Theta_j), \underline{X}_j']\,\{\mathrm{Cov}[\underline{X}_j]\}^{-1}\ E[\underline{X}_j]. \tag{1.33}$$

The general form of our optimal linear estimator then is

$$\hat{\underline{\beta}}(\Theta_j) = E[\underline{\beta}(\Theta_j)] + \mathrm{Cov}[\underline{\beta}(\Theta_j), \underline{X}_j']\,\{\mathrm{Cov}[\underline{X}_j]\}^{-1}\,\{\underline{X}_j - E[\underline{X}_j]\}. \tag{1.34}$$

Note that this general form does not depend on the assumptions (H1) and (H2). Equation (1.34) will be useful in deriving the Kalman filter in chapter 3. We now derive the credibility estimator of $\underline{\beta}(\Theta_j)$.

Theorem 1.5 (Hachemeister, 1975) The optimal linearized estimator of $\underline{\beta}(\Theta_j)$ is given by

$$\hat{\underline{\beta}}(\Theta_j) = (\mathbf{I} - \mathbf{Z}_j)\,\underline{\beta} + \mathbf{Z}_j\,\hat{\underline{\beta}}_j\,, \tag{1.35}$$

where

$$\mathbf{Z}_j = \mathbf{\Lambda}\,\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,\mathbf{Y}_j\,(\mathbf{I} + \mathbf{\Lambda}\,\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,\mathbf{Y}_j)^{-1}.$$

Proof From Lemma 1.3, (1.31) becomes

$$\mathbf{\Gamma}\,(\mathbf{\Phi}_j + \mathbf{Y}_j\,\mathbf{\Lambda}\,\mathbf{Y}_j') = \mathbf{\Lambda}\,\mathbf{Y}_j'\,. \tag{1.36}$$

If we post-multiply both sides of (1.36) by $\mathbf{\Phi}_j^{-1}\,\mathbf{Y}_j$, we find that

$$\mathbf{\Gamma}\,\mathbf{Y}_j\,(\mathbf{I} + \mathbf{\Lambda}\,\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,\mathbf{Y}_j) = \mathbf{\Lambda}\,\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,\mathbf{Y}_j\,.$$

Let $\mathbf{Z}_j = \mathbf{\Lambda}\,\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,\mathbf{Y}_j\,(\mathbf{I} + \mathbf{\Lambda}\,\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,\mathbf{Y}_j)^{-1}$, then

$$\mathbf{\Gamma}\,\mathbf{Y}_j = \mathbf{Z}_j. \tag{1.37}$$

Combining this with (1.36), we obtain

$$\mathbf{\Gamma}\,\mathbf{\Phi}_j + \mathbf{Z}_j\mathbf{\Lambda}\,\mathbf{Y}_j' = \mathbf{\Lambda}\,\mathbf{Y}_j'\,.$$

This immediately yields

$$\mathbf{\Gamma} = (\mathbf{I} - \mathbf{Z}_j)\,\mathbf{\Lambda}\,\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,. \tag{1.38}$$

We have, from (1.37) and (1.38), that

$$(\mathbf{I} - \mathbf{Z}_j)\,\mathbf{\Lambda}\,\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,\mathbf{Y}_j = \mathbf{Z}_j\,,$$

or

$$(\mathbf{I} - \mathbf{Z}_j)\,\mathbf{\Lambda} = \mathbf{Z}_j\,(\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,\mathbf{Y}_j)^{-1}.$$

Inserting this into (1.38) and post-multiplying the result by $\underline{X}_j$, we arrive at

$$\begin{aligned}\mathbf{\Gamma}\,\underline{X}_j &= \mathbf{Z}_j\,(\mathbf{Y}_j'\,\mathbf{\Phi}_j\,\mathbf{Y}_j)^{-1}\,\mathbf{Y}_j\,\mathbf{\Phi}_j^{-1}\,\underline{X}_j \\ &= \mathbf{Z}_j\,\underline{\hat{\beta}}_j.\end{aligned}$$

From (1.29),

$$\begin{aligned}\underline{\gamma} &= \underline{\beta} - \mathbf{\Lambda}\,\mathbf{Y}_j'\,(\mathbf{\Phi}_j + \mathbf{\Lambda}\,\mathbf{Y}_j')^{-1}\,\mathbf{Y}_j\,\underline{\beta} \\ &= \underline{\beta} - \mathbf{Y}_j\,\mathbf{Z}_j\,\underline{\beta} \\ &= (\mathbf{I} - \mathbf{Z}_j)\underline{\beta}\,.\end{aligned}$$

The final form of our estimator for $\underline{\beta}(\Theta_j)$ is

$$\hat{\underline{\beta}}(\Theta_j) = (\mathbf{I} - \mathbf{Z}_j)\,\underline{\beta} + \mathbf{Z}_j\,\hat{\underline{\beta}}_j, \tag{1.39}$$

with

$$\mathbf{Z}_j = \mathbf{\Lambda}\,\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,\mathbf{Y}_j\,(\mathbf{I} + \mathbf{\Lambda}\,\mathbf{Y}_j'\,\mathbf{\Phi}_j^{-1}\,\mathbf{Y}_j)^{-1},$$

and where $\hat{\underline{\beta}}_j$ is the estimator based on individual experience. □

Since the design matrix is assumed to be nonrandom, the estimator for $\mu(\Theta_j)$, for any $\mathbf{Y}_j$, is

$$\hat{\underline{\mu}}(\Theta_j) = \mathbf{Y}_j\,[(\mathbf{I} - \mathbf{Z}_j)\,\underline{\beta} + \mathbf{Z}_j\,\hat{\underline{\beta}}_j]. \tag{1.40}$$

1.6 Empirical Credibility

Our development of the credibility estimator, thus far, has been strictly theoretical. For practical application, certain parameters of the credibility estimators in, for example, equations (1.11), (1.15), and (1.40), need to be estimated. In this section we introduce estimators of the collective mean $\underline{\beta} = \mathrm{E}[\underline{\beta}(\Theta_j)]$ and the structural parameters $\mathbf{\Lambda} = \mathrm{Cov}[\underline{\beta}(\Theta_j)]$ and $\mathbf{\Phi}_j = \mathrm{E}\{\mathrm{Cov}[\underline{X}_j|\Theta_j]\}$ based on the collective data. When we recast the credibility estimator with the empirical estimators in place of the theoretical ones, we obtain an empirical credibility estimator.

An estimator of the collective mean $\underline{\beta}$ is

$$\hat{\underline{\beta}} = \left(\sum_{j=1}^{k} \mathbf{Z}_j\right)^{-1} \sum_{j=1}^{k} \mathbf{Z}_j\,\hat{\underline{\beta}}_j\,, \tag{1.41}$$

which is due to De Vylder (1981). This estimator is the solution to the minimization problem

$$\min_{\mathbf{F}_j} \mathrm{E}\left\{\left[\underline{\beta} - \sum_{j=1}^{k} \mathbf{F}_j\,\hat{\underline{\beta}}_j\right]'\mathbf{S}\left[\underline{\beta} - \sum_{j=1}^{k} \mathbf{F}_j\,\hat{\underline{\beta}}_j\right]\right\},$$

where $\sum_{j=1}^{k} \mathbf{F}_j = \mathbf{I}$ and $\mathbf{S}$ is a non-negative definite weighting matrix. The constraint that the sum of the $\mathbf{F}_j$ be equal to $\mathbf{I}$ guarantees the unbiasedness of the estimator since $\mathrm{E}[\hat{\underline{\beta}}_j] = \underline{\beta}$. Furthermore, under the restriction that $\hat{\underline{\beta}}$ converges to $\underline{\beta}$ in quadratic mean, we obtain also that $\hat{\underline{\beta}}$ is consistent. Since

$$\lim_{k\to\infty} \mathrm{E}[\hat{\underline{\beta}} - \underline{\beta}]^2 = 0,$$

Chebyshev's inequality with $\epsilon = \lambda\sqrt{\mathrm{E}[\hat{\underline{\beta}} - \underline{\beta}]^2}$ yields

$$\Pr[|\hat{\underline{\beta}} - \underline{\beta}| > \epsilon] \leq \frac{1}{\epsilon^2}\,\mathrm{E}[\hat{\underline{\beta}} - \underline{\beta}]^2.$$

Thus $\hat{\underline{\beta}}$ converges weakly to $\underline{\beta}$ as $k \to \infty$.

With different choices of the weighting matrices $\mathbf{F}_j$, we are able to obtain other estimators of $\underline{\beta}$. For example, Hachemeister selected the matrices

$$\mathbf{F}_j = \left(\sum_{j=1}^{k} \mathbf{Y}_j\,\mathbf{W}_j^{-1}\,\mathbf{Y}_j\right)^{-1} \mathbf{Y}_j\,\mathbf{W}_j^{-1}\,\mathbf{Y}_j.$$

For the parameters $\mathbf{\Lambda} = \mathrm{Cov}[\underline{\beta}(\Theta_j)]$ and $\mathbf{\Phi}_j = s^2\,\mathbf{W}_j^{-1}$, De Vylder (1981) suggested

$$\hat{\mathbf{\Lambda}} = \frac{1}{k-1}\sum_{j=1}^{k} \mathbf{Z}_j\,(\hat{\underline{\beta}}_j - \hat{\underline{\beta}})(\hat{\underline{\beta}}_j - \hat{\underline{\beta}})' \tag{1.42}$$

and $\hat{\mathbf{\Phi}}_j = \hat{s}^2\,\mathbf{W}^{-1}$, where

$$\hat{s}^2 = \frac{1}{k(t-n)}\sum_{j=1}^{k} (\underline{X}_j - \mathbf{Y}_j\,\hat{\underline{\beta}}_j)'\mathbf{W}_j^{-1}(\underline{X}_j - \mathbf{Y}_j\,\hat{\underline{\beta}}_j)\,. \tag{1.43}$$

The estimator $\hat{\mathbf{\Phi}}_j$ is the average of the individual weighted sum of squared residuals, while $\hat{\mathbf{\Lambda}}$ is the credibility-weighted average of the covariance matrices of the individual regression estimators $\hat{\underline{\beta}}_j$.

Since the estimators for the structural parameters contain parameters which have yet to be estimated, they are called *pseudo-estimators*. In practice, the empirical estimators $\hat{\underline{\beta}}$ and $\hat{\mathbf{\Lambda}}$ require an iterative procedure to obtain a numerical result as we would need to replace the $\mathbf{Z}_j$ in (1.41) and (1.42) with $\hat{\mathbf{Z}}_j$, which depend on values yet to be computed. Note that in this case, $\hat{\underline{\beta}}$ and $\hat{\mathbf{\Lambda}}$ are no longer necessarily unbiased. Also, as $\mathbf{\Lambda}$ is non-negative definite and symmetric, $(\hat{\mathbf{\Lambda}} + \hat{\mathbf{\Lambda}}')/2$ is used in place of $\hat{\mathbf{\Lambda}}$ at each iteration to yield a symmetric, but not necessarily non-negative definite, estimate.

Chapter 2

Robust Credibility Models

In credibility, the need for robust statistical methods arises due to larger than normal claims. In credibility, departure from assumptions are less of a concern than outlying claims. If excess claims occur, the variance of claims within a contract will increase, leading to a small or zero credibility factor even for contracts which did not incur an excess claim. For these contracts, the credibility premium would mostly consist of the average over the entire portfolio. In the case of the contract which did incur a large claim, since the mean of the contract is very sensitive to outliers, the effect of a large claim would be to exaggerate the expected claim amount of the next period. This large individual premium will offset the small credibility factor, leading to a credibility premium which is too high.

Early treatments of robust methods in credibility theory can be found in Gisler (1980) and Klugman (1985). In this chapter, we review some results in robust statistics and the robust credibility models of Künsch (1992), Gisler and Reinhard (1993), and Kremer (1991).

2.1 Robust Statistics

Robust statistics is an extension of classical parametric statistics. In theories of classical parametric statistics, optimal procedures are derived under exact parametric models. In robust statistics, models are assumed to be only approximately valid. Thus, procedures are developed with the intention that they be optimal in a neighbourhood of strict parametric models. In this section, we review some results from robust statistics which we will use later on. Further coverage of robust statistics can be found in Huber (1981) and Hampel et al. (1984).

Suppose we have some functional $T(F_\theta)$, where F_θ is a family of probability distributions of some parametric model. Let Δ_x be the probability measure which puts mass 1 at the point x. Let Ω be the sample space under consideration. If Δ_x is in the domain of T, we define the influence function as the following:

Definition 2.1 The influence function (IF) of T at F is given by

$$IF(x;T,F) = \lim_{\epsilon \downarrow 0} \frac{T[(1-\epsilon)F + \epsilon\Delta_x] - T(F)}{\epsilon} \tag{2.1}$$

in those $x \in \Omega$, where this limit exists.

The influence function was introduced by Hampel (1968). In (2.1), ϵ is the percentage of contamination in the population of $F(\cdot)$. Thus, as the amount of contamination approaches zero, the influence function describes the effect of an infinitesimal contamination of the point x on the estimate, divided by the mass of contamination.

The influence function is related to the Gâteaux derivative of T. The functional T is Gâteaux differentiable at F in the domain of T, if there exists a real function $h(x)$ such that for all G in the domain of T the following holds:

$$\frac{\partial}{\partial t} T[(1-\epsilon)F + \epsilon G]_{t=0} = \int h(x)\, dG(x), \tag{2.2}$$

When $G = \Delta_x$, we obtain the influence function $IF(x;T,F)$. If $G = F$, then

$$\int h(x)\, dF(x) = 0. \tag{2.3}$$

Thus, with $h(x) = IF(x; T, F)$, the first-order von Mises expansion of T at F evaluated at G is given by

$$T(G) = T(F) + \int IF(x; T, F)\, d(G - F)(x) + \text{remainder}. \tag{2.4}$$

The empirical distribution function for the random variables $X_1, \ldots, X_n$ is

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x]}(X_i), \tag{2.5}$$

where n is the number of observations and $I_{[A]}$ is the indicator function of the set A. If the observations X_i are iid, then by the Glivenko-Cantelli theorem, $\sup_x |F_n(x) - F(x)| \to 0$ with probability 1, so the empirical distribution function F_n will converge to F with probability 1. Let $G = F_n$ in (2.4), so that we obtain

$$T(F_n) = T(F) + \int IF(x; T, F)\, d(F_n - F)(x) + \text{remainder}.$$

The remainder term will tend to zero as $n \to \infty$ in most cases, hence

$$n^{-1/2} \left\{ [T(F_n) - T(F)] - \int IF(x; T, F)\, dF_n(x) \right\}$$

converges to zero in probability.

Evaluating the integral for a sample X_i, $i = 1, \ldots, n$, we obtain

$$\sqrt{n}\, [T(F_n) - T(F)] \approx n^{-1/2} \sum_{i=1}^{n} IF(X_i; T, F). \tag{2.6}$$

The expression on the right is the sum of n independent and identically distributed random variables. Therefore, by the Central Limit Theorem, the term on the right is asymptotically normal. This limiting result thus is obtained also for the left side of (2.6). From (2.3), $\int IF(x; T, F)\, dF = 0$, consequently, $\sqrt{n}[T(F_n) - T(F)]$ is asymptotically normal with mean zero and variance

$$V(T, F) = \int IF(x; T, F)^2\, dF. \tag{2.7}$$

We now define an M-estimator. The maximum likelihood estimator is defined as the value $T_n = T_n(X_1, \ldots, X_n)$ which maximizes $\prod_{i=1}^{n} f_{T_n}(X_i)$, or equivalently,

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

Huber (1964) generalized this to

$$\sum_{i=1}^{n} \rho(X_i, T) = \min!, \tag{2.8}$$

where ρ is an arbitrary function. If ρ has a derivative $\psi(x, \theta) = (\partial/\partial\theta)\rho(x, \theta)$, then T_n must satisfy

$$\sum_{i=1}^{n} \psi(X_i, T_n) = 0\,. \tag{2.9}$$

Definition 2.2 An M-estimator, T_n, is defined implicitly as the solution of either equation (2.8) or (2.9).

To derive the influence function of an M-estimator, define $T(F)$ by

$$\int \psi[x; T(F)]\, dF(x) = 0, \tag{2.10}$$

and insert $F_t = (1-\epsilon)F + \epsilon G$ for F. Then, differentiating with respect to ϵ at $\epsilon = 0$ and solving for the influence function yields

$$IF(x; T, F) = \frac{\psi[x, T(F)]}{-\int (\partial/\partial\theta)[\psi(y, \theta)]_{T(F)} dF(y)}\,. \tag{2.11}$$

From (2.7), the asymptotic variance is

$$V(T, F) = \frac{\psi[x, T(F)]^2}{-\left\{\int (\partial/\partial\theta)[\psi(y, \theta)]_{T(F)} dF(y)\right\}^2}\,. \tag{2.12}$$

2.2 Künsch's Model

Künsch's model (Künsch, 1992) is a robustified version of Bühlmann's classical model. Taking the same assumptions as the Bühlmann homogeneous case (c.f. section 1.4), Künsch proposed to replace $\bar{X}_j - \bar{X}$ in

$$\begin{aligned}\hat{\mu}(\Theta_j) &= (1-Z)\,\bar{X} + Z\bar{X}_j\,,\\ &= \bar{X} + Z\,(\bar{X}_j - \bar{X})\,,\end{aligned}$$

with $T_j - \bar{T}$, to get

$$\hat{\mu}^R(\Theta_j) = \bar{X} + Z\,(T_j - \bar{T})\,, \tag{2.13}$$

where T_j is a robust estimator of the mean for contract j, and where $\bar{T} = \frac{1}{k}\sum_{j=1}^{k} T_j$.

The robust estimator T_j is defined as the implicit solution of

$$\sum_{r=1}^{t} \psi(X_{jr}/T_j) = 0\,, \tag{2.14}$$

with $\psi(z) = \max[-c_1, \min(z-1, c_2)]$, where $0 < c_1 \leq 1$ and $c_2 > 0$.

Since $\mathrm{E}[T_j] = \mathrm{E}[\bar{T}]$, the robust credibility premium is unbiased. A scale estimator is used to take into account the non-negativity of the claim amounts and also so that larger mean values will result in larger variances. The ψ-function that we use here will result in an estimate of the claim amount at time r for contract j such that the percentage amount by which the observed claim amount X_{jr} exceeds the robust estimator T_j lies in the interval $[-c_1, c_2]$. Thus, claims are truncated at both ends. According to Künsch, the choice of c_1 and c_2 is not very crucial. He suggests $c_1 = c_2 = 1$ for small samples and $c_1 = 1, c_2 = 1.5$ or 2 for moderate samples. We notice that if $c_1 = 1$ and $c_2 = \infty$, then (2.13) reduces to the non-robust linear credibility premium.

An algorithm to solve for T_j in (2.14) can be developed by considering $\psi(z)$. We first define $\tilde{\psi}(z) = \psi(z) + 1$. Since we can also write $\psi(z) = \max[1 - c_1, \min(z, 1 + c_2)] - 1$, we have

$$\tilde{\psi}(z) = \max[1 - c_1, \min(z, 1 + c_2)].$$

Then $\sum_{r=1}^{t} \psi(X_{jr}/T_j) = \sum_{r=1}^{t} \tilde{\psi}(X_{jr}/T_j) - t = 0$, or $\frac{1}{k}\sum_{r=1}^{t} \tilde{\psi}(X_{jr}/T_j) = 1$. Hence

$$T_j^{(m+1)} = \left[\frac{1}{t}\sum_{r=1}^{t} \tilde{\psi}(X_{jr}/T_j^{(m)})\right]^{1/2} T_j^{(m)}.$$

The convergence of the iterative algorithm follows from Huber (1981) section 8.6.

The credibility factor is given by

$$Z = \frac{\mathrm{Cov}\{\mathrm{E}[T_j|\Theta_j], \mu(\Theta_j)\}}{\mathrm{E}\{\mathrm{Var}[T_j|\Theta_j]\} + \mathrm{Var}\{\mathrm{E}[T_j|\Theta_j]\}}. \tag{2.15}$$

The denominator is equal to $\mathrm{Var}[T_j]$. Therefore, in order to obtain an empirical credibility factor based on robust statistics, we require the variance of T_j. An unbiased estimator of the denominator is

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

We can estimate $\mathrm{Cov}[T_j, \bar{X}_j]$ using

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

We know that $\mathrm{Cov}[T_j, \bar{X}_j] = \mathrm{Cov}\{\mathrm{E}[T_j|\Theta_j], \mathrm{E}[\bar{X}_j|\Theta_j]\} + \mathrm{E}\{\mathrm{Cov}[T_j, \bar{X}_j]|\Theta_j\}$, so we need an estimator for $\mathrm{E}\{\mathrm{Cov}[T_j, \bar{X}_j]|\Theta_j\}$. The derivative of $\psi[x/T(F)]$ is given by

$$\frac{\partial}{\partial T}\psi[x/T(F)] = -\psi'\left[\frac{x}{T(F)}\right]\frac{x}{T(F)^2}$$

The function $\psi'[x/T(F)]$ will be equal to 1 in the interval $((1-c_1)T(F), (1+c_2)T(F))$, so from (2.11), we have

$$IF(x; T, F) = \frac{\psi[x/T(F)]T(F)^2}{\int x dF(x)} .$$

An estimator for the influence function then is

$$\widehat{IF}(X_{jr}, T_j) = \frac{\psi(X_{jr}/T_j)T_j^2\, t}{\sum_{r=1}^{t} X_{jr} I_{[(1-c_1)T_j \le X_{jr} \le (1+c_2)T_j]}},$$

The linearization (2.6) then suggests that we use

$$\frac{1}{t(t-1)k}\sum_{j=1}^{k}\sum_{r=1}^{t}\widehat{IF}(X_{jr}, T_j)(X_{jr} - \bar{X}_j)$$

to estimate $\mathrm{E}\{\mathrm{Cov}[T_j, \bar{X}_j]|\Theta_j\}$. Finally,

$$\hat{Z} = \frac{\frac{1}{k-1}\sum_{j=1}^{k}(T_j - \bar{T})(\bar{X}_j - \bar{X}) - \frac{1}{t(t-1)k}\sum_{j=1}^{k}\sum_{r=1}^{t}\widehat{IF}(X_{jr}, T_j)(X_{jr} - \bar{X}_j)}{\frac{1}{k-1}\sum_{j=1}^{k}(T_j - \bar{T})(\bar{X}_j - \bar{X})} . \tag{2.16}$$

2.3 Gisler and Reinhard's Model

Gisler and Reinhard's treatment of robustness in credibility (Gisler and Reinhard, 1993) resulted in a robustified version of Bühlmann and Straub's credibility model. The assumptions of the Bühlmann-Straub model are used again here (see section 1.5). Gisler and Reinhard proposed to partition the credibility estimation of the individual mean into two parts. The first part consists of the "ordinary part", $\mu_o(\Theta_j)$; the second, the excess (outlying) part, $\mu_{xs}(\Theta_j)$. The robust credibility premium can then be expressed as

$$\mu^R(\Theta_j) = \mu_o(\Theta_j) + \mu_{xs}(\Theta_j)\,. \tag{2.17}$$

The ordinary part of the individual premium is defined as

$$\mu_o(\Theta_j) = \mathrm{E}[T_j|\Theta_j], \tag{2.18}$$

where T_j is a robust statistic for contract j. The excess part is defined as

$$\mu_{xs}(\Theta_j) = \mu_{xs}\,.$$

Thus, all risks in the portfolio are assumed to be equally exposed to outlying events.

We write as the ordinary part of the robust credibility premium

$$\hat{\mu}_o(\Theta_j) = \mu_{T_j} + Z_j(T_j - \mu_{T_j}), \tag{2.19}$$

where $\mu_{T_j} = \mathrm{E}[T_j]$. Generalizing Künsch's method in the previous section, the robust estimator T_j is implicitly defined as

$$\sum_{r=1}^{t} w_{jr}\psi(X_{jr}/T_j) = 0, \tag{2.20}$$

where $\psi(z) = \min(z-1, cw_{jr}^{-1/2})$.

This ψ-function has a single truncation point which depends on the amount of exposure at time r for contract j. We use $cw_{jr}^{-1/2}$ since $\mathrm{Var}[X_{jr}|\Theta_j] = \sigma^2(\Theta_j)/w_{jr}$. Two choices for c are $c = \sqrt{\mathrm{median}_{j,r}(w_{jr})}$ and $c = \sqrt{\bar{w}}$ where $\bar{w} = \frac{1}{kt}\sum_{j=1}^{k}\sum_{r=1}^{t} w_{jr}$.

Now,

$$\sum_{r=1}^{t} w_{jr} \min\left(\frac{X_{jr}}{T_j} - 1, cw_{jr}^{-1/2}\right) = \sum_{j=1}^{k} w_{jr} \min\left(\frac{X_{jr}}{T_j}, 1 + cw_{jr}^{-1/2}\right) - 1$$
$$= \sum_{r=1}^{t} \{w_{jr} \min[X_{jr}, (1 + cw_{jr}^{-1/2})T_j] - T_j\}\,.$$

Hence,

$$\sum_{r=1}^{t} w_{jr} \min[X_{jr}, (1 + cw_{jr}^{-1/2})T_j] = w_j T_j,$$

with $w_j = \sum_{r=1}^{t} w_{jr}$. Let $c_{jr} = 1 + cw_{jr}^{-1/2}$, so that

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

After we have computed T_j, define as the ordinary loss ratio $T_{jr} = \min(X_{jr}, c_{jr}T_j)$, then the excess loss ratio is $XS_{jr} = X_{jr} - T_{jr}$.

The credibility factor for this model is

$$Z_j = \frac{w_j \operatorname{Var}[\mu_T(\Theta_j)]}{\operatorname{E}\{\operatorname{Var}[T_j|\Theta_j]\} + w_j \operatorname{Var}[\mu_{T_j}(\Theta_j)]}, \tag{2.21}$$

where $\mu_{T_j}(\Theta_j) = \operatorname{E}[T_j|\Theta_j]$. The asymptotic variance of T_j will be required in order to estimate $\operatorname{Var}[T_{jr}|\Theta_j]$. The empirical influence function corresponding to our ψ-function is

$$\widehat{IF}(X_{jr}, T_j) = \frac{T_{jr} - T_j}{1 - \sum_{r=1}^{t} (w_{jr}/w_j) c_{jr} I_{[T_{jr} \neq X_{jr}]}}. \tag{2.22}$$

We take $\operatorname{Var}[T_{jr}|\Theta_j] \approx w_{jr}^{-1} V(T, F_{\Theta_j})$, then

$$\hat{s}_j^2 = \frac{(n-1)^{-1} \sum_{r=1}^{t} (T_{jr} - T_j)^2}{\left[1 - \sum_{r=1}^{t} (w_{jr}/w_j) c_{jr} I_{[T_{jr} \neq X_{jr}]}\right]^2}. \tag{2.23}$$

Let $\hat{s}_T^2 = \frac{1}{k} \sum_{j=1}^{k} \hat{s}_j^2$ be the estimator for the expected variance of T_j. We can use $\hat{s}_T^2$ to estimate a_T, the variance of $\mu_{T_j}(\Theta_j)$. That is,

$$\hat{a}_T = \frac{1}{\alpha} \left\{ \sum_{j=1}^{k} \frac{w_j}{w} (T_j - T)^2 - (k-1) \frac{\hat{s}_T^2}{w} \right\}, \tag{2.24}$$

where $\alpha = w^{-1} \sum_{j=1}^{k} w_j[1-(w_j/w)]$. Note that $\hat{a}_T$ may become negative since it is defined as a difference of two quantities.

Define the robust collective mean to be

$$\hat{\mu}_T = \left(\sum_{j=1}^{k} \hat{Z}_j \right)^{-1} \sum_{j=1}^{k} \hat{Z}_j T_j, \tag{2.25}$$

where

$$\hat{Z}_j = \frac{w_j \hat{a}_T}{(w_j \hat{a}_T + \hat{s}_T^2)}. \tag{2.26}$$

The excess collective mean then is

$$\hat{\mu}_{xs} = \frac{1}{w} \sum_{j=1}^{k} w_j XS_j, \tag{2.27}$$

with $XS_j = (1/w_j) \sum_{r=1}^{t} w_{jr} XS_{jr}$. Finally,

$$\hat{\mu}^R(\Theta_j) = \hat{\mu}_{xs} + \hat{\mu}_T + \hat{Z}_j(T_j - \hat{\mu}_T) \tag{2.28}$$

is the empirical robust credibility premium for contract j.

2.4 Kremer's Robust Regression Model

The treatment of large claims in the case of regression credibility by Kremer (1991) starts by taking the credibility adjusted estimator for $\underline{\mu}(\Theta_j)$:

$$\underline{\hat{\mu}}(\Theta_j) = \mathbf{Y}_j \,[(\mathbf{I} - \mathbf{Z}_j)\, \underline{\beta} + \mathbf{Z}_j \, \underline{\hat{\beta}}_j]. \tag{2.29}$$

The weighted least-squares estimator for the individual claims $\underline{\hat{\beta}}_j$ is then replaced by a more general estimator $\underline{B}_j$, where $\underline{B}_j$ is robust. Then

$$\underline{\hat{\mu}}^R(\Theta_j) = \mathbf{Y}_j \,[(\mathbf{I} - \mathbf{Z}_j)\, \underline{\beta} + \mathbf{Z}_j \, \underline{B}_j]. \tag{2.30}$$

At this point, only assumption (H2) of Hachemeister's regression model is required. In order to determine the optimal matrices $\mathbf{Z}_j$, the risk

$$\mathrm{E}\{\mathbf{Y}_j[(\mathbf{I} - \mathbf{Z}_j)\underline{\beta} + \mathbf{Z}_j \underline{B}_j] - \underline{\mu}(\Theta_j)\}^2,$$

is minimized with respect to $\mathbf{Z}_j$. Kremer proves that

$$\mathbf{Z}_j = \mathbf{A}(\mathbf{A}+\mathbf{R}+\mathbf{N})^{-1}, \tag{2.31}$$

where $\mathbf{A} = \mathrm{Cov}[\underline{\beta}(\Theta_j), \underline{B}_j]$, $\mathbf{R} = \mathrm{E}\{\mathrm{Cov}[\underline{B}_j|\Theta_j]\}$, and $\mathbf{N} = \mathrm{E}\{\underline{B}(\Theta_j)[\underline{B}(\Theta_j) - \underline{\beta}(\Theta_j)]'\}$.

Turning now to the problem of deriving the robust estimator of the individual claim amounts $\underline{B}_j$, the following sum of squared residuals is considered

$$\mathcal{S}[\underline{\beta}(\Theta_j)] = [\underline{X}_j - \mathbf{Y}_j\underline{\beta}(\Theta_j)]\mathbf{\Phi}_j^{-1}[\underline{X}_j - \mathbf{Y}_j\underline{\beta}(\Theta_j)]', \tag{2.32}$$

where $\mathbf{\Phi} = \mathrm{E}\{\mathrm{Cov}[\underline{X}_j|\Theta_j]\}$. Let $\mathbf{\Phi}_j^{-1} = \mathbf{Q}_j'\mathbf{Q}_j$, where $\mathbf{Q}_j$ is a $t \times t$ matrix. Then, the weighted sum of squared residuals is given by

$$\mathcal{S}[\underline{\beta}(\Theta_j)] = \sum_{i=1}^{t} \left\{ \sum_{r=1}^{t} q_{ir}^{(j)} [X_{jr} - \underline{Y}_{jr}\underline{\beta}(\Theta_j)] \right\}^2 .$$

To get a robust estimator, squared deviation is replaced with a general $\rho(\cdot)$, for example, the one-sided Huber function:

$$\rho_H(x) = \begin{cases} x^2/2 & \text{if } x < c, \\ c & \text{if } x \geq c \end{cases} \tag{2.33}$$

and the corresponding ψ function:

$$\psi_H(x) = \begin{cases} x & \text{if } x < c, \\ c & \text{if } x \geq c. \end{cases} \tag{2.34}$$

Then the optimization problem becomes

$$\sum_{i=1}^{t} \rho \left[\sum_{r=1}^{t} q_{ir}^{(j)} (X_{jr} - \underline{Y}_{jr}\underline{B}_j) \right] = \min! \, . \tag{2.35}$$

If the derivative ψ' of ρ exists, then $\underline{B}_j$ must satisfy

$$\sum_{i=1}^{t} \psi \left[\sum_{r=1}^{t} q_{ir}^{(j)} (X_{jr} - \underline{Y}_{jr}\underline{B}_j) \right] \sum_{r=1}^{t} q_{ir}\underline{Y}_{jr'} = 0 \, . \tag{2.36}$$

for $r' = 1, \ldots, n$. Solving for $\underline{B}_j$, the robust estimator of $\underline{\mu}(\Theta_j)$ is

$$\hat{\underline{\mu}}^R(\Theta_j) = \mathbf{Y}_j\{\mathbf{A}(\mathbf{A}+\mathbf{R}+\mathbf{N})^{-1}\,\hat{\underline{B}}_j + [\mathbf{I} - \mathbf{A}(\mathbf{A}+\mathbf{R}+\mathbf{N})^{-1}]\underline{\beta}\}. \tag{2.37}$$

In order to implement this robust regression credibility model, some empirical estimators are required. Similar to the non-robust regression case, Kremer estimates $\underline{\beta}$ by

$$\hat{\underline{\beta}} = \sum_{j=1}^{k} \mathbf{F}_j\,\hat{\underline{B}}_j\,, \tag{2.38}$$

where $\sum_{j=1}^{k} \mathbf{F}_j = \mathbf{I}$. Then from (2.32) and assumption (H2) of the Hachemeister model, let $\mathbf{\Phi}_j = s^2\,\mathbf{W}_j^{-1}$, where s^2 and $\mathbf{W}_j$ are defined as in section 1.5. Next, the weights $\mathbf{W}_j$ are factored into

$$\mathbf{W}_j = \mathbf{P}'_j\,\mathbf{P}_j\,,$$

so that $\mathbf{\Phi}_j^{-1} = (s^{-1}\,\mathbf{P}_j)'(s^{-1}\,\mathbf{P}_j)$. Rewrite (2.35) as

$$\sum_{i=1}^{t} \rho\left[\sum_{r=1}^{t} p_{ir}^{(j)}(X_{jr} - \underline{Y}_{jr}\underline{B}_j)/s\right] = \min!\,. \tag{2.39}$$

Kremer assumes $s = 1$, and if the derivative ψ of ρ exists, we have

$$\sum_{i=1}^{t} \psi\left[\sum_{r=1}^{t} p_{ir}^{(j)}(X_{jr} - \underline{Y}_{jr}\underline{B}_j)\right] \sum_{r=1}^{t} p_{ir}\underline{Y}_{jr'} = 0\,. \tag{2.40}$$

To estimate $\mathbf{\Lambda} = \mathbf{A} + \mathbf{R} + \mathbf{N}$, note that since $\mathbf{\Lambda} = \mathrm{Cov}[\underline{B}_j]$, it can be estimated by

$$\hat{\mathbf{\Lambda}} = \sum_{j=1}^{k} \mathbf{F}_j\,(\hat{\underline{B}}_j - \hat{\underline{\beta}})(\hat{\underline{B}}_j - \hat{\underline{\beta}})'\,. \tag{2.41}$$

One notes that $\mathbf{N} \to \mathbf{0}$ when $\underline{B}(\Theta_j)$ is close to $\underline{\beta}(\Theta_j)$. Therefore, the zero matrix is taken as estimator for $\mathbf{N}$. The expected covariance of $\underline{B}_j$ given Θ_j is given by the matrix $\mathbf{R}$. Then by (2.12),

$$\hat{\mathbf{R}} = \sum_{j=1}^{k} \mathbf{F}_j\,(\mathbf{Y}'_j\,\mathbf{W}_j\,\mathbf{Y}_j)^{-1}\,\frac{\frac{1}{t-n}\sum_{i=1}^{t}\left\{\psi\left[\sum_{r=1}^{t} p_{ir}^{(j)}\,(X_{jr} - \underline{Y}_{jr}\,\hat{\underline{B}}_j)\right]\right\}^2}{\frac{1}{t}\left\{\sum_{i=1}^{t}\psi'\left[\sum_{r=1}^{t} p_{ir}^{(j)}\,(X_{jr} - \underline{Y}_{jr}\,\hat{\underline{B}}_j)\right]\right\}^2}\,, \tag{2.42}$$

where the derivative ψ' of ψ is with respect to the $\underline{B}_j$. Finally, $\mathbf{A}$ is estimated by

$$\hat{\mathbf{A}} = \hat{\mathbf{\Lambda}} - \hat{\mathbf{R}}\,.$$

Remark 2.1 The foregoing assumes a general $\mathbf{\Phi}_j$. With Hachemeister's assumptions, (2.32) becomes

$$\mathcal{S}[\underline{\beta}(\Theta_j)] = [\underline{X}_j - \mathbf{Y}_j\underline{\beta}(\Theta_j)]\mathbf{\Phi}_j^{-1}[\underline{X}_j - \mathbf{Y}_j\underline{\beta}(\Theta_j)]', \tag{2.43}$$

where $\mathbf{\Phi}_j = s^2\,\mathbf{W}_j^{-1}$, but with $\mathbf{W}_j = \mathrm{diag}(w_{j1}, \ldots, w_{jt})$. Then, (2.40) can be written more simply as

$$\sum_{r=1}^{t} \psi\left[\frac{\sqrt{w_{jr}}(X_{jr} - \underline{Y}_{jr}\,\underline{B}_j)}{s}\right] \sqrt{w_{jr}}\,\underline{Y}_{jr} = 0\,. \tag{2.44}$$

Remark 2.2 Kremer notes that if $\underline{B}_j$ is the weighted least-squares estimator, $\mathbf{R} = (\mathbf{Y}_j'\,\mathbf{V}^{-1}\,\mathbf{Y}_j)^{-1}$ and

$$\mathbf{Z}_j = \mathbf{A}\mathbf{R}^{-1}(\mathbf{I} + \mathbf{A}\mathbf{R}^{-1})^{-1}\,,$$

which is the credibility factor under Hachemeister's model.

Remark 2.3 Equation (2.41) is the general case of (1.42). Therefore, we can take equation (2.41), with possibly a different choice of $\mathbf{F}_j$, as an alternative estimator for $\mathbf{A}$. In equation (1.42), we had $\mathbf{F}_j = \mathbf{Z}_j/(k-1)$.

Chapter 3

The Kalman Filter Applied to Credibility

In this chapter, we discuss the Kalman filter and its application to credibility theory. Connections between credibility theory and the Kalman filter were first investigated by Mehra (1975). De Jong and Zehnwirth (1983) then formulated some famous credibility models as Kalman filters.

We first derive the filter from results of section 1.5. We then embed the Bühlmann-Straub and Hachemeister models within the Kalman filter framework. We also review two empirical implementations of the Kalman filter as applied to credibility models. Finally, we describe a robust Kalman filter by Cipra and Romera (1991) and present Kremer's (1994) robust credibility model based on a robust Kalman filter.

3.1 The Discrete Kalman Filter

The Kalman filter (Kalman, 1960) is a recursive technique which is used to estimate the state of a linear dynamic system from measurement data corrupted by noise. In what follows, we will consider only discrete systems, that is, we will assume that measurements are observed at equally spaced points in time. The continuous-time analog of the discrete Kalman filter is usually referred to as the Kalman-Bucy filter (Kalman and Bucy, 1961). We refer to filtering as the estimation of the state $\underline{S}_t$ when the time of the desired estimate coincides with the time of the last measurement. In other words, given the sequence of observations $\underline{X}^\tau = \{X_1, \ldots, X_\tau\}$, we wish to estimate $\underline{S}_t$ when $t = \tau$. In other cases, we may have either a smoothing problem ($t < \tau$) or a prediction problem ($t > \tau$).

The Kalman filter is based on a state space model. We regard the *state* of a system as the least amount of information about the past that is needed to predict the description of the system at a future point in time. In the following formulation, the state of the system is described by a linear difference equation. Thus, it is sufficient to know the current state of the process in order to predict the state at any other point in time.

The unknown state of the system at time t is denoted by $\underline{S}_t$, and is referred to as the *state vector*. The measurements $\underline{X}_t$ consist of linear combinations of the state variables corrupted by a sequence of uncorrelated random errors $\underline{u}_t$, which have mean $\mathrm{E}[\underline{u}_t] = \underline{0}$ and covariance matrix $\mathrm{E}[\underline{u}_t\underline{u}_t'] = \mathrm{U}_t$. In state space form, we write the measurement equation and system equation, respectively, as

$$\underline{X}_t = \mathbf{H}_t\underline{S}_t + \underline{u}_t \tag{3.1}$$

and

$$\underline{S}_t = \mathbf{A}_t\underline{S}_{t-1} + \underline{v}_t \; . \tag{3.2}$$

The matrix $\mathbf{H}_t$, which is known at time t, describes the linear combinations of the state variables which make up $\underline{X}_t$. We assume that the sequence of system errors $\underline{v}_t$, has mean vector $\mathrm{E}[\underline{v}_t] = \underline{0}$ and covariance matrix $\mathrm{E}[\underline{v}_t\underline{v}_t'] = \mathbf{V}_t$, where $\underline{v}_t$ is independent of the observation errors $\underline{u}_t$. Furthermore, from the independence assumption, $\mathrm{E}[\underline{u}_t\underline{v}_{t-s}'] = 0$ for all natural s. Finally, it is assumed that the system matrix $\mathbf{A}_t$ and the covariance matrix of both the observation errors and the system errors are known in advance.

At time t, we have observations up to time $t-1$. As above, denote the t past observations $\{\underline{X}_{t-1}, \underline{X}_{t-2}, \ldots, \underline{X}_0\}$, by $\underline{X}^{t-1}$. Also, let $\hat{\underline{S}}_{t|t-1} = \mathrm{E}[\underline{S}_t|\underline{X}^{t-1}]$ be the estimator of $\underline{S}_t$ at time t, given observations up to time $t-1$. After observing $\underline{X}_t$, we would like to update our estimate of $\underline{S}_t$; thus, we seek $\hat{\underline{S}}_{t|t}$.

We adopt here a linear Bayes approach to derive the Kalman Filter. Accordingly, let the class of estimators $\hat{\underline{S}}_{t|t}$, be restricted to affine functions of the form

$$\underline{\gamma} + \mathbf{\Gamma}\underline{X}_t. \tag{3.3}$$

If $\underline{X}_t$ is an $m \times 1$ vector, then $\underline{\gamma}$ and $\mathbf{\Gamma}$ are of dimensions $m \times 1$ and $k \times m$, respectively.

Following De Jong and Zehnwirth (1983), we derive the discrete Kalman filter by finding the linear minimum variance estimator of $\underline{S}_t$. Consider equation (3.2), given observations $\underline{X}^{t-1}$, we have

$$\mathrm{E}[\underline{S}_t|\underline{X}^{t-1}] = \mathbf{A}_t\,\mathrm{E}[\underline{S}_{t-1}|\underline{X}^{t-1}] + \mathrm{E}[\underline{v}_t|\underline{X}^{t-1}],$$

or

$$\hat{\underline{S}}_{t|t-1} = \mathbf{A}_t\hat{\underline{S}}_{t-1|t-1},$$

since $\mathrm{E}[\underline{v}_t|\underline{X}^{t-1}] = 0$. Given $\underline{X}^{t-1}$, the covariance matrix of $\underline{S}_t$ is

$$\mathrm{Cov}[\underline{S}_t|\underline{X}^{t-1}] = \mathbf{A}_t\,\mathrm{Cov}[\underline{S}_{t-1}|\underline{X}^{t-1}]\mathbf{A}_t' + \mathrm{Cov}[\underline{v}_t]\,.$$

Let us denote the above covariance matrix by $\mathbf{P}_{t|t-1}$, then

$$\mathbf{P}_{t|t-1} = \mathbf{A}_t\mathbf{P}_{t-1|t-1}\mathbf{A}_t' + \mathbf{V}_t. \tag{3.4}$$

Interpreting the foregoing from a Bayesian perspective, we see that $\hat{\underline{S}}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ are the mean and covariance, respectively, of the distribution of $\underline{S}_t$ prior to observing $\underline{X}_t$. To determine the optimal linear affine Bayes' rule for $\underline{S}_t$, we must minimize the risk function

$$\mathcal{R} = \mathrm{E}[(\underline{\gamma} + \mathbf{\Gamma}\underline{X}_t - \underline{S}_t)'(\underline{\gamma} + \mathbf{\Gamma}\underline{X}_t - \underline{S}_t)] \tag{3.5}$$

over all vectors $\underline{\gamma}$ and matrices $\mathbf{\Gamma}$ of appropriate dimensions. Here, $\underline{\gamma} + \mathbf{\Gamma}\underline{X}_t$ is the estimate of $\underline{S}_t$ *after* incorporation of the measurement $\underline{X}_t$. The contribution of the new measurement is $\mathbf{\Gamma}\underline{X}_t$. The solution to the minimization of (3.5) is given by (1.34).

Based on measurements up to time $t-1$, we find that

$$\begin{aligned}
\mathrm{E}[\underline{S}_t|\underline{X}^t] &= \mathrm{E}[\underline{S}_t|\underline{X}^{t-1}] + \mathrm{Cov}[\underline{S}_t, \underline{X}_t'|\underline{X}^{t-1}]\{\mathrm{Cov}[\underline{X}_t|\underline{X}^{t-1}]\}^{-1}\{\underline{X}_t - \mathrm{E}[\underline{X}_t|\underline{X}^{t-1}]\} \\
&= \mathrm{E}[\underline{S}_t|\underline{X}^{t-1}] + \mathrm{Cov}[\underline{S}_t, (\underline{S}_t'\mathbf{H}_t' + \underline{u}_t')|\underline{X}^{t-1}]\{\mathrm{Cov}[\mathbf{H}_t\underline{S}_t + \underline{u}_t|\underline{X}^{t-1}]\}^{-1} \\
&\quad \times \{\underline{X}_t - \mathrm{E}[\underline{X}_t|\underline{X}^{t-1}]\} \\
&= \hat{\underline{S}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{H}_t'[\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t' + \mathbf{U}_t]^{-1}\{\underline{X}_t - \mathbf{H}_t\hat{\underline{S}}_{t|t-1}\}.
\end{aligned}$$

Let

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}_t'[\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t' + \mathbf{U}_t]^{-1}, \tag{3.6}$$

where $\mathbf{K}_t$ is called the Kalman gain. The optimal inhomogeneous linear Bayes rule can be written as

$$\hat{\underline{S}}_{t|t} = \hat{\underline{S}}_{t|t-1} + \mathbf{K}_t[\underline{X}_t - \mathbf{H}_t\hat{\underline{S}}_{t|t-1}]. \tag{3.7}$$

Thus we can see that the Kalman gain matrix determines the amount by which the innovations $\underline{X}_t - \mathbf{H}_t\hat{\underline{S}}_{t|t-1}$ contribute to the prior estimate in order to obtain the updated estimate.

We now derive the covariance of $\underline{S}_t$ after observing $\underline{X}_t$. The error covariance of $\underline{S}_t$, given $\underline{X}^t$, is defined as

$$\mathbf{P}_{t|t} = \mathrm{E}[(\hat{\underline{S}}_{t|t} - \underline{S}_t)(\hat{\underline{S}}_{t|t} - \underline{S}_t)'].$$

However, given t observations, $\mathrm{E}[\underline{S}_t] = \hat{\underline{S}}_{t|t}$, so that

$$\mathbf{P}_{t|t} = \mathrm{Cov}[\hat{\underline{S}}_{t|t} - \underline{S}_t].$$

Furthermore,

$$\hat{\underline{S}}_{t|t} = \hat{\underline{S}}_{t|t-1} + \mathbf{K}_t[\mathbf{H}_t\underline{S}_t + \underline{u}_t - \mathbf{H}_t\hat{\underline{S}}_{t|t-1}]$$

or

$$\hat{\underline{S}}_{t|t} - \underline{S}_t = [\mathbf{I} - \mathbf{K}_t\mathbf{H}_t](\hat{\underline{S}}_{t|t-1} - \underline{S}_t) + \mathbf{K}_t\underline{u}_t.$$

Upon taking covariances of both sides, we see that

$$\mathbf{P}_{t|t} = [\mathbf{I} - \mathbf{K}_t\mathbf{H}_t]\mathbf{P}_{t|t-1}[\mathbf{I} - \mathbf{K}_t\mathbf{H}_t]' + \mathbf{K}_t\mathbf{U}_t\mathbf{K}_t', \tag{3.8}$$

since $\mathrm{E}[(\hat{\underline{S}}_{t|t-1} - \underline{S}_t)\underline{u}_t'] = 0$. By direct substitution of the Kalman gain matrix $\mathbf{K}_t$ from equation (3.6) into equation (3.8), and after some manipulation, we find that

$$\mathbf{P}_{t|t} = [\mathbf{I} - \mathbf{K}_t\mathbf{H}_t]\mathbf{P}_{t|t-1}. \tag{3.9}$$

We now show how some credibility models can be implemented using the Kalman filter.

Example 3.1 The credibility model of Bühlmann and Straub can be shown to be a special case of the Kalman filter. Let the risk parameter of a fixed contract be the random variable Θ. Then, from the assumptions of the Bühlmann-Straub model in Section 1.4, we have

$$\mathrm{E}[X_t|\Theta] = \mu(\Theta) \qquad \text{and} \qquad \mathrm{Var}[X_t|\Theta] = \sigma^2(\Theta)/w_t\,,$$

where all $w_t > 0$. Let $\bar{X}_t = \sum_{i=1}^t w_i X_i / w_{t.}$, $w_{t.} = \sum_{i=1}^t w_i$, and $\mu = \mathrm{E}[\mu(\Theta)]$. By Theorem 1.3, the optimal inhomogeneous linear estimator for $\mu(\Theta)$ is

$$\hat{\mu}(\Theta_j) = (1 - Z_t)\,\mu + Z_t\bar{X}_t\,, \tag{3.10}$$

where $Z_t = aw_{t.}/(aw_{t.} + s^2)$, $a = \mathrm{Var}[\mu(\Theta)]$, and $s^2 = \mathrm{E}[\sigma^2(\Theta)]$.

In the measurement and system equations (3.1) and (3.2), let $\mathbf{H}_t = 1$ and $\mathbf{A}_t = 1$. In addition, let $\mathbf{U}_t = s^2/w_t$ and $\mathbf{V}_t = 0$. To start the Kalman filter recursions, we use the initial values $\hat{\underline{S}}_{0|0} = \mu$ and $\mathbf{P}_{0|0} = a$. We note that the above variables are all scalars, so in the sequel, we drop the bold-face and underline notation.

At time t, from equation (3.7), we have

$$\hat{S}_{t|t} = (1 - K_t)\hat{S}_{t|t-1} + K_t X_t\,. \tag{3.11}$$

And since $A_t = 1$, we have

$$\hat{S}_{t|t-1} = \hat{S}_{t-1|t-1} = (1 - K_{t-1})\hat{S}_{t-1|t-1} + K_{t-1}X_{t-1}\,, \tag{3.12}$$

we can apply (3.11) repeatedly for $t, t-1, ..., 1$ to get

$$\begin{aligned}\hat{S}_{t|t} &= (1-K_t)(1-K_{t-1})\cdots(1-K_1)\,\mu \\ &\quad + (1-K_t)(1-K_{t-1})\cdots(1-K_2)K_1X_1 \\ &\quad + \cdots + (1-K_t)(1-K_{t-1})K_{t-2}X_{t-2} \\ &\quad + (1-K_t)K_{t-1}X_{t-1} + K_tX_t\,.\end{aligned} \tag{3.13}$$

It can also be shown that

$$K_t = \frac{aw_t}{a\sum_{i=1}^t w_i + s^2} \tag{3.14}$$

and

$$P_{t|t} = \frac{as^2}{a\sum_{i=1}^t w_i + s^2}\,. \tag{3.15}$$

From (3.14), we find that

$$1 - K_t = \frac{a\sum_{i=1}^{t-1} w_i + s^2}{a\sum_{i=1}^t w_i + s^2}\,,$$

for $t > 1$, thus (3.13) can be recast as

$$\hat{S}_{t|t} = \frac{s^2}{a\sum_{i=1}^t w_i + s^2}\,\mu + \frac{a}{a\sum_{i=1}^t w_i + s^2}w_1X_1 + \cdots + \frac{a}{a\sum_{i=1}^t w_i + s^2}w_tX_t\,.$$

We let $Z_t = aw_{t.}/(aw_{t.} + s^2)$, then

$$\begin{aligned}\hat{S}_{t|t} &= (1-Z_t)\,\mu + \frac{Z_t}{w_{t.}}(w_1X_1 + \cdots + w_tX_t) \\ &= (1-Z_t)\,\mu + Z_t\bar{X}_t\,.\end{aligned}$$

This form is equivalent to the Bühlmann-Straub credibility formula in (3.10).

Example 3.2 Hachemeister's model can also be shown to be a special case of the Kalman filter. Let the risk parameter for contract j be Θ. Then, from the assumptions of the Hachemeister model in Section 1.5, we have for contract j,

$$\mathrm{E}[X_t|\Theta] = \underline{Y}_t\,\underline{\beta}(\Theta) \qquad \text{and} \qquad \mathrm{Var}[X_t|\Theta] = \sigma^2(\Theta)/w_t\,,$$

where $\underline{Y}_t$ is a $1 \times n$ design matrix and $w_t > 0$. Let $\mathbf{A}_t = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$, $\mathbf{V}_t = \mathbf{0}$, and $\mathbf{U}_t = s^2/w_t$. Also, let $\underline{H}_t = \underline{Y}_t = \left[\begin{smallmatrix} 1 & 0 \end{smallmatrix}\right]$. The filter is started with the initial conditions $\underline{\hat{S}}_{0|0} = \mathrm{E}[\underline{\beta}(\Theta)] = \underline{\beta}$ and $\mathbf{P}_{0|0} = \mathrm{Cov}[\underline{\beta}(\Theta)] = \mathbf{\Lambda}$. The Kalman recursions for the Hachemeister model then are

$$\underline{\hat{S}}_{t|t} = \underline{\hat{S}}_{t|t-1} + \underline{K}_t\,[X_t - \underline{H}_t\,\underline{\hat{S}}_{t|t-1}]]\ ;,$$
$$\underline{K}_t = \mathbf{P}_{t|t-1}\,\underline{H}_t'\,(\underline{H}_t\,\mathbf{P}_{t|t-1}\,\underline{H}_t' + s^2/w_t)^{-1}\,,$$

and

$$\mathbf{P}_{t|t-1} = (1 - \underline{K}_t\,\underline{H}_t)\mathbf{P}_{t|t-1}\,.$$

In this example we have $\underline{S}_t = (S_{0,t}, S_{1,t})'$. So for the choice of the matrix $\mathbf{A}_t$ given in this example, the slope $\hat{S}_{1,t|t-1}$ and intercept $\hat{S}_{0,t|t-1}$ in $\underline{\hat{S}}_{t|t-1}$ will change over time. With this $\mathbf{A}_t$, the estimate at time t of $\underline{S}_t$ will be $\underline{\hat{S}}_{t|t} = (\hat{S}_{0,t|t} + \hat{S}_{1,t|t}, \hat{S}_{1,t|t})'$. The choice of $\underline{H}_t$ ensures that the estimate of $\underline{X}_t$ be equal to the first element only of $\underline{\hat{S}}_{t|t}$ for each t. To interpret $\mathbf{V}_t$ and U_t, note that in taking $\mathbf{V}_t = \mathbf{0}$, we are assuming that no system errors are present so that the evolution of the states is deterministic. On the other hand, the error in observing the true state of the system is given by U_t.

3.2 Empirical Credibility with the Kalman Filter

In this section, we describe two implementations of the Kalman filter in credibility theory. We have seen that we can embed a credibility model within a Kalman filter framework. However, the Kalman filter is composed of certain parameters which require estimation in practice. The parameters are $\mathbf{P}_{t|t-1}$, the covariance of $\underline{S}_t$ prior to the t-th observation; $\mathbf{P}_{t|t}$, the error covariance of $\hat{\underline{S}}_{t|t}$; $\mathbf{U}_t$, the covariance matrix of the random errors $\underline{u}_t$; and $\mathbf{V}_t$, the covariance matrix of the random errors $\underline{v}_t$. Along with an estimate for the collective mean, we may then derive an empirical credibility estimate of $\hat{\underline{S}}_{t|t}$.

The first implementation we discuss is due to Ledolter, Klugman, and Lee (1991). We consider first, an individual series of observations. For a single series $X_{j1}, \ldots, X_{jr}$ for $r = 1, \ldots, t$ and $j = 1, \ldots, k$, which corresponds to realizations of the risk parameter Θ_j, the following measurement model is observed:

$$X_{jt} = \underline{H}'_{jt}\, \underline{S}_t^{(j)} + u_{jt}\,, \tag{3.16}$$

where $\underline{H}_{jt}$ is an n-dimensional column vector. For the starting values to begin the iterative Kalman process, Ledolter et al. suggest using a vector of zeros for $\hat{\underline{S}}_{0|0}^{(j)}$ and a diagonal matrix with large (but finite) diagonal elements for $\mathbf{P}_{0|0}^{(j)}$. This is similar to using a non-informative prior in a Bayesian framework since for moderate values of t, the initial choice for $\hat{\underline{S}}_{0|0}^{(j)}$ will be dominated by the data.

In order to compute a value for $\mathbf{P}_{t|t-1}^{(j)}$, we require an estimate of $\mathbf{V}_{jt}$. The covariance matrix $\mathbf{V}_{jt}$ is considered to be time-invariant; thus, $\mathbf{V}_{jt} = \mathbf{V}_j$. Following the paper by Ledolter et al., we assume that u_{jt} and $\underline{v}_{jt}$ are both normally distributed. Thus we have $u_{jt} \sim N(0, U_{jt})$ and $\underline{v}_{jt} \sim N(0, \mathbf{V}_j)$. For carrying out the maximum likelihood estimation of $\mathbf{V}_j$, we assume further that $U_{jt} = \sigma_t^2(\Theta_j)/w_{jt}$ and that the covariance matrix of $\hat{\underline{S}}_{t|t-1}^{(j)}$ is proportional to $\sigma_t^2(\Theta_j)$. Then the log-likelihood function is

$$\ell(\sigma_t^2(\Theta_j), \mathbf{V}_j) = -\frac{n}{2}\log \sigma_t^2(\Theta_j) - \frac{1}{2}\sum_{r=1}^{t} f_{jr} - \frac{1}{2\sigma_t^2(\Theta_j)}\sum_{r=1}^{t}(X_{jr} - \underline{H}_{jr}\underline{S}_{r|r-1}^{(j)})^2/f_{jr}\,,$$

where $f_{jr} = \underline{H}_{jr}\,\mathbf{P}_{r|r-1}^{(j)}\,\underline{H}'_{jr} + 1/w_{jr}$.

Maximization of ℓ with respect to σ_t^2 yields the following maximum likelihood estimate

$$\hat{\sigma}_t^2(\Theta_j) = \frac{1}{n}\sum_{r=1}^{t}(X_{jr} - \underline{H}_{jr}\underline{S}^{(j)}_{r|r-1})^2/f_{jr}\,.$$

We then replace $\sigma_t^2(\Theta_j)$ by $\hat{\sigma}_t^2(\Theta_j)$ to get the *concentrated* log-likelihood function

$$\ell_c(\mathbf{V}_j) = \frac{1}{2}\sum_{r=1}^{t}\log f_{jr} - \frac{n}{2}\log\hat{\sigma}_t^2(\Theta_j)\,.$$

Since we have k risk classes from which k concentrated log-likelihood functions arise, we can pool the information to form a better estimate of $\mathbf{V}_j$. Accordingly, Ledolter et al. assume that the $\mathbf{V}_j$'s are identical across the k risk groups and that the k groups are independent of each other. Then the k concentrated log-likelihood functions are added and a common $\mathbf{V}$ which maximizes the collected concentrated log-likelihood is determined numerically. The maximum likelihood estimate of each $\sigma_t^2(\Theta_j)$ from all risk classes then is an average of the $\sigma_t^2(\Theta_j)$,

$$\hat{s}_t^2 = \frac{1}{kt}\sum_{j=1}^{k}\sum_{r=1}^{t}(X_{jr} - \underline{H}_{jr}\hat{\underline{S}}^{(j)}_{r|r-1})^2/f_{jr}\,.$$

This method derives the individual estimators of $\underline{\mu}(\Theta_j)$ by applying the Kalman filter to the k separate series of observations. If we assume that $\mathbf{V}$ equals zero, we obtain the weighted least squares estimates used in Hachemeister's model.

From the k separate Kalman filters, we have $\mathrm{E}[\hat{\underline{S}}^{(j)}_{t|t}] = \underline{S}^{(j)}_t$ and $\mathrm{Cov}[\hat{\underline{S}}^{(j)}_{t|t} - \underline{S}^{(j)}_t] = \sigma_t^2\,\mathbf{P}^{(j)}_{t|t}$. From this, we introduce

$$\hat{\underline{S}}^{(j)}_{t|t} = \underline{S}^{(j)}_t + \underline{e}_{jt}\,, \tag{3.17}$$

where $\underline{e}_{jt} \sim N(0, \sigma_t^2\,\mathbf{P}^{(j)}_{t|t})$. The $\underline{e}_{jt}$ are independent across the k risk classes. To get the credibility estimators, Ledolter et al. introduce a second equation

$$\underline{S}^{(j)}_t = \underline{B}_t + \xi^{(j)}_t \tag{3.18}$$

at time t. The random term $\xi^{(j)}_t$ is normal with mean vector zero and covariance matrix $\sigma_t^2\,\mathbf{\Omega}_t$.

Since $\underline{\hat{S}}_{t|t}^{(j)}$ has a normal distribution by (3.17), and since the mean $\underline{S}_t^{(j)}$ is also normal (by (3.18)), we obtain the Bayes shrinkage estimator of $\underline{S}_t^{(j)}$ as

$$\underline{\tilde{S}}_t^{(j)} = \mathbf{Z}_j \, \underline{\hat{S}}_{t|t}^{(j)} + (\mathbf{I} - \mathbf{Z}_j) \, \underline{B}_t \,, \tag{3.19}$$

where $\mathbf{Z}_j = \mathbf{\Omega}_t \, (\mathbf{\Omega}_t + \mathbf{P}_{t|t}^{(j)})^{-1}$. To estimate $\underline{B}_t$ and $\mathbf{\Omega}_t$, we may use the methods described in Section 1.6, in particular, (1.41) and (1.42). In this case, an iterative approach will be required.

The second implementation of the Kalman filter applied to credibility is from Kremer (1995). In this paper, the Kalman recursions for a single contract are written in the following form:

$$\underline{\hat{S}}_{t+1|t} = \mathbf{A}_t \, [\mathbf{K}_t \, X_t + (\mathbf{I} - \mathbf{K}_t \, \mathbf{H}_t) \underline{\hat{S}}_{t|t-1}], \tag{3.20}$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \, \mathbf{H}_t' \, [\mathbf{H}_t \, \mathbf{P}_{t|t-1} \, \mathbf{H}_t' + U_t]^{-1}, \tag{3.21}$$

and

$$\mathbf{P}_{t+1|t} = \mathbf{A}_{t+1} \, (\mathbf{I} - \mathbf{K}_t \, \mathbf{H}_t) \, \mathbf{P}_{t|t-1} \, \mathbf{A}_{t+1}' + \mathbf{V}_{t+1} \,. \tag{3.22}$$

The recursions are initialized with $\underline{S}_{1|0} = \underline{B}_1$ and $\mathbf{P}_{1|0} = \mathbf{\Lambda}_1$. If we have k risk classes, we estimate $\underline{B}_1$ with the simple regression estimator

$$\underline{\hat{B}}_1 = (\mathbf{H}' \, \mathbf{H})^{-1} \, \mathbf{H}' \, \underline{X} \,,$$

where $\mathbf{H} = (\underline{H}_{11}', \underline{H}_{21}', \ldots, \underline{H}_{k1}')'$ and $\underline{X} = (X_{11}, X_{21}, \ldots, X_{k1})'$. The $1 \times n$ matrix $\underline{H}_{j1}$ is from (3.16).

Kremer assumes that $\mathbf{\Lambda}_1$ is proportional to the identity matrix, that is, $\mathbf{\Lambda}_1 = f_1 \, \mathbf{I}$. Let $U_{jt} = s_t^2 / w_{jt}$, where $s_t^2 = \mathrm{E}[\sigma_t^2(\Theta_j)]$. Thus, if the variance of X_{jt} is given by $\mathrm{E}[X_{jt} - \underline{H}_{jt} \, \underline{S}_t^{(j)}]^2$, we obtain that $U_t = \mathrm{E}\{\mathrm{Var}[X_{jt}|\Theta_j]\}$ equals

$$\mathrm{E}[X_{jt} - \underline{H}_{jt} \, \underline{S}_t^{(j)}]^2 - \underline{H}_{jt} \, \mathbf{\Lambda}_t \, \underline{H}_{jt}' \,.$$

From this, we get

$$\mathrm{E}[X_{j1} - \underline{H}_{j1} \, \underline{S}_1^{(j)}]^2 = \frac{s_1^2}{w_{j1}} + f_1 \, \underline{H}_{j1} \, \underline{H}_{j1}' \,.$$

Let $q_j = (X_{j1} - \underline{H}_{j1}\,\hat{\underline{B}}_1)^2$ and $\underline{Q} = (q_1, q_2, \ldots, q_k)'$. If we define $\mathbf{M}$ as

$$\begin{bmatrix} 1/w_{11} & \underline{H}_{11}\,\underline{H}'_{11} \\ 1/w_{21} & \underline{H}_{21}\,\underline{H}'_{21} \\ \vdots & \\ 1/w_{k1} & \underline{H}_{k1}\,\underline{H}'_{k1} \end{bmatrix},$$

then

$$\underline{Q} = \mathbf{M} \begin{bmatrix} s_1^2 \\ f_1 \end{bmatrix}.$$

The linear regression estimator of $\underline{\zeta} = (s_1^2, f_1)'$ is then given by

$$\hat{\underline{\zeta}} = (\mathbf{M}'\,\mathbf{M})^{-1}\,\mathbf{M}'\,\underline{Q}. \tag{3.23}$$

With the initial values $\hat{s}_1^2$, $\hat{\underline{B}}_1$, and $\hat{\mathbf{\Lambda}}_1$, suppose that at time $t \geq 1$, we have the estimators $\hat{\underline{S}}^{(j)}_{t|t-1}$, $\hat{U}_{jt}$, $\hat{\mathbf{\Lambda}}_t$, and $\mathbf{P}^{(j)}_{t|t-1}$, arrived at by application of the Kalman filter equations (3.20), (3.21), and (3.22). We then proceed as follows.

We determine the value of $\hat{\underline{S}}^{(j)}_{t+1|t}$ using equations (3.20) and (3.21). After the values $\hat{\underline{S}}^{(j)}_{t+1|t}$ are available for each $j = 1, \ldots, k$, we can compute an estimate of the collective mean by

$$\hat{\underline{B}}_{t+1} = \sum_{j=1}^{k} \mathbf{F}_{jt}\,\hat{\underline{S}}^{(j)}_{t+1|t}, \tag{3.24}$$

with $\sum_{j=1}^{k} \mathbf{F}_{jt} = \mathbf{I}$. We refer to section 1.6 for a discussion on possible choices of the matrix $\mathbf{F}_{jt}$.

We may now estimate $\mathbf{\Lambda}_{t+1}$ by

$$\hat{\mathbf{\Lambda}}_{t+1} = \sum_{j=1}^{k} \tilde{\mathbf{F}}_{jt}\,(\hat{\underline{S}}^{(j)}_{t+1|t} - \hat{\underline{B}}_{t+1})(\hat{\underline{S}}^{(j)}_{t+1|t} - \hat{\underline{B}}_{t+1})', \tag{3.25}$$

where the $\tilde{\mathbf{F}}_{jt}$ are weighting matrices such that $\sum_{j=1}^{k} \tilde{\mathbf{F}}_{jt} = \mathbf{I}$. The choices for $\tilde{\mathbf{F}}_{jt}$ are similar to those for $\mathbf{F}_{jt}$.

To estimate $U_{j,t+1}$, we can use

$$\hat{U}_{j,t+1} = \hat{s}^2_{t+1}/w_{j,t+1}$$

where $\hat{s}^2_{t+1}$ is an estimator of s^2_{t+1}. Starting with $\hat{s}^2_1$ from (3.23), we can determine $\hat{s}^2_{t+1}$ recursively through

$$\hat{s}^2_{t+1} = \max(\tilde{s}^2_{t+1}, 0) , \tag{3.26}$$

where

$$\tilde{s}^2_{t+1} = \frac{1}{k}\sum_{j=1}^{k} w_{j,t+1}(X_{j,t+1} - \underline{H}_{j,t+1}\,\hat{\underline{B}}_{t+1})^2 - \frac{1}{k}\sum_{j=1}^{k} w_{j,t+1}\,\underline{H}_{j,t+1}\,\hat{\mathbf{\Lambda}}_{t+1}\,\underline{H}'_{j,t+1}\,. \tag{3.27}$$

With starting point $\hat{\mathbf{\Lambda}} = \hat{f}_1\,\mathbf{I}$ and equation (3.25), we estimate $\mathbf{V}_{t+1}$ with

$$\hat{\mathbf{V}}_{t+1} = \hat{\mathbf{\Lambda}}_{t+1} - \mathbf{A}_t\,\hat{\mathbf{\Lambda}}_t\,\mathbf{A}'_t\,. \tag{3.28}$$

Through repeated application of the equations (3.20), (3.21), and (3.22), at time $t+1$, we can arrive at the empirical credibility estimator $\hat{\underline{S}}^{(j)}_{t+1|t}$ of $\underline{S}^{(j)}_t$.

Remark 3.1 In practice, one may have difficulties with the initial value $\hat{\underline{B}}_1$. From the definition of $\mathbf{H}$, if the design matrices $\underline{H}_{j1}$ at time 1 for each contract j are identical, the matrix $\mathbf{H}'\,\mathbf{H}$ will be singular. An alternative estimator for the initial value $\hat{\underline{S}}^{(j)}_{0|0}$ can be determined by taking a Bayesian approach where the initial estimate is obtained by relying on prior knowledge of contract j. In the absence of clear prior knowledge, one may wish to use last year's claims data or data from a similar contract.

If data on the contract of interest already exists, another initial value may be the estimate, based on existing data, one obtains by using a non-evolutionary credibility model (e.g., Hachemeister's model). For example, if there are t observations on contract j, we may estimate, by Hachemeister's model, the credibility estimator based on the first $r \leq t$ observations. This credibility estimator would then be used, via the Kalman filter, to get the credibility estimator at time $t \geq r$.

Remark 3.2 In equations (3.24) and (3.25), the estimation is done using the updated estimates of $\underline{S}_t^{(j)}$. However, these estimates represent the credibility adjusted estimates of $\underline{S}_t^{(j)}$. We note that in the classical credibility models, the individual estimators are normally used. However, the advantage of Kremer's estimators $\hat{\underline{B}}_{t+1}$ and $\hat{\Lambda}_{t+1}$ is that they share the same recursive structure as $\hat{\underline{S}}_{t|t}^{(j)}$.

3.3 Robust Kalman Filtering in Credibility

In Section 2 of this chapter, we showed how the Kalman filter could be applied to credibility theory. In particular, it was shown that both the Bühlmann-Straub and Hachemeister models were special cases of the Kalman filter. However, both these credibility models are sensitive to large claims; therefore, in order to robustify these models, we will need a robust Kalman filter. The robust Kalman filter that we next describe is due to Cipra and Romera (1991). In their paper, a robust Kalman filter is developed by using M-estimators.

Following the notation of Section 2, if

$$\begin{bmatrix} \underline{d}_t \\ \underline{e}_t \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{t|t-1}^{-1/2}\,(\hat{\underline{S}}_{t|t-1} - \underline{S}_t) \\ \mathbf{U}_t^{-1/2}\,(\underline{X}_t - \mathbf{H}_t\underline{S}_t) \end{bmatrix}, \tag{3.29}$$

then $\underline{S}_t = \hat{\underline{S}}_{t|t}$ minimizes

$$\begin{bmatrix} \underline{d}_t' & \underline{e}_t' \end{bmatrix} \begin{bmatrix} \underline{d}_t \\ \underline{e}_t \end{bmatrix} = (\hat{\underline{S}}_{t|t-1} - \underline{S}_t)'\mathbf{P}_{t|t-1}(\hat{\underline{S}}_{t|t-1} - \underline{S}_t) + (\underline{X}_t - \mathbf{H}_t\underline{S}_t)'\mathbf{U}_t^{-1}(\underline{X}_t - \mathbf{H}_t\underline{S}_t). \tag{3.30}$$

Define

$$\mathbf{P}_{t|t-1}^{-1/2}\,\hat{\underline{S}}_{t|t-1} = \begin{bmatrix} p_{1t} \\ \vdots \\ p_{nt} \end{bmatrix}, \quad \mathbf{P}_{t|t-1}^{-1/2} = \begin{bmatrix} \underline{a}_{1t} \\ \vdots \\ \underline{a}_{nt} \end{bmatrix}$$

and

$$\mathbf{U}_t^{-1/2}\,\underline{X}_t = \begin{bmatrix} g_{1t} \\ \vdots \\ g_{mt} \end{bmatrix}, \quad \mathbf{U}_t^{-1/2}\,\mathbf{H}_t = \begin{bmatrix} \underline{b}_{1t} \\ \vdots \\ \underline{b}_{mt} \end{bmatrix},$$

where $\underline{a}_{qt}$, $q = 1, \ldots, n$ and $\underline{b}_{rt}$, $r = 1, \ldots, m$ are n dimensional row vectors. Then, with

$$\underline{d}_t = \begin{bmatrix} d_{1t} \\ \vdots \\ d_{nt} \end{bmatrix} \quad \text{and} \quad \underline{e}_t = \begin{bmatrix} e_{1t} \\ \vdots \\ e_{mt} \end{bmatrix},$$

we can reformulate (3.29) as

$$d_{qt} = p_{qt} - \underline{a}_{qt}\underline{S}_t \quad \text{for} \quad q = 1, \ldots, l,$$
$$e_{rt} = g_{rt} - \underline{b}_{rt}\underline{S}_t \quad \text{for} \quad r = 1, \ldots, m.$$

It is clear that

$$(\underline{\hat{S}}_{t|t-1} - \underline{S}_t)'\,\mathbf{P}_{t|t-1}^{-1}\,(\underline{\hat{S}}_{t|t-1} - \underline{S}_t) = \sum_{q=1}^{n} (p_{qt} - \underline{a}_{qt}\underline{S}_t)^2 \tag{3.31}$$

and

$$(\underline{X}_t - \mathbf{H}_t\underline{S}_t)'\,\mathbf{U}_t^{-1}\,(\underline{X}_t - \mathbf{H}_t\underline{S}_t) = \sum_{r=1}^{m} (s_{rt} - \underline{b}_{rt}\underline{S}_t)^2. \tag{3.32}$$

In order to robustify the Kalman filter, we replace quadratic loss with a general function ρ. Then $\underline{\hat{S}}_{t|t}$ is the value of $\underline{S}_t$ which satisfies

$$\sum_{q=1}^{n} \rho_{1q}(p_{qt} - \underline{a}_{qt}\underline{S}_t) + \sum_{r=1}^{m} \rho_{2r}(g_{rt} - \underline{b}_{rt}\underline{S}_t) = \min!\,. \tag{3.33}$$

Alternatively, if the derivative ψ of ρ exists,

$$\sum_{q=1}^{n} \underline{a}'_{qt}\psi_{1q}(p_{qt} - \underline{a}_{qt}\underline{\hat{S}}_{t|t}) + \sum_{r=1}^{m} \underline{b}'_{rt}\psi_{2r}(g_{rt} - \underline{b}_{rt}\underline{\hat{S}}_{t|t}) = 0. \tag{3.34}$$

In general, the normal equations indicated by (3.34) can not be solved explicitly. However, an approximation is available. From (3.34), we have

$$\sum_{q=1}^{n} \underline{a}'_{qt}\psi_{1q}(p_{qt} - \underline{a}_{qt}\hat{\underline{S}}_{t|t})\frac{(p_{qt} - \underline{a}_{qt}\hat{\underline{S}}_{t|t})}{p_{qt} - \underline{a}_{qt}\hat{\underline{S}}_{t|t}} + \sum_{r=1}^{m} \underline{b}'_{rt}\psi_{2r}(g_{rt} - \underline{b}_{rt}\hat{\underline{S}}_{t|t})\frac{(g_{rt} - \underline{b}_{rt}\hat{\underline{S}}_{t|t})}{g_{rt} - \underline{b}_{rt}\hat{\underline{S}}_{t|t}} = 0. \tag{3.35}$$

We approximate $\hat{\underline{S}}_{t|t}$ with $\hat{\underline{S}}_{t|t-1}$, then

$$\sum_{q=1}^{n} \gamma_{1qt}\underline{a}'_{qt}(p_{qt} - \underline{a}_{qt}\hat{\underline{S}}_{t|t}) + \sum_{r=1}^{m} \gamma_{2rt}\underline{b}'_{rt}(g_{rt} - \underline{b}_{rt}\hat{\underline{S}}_{t|t}) = 0, \tag{3.36}$$

where

$$\gamma_{1qt} = \frac{\psi_{1q}(p_{qt} - \underline{a}_{qt}\hat{\underline{S}}_{t|t-1})}{p_{qt} - \underline{a}_{qt}\hat{\underline{S}}_{t|t-1}} \quad \text{and} \quad \gamma_{2rt} = \frac{\psi_{2r}(g_{rt} - \underline{b}_{rt}\hat{\underline{S}}_{t|t-1})}{g_{rt} - \underline{b}_{rt}\hat{\underline{S}}_{t|t-1}}. \tag{3.37}$$

With

$$\mathbf{\Gamma}_{1t} = \mathrm{diag}(\gamma_{11t}, \ldots, \gamma_{1nt}), \quad \mathbf{\Gamma}_{2t} = \mathrm{diag}(\gamma_{21t}, \ldots, \gamma_{2mt}),$$

the normal equation (3.36) can be written as

$$\mathbf{P}_{t|t-1}^{-1/2}\,\mathbf{\Gamma}_{1t}\,\mathbf{P}_{t|t-1}^{-1/2}\,(\hat{\underline{S}}_{t|t-1} - \hat{\underline{S}}_{t|t}) + \mathbf{H}'_t\,\mathbf{U}_t^{-1/2}\,\mathbf{\Gamma}_{2t}\,\mathbf{U}_t^{-1/2}\,(\underline{X}_t - \mathbf{H}_t\,\hat{\underline{S}}_{t|t}) = 0. \tag{3.38}$$

Solving for $\hat{\underline{S}}_{t|t}$, we obtain the recursive formula for the robust Kalman filter

$$\begin{aligned}\hat{\underline{S}}_{t|t} = \hat{\underline{S}}_{t|t-1} &+ \mathbf{P}_{t|t-1}^{1/2}\,\mathbf{\Gamma}_{1t}^{-1}\,\mathbf{P}_{t|t-1}^{-1/2}\,\mathbf{H}'_t\,(\mathbf{U}_t^{1/2}\,\mathbf{\Gamma}_{2t}^{-1}\,\mathbf{U}_t^{1/2} \\ &+ \mathbf{H}_t\,\mathbf{P}_{t|t-1}^{1/2}\,\mathbf{\Gamma}_{1t}^{-1}\,\mathbf{P}_{t|t-1}^{1/2}\,\mathbf{H}'_t)^{-1}\,(\underline{X}_t - \mathbf{H}_t\,\hat{\underline{S}}_{t|t-1}).\end{aligned} \tag{3.39}$$

The robust error covariance is given by

$$\begin{aligned}\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1}^{1/2}\,\mathbf{\Gamma}_{1t}^{-1}\,\mathbf{P}_{t|t-1}^{1/2} &- \mathbf{P}_{t|t-1}^{1/2}\,\mathbf{\Gamma}_{1t}^{-1}\,\mathbf{P}_{t|t-1}^{1/2}\,\mathbf{H}'_t\,(\mathbf{U}_t^{1/2}\,\mathbf{\Gamma}_{2t}^{-1}\,\mathbf{U}_t^{1/2} \\ &+ \mathbf{H}_t\,\mathbf{P}_{t|t-1}^{1/2}\,\mathbf{\Gamma}_{1t}^{-1}\,\mathbf{P}_{t|t-1}^{1/2}\,\mathbf{H}'_t)^{-1}\,\mathbf{H}_t\,\mathbf{P}_{t|t-1}\,\mathbf{\Gamma}_{1t}^{-1}\,\mathbf{P}_{t|t-1}^{1/2}.\end{aligned} \tag{3.40}$$

Example 3.3 A robust Bühlmann-Straub model can be defined by using our robust Kalman filter. For a contract j, let $\mathbf{H}_t = 1$, $\mathbf{U}_t = s_t^2/w_t$, $\mathbf{A}_t = 1$, and $\mathbf{V}_t = 0$, where $s_t^2 = \mathrm{E}[\sigma_t^2(\Theta)]$. Assume that the errors $\underline{d}_t$ do not produce any outliers, consequently, we take

$$\psi_{1q}(x) = x.$$

We also assume that ψ_{2r} is independent of r, that is,

$$\psi_{2r}(\cdot) = \psi(\cdot).$$

Let

$$\begin{aligned}\gamma_t &= \frac{\psi(g_{1t} - b_{1t}\,\hat{S}_{t|t})}{g_{1t} - b_{1t}\,\hat{S}_{t|t}} \\ &= \frac{\psi[w_t^{1/2}\,(X_t - \hat{S}_{t|t})/s_t]}{w_t^{1/2}\,(X_t - \hat{S}_{t|t})/s_t}\,.\end{aligned}$$

Then for a general $\psi(\cdot)$, we have from equation (3.36),

$$a_{1t}\,(p_{1t} - a_{1t}\,\hat{S}_{t|t}) + b_{1t}\,\psi(g_{1t} - b_{1t}\,\hat{S}_{t|t}) = 0. \tag{3.41}$$

Rewriting this as

$$P_{t|t-1}^{-1}\,(\hat{S}_{t|t-1} - \hat{S}_{t|t}) + \gamma_t\,(s_t^2/w_t)^{-1}\,[X_t - \hat{S}_{t|t}] = 0, \tag{3.42}$$

we obtain

$$\hat{S}_{t|t} = \hat{S}_{t|t-1} + \frac{P_{t|t-1}\,\gamma_t\,w_t}{P_{t|t-1}\,\gamma_t\,w_t + s_t^2}(X_t - \hat{S}_{t|t-1}) \tag{3.43}$$

as the updated credibility estimate of S_t at time t. The variance of $S_t - \hat{S}_{t|t}$ is given by

$$\begin{aligned}P_{t|t} &= P_{t|t-1}\,[P_{t|t-1} + (s_t^2/w_t)/\gamma_t]^{-1}\,P_{t|t-1} \\ &= P_{t|t-1} - \frac{P_{t|t-1}^2\,\gamma_t\,w_t}{P_{t|t-1}\,\gamma_t\,w_t + s_t^2}\,.\end{aligned} \tag{3.44}$$

Example 3.4 For the Hachemeister case, let $\mathbf{U}_t = s_t^2/w_t$, where as in the previous example, $s_t^2 = \mathrm{E}[\sigma_t^2(\Theta)]$. Let $\mathbf{A}_t$ equal $\mathbf{I}$, an $n \times n$ identity matrix, and $\mathbf{V}_t = 0$. We further assume that $\mathbf{H}_t$ is now an n-dimensional row vector. As before, we assume that the errors $\underline{d}_t$ do not produce any outliers and that $\psi_{2r} = \psi$. Finally, $\mathbf{\Gamma}_{2t}$ can now be replaced by $\gamma_t\,\mathbf{I}$.

With $m = 1$, equation (3.36) becomes

$$\sum_{q=1}^{n} \underline{a}'_{qt}(p_{qt} - \underline{a}_{qt}\underline{\hat{S}}_{t|t}) + \gamma_t\,\underline{b}'_{1t}(g_{1t} - \underline{b}_{1t}\underline{\hat{S}}_{t|t}) = 0, \tag{3.45}$$

or

$$\mathbf{P}_{t|t-1}^{-1}\,(\hat{\underline{S}}_{t|t-1}-\hat{\underline{S}}_{t|t})+\gamma_t\,(s_t^2/w_t)^{-1}\,\underline{H}_t'\,(X_t-\underline{H}_t\,\hat{\underline{S}}_{t|t})=0. \tag{3.46}$$

Then, from equation (3.39), we have

$$\begin{aligned}\hat{\underline{S}}_{t|t}&=\hat{\underline{S}}_{t|t-1}+\mathbf{P}_{t|t-1}\,\underline{H}_t'\,[\underline{H}_t\,\mathbf{P}_{t|t-1}\,\underline{H}_t'+\gamma_t^{-1}\,(s_t^2/w_t)]^{-1}\,(X_t-\underline{H}_t\,\hat{\underline{S}}_{t|t-1})\\&=\hat{\underline{S}}_{t|t-1}+\frac{\mathbf{P}_{t|t-1}\,\underline{H}_t'\,\gamma_t\,w_t}{\underline{H}_t\,\mathbf{P}_{t|t-1}\,\underline{H}_t'\,\gamma_t\,w_t+s_t^2}(X_t-\underline{H}_t\,\hat{\underline{S}}_{t|t})\,,\end{aligned} \tag{3.47}$$

where

$$\gamma_t=\frac{\psi[w_t^{1/2}\,(X_t-\underline{H}_t\,\hat{\underline{S}}_{t|t})/s_t]}{w_t^{1/2}\,(X_t-\underline{H}_t\,\hat{\underline{S}}_{t|t})/s_t}.$$

The robust error covariance matrix is given by equation (3.40), where

$$\begin{aligned}\mathbf{P}_{t|t}&=\mathbf{P}_{t|t-1}-\mathbf{P}_{t|t-1}\,\underline{H}_t'\,[\underline{H}_t\,\mathbf{P}_{t|t-1}\,\underline{H}_t'+\gamma_t^{-1}\,(s_t^2/w_t)]^{-1}\,\underline{H}_t\,\mathbf{P}_{t|t-1}\\&=\mathbf{P}_{t|t-1}-\frac{\mathbf{P}_{t|t-1}\,\underline{H}_t'\,\underline{H}_t\,\mathbf{P}_{t|t-1}\,\gamma_t\,w_t}{\underline{H}_t\,\mathbf{P}_{t|t-1}\,\underline{H}_t'\,\gamma_t\,w_t+s_t^2}\,.\end{aligned} \tag{3.48}$$

3.4 An Empirical Robust Kalman Filter Credibility Model

In this section we present an implementation of a Hachemeister's regression model via robust Kalman filtering. For more general regression problems, one may want to adapt Kremer's procedure from the second part of section 3.2.

Consider the state space model (3.1) and (3.2). The observations of class j at time t are scalar, thus $\underline{X}_{jt}=X_{jt}$ is a scalar, $\mathbf{H}_{jt}=\underline{H}_{jt}$ is a $1\times n$ vector, and $\underline{u}_{jt}=u_{jt}$ is a scalar. The state space model becomes

$$X_{jt}=\underline{H}_{jt}\,\underline{S}_t^{(j)}+u_{jt}\,, \tag{3.49}$$

$$\underline{S}_t^{(j)}=\mathbf{A}_t\,\underline{S}_{t-1}^{(j)}+\underline{v}_{jt}\,. \tag{3.50}$$

The recursions are initialized for each risk class j, for $j = 1, \ldots, k$. Thus, we have for the j-th risk, at time $t = 1$, $\hat{\underline{S}}^{(j)}_{0|0} = \underline{B}_1$ and $\mathbf{P}^{(j)}_{0|0} = \mathbf{\Lambda}_1$. Further, we assume $\mathbf{V}_t = 0$, and $\mathbf{A}_t = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$.

We attempt to employ a Bayesian approach to begin the Kalman recursions. Assuming t observations are available, we use these points to determine our initial estimates. For our initial value of $s_t^2 = \mathrm{E}[\sigma_t^2(\Theta_j)]$, we use the $\hat{s}^2$ computed from Hachemeister's model, in particular, from equation (1.43). Our initial estimate of $\mathbf{P}_{t|t}$ is the $\hat{\mathbf{\Lambda}}$ from (1.42). As before, $\underline{B}_1$ represents the estimator for the collective mean, and $\mathbf{\Lambda}_1$ denotes the covariance of the $\underline{S}^{(j)}_t$. We estimate $\underline{B}_1$ by $\hat{\underline{\beta}}$ from Hachemeister's model, for example, equation (1.41).

We now can proceed to the next recursion of the filter. However, we have not yet specified the robustifying functions. In equation (3.34), we take ψ_{1r} to be

$$\psi_{1r}(x) = x\,.$$

For the robustifying function corresponding to outliers arising from the measurement errors, we use the one-sided Huber function. Letting $\psi_{2r} = \psi_H$ where ψ_H is from (2.34). Then from (3.34),

$$\sum_{q=1}^{n} \underline{a}'_{qjt}\,(p_{qjt} - \underline{a}_{qjt}\hat{\underline{S}}^{(j)}_{t|t}) + \underline{b}'_{1jt}\psi_H(g_{1jt} - \underline{b}_{1jt}\hat{\underline{S}}^{(j)}_{t|t}) = 0, \tag{3.51}$$

or

$$(\mathbf{P}^{(j)}_{t|t-1})^{-1}\,(\hat{\underline{S}}^{(j)}_{t|t-1} - \hat{\underline{S}}^{(j)}_{t|t}) + \underline{H}'_{jt}\,U_{jt}^{-1/2}\,\psi_H(X_{jt} - \underline{H}_{jt}\,\hat{\underline{S}}^{(j)}_{t|t}) = 0. \tag{3.52}$$

Let $U_{jt} = s_t^2/w_{jt}$. Our robust Kalman filter then is

$$\hat{\underline{S}}^{(j)}_{t|t} = \hat{\underline{S}}^{(j)}_{t|t-1} + \mathbf{P}^{(j)}_{t|t-1}\,\underline{H}'_t\left(\frac{w_{jt}^{1/2}}{s_t}\right)\psi_H\left[\frac{s_t\,w_{jt}^{1/2}\,(X_{jt} - \underline{H}_{jt}\,\hat{\underline{S}}^{(j)}_{t|t})}{\underline{H}_{jt}\,\mathbf{P}^{(j)}_{t|t-1}\,\underline{H}'_{jt}\,w_{jt} + s_t^2}\right]\,, \tag{3.53}$$

and

$$\mathbf{P}^{(j)}_{t|t} = \mathbf{P}^{(j)}_{t|t-1} - \frac{\mathbf{P}^{(j)}_{t|t-1}\,\underline{H}'_{jt}\,\underline{H}_{jt}\,\mathbf{P}^{(j)}_{t|t-1}\,w_{jt}}{\underline{H}_{jt}\,\mathbf{P}^{(j)}_{t|t-1}\,\underline{H}'_{jt}\,w_{jt} + s_t^2}\,. \tag{3.54}$$

In order to compute $\underline{\hat{S}}^{(j)}_{t|t}$ from (3.53), we need an estimate of s_t^2. We can use (1.43) as a function of t, that is,

$$\hat{s}_t^2 = \frac{1}{k(t-n)} \sum_{j=1}^{k} \sum_{r=1}^{t} \frac{(X_{jr} - \mathbf{H}_{jr}\,\underline{\hat{\beta}}_j)^2}{w_{jr}}\,, \tag{3.55}$$

where $\underline{\hat{\beta}}_j$ is the weighted least-squares estimate based on t observations. For additional robustness, one may prefer to use a robust estimate for $\underline{\hat{\beta}}_j$. This estimator will require that we begin the recursions at time $t = n+1$.

The following describes a procedure for implementing Kremer's method for finding the preceding estimators via the Kalman filter.

1. Define $\mathbf{A}_t$ and $\underline{H}_{jt}$.

2. Determine $\underline{\hat{S}}^{(j)}_{0|0} = \underline{\hat{B}}_1$, $\mathbf{P}^{(j)}_{0|0} = \hat{\mathbf{\Lambda}}$, and $\hat{s}_1^2$.

3. For t = 1, 2, ... , do steps 4–6.

4. For j = 1, 2, ... , k, do steps i–iv.

   (i) Compute $\underline{\hat{S}}^{(j)}_{t|t-1} = \mathbf{A}_t\,\underline{\hat{S}}^{(j)}_{t-1|t-1}$.

   (ii) Compute $\mathbf{P}^{(j)}_{t|t-1} = \mathbf{A}_t\,\mathbf{P}^{(j)}_{t-1|t-1}\,\mathbf{A}'_t$.

   (iii) Compute $\underline{\hat{S}}^{(j)}_{t|t}$ from (3.53).

   (iv) Compute $\mathbf{P}^{(j)}_{t|t}$ from (3.54).

5. Compute $\underline{\hat{B}}_t = \sum_{j=1}^{k} \mathbf{F}_{jt}\,\underline{\hat{S}}^{(j)}_{t|t}$, where $\mathbf{F}_{jt} = \frac{1}{k}\,\mathbf{I}$.

6. Compute $\hat{s}_t^2$ from (3.55).

With small modifications, this is the algorithm that was followed to get the Kalman filter credibility estimates in the next chapter. In addition to the difficulties that were mentioned in the Remark 3.1, an additional problem may arise in connection with Hachemeister's model. If the estimator in equation (3.26) is used to estimate s^2, we will encounter difficulties when $\hat{s}^2$ is taken to be zero since equation

(3.53) requires division by s^2. This problem can be avoided by using (3.55). However, (3.55) is neither robust nor in recursive form. Clearly, further work is required in finding more adequate empirical estimators.

Chapter 4

Numerical Illustrations

In this chapter, we present results from the credibility models previously discussed. The data that is analyzed is from Hachemeister (1975). There are five contracts and twelve periods for each contract. We first present the estimates based on the classical credibility models of Bühlmann, Bühlmann and Straub, and Hachemeister. These estimates will be based on Hachemeister's original data set and on Hachemeister's data set with the twelfth observation of contract five ($X_{5,12}$) replaced with an outlier. We then report estimates based on the corrupted data using the robust credibility models of Künsch, Gisler and Reinhard, and Kremer. Finally, the results using the robust Kalman filter are presented.

4.1 Classical Credibility Estimates

The estimates for the Bühlmann, Bühlmann and Straub, and Hachemeister models are shown in Tables 4.1, 4.2, and 4.3, respectively. These estimates are based on the data which is uncorrupted by outliers.

The estimators for the structural parameters for the Bühlmann and Bühlmann and Straub model are taken from Goovaerts and Hoogstad (1987). For the Hachemeister model, we use the estimator

$$\frac{1}{k}\sum_{j=1}^{k}\hat{\underline{\beta}}_j$$

for $\underline{\beta}$, and

$$(\frac{1}{k}\mathbf{I})\sum_{j=1}^{k}(\hat{\underline{\beta}}_j-\hat{\underline{\beta}})(\hat{\underline{\beta}}_j-\hat{\underline{\beta}})'$$

for $\hat{\mathbf{\Lambda}}$. Additionally, we use the design matrix

$$\mathbf{Y}_j=\begin{bmatrix}1 & 1\\ \vdots & \vdots\\ 1 & 12\end{bmatrix}$$

rather than the one used in Goovaerts and Hoogstad (1987).

In Table 4.1, the credibility estimates of Bühlmann's model are shown. The credibility factor is $Z = 0.95$. The estimate of the collective mean is $\bar{X} = 1{,}671$. Estimates of the structural parameters are $\hat{a} = 72{,}310$ and $\hat{s}^2 = 46{,}040$. The credibility factor is quite high using the Bühlmann model. This is due to the fact that the "within contracts" variance $\hat{s}^2$ is small compared to the "between contracts" variance $\hat{a}$. With less heterogeneity within a contract, more credibility is assigned to the individual data.

j	1	2	3	4	5
$\bar{X}_j$	2,064	1,511	1,822	1,360	1,599
$\hat{\mu}(\Theta_j)$	2,044	1,519	1,814	1,376	1,602

Table 4.1: Bühlmann's Model

Table 4.2 shows estimates of the Bühlmann and Straub model. The estimate of the collective mean is $X_{zw} = 1{,}684$. Estimates of the structural parameters are $\hat{a} = 89{,}639$, and $\hat{s}^2 = 139{,}120{,}026$. The results here are similar to the results in

Table 4.1. However, the credibility factor for contract 4 is low compared to the other contracts. Since we now take the number of claims into account, we can see from table A.2 that the number of claims for each time period for contract 4 is quite low compared with the rest of the portfolio. The relatively small amount of experience of contract 4 leads to a smaller credibility factor.

j	1	2	3	4	5
X_{jw}	2,061	1,511	1,806	1,353	1,600
$\hat{\mu}(\Theta_j)$	2,055	1,524	1,793	1,443	1,603
Z_j	0.98	0.93	0.90	0.73	0.96

Table 4.2: Bühlmann and Straub's Model

Table 4.3 gives the results from Hachemeister's model. Figure 4.1 shows the actual trend of contract 5 resulting from Hachemeister's model. The estimates of the collective regression parameters are $\hat{\underline{\beta}} = \begin{bmatrix} 1{,}458 \\ 32 \end{bmatrix}$. Estimates of the structural parameters are $\hat{a} = \begin{bmatrix} 26{,}517 & 1{,}544 \\ 1{,}544 & 339 \end{bmatrix}$ and $\hat{s}^2 = 49{,}870{,}187$. We can see from Figure 4.1 that the credibility line for contract 5 is closer to the individual least-squares line than to the collective line. Looking at the number of claims for contract 5 in Table A.2, we note that the number of claims at each time period is quite extensive. Adding up the number of claims for each contract, we find that $w_5 = 36{,}110$, which makes up 20.7% of the aggregate number of claims of the total portfolio. Therefore there seems to be adequate experience to assign high credibility to the individual data of contract 5.

j	1	2	3	4	5
$\hat{\underline{\beta}}_j$	$\begin{bmatrix} 1{,}658 \\ 62 \end{bmatrix}$	$\begin{bmatrix} 1{,}398 \\ 17 \end{bmatrix}$	$\begin{bmatrix} 1{,}533 \\ 43 \end{bmatrix}$	$\begin{bmatrix} 1{,}177 \\ 28 \end{bmatrix}$	$\begin{bmatrix} 1{,}522 \\ 12 \end{bmatrix}$
$\hat{\underline{\beta}}(\Theta_j)$	$\begin{bmatrix} 1{,}667 \\ 61 \end{bmatrix}$	$\begin{bmatrix} 1{,}377 \\ 21 \end{bmatrix}$	$\begin{bmatrix} 1{,}537 \\ 42 \end{bmatrix}$	$\begin{bmatrix} 1{,}297 \\ 18 \end{bmatrix}$	$\begin{bmatrix} 1{,}464 \\ 20 \end{bmatrix}$
$\mathbf{Z}_j$	$\begin{bmatrix} 0.87 & 1.13 \\ 0.02 & 0.83 \end{bmatrix}$	$\begin{bmatrix} 0.69 & 2.60 \\ 0.04 & 0.61 \end{bmatrix}$	$\begin{bmatrix} 0.65 & 2.78 \\ 0.04 & 0.55 \end{bmatrix}$	$\begin{bmatrix} 0.52 & 3.01 \\ 0.04 & 0.42 \end{bmatrix}$	$\begin{bmatrix} 0.76 & 2.06 \\ 0.03 & 0.70 \end{bmatrix}$

Table 4.3: Hachemeister's Model

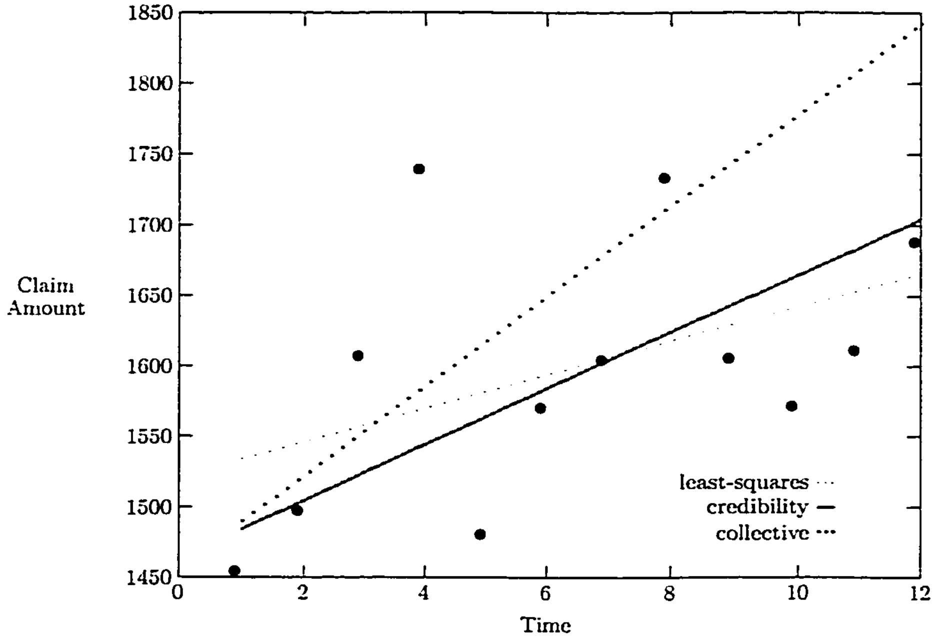

Figure 4.1: Hachemeister's Model for Contract No. 5

4.2 Robust Credibility Estimates

In this section, we compare the estimates given by the classical models of Bühlmann, Bühlmann and Straub, Hachemeister with the robust credibility models of Künsch, Gisler and Reinhard, and Kremer. Without outliers, the robust models give the same results as their non-robust counterparts. However, by introducing an outlier into Hachemeister's data set, we can see that the robust credibility models greatly mitigate the influence of a single outlying observation.

The last observation of the fifth contract ($X_{5,12}$) in the uncontaminated data set is 1,690. We wish to observe the effects on the credibility estimates when we replace that observation by $X_{5,12} = 7{,}000$.

Table 4.4 list the results of Bühlmann's model. The credibility factor is $Z = 0.55$. The collective estimate is $\bar{X} = 1,760$. The structural parameters are $\hat{a} = 54,813$ and $\hat{s}^2 = 533,627$. As expected, with an outlying claim in contract 5, the expected variance of the claims has greatly increased. Accordingly, the credibility factor has decreased significantly. Each of the individual estimates have been "pulled" towards the collective mean.

j	1	2	3	4	5
$\bar{X}_j$	2,064	1,511	1,822	1,360	2,041
$\hat{\mu}(\Theta_j)$	1,927	1,622	1,794	1,539	1,915

Table 4.4: Bühlmann's Model with $X_{5,12} = 7{,}000$

Table 4.5 are the results of Künsch's model with truncation points $c_1 = 1$ and $c_2 = 1.5$. The credibility factor is $Z = 0.67$. The average of the robust estimates is $\bar{T} = 1{,}720$. We also have

$$\frac{1}{k-1}\sum_{j=1}^{k}(T_j - \bar{T})(\bar{X}_j - \bar{X}) = 52{,}770,$$

$$\frac{1}{t(t-1)k}\sum_{j=1}^{k}\sum_{r=1}^{t}\widehat{IF}(X_{jr}, T_j)(X_{jr} - \bar{X}_j) = 32{,}453,$$

and

$$\frac{1}{k-1}\sum_{j=1}^{k}(T_j - \bar{T})^2 = 79{,}142.$$

The credibility factor using Künsch's model is greater than in the non-robust Bühlmann case. Also, each of the credibility estimates have moved closer to their individual estimates indicating that robustifying Bühlmann's model has helped in mitigating the effect of the large claim. However, in Künsch's model the sample mean of the portfolio is still used as the estimator for the collective mean. Since the sample mean is not robust, the large claim in contract 5 still has a substantial effect on the estimates.

j	1	2	3	4	5
$\bar{X}_j$	2,064	1,511	1,822	1,360	2,041
T_j	2,064	1,511	1,822	1,360	1,841
$\hat{\mu}^R(\Theta_j)$	1,989	1,620	1,828	1,520	1,841

Table 4.5: Künsch's Model with $X_{5,12} = 7{,}000$

Results from Bühlmann and Straub's model are in Table 4.6. The estimate of the collective estimator is $X_{zw} = 1{,}959$. Estimates of the structural parameters are given by $\hat{a} = 4{,}336$ and $\hat{s}^2 = 1{,}788{,}061{,}134$. The large value of $\hat{s}^2$ has caused the credibility factor to be virtually zero for each contract. Only the large number of claims for contract 1 has allowed that credibility factor to remain far from zero. However, the credibility factor of 0.20 for contract 1 is still a big decrease from 0.98 in the case with no large claims. Each credibility premium is now mainly the premium based on the entire portfolio.

Values from Gisler and Reinhard's model are in Table 4.6. The estimate of the excess mean is $\hat{\mu}_{xs} = 74$. The estimate of the ordinary mean is $\hat{\mu}_T = 1{,}733$. Estimates of the structural parameters are $\hat{a}_T = 65{,}396$ and $\hat{s}^2_T = 378{,}151{,}153$. The credibility factors are now in a more reasonable range. The credibility factor for contract 1 is now closer to the credibility factor in Bühlmann-Straub model without outliers. Contract 5, with its large number of claims, also enjoys a high credibility factor. We can explain the low credibility factor for contract 4 by recalling that the exposure for this contract is relatively low.

j	1	2	3	4	5
X_{jw}	2,061	1,511	1,806	1,353	2,103
$\hat{\mu}(\Theta_j)$	1,979	1,938	1,954	1,953	1,971
Z_j	0.20	0.05	0.03	0.01	0.08

Table 4.6: Bühlmann and Straub's Model with $X_{5,12} = 7{,}000$

j	1	2	3	4	5
T_j	2,061	1,511	1,806	1,353	1,698
$\hat{\mu}^R(\Theta_j)$	2,117	1,635	1,858	1,648	1,776
Z_j	0.95	0.77	0.70	0.42	0.86

Table 4.7: Gisler and Reinhard's Model with $X_{5,12} = 7{,}000$

In Table 4.8, results from Hachmeister's model are shown. The collective estimate is $\underline{\hat{\beta}} = \begin{bmatrix} 1{,}267 \\ 77 \end{bmatrix}$. The estimates of the structural parameters are $\hat{a} = \begin{bmatrix} 147{,}917 & -25{,}624 \\ -25{,}624 & 6{,}409 \end{bmatrix}$ and $\hat{s}^2 = 1{,}377{,}156{,}010$.

j	1	2	3	4	5
$\underline{\hat{\beta}}_j$	$\begin{bmatrix} 1{,}658 \\ 62 \end{bmatrix}$	$\begin{bmatrix} 1{,}398 \\ 17 \end{bmatrix}$	$\begin{bmatrix} 1{,}533 \\ 43 \end{bmatrix}$	$\begin{bmatrix} 1{,}177 \\ 28 \end{bmatrix}$	$\begin{bmatrix} 567 \\ 234 \end{bmatrix}$
$\underline{\hat{\beta}}(\Theta_j)$	$\begin{bmatrix} 1{,}497 \\ 77 \end{bmatrix}$	$\begin{bmatrix} 1{,}374 \\ 35 \end{bmatrix}$	$\begin{bmatrix} 1{,}332 \\ 67 \end{bmatrix}$	$\begin{bmatrix} 1{,}307 \\ 56 \end{bmatrix}$	$\begin{bmatrix} 858 \\ 186 \end{bmatrix}$
$\mathbf{Z}_j$	$\begin{bmatrix} 0.55 & -0.98 \\ 0.04 & 1.01 \end{bmatrix}$	$\begin{bmatrix} 0.11 & -1.56 \\ 0.06 & 0.83 \end{bmatrix}$	$\begin{bmatrix} 0.07 & -1.35 \\ 0.05 & 0.70 \end{bmatrix}$	$\begin{bmatrix} -0.01 & -0.82 \\ 0.03 & 0.37 \end{bmatrix}$	$\begin{bmatrix} 0.25 & -1.50 \\ 0.05 & 0.94 \end{bmatrix}$

Table 4.8: Hachemeister's Model with $X_{5,12} = 7{,}000$

Table 4.9 describes the results from Kremer's robust regression credibility model. We show here the special case of Hachemeister's model. The computations were done assuming a one-sided Huber function. The M-estimator $\underline{\hat{B}}_j$ was computed by *iterated re-weighted least-squares.*

The tuning constant c for the one-sided Huber function is usually taken to be 1.345, since for this value, the M-estimator will be 95% efficient at the normal distribution. We retain this convention, however, we note that since we have not estimated the scale parameter in (2.44), we need to multiply 1.345 by an estimate of scale. For this, we use $\hat{\sigma}_T$, the expected variance of the robust means from the Künsch model. Finally, to adjust the truncation point for claims volume, we use $\frac{1}{kt}\sqrt{\sum_{j=1}^{k}\sum_{r=1}^{t} w_{jr}}$. Our ad hoc procedure for calculating the truncation point c then is

$$c = (1.345\hat{\sigma}_T/kt)\sqrt{\sum_{j=1}^{k}\sum_{r=1}^{t} w_{jr}}\,,$$

which for our example is $c = 19{,}982$.

The collective estimate is $\underline{\hat{\beta}} = \begin{bmatrix} 1{,}440 \\ 36 \end{bmatrix}$. The estimates of the structural parameters are $\hat{a} = \begin{bmatrix} 25{,}488 & 1{,}879 \\ 1{,}879 & 238 \end{bmatrix}$ and $\hat{s}^2 = 64{,}618{,}028$. Figure 4.2 compares the Hachemeister non-robust trend estimate with the robust estimates from Kremer's robust regression model. It is clear that truncating the claims has a big effect on the estimate of the trend. The line provided by the Hachemeister model can be seen to be pulled towards the large claim. The line from Kremer's robust regression credibility model, fits the data quite well.

j	1	2	3	4	5
$\underline{\hat{B}}_j$	$\begin{bmatrix} 1{,}658 \\ 62 \end{bmatrix}$	$\begin{bmatrix} 1{,}398 \\ 17 \end{bmatrix}$	$\begin{bmatrix} 1{,}533 \\ 43 \end{bmatrix}$	$\begin{bmatrix} 1{,}177 \\ 28 \end{bmatrix}$	$\begin{bmatrix} 1{,}438 \\ 31 \end{bmatrix}$
$\underline{\hat{\beta}}^R(\Theta_j)$	$\begin{bmatrix} 1{,}670 \\ 60 \end{bmatrix}$	$\begin{bmatrix} 1{,}345 \\ 25 \end{bmatrix}$	$\begin{bmatrix} 1{,}520 \\ 43 \end{bmatrix}$	$\begin{bmatrix} 1{,}282 \\ 22 \end{bmatrix}$	$\begin{bmatrix} 1{,}422 \\ 33 \end{bmatrix}$
$\mathbf{Z}_j$	$\begin{bmatrix} 0.77 & 2.40 \\ 0.03 & 0.65 \end{bmatrix}$	$\begin{bmatrix} 0.62 & 3.60 \\ 0.05 & 0.46 \end{bmatrix}$	$\begin{bmatrix} 0.60 & 3.54 \\ 0.05 & 0.42 \end{bmatrix}$	$\begin{bmatrix} 0.50 & 3.28 \\ 0.04 & 0.33 \end{bmatrix}$	$\begin{bmatrix} 0.67 & 3.28 \\ 0.42 & 0.52 \end{bmatrix}$

Table 4.9: Kremer's Robust Regression Model with $X_{5,12} = 7{,}000$

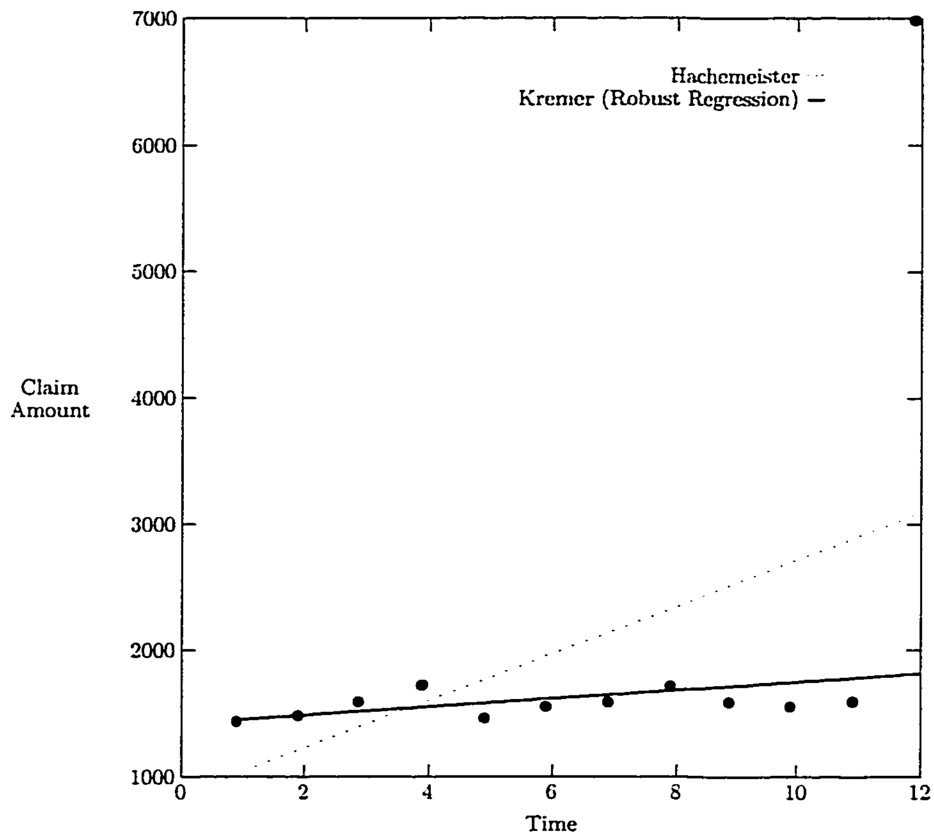

Figure 4.2: Kremer's Robust Regression Model with $X_{5,12} = 7{,}000$

4.3 Kalman Filter Estimates

We conclude our examples with estimates from the non-robust and robust versions of the Kalman filter applied to the Hachemeister dataset with and without an outlier. The implementation of the filter is based on section 3.4.

We begin with a non-robust version of the Kalman filter applied to the data which contains no outliers. We can then compare these estimates to the robust Kalman filter in the presence of a large claim.

Table 4.10 shows the estimates of the non-robust Kalman filter. The collective estimator is $\underline{\beta} = \begin{bmatrix} 1{,}483 \\ 32 \end{bmatrix}$. The estimate for s^2 is 49,870,186. We omit the estimates for a and $\mathbf{Z}_j$ as the computations of these matrices are embedded in the recursions and were not computed explicitly. Figure 4.3 shows the results of the Kalman filter graphically. The estimates of $\beta(\Theta_j)$ given by Table 4.10 are similar to the estimates in Table 4.3. Looking at Figure 4.3, we can see a similar relationship between the credibility line and the collective line that we saw in Figure 4.1.

j	1	2	3	4	5
$\underline{\hat{\beta}}(\Theta_j)$	$\begin{bmatrix} 1{,}745 \\ 54 \end{bmatrix}$	$\begin{bmatrix} 1{,}370 \\ 22 \end{bmatrix}$	$\begin{bmatrix} 1{,}479 \\ 47 \end{bmatrix}$	$\begin{bmatrix} 1{,}301 \\ 22 \end{bmatrix}$	$\begin{bmatrix} 1{,}507 \\ 17 \end{bmatrix}$

Table 4.10: Kalman Filter (Hachemeister) Model for Contract No. 5

Finally, we compare the robust and non-robust Kalman filter as applied to the data with $X_{5,12} = 7{,}000$, an outlier. The collective estimator in the non-robust case is $\underline{\beta} = \begin{bmatrix} 1{,}155 \\ 92 \end{bmatrix}$, and in the robust case, $\underline{\beta} = \begin{bmatrix} 1{,}267 \\ 38 \end{bmatrix}$. In both cases, $s^2 = 1{,}377{,}156{,}010$. Figure 4.4 compares the regression lines from the two models. In the robust case, we use for the truncation point $c = 0$. This is to recognize that at time t, we expect convergence to an estimate, and so the residual $X_{jt} - \underline{H}_{jt}\,\hat{\underline{S}}_{t|t-1}$ should be small. The graphs in Figure 4.4 are similar to Figure 4.2. In Figure 4.4, we see that truncating

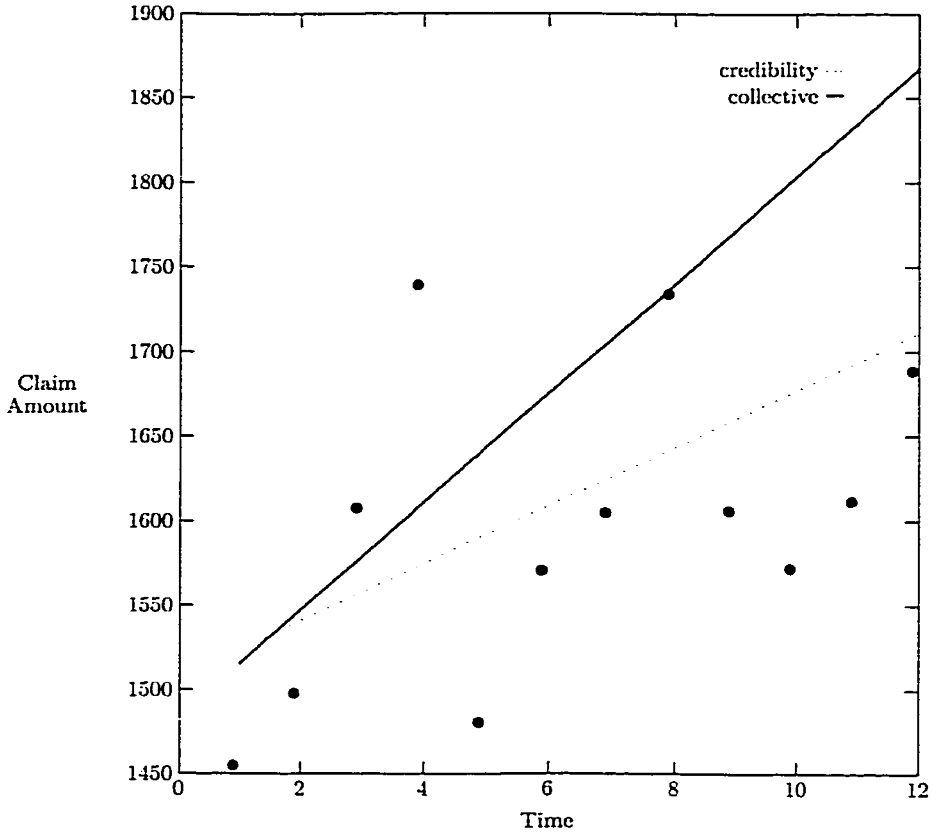

Figure 4.3: Kalman Filter (Hachemeister) Model for Contract No. 5

the claims have improved the fit greatly. It appears that a robust Kalman filter can be useful in reducing the difficulties in credibility estimation that are caused by large claims.

j	1	2	3	4	5
$\underline{\hat{\beta}}(\Theta_j)$	$\begin{bmatrix}1,837\\43\end{bmatrix}$	$\begin{bmatrix}1,360\\22\end{bmatrix}$	$\begin{bmatrix}1,453\\52\end{bmatrix}$	$\begin{bmatrix}1,334\\19\end{bmatrix}$	$\begin{bmatrix}-197\\323\end{bmatrix}$
$\underline{\hat{\beta}}^R(\Theta_j)$	$\begin{bmatrix}1,112\\77\end{bmatrix}$	$\begin{bmatrix}1,338\\18\end{bmatrix}$	$\begin{bmatrix}1,210\\53\end{bmatrix}$	$\begin{bmatrix}1,342\\11\end{bmatrix}$	$\begin{bmatrix}1,345\\30\end{bmatrix}$

Table 4.11: Robust Kalman Filter (Hachemeister) Model for Contract No. 5 with $X_{5,12} = 7{,}000$

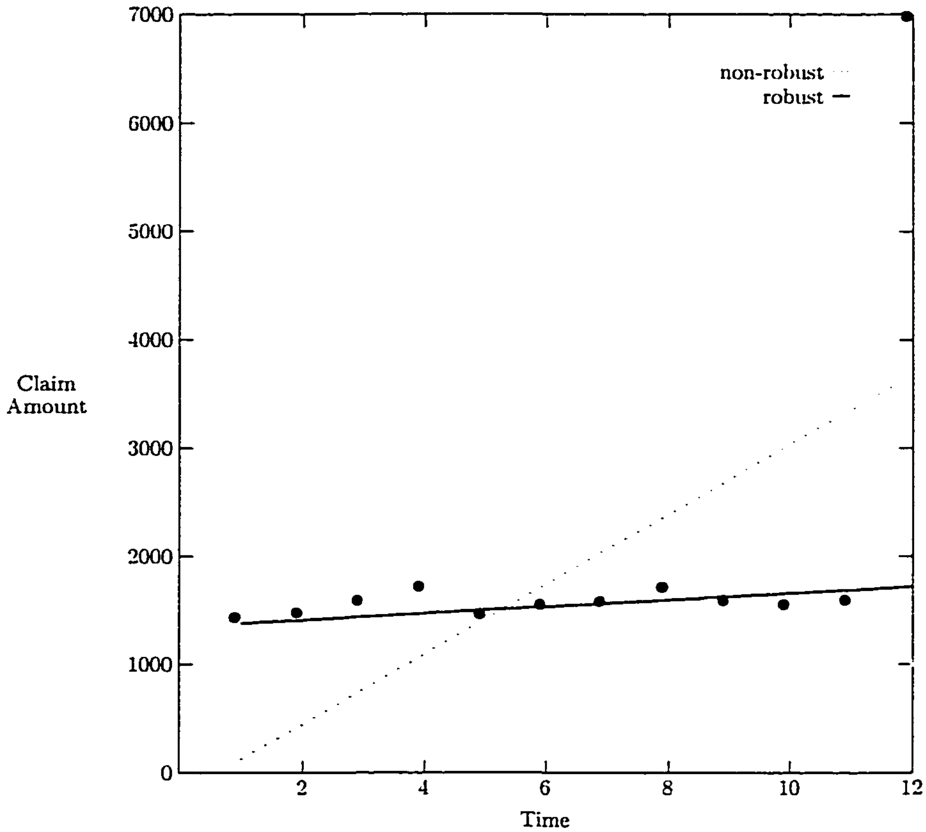

Figure 4.4: Robust Kalman Filter (Hachemeister) Model for Contract No. 5 with $X_{5,12} = 7{,}000$

Conclusion

In this thesis, we have reviewed some of the developments in credibility theory leading up to the treatment of large claims using robust Kalman filtering methods. The Kalman filter is considered useful in credibility theory as it also provides a unified framework in which to apply credibility. In chapter 2, we stated why robustness in credibility should be pursued and we saw in chapter 4 that, indeed, large claims effects can be detrimental to credibility estimation. Robust methods which have already been applied to a variety of credibility models was reviewed in chapter 2. Thus, robustification of the Kalman filter allows for efficient processing of credibility estimates which are robust against large claims.

Parameter estimation is also very important in credibility theory as the credibility formulas cannot be applied until the structural parameters have been estimated. Parameter estimation was discussed in the classical, robust, and Kalman filter approaches to credibility. Some difficulties in estimating parameters when applying the Kalman filter were noted.

In conclusion, we point out some areas where further work may be done. We have seen that clear solutions to the estimation of the structural parameters and the choice of initial values in the robust empirical Kalman filter credibility model have not yet been developed. Estimators have been proposed in the case of the non-robust Kalman credibility model; however, the question of empirical estimators based on data for the robust case has yet to be resolved. The choice of the initial estimate to start the recursive procedure is also an important question. Finally, we note that in both estimating the initial value and estimating the model parameters, one should seek estimators which are robust.

Appendix A

Hachemeister's Dataset

The following tables contain private passenger automobile data used by Hachemeister in his 1975 article. The first table contains the claim amounts while the second table contains the number of claims. Both tables are split by state with each column representing a state. The rows correspond to the time periods. For this data, each time period is equal to three months.

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

Table A.1: Claim amounts from private passenger bodily injury (Hachemeister, 1975)

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

Table A.2: Number of claims from private passenger bodily injury (Hachemeister, 1975)

References

Bailey, A.L. (1950). Credibility procedures, Laplace's generalization of Bayes' rule and the combination of collateral knowledge with observed data. *Proceedings of the Casualty Actuarial Society*, **XXXVII**, 7-23.

Bühlmann, H. (1967). Experience rating and credibility. *Astin Bulletin*, **4**, 199-207.

Bühlmann, H. (1969). Experience rating and credibility. *Astin Bulletin*, **5**, 157-165.

Bühlmann, H. and Straub, E. (1970). Glaubwürdigkeit für schadensätze. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, **70**, 111-133.

Cipra, T. and Romera, R. (1991). Robust Kalman filter and its application in time series analysis. *Kybernetika*, **27**, 481-494.

De Jong, P. and Zehnwirth, B. (1983). Credibility theory and the Kalman filter. *Insurance: Mathematics and Economics*, **2**, 281-286.

De Vylder, F. (1981). Practical credibility theory with emphasis on optimal parameter estimation. *Astin Bulletin*, **12**, 115-131.

Gisler, A. (1980). Optimum trimming of data in the credibility model. *Bulletin of the Association of Swiss Actuaries*, **80**, 313-325.

Gisler, A. and Reinhard, P. (1993). Robust credibility. *Astin Bulletin*, **23**, 117-143.

Goovaerts, M.J. and Hoogstad, W.J. (1987). *Credibility Theory, Surveys in Actuarial Studies No. 4*. Research Department, Nationale-Nederlanden N.V.

Goovaerts, M.J., Kaas, R., Van Heerwaarden, A.E., and Bauwelinckx, T. (1990). *Effective Actuarial Methods.* North-Holland.

Hachemeister, C.A. (1975). Credibility for regression models with applications to trend. In P. Kahn (Ed.), *Credibility, Theory and Applications. Proceedings of the Berkeley Actuarial Research Conference on Credbility*, 129-163. Academic Press, New York.

Hampel, F.R. (1968). *Contributions to the Theory of Robust Estimation.* Ph.D. thesis, University of California, Berkeley.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1984). *Robust Statistics: The Approach Based on Influence Functions.* John Wiley & Sons, New York.

Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**, 73-101.

Huber, P.J. (1981). *Robust Statistics.* John Wiley & Sons, New York.

Jewell, W. (1974). Credible means are exact bayesian for exponential families. *Astin Bulletin*, **8**, 77-90.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, **82D**, 35-45.

Kalman, R.E. and Bucy, R.S. (1961). New results in linear filtering and prediction. *Journal of Basic Engineering*, **83D**, 95-108.

Klugman, S.A. (1985). Robust credibility. In *Proceedings. Nineteenth Annual Actuarial Research Conference, 1985.2 Issue*, 209-220. Society of Actuaries, Schaumburg.

Kremer, E. (1991). Large claims in credibility. *Blätter der deutschen Gesellschaft für Versicherungsmathematik*, **XX**, 123-150.

Kremer, E. (1994). Robust credibility via robust Kalman filtering. *Astin Bulletin*, **24**, 221-233.

Kremer, E. (1995). Empirical Kalman-credibility. *Blätter der Deutsche Gesellschaft für Versicherungsmathematik*, **XXII**, 1-11.

Künsch, H.R. (1992). Robust methods for credibility. *Astin Bulletin*, **22**, 33-49.

Ledolter, J., Klugman, S.A., and Lee, C. S. (1991). Credibility models with time-varying trend components. *Astin Bulletin*, **21**, 73-91.

Mayerson, A.L. (1964). A Bayesian view of credibility. *Proceedings of the Casualty Actuarial Society*, **LI**, 85-104.

Mehra, R.K. (1975). *Credibility theory and Kalman filtering with extensions* (Research Memorandum No. RM-75-64). International Institute for Applied Systems Analysis.

Meinhold, R.J. and Singpurwalla, N.D. (1983). Understanding the Kalman filter. *The American Statistician*, **37**, 123-127.

Mowbray, A.H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium. *Proceedings of the Casualty Actuarial Society*, **I**, 24-30.

Norberg, R. (1979). The credibility approach to experience rating. *Scandinavian Actuarial Journal*, 181-221.

Norberg, R. (1980). Empirical Bayes credibility. *Scandinavian Actuarial Journal*, 177-194.

Whitney, A.W. (1918). The theory of experience rating. *Proceedings of the Casualty Actuarial Society*, IV, 274-292.

Zehnwirth, B. (1984). Credibility: estimation of structural parameters. In F. de Vylder, M. Goovaerts, & J. Haezendonck (Eds.), *Premium Calculation in Insurance*, 347-359. D. Reidel Publishing Company.