Stochastic Loss Reserving with Emphasis on the Bornhuetter-Ferguson Method

Diss. ETH No. 22267

Stochastic Loss Reserving with Emphasis on the Bornhuetter-Ferguson Method

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

ANNINA SALUZ

MSc ETH Mathematics, ETH Zurich
born March 27, 1986
citizen of Lavin (GR), Switzerland

accepted on the recommendation of

Prof. Dr. Alois Gisler, examiner
Prof. Dr. Paul Embrechts, co-examiner
Prof. Dr. Franco Moriconi, co-examiner
Prof. Dr. Mario V. Wüthrich, co-examiner

2014

DOI: 10.3929/ethz-a-010358837
ISBN: 978-3-906031-88-0

Acknowledgements

I would like to express my sincere gratitude to Paul Embrechts, Alois Gisler and Mario Wüthrich for giving me the opportunity to do my PhD at RiskLab. First and foremost I am indebted to my supervisor Alois Gisler for constant advice, encouragement and for sharing his invaluable practical experience with me. He always showed great interest in my work and he also supported me on personal matters. I am most grateful to Hans Bühlmann and Mario Wüthrich for excellent collaboration and numerous interesting discussions. Further I would like to thank Paul Embrechts for creating an inspiring and familiar atmosphere at RiskLab and for encouraging me to attend different conferences to meet various researchers. Apart from work I have many great memories from our social activities at RiskLab.

A big thank you goes to Franco Moriconi for valuable collaboration and for organising a memorable research visit for me at Università di Perugia. Moreover, I would like to thank him for his willingness to be co-examiner of this thesis.

During my PhD I had the opportunity to work in the reserving team at AXA Winterthur. I am grateful to Philipp Reinmann, Lukas Meier, Ursin Mayer and Francesco Pagliari for introducing me to applications and problems occurring in everyday life of an actuary.

I would like to thank all my colleagues and friends for the wonderful time I had at ETH Zurich and outside.

Special thanks go to: Philippe Deprez, Tara Huber, Mirjana Vukelja, Laurent Huber, Edgars Jakobsons, Matthias Kirchner, David Stefanovits, Philipp Arbenz, Robert Salzmann, Erwan Koch, Anne McKay, Galit Shoham, as well as the whole D-MATH Group 3. In particular, I would like to thank Philippe for all his help and patience during the final steps of my PhD.

Finally, I thank my family for their confidence and constant support during all those years.

Annina Saluz

Abstract

Claims reserving is one of the most important tasks in any non-life insurance company. Accurate claims reserves are essential for the pricing of insurance products and consequently they have a large impact on the profitability of a company. A change of the reserves by a small percentage can change the yearly result from very positive to rather negative and vice versa.

Historically, most claims reserving methods were established as deterministic algorithms which give a point estimate for the outstanding loss liabilities. Such a deterministic assessment was sufficient to satisfy the Solvency I requirements, which was the EU regulation at that time. However, in order to quantify the uncertainties in the reserve estimates, the claims reserving problem needs to be embedded in a stochastic framework. Moreover, the stock market crash following the terrorist attacks of 9/11 clearly showed that the rule-based regulation of Solvency I was not sufficient. These experiences lead to the development of new risk-adjusted regulatory frameworks, such as the Swiss Solvency Test in Switzerland and Solvency II in the EU. The new risk-based regulations require a quantification of the uncertainties in claims reserves. Therefore, interest in stochastic claims reserving was reinforced in recent years. In this thesis we contribute to problems in stochastic claims reserving.

The Bornhuetter-Ferguson (BF) method is one of the most popular reserving methods. In Paper A we consider distributional models for the BF method. Within these models we derive parameter estimates and quantify the prediction uncertainty.

In Paper B we introduce the BF method with repricing. The repricing allows to incorporate claims experience in order to update a priori estimates used in the BF method. Such repricing procedures are common in practice and we show how repricing can be done in a mathematically consistent way by using credibility theory.

In Paper C we show that claims development results with the same sign over several consecutive calendar years are not a contradiction to best estimate reserves. In contrast, such a phenomenon is to be expected in certain situations of change.

In Paper D we consider the Cape Cod method in a stochastic framework and we derive estimates for the prediction uncertainty and for the uncertainty in the claims development result.

Kurzfassung

Schadenreservierung ist eine der wichtigsten Aufgaben in einer Nichtleben-Versicherungsunternehmung. Angemessene Reserven sind essentiell für die Tarifierung von Versicherungsprodukten und somit auch für die Rentabilität der Unternehmung. Eine Veränderung der Reserven um einen kleinen Prozentsatz kann ein positives Jahresergebnis in ein Negatives ändern und umgekehrt.

Die meisten Schadenreservierungsmethoden wurden als deterministische Algorithmen etabliert, welche eine Punktschätzung für die ausstehenden Schadensverpflichtungen liefern. Eine solche deterministische Betrachtung war ausreichend, um die damaligen EU Solvenzanforderungen aus Solvency I zu erfüllen. Um die Unsicherheiten in den Schadenreserven zu quantifizieren, muss das Schadenreservierungsproblem jedoch in einen stochastischen Rahmen eingebettet werden. Der Börsencrash nach den Anschlägen vom 11. September 2001 zeigte zudem auf, dass die regelbasierten regulatorischen Anforderungen aus Solvency I nicht ausreichend waren. Diese Erfahrungen führten zur Entwicklung von neuen risikobasierten regulatorischen Anforderungen wie dem Schweizer Solvenztest und Solvency II in der EU. Die neuen risikobasierten Regelungen verlangen eine Quantifizierung der Unsicherheiten in den Schadensreserven und daher ist das Interesse an stochastischen Schadenreservierungsmethoden in den letzten Jahren stark gewachsen. In dieser Arbeit behandeln wir Probleme aus der stochastischen Schadenreservierung.

Die Bornhuetter-Ferguson (BF) Methode ist eine der bekanntesten und meistbenutzten Reservierungsmethoden. In Artikel A betrachten wir stochastische Modelle mit parametrischen Verteilungsannahmen für die BF Methode. In diesen Modellen leiten wir Parameterschätzer her und wir quantifizieren die Prognoseunsicherheit.

In Artikel B führen wir die BF Methode mit Neukalkulation (repricing) ein. Die Neukalkulation ermöglicht die Einbindung von Schadenserfahrung, um die in der BF Methode benutzten a priori Schätzer zu aktualisieren. Solche Neukalkulationen kommen in der Praxis häufig vor. Wir zeigen, wie eine Neukalkulation mit Kredibilitätstheorie auf eine mathematisch konsistente Art möglich ist.

In Artikel C zeigen wir, dass Abwicklungsergebnisse mit demselben Vorzeichen über mehrere aufeinanderfolgende Kalenderjahre keinen Widerspruch zu 'best estimate' Reserven sind. Im Gegenteil sind solche Phänomene in gewissen Situationen von Veränderung zu erwarten.

In Artikel D betrachten wir die Cape Cod Methode in einem stochastischen Rahmen und wir leiten Schätzer für die Prognoseunsicherheit und für die Unsicherheit im Abwicklungsergebnis her.

Contents

Accompanying Papers

A Annina Saluz, Alois Gisler, Mario V. Wüthrich.
Development Pattern and Prediction Error for the Stochastic Bornhuetter-Ferguson Claims Reserving Method.
ASTIN Bulletin, **41**(2), 2011, 279–313.

B Annina Saluz, Hans Bühlmann, Alois Gisler, Franco Moriconi.
The Bornhuetter-Ferguson Method with Repricing.
Submitted.

C Annina Saluz, Alois Gisler.
Best Estimate Reserves and the Claims Development Results in Consecutive Calendar Years.
Annals of Actuarial Science, **8**(2), 2014, 351–373.

D Annina Saluz.
Prediction Uncertainties in the Cape Cod Reserving Method.
Annals of Actuarial Science, forthcoming in 2015.

1. Introduction

In this thesis we study the claims reserving problem in non-life insurance. This branch of insurance is known as Property and Casualty Insurance in the United States and as General Insurance in Great Britain. The separation from life insurance is due to the different nature in the modelling of life insurance and non-life insurance cash flows and products. In Switzerland and in many other countries non-life and life insurance business need to be split into different legal entities by law, too.

In a non-life insurance contract the insurer provides the insured a guarantee of financial coverage against loss resulting from a well-specified random event. The insured pays a fixed amount for this guarantee. Hence, an insurance contract is an exchange of risk: the insured exchanges the risk of a random loss against a deterministic premium. The precise conditions of insured risks or people and the coverage are settled in the insurance contract. The amount that the insurer is obliged to pay to the insured in case of a loss event is called claim amount or loss amount. The time-lag between the occurrence of a claim and its reporting to the insurer is called reporting-delay. In certain lines of business the reporting-delay can amount to several years. For instance, in the case of asbestos claims the claim (disease) is usually discovered by the insured many years after it was caused. After a claim is reported, the settlement of the claim normally takes some time. Depending on the line of business the settlement can take several years (typically in liability lines of business) or it can be rather fast (typically in property insurance). It is also possible that closed claims need to be reopened due to new information, relapses or other changes. Due to the time-lag between occurrence and settlement of a claim and due to the uncertainties in the claims development, insurance companies need to set adequate reserves for future payments. Claims reserves are typically the most important item on the liability side of the balance sheet of an insurance company. Moreover, the reserves are central for the determination of the premium and hence have a decisive impact on the profitability of an insurance company.

Many reserving methods evolved as deterministic algorithms. However, it is not sufficient to predict outstanding claims, but it is also necessary to

assess the uncertainties in these predictions. In order to quantify uncertainties in claims reserves one has to model the stochastic nature of the claims process. Such a risk-based assessment of claims reserves is called *stochastic claims reserving*. New solvency regulations require stochastic claims reserving methods and, hence, this field received considerable attention in recent years. However, it should be noted that many important contributions to stochastic claims reserving appeared already before the establishment of the new solvency regulations, see for instance Taylor (2000).

The focus of this thesis lies on the Bornhuetter-Ferguson (BF) method, which is one of the most popular claims reserving methods. Like many other methods, the BF method evolved as a deterministic algorithm. In this thesis we study the BF method in a stochastic framework.

In the next section we will describe the claims reserving problem in more detail and we will introduce the most commonly used methods. The subsequent sections summarise the contributions in the author's papers. The papers in full length are given in the Appendix.

1.1 Claims Reserving Notation and Data Structure

Claims reserving data is usually given in the form of a triangle or trapezoid. The rows represent the accident years (years of occurrence), which are indexed by $i \in \{0, 1, \ldots, I\}$. The columns represent the development years, which are indexed by $j \in \{0, 1, \ldots, J\}$. Thereby I denotes the most recent accident year and $J \leq I$ denotes the last development year. The diagonals $i + j = k$ in a claims development trapezoid describe the accounting years.

In many insurance companies claims adjusters provide an estimate of the outstanding loss liabilities for each reported claim, called case reserves. Sometimes default values are used for the small claims instead of estimating them separately. The aggregated amount of all case reserves in a portfolio is called claims incurred. Claims incurred provide an estimate for the outstanding loss liabilities of reported claims. Additionally to the reported claims one has to consider incurred but not reported (IBNR) claims. For the IBNR claims there is no information on individual claims basis available. However, by modelling the aggregated incurred claims amount we can estimate reserves for reported claims and for IBNR claims at the same time.

The data in a claims development triangle/trapezoid can be from claim payments, number of claims or claims incurred. In this thesis we consider data from payments or incurred claim amounts and the case of claim numbers is not treated. We distinguish between incremental claim amounts denoted by

$X_{i,j}$ and cumulative claim amounts denoted by

$$C_{i,j} = \sum_{l=0}^{j} X_{i,l}.$$

For instance, if we consider payments, then $X_{i,j}$ denotes all payments in development year j for claims that occurred in accident year i, and $C_{i,j}$ is the total nominal amount paid up to development year j for claims with accident year i. Similarly, if $X_{i,j}$ denotes the change in reported claim amount in development year j for claims with accident year i, then $C_{i,j}$ is the total amount of case estimates for claims of accident year i at the end of development year j, that is, the claims incurred at the end of development year j. We assume that all claims are settled after $J \leq I$ years, that is, $X_{i,j} = 0$ for $j > J$, and hence $C_{i,J}$ is called the ultimate claim amount.

accident year i	development year j
	$0 \quad 1 \quad \dots \quad j \quad \dots \quad J$
0	
1	observations $\mathcal{D}_I$
$\vdots$	
i	
$\vdots$	$\mathcal{D}_I^c$
I	

Table 1.1: Claims development trapezoid: the observations are given in the upper left trapezoid $\mathcal{D}_I$. The lower right triangle $\mathcal{D}_I^c$ needs to be predicted.

At time I we have the observations in the upper triangle/trapezoid

$$\mathcal{D}_I = \{X_{i,j};\ i + j \leq I,\ j \leq J\},$$

and our goal is to predict the lower right triangle $\mathcal{D}_I^c = \{X_{i,j};\ i + j > I,\ i \leq I,\ j \leq J\}$. The data structure is illustrated in Table 1.1.

The outstanding loss liabilities for accident year $i > I - J$ at time I are given by

$$R_i = C_{i,J} - C_{i,I-i} = \sum_{j=I-i+1}^{J} X_{i,j}, \tag{1.1}$$

and the total outstanding loss liabilities of all accident years are given by $R = \sum_{i=I-J+1}^{I} R_i$.

Remark 1.1 The 'true' outstanding loss liabilities are given by formula (1.1) only if $C_{i,j}$ denote cumulative payments. For incurred losses $C_{i,j}$ the outstanding loss liabilities are given by

$$R_i = C_{i,J} - C_{i,I-i} + C_{i,I-i} - C^{paid}_{i,I-i}, \tag{1.2}$$

with $C^{paid}_{i,j}$ denoting the cumulative payments of accident year i up to development year j. Note that the additional term $C_{i,I-i} - C^{paid}_{i,I-i}$ is observable at time I and has no impact on the claims prediction problem and the corresponding uncertainty. Therefore, we only consider the outstanding loss liabilities as defined in (1.1).

The fundamental tasks in claims reserving are the prediction of the outstanding loss liabilities R_i and R, and the quantification of the uncertainties in these predictions. A predictor $\hat{R}_i$ of R_i is called claims reserve for accident year i and

$$\hat{R} = \sum_{i=I-J+1}^{I} \hat{R}_i$$

is the total reserve at time I.

The solvency regulation in Switzerland is described in the Swiss Solvency Test (SST), see FINMA (2004; 2006). For the valuation of the liabilities in the SST a best estimate principle is applied: best estimate reserves are reserves that do not contain any implicit or explicit safety or fluctuation loadings. In addition to best estimate reserves, the SST requires a risk margin (market value margin) corresponding to the capital cost for the additional capital needed for protection against adverse developments. For the calculation of the risk margin a market-consistent valuation of the liabilities is prescribed in the SST. Similar rules apply in the EU regulation Solvency II, see European Union (2009). In this thesis we focus on the calculation of best estimate reserves. For considerations on the risk margin we refer to Salzmann-Wüthrich (2010) and Wüthrich et al. (2011). Details on market-consistent valuation are given in Wüthrich et al. (2010) and Wüthrich-Merz (2013).

1.2 Uncertainties: Short-term and Long-term View

The prediction of the outstanding loss liabilities can be based on different information. At time I we have information $\mathcal{D}_I$ from the data. Possibly, there is additional information available such as external information from

experts or from other data triangles/trapezoids. Let $\mathcal{I}_I$ denote the σ-algebra generated by all information that is taken into account for the calculation of the reserves at time I. Typically, $\mathcal{I}_I$ will consist of the information $\mathcal{D}_I$ from the triangle/trapezoid and some external information on some parameters. We assume that $\hat{R}_i$ are $\mathcal{I}_I$-measurable predictors for R_i and

$$\hat{C}_{i,J} = C_{i,I-i} + \hat{R}_i$$

are the corresponding predictions of the ultimate claim amount for accident year i, where we assume $\mathcal{D}_I \subset \mathcal{I}_I$. The prediction of outstanding loss liabilities naturally raises the question of the accuracy of this prediction: we are not only interested in the estimation of claims reserves but also in the variability of these reserves. Moreover, an assessment of adverse developments is required by new solvency regimes such as the SST. In order to quantify the uncertainties of the estimated claims reserves we study second moments. More precisely, for an $\mathcal{I}_I$-measurable predictor $\hat{C}_{i,J}$ we consider the conditional mean square error of prediction (MSEP) given by

$$\begin{aligned}\operatorname{msep}_{C_{i,J}|\mathcal{I}_I}\left(\hat{C}_{i,J}\right) &= E\left[\left(\hat{C}_{i,J} - C_{i,J}\right)^2 \middle| \mathcal{I}_I\right] = E\left[\left(\hat{R}_i - R_i\right)^2 \middle| \mathcal{I}_I\right] \\ &= \operatorname{msep}_{R_i|\mathcal{I}_I}\left(\hat{R}_i\right),\end{aligned}$$

and we observe that the conditional MSEP for the reserve $\hat{R}_i$ and the predictor $\hat{C}_{i,J} = C_{i,I-i} + \hat{R}_i$ coincide. Due to the $\mathcal{I}_I$-measurability of $\hat{C}_{i,J}$ we can decouple the conditional MSEP into a variance term and an estimation error term:

$$\operatorname{msep}_{C_{i,J}|\mathcal{I}_I}\left(\hat{C}_{i,J}\right) = \underbrace{\operatorname{Var}\left(C_{i,J}|\mathcal{I}_I\right)}_{=\mathrm{PV}_i} + \underbrace{\left(\hat{C}_{i,J} - E\left[C_{i,J}|\,\mathcal{I}_I\right]\right)^2}_{=\mathrm{PEE}_i}.$$

The first term is the conditional process variance, which describes the pure randomness in the model. The second summand describes the uncertainties in the parameter estimates and is called parameter estimation error.

The study of the conditional MSEP of the ultimate claim is a long-term view; it considers the full run-off. A long-term view is important in order to remain solvent over a longer time.

On the other hand a short-term view is important in solvency regulations, for instance, in the SST and in Solvency II a time horizon of one year is considered. An early contribution to this one-year view is given in De Felice-Moriconi (2003). After one year new data and possibly new external information are available and the prediction of the outstanding loss liabilities is updated according to this new information. Let $\hat{C}_{i,J}^{(t)}$ denote the predictor of

$C_{i,J}$ at time t. The claims development result (CDR) considers the difference between two successive predictions of the ultimate claim. At time $t+1 > I$, the CDR for accident year i is defined as

$$\mathrm{CDR}_i^{(t+1)} = \hat{C}_{i,J}^{(t)} - \hat{C}_{i,J}^{(t+1)},$$

and the total CDR is defined as $\mathrm{CDR}^{(t+1)} = \sum_{i=t-J+1}^{t} \mathrm{CDR}_i^{(t+1)}$. The CDR was for instance studied in Merz-Wüthrich (2008) and Ohlsson-Lauzeningks (2009). Ohlsson-Lauzeningks (2009) give a general simulation approach for the one-year reserve risk, whereas Merz-Wüthrich (2008) derive analytical formulas for the conditional MSEP in the CDR for the chain ladder (CL) method which is one of the most used claims reserving methods in practice. In the context of the CL method also higher moments of the CDR are studied in Salzmann et al. (2012). For other claims reserving methods there is less literature on the CDR. In Paper D we derive a formula for the conditional MSEP of the CDR in the Cape Cod method.

As outlined in Merz-Wüthrich (2008) the CDR has a direct impact on the profit & loss statement of an insurance company. Moreover in Merz-Wüthrich (2008) the importance of the short-term view for management, regulators, investors and clients is explained. For instance, many management actions in a non-life insurance company need to be taken on a yearly basis. Management often considers the CDR as an indicator whether the reserves are accurate. It is argued that with 'best estimate' reserves, that is, reserves that are neither on the prudent nor on the aggressive side, the CDR should fluctuate around zero. However, if the CDR is for instance positive over several accounting years, this is not necessarily an indicator that the reserves are too conservative. We analyse this issue in more detail in Paper C and aim to provide a mathematical argument to solve a common misunderstanding between actuaries and management.

1.3 Basic Claims Reserving Methods

In this section we briefly present some basic claims reserving methods. The most popular claims reserving methods are the CL and the Bornhuetter-Ferguson (BF) method (Bornhuetter-Ferguson, 1972). These methods were both established as pure deterministic algorithms which deliver an estimate for the claims reserves. Whereas for the CL method the uncertainties in these estimates were studied already in 1993 by Mack (1993), such considerations are rather new for the BF method.

The methods we consider in this thesis are all based on one triangle (or trapezoid) of either cumulative payments or incurred claims data. In the

literature there are also methods that incorporate both sources of information. Examples for such methods are the Munich CL method by Quarg-Mack (2004) and the paid-incurred chain claims reserving method by Merz-Wüthrich (2010).

1.3.1 Chain Ladder Method

There are different stochastic models underlying the CL method. The following distribution-free model was introduced in Mack (1993).

Model Assumptions 1.2 (Distribution-free CL Model)

- *Cumulative claims $C_{i,j}$ of different accident years i are independent.*
- *There exist positive development factors $f_0, f_1, \ldots, f_{J-1}$ and positive parameters σ_0^2, σ_1^2, $\ldots$, σ_{J-1}^2 such that for all $0 \le i \le I$ and all $1 \le j \le J$*

$$\begin{aligned} E[C_{i,j}|C_{i,0}, C_{i,1}, \ldots, C_{i,j-1}] &= f_{j-1} C_{i,j-1}, \\ \mathrm{Var}(C_{i,j}|C_{i,0}, C_{i,1}, \ldots, C_{i,j-1}) &= \sigma_{j-1}^2 C_{i,j-1}. \end{aligned}$$

The CL method estimates the development factors f_j at time I by

$$\hat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}} = \frac{C_{[I-j-1],j+1}}{C_{[I-j-1],j}}, \quad 0 \le j \le J-1, \tag{1.3}$$

where here and in the following a square bracket in the index denotes a summation over this index starting from 0. The ultimate claims amount $C_{i,J}$ is then predicted by

$$\hat{C}_{i,J}^{CL} = C_{i,I-i} \prod_{j=I-i}^{J-1} \hat{f}_j, \quad I-J+1 \le i \le I. \tag{1.4}$$

Therefore the CL reserve for accident year i, $I-J+1 \le i \le I$, is given by

$$\hat{R}_i = C_{i,I-i} \left(\prod_{j=I-i}^{J-1} \hat{f}_j - 1 \right) = C_{i,I-i} \prod_{j=I-i}^{J-1} \hat{f}_j \left(1 - \prod_{j=I-i}^{J-1} \hat{f}_j^{-1} \right).$$

The CL development pattern is estimated by

$$\hat{\beta}_j^{CL} = \prod_{k=j}^{J-1} \hat{f}_k^{-1}, \quad 0 \le j \le J-1, \quad \hat{\beta}_J^{CL} = 1, \tag{1.5}$$

that is, $\hat{\beta}_j^{CL}$ is an estimate of the proportion of the expected ultimate claim which emerges up to development year j. From Model Assumptions 1.2 it follows that for $I - J + 1 \leq i \leq I$

$$E[C_{i,J}|\mathcal{D}_I] = E[C_{i,J}|C_{i,0}, C_{i,1}, \ldots, C_{i,I-i}] = C_{i,I-i} \prod_{j=I-i}^{J-1} f_j.$$

Mack (1993) shows that under Model Assumptions 1.2 the estimators $\hat{f}_j$ given in (1.3) are uncorrelated and unbiased for f_j, $0 \leq j \leq J-1$. As stated in Mack (1993) the uncorrelatedness of the $\hat{f}_j$ is surprising, as $\hat{f}_{j-1}$ and $\hat{f}_j$ depend on the same data $C_{[I-j-1],j}$. In order to estimate the conditional MSEP we additionally need to estimate the variance parameters σ_j^2. The estimators

$$\hat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2, \quad 0 \leq j \leq J-1,\ j \neq I-1,$$

are unbiased for σ_j^2. If $J = I$ we need an extrapolation to obtain an estimator for σ_{J-1}^2. Mack (1993) states that the series $\hat{\sigma}_0^2, \ldots, \hat{\sigma}_{J-2}^2$ is usually exponentially decreasing and suggests to use the extrapolation

$$\hat{\sigma}_{J-1}^2 = \min\left(\hat{\sigma}_{J-2}^4/\hat{\sigma}_{J-3}^2, \hat{\sigma}_{J-3}^2, \hat{\sigma}_{J-2}^2\right) = \min\left(\hat{\sigma}_{J-2}^4/\hat{\sigma}_{J-3}^2, \hat{\sigma}_{J-3}^2\right). \tag{1.6}$$

Under Model Assumptions 1.2 Mack (1993) derived the following estimate for the conditional MSEP of accident year $i > I - J$

$$\widehat{\text{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\hat{C}_{i,J}^{CL}\right) = \left(\hat{C}_{i,J}^{CL}\right)^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left(\frac{1}{\hat{C}_{i,j}^{CL}} + \frac{1}{C_{[I-j-1],j}} \right). \tag{1.7}$$

Note that here we assume that $\mathcal{D}_I = \mathcal{I}_I$ since the CL method does not incorporate any information apart from the data $\mathcal{D}_I$. In (1.7), the process variance $\text{PV}_i = \text{Var}(C_{i,J}|\mathcal{D}_I)$ is estimated by

$$\widehat{\text{PV}}_i = \left(\hat{C}_{i,J}^{CL}\right)^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2/\hat{f}_j^2}{\hat{C}_{i,j}^{CL}},$$

and

$$\widehat{\text{PEE}}_i = \left(\hat{C}_{i,J}^{CL}\right)^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2/\hat{f}_j^2}{C_{[I-j-1],j}}$$

is an estimate of the parameter estimation error

$$\text{PEE}_i = \left(\hat{C}_{i,J}^{CL} - E\left[C_{i,J}|\,\mathcal{D}_I\right]\right)^2.$$

Similarly, for aggregated accident years Mack (1993) derived the following estimator for the conditional MSEP under Model Assumptions 1.2

$$\widehat{\text{msep}}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I}\hat{C}_{i,J}^{CL}\right)$$
$$=\sum_{i=I-J+1}^{I}\widehat{\text{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\hat{C}_{i,J}^{CL}\right)+2\sum_{I-J+1\le i<k\le I}\hat{C}_{i,J}^{CL}\hat{C}_{k,J}^{CL}\sum_{j=I-i}^{J-1}\frac{\hat{\sigma}_j^2/\hat{f}_j^2}{C_{[I-j-1],j}}.$$

In the distribution-free model given in Model Assumptions 1.2, the assumption on the conditional expectation is natural if one considers the CL formula. Due to the proportionality of the CL prediction $\hat{C}_{i,J}^{CL}$ to the last observed claim $C_{i,I-i}$ one expects a dependence between claims of the same accident year. However, the CL algorithm can also be derived under a completely different model. In the literature it is shown that the following model gives a justification of the CL algorithm:

Model Assumptions 1.3 (Over-dispersed Poisson (ODP) Model) *There exist positive parameters $\mu_0, \mu_1, \ldots, \mu_I$, $\gamma_0, \gamma_1, \ldots, \gamma_J$ with $\sum_{j=0}^{J}\gamma_j = 1$ and ϕ such that $Y_{i,j} = X_{i,j}/\phi$ are independent and Poisson distributed with*

$$E[Y_{i,j}] = \frac{\mu_i\gamma_j}{\phi}, \quad 0 \le i \le I, \quad 0 \le j \le J.$$

The ODP Model 1.3 is, for instance, studied in England-Verrall (2002). If $Y_{i,j} = X_{i,j}/\phi$ is Poisson distributed then we say that $X_{i,j}$ is ODP distributed and we have $\text{Var}(X_{i,j}) = \phi E[X_{i,j}]$. In the case where the dispersion parameter $\phi = 1$ we have a Poisson model and hence $E[X_{i,j}] = \text{Var}(X_{i,j})$. However, in practical applications a model with $\phi > 1$ is usually more appropriate. Note that due to the distributional assumption, the ODP Model 1.3 only allows for non-negative incremental claims. In the ODP Model 1.3 we need to calculate estimates $\hat{\mu}_i$ and $\hat{\gamma}_j$ for μ_i and γ_j. For $i \ge I - J + 1$ the ultimate claim is then predicted by

$$\hat{C}_{i,J} = C_{i,I-i} + \hat{\mu}_i\left(1 - \sum_{j=0}^{I-i}\hat{\gamma}_j\right).$$

If the parameters μ_i and γ_j are estimated by maximum likelihood (ML) estimation, then the predictions in the ODP Model 1.3 coincide with the CL predictors that is,

$$\hat{C}_{i,J} = C_{i,I-i} + \hat{\mu}_i\left(1 - \sum_{j=0}^{I-i}\hat{\gamma}_j\right) = \hat{C}_{i,J}^{CL},$$

where $\hat{C}_{i,J}^{CL}$ is defined in (1.4). In the Poisson case that is, if $\phi = 1$, this result goes back to Hachemeister-Stanard (1975) and it is published for instance in Kremer (1985), Mack (1991) or Renshaw-Verrall (1998). In the ODP Model 1.3 the ML equations are the same as in the Poisson case (see for instance Mack-Venter, 2000). Hence in this case the ML estimates also coincide with the CL estimates.

The drawback of the Poisson or ODP assumption is that all increments need to be non-negative. Mack-Venter (2000) mention that this problem can be overcome if one relaxes the distributional assumption and only assumes that the distribution of the incremental claims belongs to the exponential family with $\text{Var}(X_{i,j}) = \phi E[X_{i,j}] = \phi\mu_i\gamma_j$. In Mack (1991) it is shown that the unique solution of the resulting quasi-likelihood equations is given by the CL estimates.

Other models which reproduce the CL estimates are introduced in Verrall (2000) and compared to the models given in Renshaw-Verrall (1998) and to the Distribution-free Model 1.2. Mack-Venter (2000) argue that the ODP Model 1.3 differs from the Distribution-free Model 1.2 and that the latter is the only model underlying the CL method. Mack-Venter (2000) also show that in the case where some data is missing, the estimators derived from the ODP Model 1.3 differ from the CL estimates.

In the literature additional models, which are based on a lognormal assumption for the distribution of the increments, were referred to as CL models (see for instance Renshaw, 1989 and Verrall, 1990; 1991). This is criticised in Mack (1994), because these models lead to estimators different from the CL estimators.

A further model in the literature is a time series model for the CL method (see for instance Murphy, 1994 and Buchwalder et al. 2006). The time series model satisfies Model Assumptions 1.2 and is more restrictive. We do not consider this model in this thesis.

1.3.2 Bornhuetter-Ferguson (BF) Method

In this section we give a brief introduction to the BF method which goes back to Bornhuetter-Ferguson (1972). The BF method will be discussed in more detail in Chapter 2. The BF reserve is given by

$$\hat{R}_i^{BF} = \hat{\mu}_i \left(1 - \hat{\beta}_{I-i}\right),$$

where $\hat{\mu}_i$ is an a priori estimate of the expected ultimate claim $E[C_{i,J}]$ and $\hat{\beta}_{I-i}$ is the estimated development pattern. This means that $1 - \hat{\beta}_{I-i}$ is the estimated still to come percentage of accident year i at time I. The a priori estimates incorporate the information from the premium that is,

$\hat{\mu}_i = \nu_i \hat{q}_i$, where ν_i denotes the earned premium for accident year i and $\hat{q}_i$ is an a priori estimate for the expected loss ratio of accident year i. The $\hat{\mu}_i$ are assumed to be specified by experts and they should not depend on the data. The incorporation of the additional information from the premium is a fundamental difference compared to the CL method, which is purely based on the observations in the triangle. The BF predictor of the ultimate claim

$$\hat{C}_{i,J}^{BF} = C_{i,I-i} + \hat{\mu}_i \left(1 - \hat{\beta}_{I-i}\right)$$

establishes an additive relation between the last observation $C_{i,I-i}$ and the predictor of the ultimate claim. This additive connection reflects an independence assumption of the past claims and the outstanding claims, which underlies the BF method (see Mack, 2006). In contrast, the CL predictor is proportional to the current claims amount $C_{i,I-i}$, that is, the CL method assumes a multiplicative relation. The estimation of the development pattern and prediction uncertainty will be discussed in detail in Chapter 2 and in Paper A. The CL method and the BF method are in some sense two extreme cases: the CL method does not incorporate any prior information, whereas the BF method assumes perfect prior information and does not incorporate the data in the calculation of the estimate $\hat{\mu}_i$. In order to highlight this difference we rewrite the CL predictor (1.4) as follows

$$\hat{C}_{i,J}^{CL} = C_{i,I-i} + \hat{C}_{i,J}^{CL} \left(1 - \hat{\beta}_{I-i}^{CL}\right),$$

where the CL development pattern $\hat{\beta}_{I-i}^{CL}$ is given in (1.5).

1.3.3 Cape Cod Method

From (1.4) we see that in the CL method the predictor of the ultimate claim of an accident year is directly proportional to the last observation on the diagonal. If this observation is zero or an outlier the CL method gives unsatisfactory results. Moreover, the CL method is very sensitive to changes of individual numbers. A change of $C_{0,J}$ has a multiplicative effect on the reserves of all accident years. Finally, as mentioned in Section 1.3.1, the CL method is purely based on the observations in the claims triangle and it disregards the information in the earned premium. The Cape Cod method is a more robust method and it incorporates the earned premium. The Cape Cod method goes back to Bühlmann-Straub (1983), who designed the method in order to overcome the above mentioned deficiencies of the CL method. A derivation of the method is given in Straub (1988). We denote the earned premium for accident year i by ν_i, $0 \leq i \leq I$. More precisely, we assume that ν_i is an on-level premium, that is, the earned premium adjusted to business

cycles such that the expected loss ratio is the same for all accident years. As in the previous methods, the Cape Cod method assumes that the development pattern is the same for all accident years. In the following let $\iota(i) = (I-i) \wedge J$ denote the last observed development year for accident year i. The Cape Cod predictor of the ultimate claim for accident year $i \geq I - J + 1$ is given by

$$\hat{C}_{i,J}^{CC} = C_{i,\iota(i)} + \nu_i \frac{\sum_{l=0}^{I} C_{l,\iota(l)}}{\sum_{l=0}^{I} \nu_l \hat{\beta}_{\iota(l)}} \left(1 - \hat{\beta}_{\iota(i)}\right),$$

where $\hat{\beta}_{\iota(i)}$ is an estimate for the development pattern, and we define

$$\hat{q} = \frac{\sum_{l=0}^{I} C_{l,\iota(l)}}{\sum_{l=0}^{I} \nu_l \hat{\beta}_{\iota(l)}}, \tag{1.8}$$

the estimated loss ratio. If the development pattern is estimated by the CL development pattern (1.5) we have

$$\hat{C}_{i,J}^{CC} = C_{i,\iota(i)} + \nu_i \hat{\beta}_{\iota(i)}^{CL} \frac{\sum_{l=0}^{I} C_{l,\iota(l)}}{\sum_{l=0}^{I} \nu_l \hat{\beta}_{\iota(l)}^{CL}} \left(\prod_{j=\iota(i)}^{J-1} \hat{f}_j - 1 \right).$$

Therefore the Cape Cod predictor is in this case obtained from the CL predictor by replacing the diagonal element $C_{i,\iota(i)}$ by the more robust value

$$\nu_i \hat{\beta}_{\iota(i)}^{CL} \hat{q} = \nu_i \hat{\beta}_{\iota(i)}^{CL} \frac{\sum_{l=0}^{I} C_{l,\iota(l)}}{\sum_{l=0}^{I} \nu_l \hat{\beta}_{\iota(l)}^{CL}},$$

see (1.4). Due to its simplicity and its advantages over the CL method, the Cape Cod method is another well-established method in practice. However, in contrast to the CL and BF method, in current literature there seems to be no analytical formula for the conditional MSEP of the Cape Cod predictor. Paper D aims to fill that gap. We will consider stochastic models for the Cape Cod method and derive an estimate for the prediction uncertainty in Chapter 5.

1.3.4 Benktander-Hovinen Method

In practice different reserving methods are often combined. Since the CL method completely relies on the last observation of the diagonal, it is common to use the BF method for more recent accident years, where there is only few information in the data. On the other hand, the BF method relies mainly on the a priori estimate, and hence for older accident years, where a lot

of data is available, the CL method is often preferred. Benktander (1976) and Hovinen (1981) proposed, independently from each other, a method that allows to combine the CL and BF predictors. Let $\hat{\mu}_i$ denote a prior estimate of the expected ultimate claim and let β_j, $0 \leq j \leq J$, denote the claims development pattern. The Benktander-Hovinen predictor of the ultimate claim is given by

$$\hat{C}_{i,J}^{BH} = C_{i,I-i} + (1 - \beta_{I-i}) \left(\beta_{I-i} \hat{C}_{i,J}^{CL} + (1 - \beta_{I-i}) \hat{\mu}_i \right), \quad I - J + 1 \leq i \leq I.$$

Contrary to the BF method, the Benktander-Hovinen method uses a weighted average between the a priori estimate and the CL predictor. The weights are chosen such that with increasing development of the data, more weight is given to the data based CL predictor. The Benktander-Hovinen method can be interpreted as a BF method in which the a priori estimate is adjusted in order to take information from the claims development into account. We call these adjustments repricing and we introduce a BF method with repricing in Paper B, where the repricing is done by using credibility theory. In Chapter 3 we give an overview of the method.

1.4 Summary

So far, we gave a brief overview of basic methods and problems in claims reserving. For a more comprehensive review we refer to Taylor (2000) and Wüthrich-Merz (2008). In the following sections we summarise the contributions of the author to the above mentioned problems.

- **Paper A** suggests different stochastic models for the BF method and derives estimates for the development pattern and for prediction uncertainty.
- **Paper B** introduces the BF method with repricing that allows to adjust the a priori estimate in the BF method over time.
- **Paper C** studies aspects of the claims development result and introduces models that allow to incorporate changes in expert opinion.
- **Paper D** discusses stochastic models for the Cape Cod method and derives parameter estimates and estimators for the prediction uncertainty.

The complete papers are given in the Appendix.

2. Bornhuetter-Ferguson (BF) Method

The BF method was briefly outlined in Section 1.3.2. Recall the BF reserve

$$\hat{R}_i^{BF} = \hat{\mu}_i \left(1 - \hat{\beta}_{I-i}\right),$$

where $\hat{\mu}_i$ is an a priori estimate of the expected ultimate claim and $1 - \hat{\beta}_{I-i}$ is the estimated still to come percentage. So far, we did not elaborate on how to estimate the development pattern. In the past, the development pattern was often simply estimated by the CL development pattern (1.5). However, due to different assumptions underlying the CL and the BF method, the use of the CL development pattern for the BF method is criticised and different estimates are proposed for instance in Mack (2006).

Stochastic models for the BF method were introduced much later than for the CL method. Mack (2008) introduced a distribution-free model for the BF method and derived a formula for the prediction error of the BF reserve. Alai et al. (2009; 2010) study the BF method within an ODP model and derive an estimate for the conditional MSEP. Alai et al. (2009; 2010) assume that the development pattern is estimated by the CL development pattern. In Paper A we consider different stochastic models for the BF method with parametric assumptions on the distribution of the incremental claims. Within these models we derive estimators of the development pattern based on ML considerations. We also derive estimators for the conditional MSEP within these models.

In this chapter we give an overview of the findings in Mack (2008), Alai et al. (2010) and Paper A.

2.1 Distribution-free Model

From the BF reserve formula Mack (2008) derives a distribution-free stochastic model for the BF method. The model is probably the most general stochastic model underlying the BF method. In contrast to the CL method, the independence between past and future claims amounts was fundamental for the

origin of the BF method (see Mack, 2006). Therefore Mack (2008) introduces an additional independence assumption compared to the Distribution-free CL Model 1.2. Moreover, the BF reserve formula suggests a cross-classified structure of the type

$$E[C_{i,j}] = \mu_i \beta_j \quad \text{or equivalently} \quad E[X_{i,j}] = \mu_i \gamma_j,$$

with $\beta_j = \sum_{k=0}^{j} \gamma_k$. Since $\mu_i \gamma_j = (\mu_i c)(\gamma_j / c)$ for any constant $c > 0$, the parameters are only unique up to a constant factor. Thus, without loss of generality, one can assume that $\gamma_0 + \ldots + \gamma_J = 1$. Additionally, Mack (2008) incorporates the tail development with an additional parameter γ_{J+1} and requires that $\sum_{j=0}^{J+1} \gamma_j = 1$. For the tail no data is available and expert judgement is required. In this thesis we assume that all claims are settled after development year J, and hence we omit considerations for the tail. Note that

$$E[C_{i,J}] = \mu_i,$$

and hence μ_i can be considered as a measure of volume. Thus, Mack suggests the variance assumption

$$\mathrm{Var}(X_{i,j}) = \mu_i s_j^2.$$

The ODP variance assumption $\mathrm{Var}(X_{i,j}) = \phi \mu_i \gamma_j$ is criticised in Taylor (2002) as it seems to contradict to what one observes in practice. Additionally, this variance assumption implies that all γ_j are positive, which is often not the case for incurred claims. The model assumptions of Mack (2008) are summarised as follows:

Model Assumptions 2.1 (Distribution-free BF Model)

- *All incremental claims $X_{i,j}$ are independent.*
- *There are parameters μ_i, γ_j, $0 \leq i \leq I$ and $0 \leq j \leq J$, with $E[X_{i,j}] = \mu_i \gamma_j$ and $\gamma_0 + \ldots + \gamma_J = 1$.*
- *There are proportionality constants s_j^2, $0 \leq j \leq J$, with $\mathrm{Var}(X_{i,j}) = \mu_i s_j^2$.*
- *There are given unbiased a priori estimates $\hat{\mu}_i$ for μ_i, $0 \leq i \leq I$.*

From Model Assumptions 2.1 it follows for $i \geq I - J + 1$

$$E[R_i | \mathcal{D}_I] = \mu_i (1 - \beta_{I-i}), \quad \text{where} \quad \beta_{I-i} = \sum_{j=0}^{I-i} \gamma_j,$$

and thus the assumptions are consistent with the BF reserve formula. Mack (2006; 2008) criticises the use of the CL development pattern in the BF method due to the fundamental assumption of independence between past and future claims, which underlies the BF method. Due to the proportionality of the CL reserve to the current claims amount, the CL reserve for accident year i is the smaller, the smaller the current claims amount $C_{i,I-i}$ is. If the reserve was estimated by $\tilde{R}_i = \hat{\mu}_i - C_{i,I-i}$, then we would have an opposite effect. However, Mack (2008) points out that the BF method takes a neutral position: the BF reserve for accident year i does not directly depend on $C_{i,I-i}$. Further, Mack (2008) states that 'the systematic use of the CL link ratios assumes that the outstanding claims part is a direct multiple of the already known claims part at each point of the development'. Therefore, the development pattern should be estimated differently from the CL development pattern (1.5). As a starting point for the estimation of the development pattern note that

$$\tilde{\gamma}_j = \frac{\sum_{i=0}^{I-j} X_{i,j}}{\sum_{i=0}^{I-j} \mu_i} = \frac{X_{[I-j],j}}{\mu_{[I-j]}}$$

is the best linear unbiased estimator for γ_j. As the μ_i are unknown, Mack (2008) suggests to start with

$$\hat{\tilde{\gamma}}_j = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}},$$

and to apply some manual smoothing to select a final pattern $\hat{\gamma}_0^*, \ldots, \hat{\gamma}_J^*$ with $\sum_{j=0}^{J} \hat{\gamma}_j^* = 1$ and $\hat{\beta}_j^* = \sum_{k=0}^{j} \hat{\gamma}_k^*$, $0 \le j \le J$. For the estimation of the variance parameters observe that

$$\tilde{s}_j^2 = \frac{1}{I-j} \sum_{i=0}^{I-j} \frac{(X_{i,j} - \mu_i \tilde{\gamma}_j)^2}{\mu_i}$$

is an unbiased estimate of s_j^2, $0 \le j \le J$, $j \ne I$. Thus, one can start with

$$\hat{\tilde{s}}_j^2 = \frac{1}{I-j} \sum_{i=0}^{I-j} \frac{(X_{i,j} - \hat{\mu}_i \hat{\gamma}_j^*)^2}{\hat{\mu}_i}, \quad 0 \le j \le J,\ j \ne I,$$

and apply some manual smoothing and, if $J = I$ an extrapolation, to obtain the final estimates $\hat{s}_0^{2*}, \ldots, \hat{s}_J^{2*}$. For the case of rather stable data Mack (2008) also gives a more formal way for the estimation of the parameters γ_j and s_j^2 with smoothing regressions.

The BF reserve estimate for accident year $i > I - J$ is then given by

$$\hat{R}_i = \hat{\mu}_i \left(1 - \hat{\beta}_{I-i}^*\right).$$

In order to estimate the conditional MSEP of $\hat{R}_i$ under Model Assumptions 2.1, Mack (2008) additionally assumes that

- the estimates $\hat{\gamma}_0^*, \ldots, \hat{\gamma}_J^*$ are independent from the a priori estimates $\hat{\mu}_0, \ldots, \hat{\mu}_I$,
- the estimates $\hat{\gamma}_j^*$, $\hat{s}_j^{2*}$ are unbiased for γ_j and s_j^2, respectively.

Mack (2008) considers the conditional MSEP, given the data of accident year i only, that is he defines

$$\operatorname{msep}_{R_i|X_{i,0},\ldots,X_{i,I-i}}\left(\hat{R}_i\right) = E\left[\left(\hat{R}_i - R_i\right)^2 \middle| X_{i,0}, \ldots, X_{i,I-i}\right].$$

Further, Mack (2008) neglects dependencies between $\hat{R}_i$ and $X_{i,0}, \ldots, X_{i,I-i}$ and derives an estimate for the unconditional MSEP

$$\operatorname{msep}_{R_i}\left(\hat{R}_i\right) = E\left[\left(\hat{R}_i - R_i\right)^2\right] = \operatorname{Var}(R_i) + \operatorname{Var}\left(\hat{R}_i\right).$$

The process variance $\mathrm{PV}_i = \operatorname{Var}(R_i)$ is estimated by inserting the parameter estimates which yields

$$\widehat{\mathrm{PV}}_i = \widehat{\operatorname{Var}}(R_i) = \hat{\mu}_i \sum_{j=I-i+1}^{J} \hat{s}_j^{2*}.$$

In order to estimate the variance of $\hat{R}_i$, estimators $\widehat{\operatorname{Var}}\left(\hat{\mu}_i\right)$ for $\operatorname{Var}\left(\hat{\mu}_i\right)$ and $\widehat{\operatorname{Var}}(\hat{\beta}_j^*)$ for $\operatorname{Var}(\hat{\beta}_j^*)$ are derived in Mack (2008). The variance $\operatorname{Var}(\hat{R}_i)$ is then estimated by

$$\widehat{\operatorname{Var}}\left(\hat{R}_i\right) = \left(\hat{\mu}_i^2 + \widehat{\operatorname{Var}}\left(\hat{\mu}_i\right)\right)\widehat{\operatorname{Var}}\left(\hat{\beta}_j^*\right) + \widehat{\operatorname{Var}}\left(\hat{\mu}_i\right)\left(1 - \hat{\beta}_{I-i}^*\right)^2,$$

which is an estimate for the parameter estimation error.

2.2 BF Method with CL Pattern

Despite of criticism, the CL development pattern (1.5) is often used in the BF method. For this reason Alai et al. (2010) derive an estimate for the conditional MSEP of the reserve

$$\hat{R}_i = \hat{\mu}_i\left(1 - \hat{\beta}_{I-i}^{CL}\right), \tag{2.1}$$

where $\hat{\mu}_i$ is an a priori estimate of the expected ultimate claim $\mu_i = E[C_{i,J}]$ and $\hat{\beta}_{I-i}^{CL}$ denotes the CL development pattern given in (1.5). In contrast to the Distribution-free Model 2.1, Alai et al. (2010) assume an ODP distribution for the incremental claims $X_{i,j}$:

Model Assumptions 2.2 (BF ODP Model)

- *Incremental claims $X_{i,j}$ are independent and ODP distributed and there exist positive parameters ϕ, $\mu_0, \mu_1, \ldots, \mu_I$ and $\gamma_0, \gamma_1, \ldots, \gamma_J$ with*

$$E[X_{i,j}] = \mu_i \gamma_j, \quad \operatorname{Var}(X_{i,j}) = \phi \mu_i \gamma_j, \quad 0 \le i \le I,\ 0 \le j \le J,$$

  and $\sum_{j=0}^{J} \gamma_j = 1$.

- *$\hat{\mu}_i$ are independent random variables that are unbiased estimators of the expected ultimate claim $\mu_i = E[C_{i,J}]$, $0 \le i \le I$.*

- *$X_{i,j}$ and $\hat{\mu}_k$ are independent for all i, j, k.*

Recall that under Model Assumptions 2.2 the ML estimates for μ_i and γ_j coincide with the CL estimates (see Section 1.3.1). It is important to note that the CL development pattern (1.5) is only obtained if the μ_i are also estimated by ML. In the BF method, we are given external a priori estimates $\hat{\mu}_i$ and therefore the use of the ML estimates for μ_i in the estimation of the development pattern is in some sense inconsistent. However, the use of the CL development pattern is close to practitioners implementations of the BF method.

Alai et al. (2010) derive an estimate for the conditional MSEP, given $\mathcal{D}_I$, that is,

$$\begin{aligned}
&\operatorname{msep}_{R_i|\mathcal{D}_I}\left(\hat{R}_i\right) = E\left[\left(\hat{R}_i - R_i\right)^2 \middle| \mathcal{D}_I\right] \\
&= \underbrace{\sum_{j=I-i+1}^{J} \operatorname{Var}(X_{i,j})}_{=\mathrm{PV}_i} + \underbrace{\left(1 - \hat{\beta}_{I-i}^{CL}\right)^2 \operatorname{Var}\left(\hat{\mu}_i\right) + \mu_i^2 \left(\sum_{j=I-i+1}^{J} \left(\hat{\gamma}_j^{CL} - \gamma_j\right)\right)^2}_{=\mathrm{PEE}_i}.
\end{aligned}$$

The process variance can be estimated by inserting the parameter estimates which yields

$$\widehat{\mathrm{PV}}_i = \sum_{j=I-i+1}^{J} \hat{\phi} \hat{\mu}_i \hat{\gamma}_j^{CL} = \hat{\phi} \hat{\mu}_i \left(1 - \hat{\beta}_{I-i}^{CL}\right),$$

where $\hat{\gamma}_j^{CL}$ denotes the CL development pattern and $\hat{\phi}$ is estimated with the help of Pearson residuals:

$$\hat{\phi} = \frac{1}{|\mathcal{D}_I| - (I + J + 1)} \sum_{\substack{0 \le i+j \le I, \\ j \le J}} \frac{\left(X_{i,j} - \hat{C}_{i,J}^{CL} \hat{\gamma}_j^{CL}\right)^2}{\hat{C}_{i,J}^{CL} \hat{\gamma}_j^{CL}}.$$

The uncertainty $\mathrm{Var}(\hat{\mu}_i)$ in the prior estimate $\hat{\mu}_i$ is estimated by using an external estimate for the coefficient of variation $\mathrm{CoVa}(\hat{\mu}_i)$. Such estimates are for example studied in the SST, see FINMA (2006), and it is suggested that 5% to 10% is a reasonable range for $\widehat{\mathrm{CoVa}}(\hat{\mu}_i)$. The variance $\mathrm{Var}(\hat{\mu}_i)$ is then estimated by

$$\widehat{\mathrm{Var}}(\hat{\mu}_i) = \hat{\mu}_i^2 \widehat{\mathrm{CoVa}}(\hat{\mu}_i)^2 .$$

The term

$$\left(\sum_{j=I-i+1}^{J} \left(\hat{\gamma}_j^{CL} - \gamma_j \right) \right)^2$$

is estimated by its unconditional expectation and neglecting a possible bias term the following approximation is used:

$$\begin{aligned} E\left[\left(\sum_{j=I-i+1}^{J} \left(\hat{\gamma}_j^{CL} - \gamma_j \right) \right)^2 \right] &= \sum_{j,l=I-i+1}^{J} E\left[\left(\hat{\gamma}_j^{CL} - \gamma_j \right) \left(\hat{\gamma}_l^{CL} - \gamma_l \right) \right] \\ &\approx \sum_{j,l=I-i+1}^{J} \mathrm{Cov}\left(\hat{\gamma}_j^{CL}, \hat{\gamma}_l^{CL} \right) . \end{aligned}$$

Due to the asymptotic properties of ML estimates the covariance matrix of the $\hat{\gamma}_j^{CL}$ is estimated by the inverse Fisher information matrix. The Fisher information matrix $H(\boldsymbol{\zeta}, \phi)$ is given by

$$H(\boldsymbol{\zeta}, \phi)_{r,s} = -E_{\zeta}\left[\frac{\partial^2}{\partial \zeta_r \partial \zeta_s} l_{\mathcal{D}_I}(\boldsymbol{\zeta}, \phi) \right], \quad r, s = 1, \ldots, I + J + 1,$$

where $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{I+J+1}) = (\mu_0, \ldots, \mu_I, \gamma_0, \ldots, \gamma_{J-1})$ and $l_{\mathcal{D}_I}$ is the log-likelihood function. An estimate for the Fisher information matrix is obtained by replacing the parameters μ_i, γ_j and ϕ by the estimates $\hat{C}_{i,J}^{CL}$, $\hat{\gamma}_j^{CL}$ and $\hat{\phi}$. The covariances of the $\hat{\gamma}_j^{CL}$ are then estimated by

$$\begin{aligned} \widehat{\mathrm{Cov}}\left(\hat{\gamma}_j^{CL}, \hat{\gamma}_l^{CL} \right) &= H(\hat{\boldsymbol{\zeta}}, \hat{\phi})^{-1}_{I+2+j, I+2+l}, \quad 0 \le j, l \le J-1, \\ \widehat{\mathrm{Cov}}\left(\hat{\gamma}_j^{CL}, \hat{\gamma}_J^{CL} \right) &= -\sum_{l=0}^{J-1} H(\hat{\boldsymbol{\zeta}}, \hat{\phi})^{-1}_{I+2+j, I+2+l}, \quad 0 \le j \le J-1, \\ \widehat{\mathrm{Var}}\left(\hat{\gamma}_J^{CL} \right) &= \sum_{0 \le j,l \le J-1} H(\hat{\boldsymbol{\zeta}}, \hat{\phi})^{-1}_{I+2+j, I+2+l}, \quad 0 \le j, l \le J-1. \end{aligned}$$

Under Model Assumptions 2.2 Alai et al. (2010) arrive at the following estimator for the conditional MSEP of $\hat{R}_i$ given in (2.1)

$$\begin{aligned}&\widehat{\operatorname{msep}}_{R_i|\mathcal{D}_I}\left(\hat{R}_i\right)\\&=\hat{\phi}\hat{\mu}_i\left(1-\hat{\beta}_{I-i}^{CL}\right)+\left(1-\hat{\beta}_{I-i}^{CL}\right)^2\widehat{\operatorname{Var}}\left(\hat{\mu}_i\right)+\hat{\mu}_i^2\sum_{j,l=I-i+1}^{J}\widehat{\operatorname{Cov}}\left(\hat{\gamma}_j^{CL},\hat{\gamma}_l^{CL}\right),\end{aligned}$$

and for aggregated accident years the corresponding conditional MSEP is estimated by

$$\begin{aligned}&\widehat{\operatorname{msep}}_{\sum_{i=I-J+1}^{I}R_i|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I}\hat{R}_i\right)\\&=\sum_{i=I-J+1}^{I}\widehat{\operatorname{msep}}_{R_i|\mathcal{D}_I}\left(\hat{R}_i\right)+2\sum_{i<k}\hat{\mu}_i\hat{\mu}_k\sum_{\substack{j>I-i,\\l>I-k}}\widehat{\operatorname{Cov}}\left(\hat{\gamma}_j^{CL},\hat{\gamma}_l^{CL}\right).\end{aligned}$$

The formulas in Alai et al. (2010) are quite simple to implement in a spreadsheet. The disadvantage is that the method can only be applied for non-negative incremental claims.

2.3 Distributional Models

In Paper A we consider three distributional models for the BF method, which satisfy Model Assumptions 2.1. In these models we estimate the development pattern differently from the CL development pattern. The motivation for the estimation of the development pattern comes from another essential difference between the CL and the BF philosophy. Whereas the CL method is purely based on the claims data in the development triangle, the BF method additionally incorporates the information from the a priori estimates $\hat{\mu}_i$. The CL development pattern disregards this additional information and is therefore not consistent with the BF philosophy. As a starting point for the estimation of the development pattern assume that we are given a full claims rectangle. Then, an obvious estimator is

$$\tilde{\gamma}_j=\frac{X_{[I],j}}{C_{[I],J}},$$

and $\sum_{j=0}^{J}\tilde{\gamma}_j=1$. Since the lower right triangle $\mathcal{D}_I^c$ is unknown, a natural idea is to replace the unknown $X_{i,j}$ by the BF predictors $\hat{\mu}_i\hat{\gamma}_j$ resulting in the

following system of equations

$$\hat{\gamma}_j = \frac{X_{[I-j],j} + \sum_{i=I-j+1}^{I} \hat{\mu}_i \hat{\gamma}_j}{C_{[I-J],J} + \sum_{i=I-J+1}^{I} \left(C_{i,I-i} + \sum_{l=I-i+1}^{J} \hat{\mu}_i \hat{\gamma}_l \right)}, \quad 0 \leq j \leq J. \tag{2.2}$$

In the three models presented in Paper A we derive ML estimates of the development pattern. Moreover, we find explicit formulas for the estimation of the correlation matrices of these estimators. These correlations are needed for the estimation of the conditional MSEP of the BF reserves. We start with an ODP model, which is similar to Model Assumptions 2.2, but the a priori estimates are allowed to be correlated. In practice it is likely that the a priori estimates of consecutive accident years are strongly correlated. Under the resulting ODP model the ML estimates for the γ_j's coincide with the estimators (2.2) if we replace the μ_i's by the a priori estimates $\hat{\mu}_i$. A proof of this result is given in Paper A. For the ML estimation we first assume that the μ_i's are given and in a second step we replace the unknown μ_i's by the a priori estimates $\hat{\mu}_i$. If the μ_i's are estimated by the ML method, too, then the resulting pattern is the CL development pattern as we have seen in Section 1.3.1 and Section 2.2.

Due to the restrictive assumption $\mathrm{Var}(X_{i,j}) = \phi \mu_i \gamma_j$, which is often not appropriate in practical applications, we consider a second model in Paper A, which allows for a varying dispersion parameter ϕ_j in the development years. In this more general ODP model we still have the positivity constraint for incremental claims amounts due to the distributional assumptions. The following model does not have this restriction:

Model Assumptions 2.3 (Normal Model)

- *Incremental claims $X_{i,j}$ are independent and normally distributed and there exist parameters $\gamma_0, \ldots, \gamma_J$ and positive parameters $\mu_0, \ldots, \mu_I$ and $\sigma_0^2, \ldots, \sigma_J^2$ such that*

$$E[X_{i,j}] = \mu_i \gamma_j, \quad \mathrm{Var}(X_{i,j}) = \mu_i \sigma_j^2, \quad 0 \leq i \leq I,\ 0 \leq j \leq J,$$

  and $\sum_{j=0}^{J} \gamma_j = 1$.

- *The a priori estimates $\hat{\mu}_i$ for $\mu_i = E[C_{i,J}]$ are unbiased and independent from $X_{l,j}$ for $0 \leq l \leq I$, $0 \leq j \leq J$.*

The ideas for the estimation of the conditional MSEP are the same in the three models. In the following we focus on the Normal Model 2.3. In order to calculate the estimates of the development pattern we assume first

that the μ_i's are known and calculate the ML estimates under the constraint $\sum_{j=0}^{J} \gamma_j = 1$. With the side constraint we obtain the following ML estimates

$$\tilde{\gamma}_j = \frac{X_{[I-j],j}}{\mu_{[I-j]}} + \frac{\sigma_j^2/\mu_{[I-j]}}{\sum_{l=0}^{J} \sigma_l^2/\mu_{[I-l]}} \left(1 - \sum_{l=0}^{J} \frac{X_{[I-l],l}}{\mu_{[I-l]}}\right), \quad 0 \leq j \leq J. \tag{2.3}$$

Note that the first term in $\tilde{\gamma}_j$ is the best linear unbiased estimator for γ_j and the second term is an additive normalisation correction. The μ_i and σ_j^2 in (2.3) are unknown and have to be replaced by estimators. For μ_i we insert the a priori estimates $\hat{\mu}_i$. If the σ_j^2 are estimated by ML, the resulting system of equations is only iteratively solvable. Therefore we use the estimators

$$\hat{\sigma}_j^2 = \frac{1}{I-j} \sum_{i=0}^{I-j} \hat{\mu}_i \left(\frac{X_{i,j}}{\hat{\mu}_i} - \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}}\right)^2, \quad 0 \leq j \leq J,\ j \neq I,$$

which are unbiased estimates for the σ_j^2's if $\hat{\mu}_i = \mu_i$, $0 \leq i \leq I$. If $J = I$ we use an extrapolation to obtain an estimate for σ_J^2, analogously to (1.6). In Paper A it is shown that the final estimate

$$\hat{\gamma}_j = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}} + \frac{\hat{\sigma}_j^2/\hat{\mu}_{[I-j]}}{\sum_{l=0}^{J} \hat{\sigma}_l^2/\hat{\mu}_{[I-l]}} \left(1 - \sum_{l=0}^{J} \frac{X_{[I-l],l}}{\hat{\mu}_{[I-l]}}\right), \tag{2.4}$$

coincides with the estimate derived in the more general ODP model. The covariance matrix of the estimated development pattern is approximated with the inverse Fisher information matrix. The log-likelihood function is given by

$$\begin{aligned} &l_{\mathcal{D}_I}(\gamma_0, \ldots, \gamma_{J-1}) \\ &= -\sum_{\substack{i+j \leq I,\\ j \leq J-1}} \frac{(X_{i,j} - \mu_i \gamma_j)^2}{2\mu_i \sigma_j^2} - \sum_{i=0}^{I-J} \frac{\left(X_{i,J} - \mu_i \left(1 - \sum_{j=0}^{J-1} \gamma_j\right)\right)^2}{2\mu_i \sigma_J^2} + r, \end{aligned}$$

where r contains all remaining terms which do not depend on $(\gamma_0, \ldots, \gamma_{J-1})$, and for the entries of the Fisher information matrix we have

$$H(\boldsymbol{\gamma})_{j,k} = E\left[\frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_j} \frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_k}\right], \quad 0 \leq j, k \leq J-1.$$

The inverse of the Fisher information matrix can explicitly be calculated and the resulting estimates for the covariances of the $\hat{\gamma}_j$ are given by

$$\widehat{\mathrm{Cov}}\left(\hat{\gamma}_j, \hat{\gamma}_k\right) = \frac{\hat{\sigma}_j^2}{\hat{\mu}_{[I-j]}} \left(\mathbb{1}_{j=k} - \frac{\hat{\sigma}_k^2/\hat{\mu}_{[I-k]}}{\sum_{l=0}^{J} \hat{\sigma}_l^2/\hat{\mu}_{[I-l]}}\right), \quad 0 \leq j, k \leq J. \tag{2.5}$$

Due to the side constraint $\sum_{j=0}^{J} \hat{\gamma}_j = 1$, the $\hat{\gamma}_j$'s are negatively correlated.

In Paper A we derive an estimate for the conditional MSEP of the BF predictor

$$\hat{C}_{i,J} = C_{i,I-i} + \hat{\mu}_i(1 - \hat{\beta}_{I-i}),$$

under the three different models. Let $\hat{\boldsymbol{\mu}} = (\hat{\mu}_0, \ldots, \hat{\mu}_I)$ and let $\mathcal{I}_I = \sigma(\mathcal{D}_I, \hat{\boldsymbol{\mu}})$ denote the σ-field containing the information at time I. The conditional MSEP of $\hat{C}_{i,J}$ given $\mathcal{I}_I$ is given by

$$\operatorname{msep}_{C_{i,J}|\mathcal{I}_I}\left(\hat{C}_{i,J}\right) = \underbrace{\sum_{j=I-i+1}^{J} \operatorname{Var}(X_{i,j})}_{=\mathrm{PV}_i} + \underbrace{\left(\hat{\mu}_i\left(1 - \hat{\beta}_{I-i}\right) - \mu_i(1 - \beta_{I-i})\right)^2}_{=\mathrm{PEE}_i}.$$

The parameter estimation error PEE_i can be rewritten as

$$\begin{aligned}\mathrm{PEE}_i = &\left(1 - \hat{\beta}_{I-i}\right)^2 (\hat{\mu}_i - \mu_i)^2 + \mu_i^2 \left(\beta_{I-i} - \hat{\beta}_{I-i}\right)^2 \\ &+ 2\left(1 - \hat{\beta}_{I-i}\right) \mu_i \left(\hat{\mu}_i - \mu_i\right)\left(\beta_{I-i} - \hat{\beta}_{I-i}\right).\end{aligned}$$

In order to estimate the unknown terms

$$(\hat{\mu}_i - \mu_i)^2, \quad \left(\beta_{I-i} - \hat{\beta}_{I-i}\right)^2 \quad \text{and} \quad (\hat{\mu}_i - \mu_i)\left(\beta_{I-i} - \hat{\beta}_{I-i}\right),$$

we study the fluctuations of $\hat{\mu}_i$ and $\hat{\beta}_{I-i}$ around the true values μ_i and β_{I-i}, respectively, that is, we take the mean over all possible $\mathcal{I}_I$. If we neglect the dependencies between the $\hat{\mu}_i$'s and the $\hat{\beta}_j$'s we obtain with the unbiasedness of the $\hat{\mu}_i$'s and the (approximate) asymptotic unbiasedness of the $\hat{\beta}_j$'s the following estimator for the parameter estimation error

$$\widehat{\mathrm{PEE}}_i = (1 - \hat{\beta}_{I-i})^2 \widehat{\operatorname{Var}}(\hat{\mu}_i) + \hat{\mu}_i^2 \widehat{\operatorname{Var}}(\hat{\beta}_{I-i}),$$

where $\widehat{\operatorname{Var}}(\hat{\mu}_i)$ and $\widehat{\operatorname{Var}}(\hat{\beta}_{I-i})$ are estimates for the variances $\operatorname{Var}(\hat{\mu}_i)$ and $\operatorname{Var}(\hat{\beta}_{I-i})$, respectively. The covariance matrix of the $\hat{\beta}_{I-i}$'s is estimated using the Fisher information matrix

$$\widehat{\operatorname{Var}}\left(\hat{\beta}_{I-i}\right) = \widehat{\operatorname{Cov}}\left(\hat{\beta}_{I-i}, \hat{\beta}_{I-i}\right), \quad \widehat{\operatorname{Cov}}\left(\hat{\beta}_{I-i}, \hat{\beta}_{I-k}\right) = \sum_{j=0}^{I-i} \sum_{l=0}^{I-k} \widehat{\operatorname{Cov}}(\hat{\gamma}_j, \hat{\gamma}_l).$$

Under the Normal Model 2.3 the covariances $\widehat{\operatorname{Cov}}(\hat{\gamma}_j, \hat{\gamma}_k)$ are given in (2.5).

It remains to estimate the variances of the a priori estimates $\hat{\mu}_i$. An actuary being able to deliver a priori estimates $\hat{\mu}_i$ is often also able to make a statement about the precision of these estimates. In this case beside the

$\hat{\mu}_i$'s there are also given a priori estimates of $\mathrm{Var}(\hat{\mu}_i)$. If this is not the case we show in Paper A how $\mathrm{Var}(\hat{\mu}_i)$ can be estimated from the data under the assumption that the $\hat{\mu}_i$ have a constant coefficient of variation $\mathrm{CoVa}(\hat{\mu}_i) \equiv c$ and that the correlations $\mathrm{Corr}(\hat{\mu}_i, \hat{\mu}_k)$ can be estimated by

$$\widehat{\mathrm{Corr}}(\hat{\mu}_i, \hat{\mu}_k) = \begin{cases} \frac{n-|i-k|}{n}, & 0 \leq i,k \leq I \text{ and } |i-k| < n, \\ 0, & \text{else}, \end{cases} \tag{2.6}$$

where $0 \leq n \leq I$. Formula (2.6) is obtained if the a priori loss ratios are estimated with a moving average of the ultimate loss ratios over n years under the assumption that the premium ν_i and the variances of the ultimate loss ratios are constant.

Thus, we have derived the following estimate for the conditional MSEP of the BF predictor $\hat{C}_{i,J}$ under the Normal Model 2.3

$$\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) = \hat{\mu}_i \sum_{j=I-i+1}^{J} \hat{\sigma}_j^2 + \widehat{\mathrm{Var}}(\hat{\mu}_i)(1-\hat{\beta}_{I-i})^2 + \hat{\mu}_i^2 \widehat{\mathrm{Var}}(\hat{\beta}_{I-i}).$$

For aggregated accident years we obtain in the Normal Model 2.3

$$\begin{aligned} \widehat{\mathrm{msep}}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{I}_I}\left(\sum_{i=I-J+1}^{I} \widehat{C}_{i,J}\right) &= \sum_{i=I-J+1}^{I} \widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{I}_I}\left(\widehat{C}_{i,J}\right) \\ &+ 2 \sum_{I-J+1 \leq i < k \leq I} \left((1-\hat{\beta}_{I-i})(1-\hat{\beta}_{I-k}) \widehat{\mathrm{Cov}}(\hat{\mu}_i, \hat{\mu}_k) + \hat{\mu}_i \hat{\mu}_k \widehat{\mathrm{Cov}}(\hat{\beta}_{I-i}, \hat{\beta}_{I-k}) \right). \end{aligned}$$

2.4 Comparison

In Paper A we calculate the different estimators for two data sets and we compare the results to the results from Mack's distribution-free model given in Mack (2008). Additionally we give the results from the CL method using the distribution-free model given in Mack (1993) and we calculate the BF reserves under the assumption that the CL development pattern is used. One of the examples with incurred claims data from motor liability clearly shows the differences resulting from the different estimation procedures for the development patterns. In Figure 2.1 the CL development pattern (1.5) and the development pattern calculated from (2.4) are compared. In this example we have

$$\frac{\hat{\sigma}_j^2/\hat{\mu}_{[I-j]}}{\sum_{l=0}^{J} \hat{\sigma}_l^2/\hat{\mu}_{[I-l]}} \left(1 - \sum_{l=0}^{J} \frac{X_{[I-l],l}}{\hat{\mu}_{[I-l]}}\right) > 0, \quad 0 \leq j \leq J,$$

and these additive correction terms add up to 7.62%. The smoothing for each development year depends on the variance terms $\hat{\sigma}_j^2/\hat{\mu}_{[I-j]}$. In this case this leads to a longer tailed development pattern compared to the CL development pattern.

The corresponding data and more details are given in Paper A. As pointed out previously the use of the CL development pattern is not consistent with the BF philosophy.

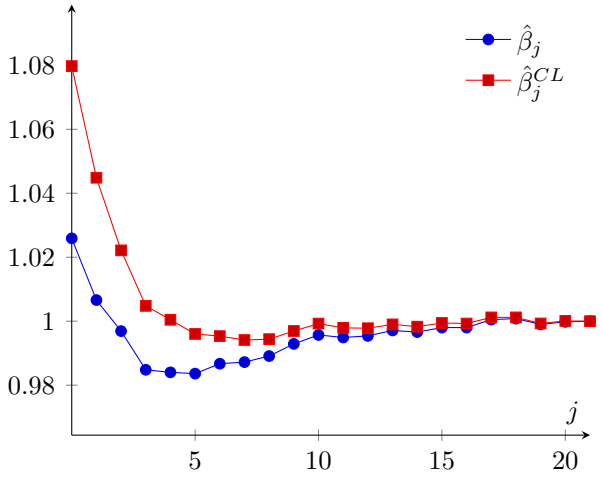

Figure 2.1: Estimated pattern $\hat{\beta}_j$ under Model Assumptions 2.3 and CL development pattern $\hat{\beta}_j^{CL}$ as given in (1.5).

The BF reserves strongly depend on the quality of the a priori estimates. If the CL predictors are either higher or lower than the BF predictors for most accident years, this is a strong indicator that the a priori estimates might be biased. Sometimes one then adjusts the a priori estimates. In the next chapter we present a method that allows to modify the a priori estimates.

3. Bornhuetter-Ferguson Method with Repricing

In the previous sections we have seen that the BF reserve completely relies on the a priori estimate. However, in some situations it might be reasonable to incorporate claims experience in order to update this a priori estimate. We call this process repricing. The combination of a priori information with observations is typically done in the framework of Bayesian statistics. For references on Bayesian methods for actuarial science we refer to Klugman (1992) and Makov et al. (1996). In Bayesian claims reserving methods expert information is incorporated with a prior distribution for a model parameter. In the case of the BF method one could for instance introduce a prior distribution around the a priori estimates. This prior distribution reflects uncertainty in the expert opinion. Based on the observations in the triangle/trapezoid this prior information is then updated with the use of Bayes' theorem. Bayesian predictors of the ultimate claim minimize the quadratic loss in the class of all $\mathcal{D}_I$-measurable and square integrable estimators. Moreover, parameter uncertainty is naturally incorporated in the posterior distributions. For the BF method a Bayesian model is studied in Verrall (2004). The ODP model from Verrall (2004) is investigated in more detail in England et al. (2012).

It is not always possible to calculate the posterior distribution of the ultimate claim explicitly. In such cases one can for instance use Markov Chain Monte Carlo (MCMC) simulation methods. For the implementation of MCMC methods in claims reserving we refer to Scollnik (2001). Another approach is to use credibility estimators (see Bühlmann-Gisler, 2005). An advantage of credibility methods is that one only needs to specify the first two moments instead of making full distributional assumptions. For credibility methods in the context of the CL method we refer to Gisler-Wüthrich (2008) and Bühlmann et al. (2009).

The credibility model in De Vylder (1982) is closer to the BF method. Mack (1990) generalized the model of De Vylder and embedded the model in the Bühlmann-Straub framework. The model in Paper B is a further generalisation of the model in Mack (1990). Moreover, in contrast to De

Vylder (1982) and Mack (1990) the estimation of the development pattern in Paper B incorporates the updated a posteriori estimates of the expected ultimate claims. Finally, Paper B also provides an estimate for the prediction uncertainty in the resulting method.

A second motivation for the credibility reserving method in Paper B comes from the fact that in practice different reserving methods are often combined. Recall that the BF reserve ignores the information from the claims triangle/trapezoid, whereas the CL method purely relies on the data and completely ignores any prior information. In a special case the repricing method given in Paper B can be interpreted as a weighted average between the CL and the BF method, similarly to the Benktander-Hovinen method (see Benktander, 1976; Hovinen, 1981). The optimal additive mixture between the CL and BF method is discussed in Mack (2000) and Wüthrich-Merz (2008).

3.1 Stochastic Model for the BF Method with Repricing

In Paper B the $\hat{\mu}_i$, which are used in the BF reserve formula

$$\hat{R}_i = \hat{\mu}_i \left(1 - \hat{\beta}_{I-i}\right), \tag{3.1}$$

are a posteriori adjusted estimates. For the repricing we switch from the basic BF assumptions formulated in the Distribution-free BF Model 2.1 to a Bayesian framework in which the unknown loss ratio parameter is modelled as a random variable. We assume that the true expected ultimate claim amount is given by $\mu_i(\Theta_i) = p_i\Theta_i$, where p_i are assumed to be known (earned premium) and Θ_i is a random variable describing the unknown expected loss ratio of accident year i. The distribution of Θ_i thus reflects the uncertainty of our a priori knowledge with regard to the 'true' individual expected loss ratio. The assumptions are formulated for $Y_{i,j} = X_{i,j}/p_i$. We denote the modified observations of accident year i by $\boldsymbol{Y}_i = (Y_{i,0}, \ldots, Y_{i,\iota(i)})'$, where we recall that $\iota(i) = \min(I - i, J)$ is the index of the last observed development year for accident year i.

Moreover, we assume in the following that the premium p_i is an on-level premium that is, earned premium adjusted for premium-cycles such that the a priori expected loss ratio θ_0 is the same for all accident years.

Model Assumptions 3.1 (BF Model with Repricing)
Each accident year i is characterised by its true expected loss ratio Θ_i and $\Theta_i > 0$ almost surely.

- *Given* $\boldsymbol{\Theta} = (\Theta_0, \ldots, \Theta_I)$*, all* $Y_{i,j}$ *are conditionally independent and there exist parameters* $\gamma_1, \ldots, \gamma_J$ *and positive parameters* γ_0*,* σ^2 *and* $\delta_0^2, \ldots, \delta_J^2$ *with* $\sum_{j=0}^{J} \gamma_j = \sum_{j=0}^{J} \delta_j^2 = 1$ *such that for* $0 \leq i \leq I$ *and* $0 \leq j \leq J$

$$E[Y_{i,j}|\boldsymbol{\Theta}] = \Theta_i \gamma_j, \quad \mathrm{Var}(Y_{i,j}|\,\boldsymbol{\Theta}) = \Theta_i \frac{\delta_j^2 \sigma^2}{p_i}.$$

- *We assume*

$$E[\Theta_i] = \theta_0 \quad \textit{and} \quad \mathrm{Cov}(\boldsymbol{\Theta}) = T,$$

where T *is symmetric and positive definite with* $(T)_{i,k} = \tau_{i,k}$ *and* $\tau_{i,i} = \tau^2$ *for all* i*.*

Remark 3.2

- Model Assumptions 3.1 should be compared with the basic BF assumptions formulated in the Distribution-free BF Model 2.1. In the Distribution-free BF Model 2.1 the μ_i's are unknown and it is assumed that there are unbiased a priori estimates for μ_i, $0 \leq i \leq I$. In the BF Model with Repricing 3.1 this a priori knowledge is given by the a priori expected value $\mu_i = p_i \theta_0$ and the corresponding uncertainties are modelled with the random variables Θ_i. In Paper B we also consider the case where θ_0 is not known. In this case the a priori expected ultimate claim amount is not specified by an expert, but it needs to be estimated from the data.

- In the special case $\delta_j^2 = \gamma_j$ we have $\mathrm{Var}(\,X_{i,j}|\,\Theta_i) = \sigma^2 E[X_{i,j}|\Theta_i]$, which is referred to as the ODP case. Note that the ODP case is not appropriate if negative incremental claims occur.

- If we assume that $\Theta_i \equiv \theta_0$ we distinguish the following cases. If θ_0 is known then the expected values $E[C_{i,J}]$ are exactly known and a repricing makes no sense and one should use $p_i \theta_0$ in the BF reserving formula. On the other hand, if θ_0 is unknown it needs to be estimated from the data, which is similar to the Cape Cod method of Bühlmann-Straub (1983), see also Section 1.3.3.

- If all $\gamma_j \neq 0$ we define $Z_{i,j} = Y_{i,j}/\gamma_j$ and we have

$$E[Z_{i,j}|\boldsymbol{\Theta}] = \Theta_i = \mu(\Theta_i),$$
$$\mathrm{Var}(Z_{i,j}|\boldsymbol{\Theta}) = \frac{\sigma^2 \Theta_i}{p_i \gamma_j^2/\delta_j^2} = \frac{\sigma^2(\Theta_i)}{w_{i,j}},$$

with $\mu(\Theta_i) = \Theta_i$, $\sigma^2(\Theta_i) = \Theta_i\sigma^2$ and $w_{i,j} = p_i\gamma_j^2/\delta_j^2$. For independent Θ_i's this gives the Bühlmann-Straub model. Under Model Assumptions 3.1 some γ_j's can be zero. Moreover, the Θ_i's can be dependent.

3.2 Credibility Estimates

The $\hat{\mu}_i$ in the BF reserve formula (3.1) are estimates of the true expected value of the ultimate claim $C_{i,J}$. Under Model Assumptions 3.1, this expected value depends on the risk characteristics Θ_i of year i, which means that the $\hat{\mu}_i$ in (3.1) should be estimates of $\mu_i(\Theta_i) = p_i\Theta_i$. We estimate the Θ_i's by credibility estimators.

Theorem 3.3 *Under Model Assumptions 3.1 the credibility estimators of* $\boldsymbol{\Theta} = (\Theta_0, \ldots, \Theta_I)'$ *based on all observations* $Y_{i,j}$, $i + j \leq I$, $0 \leq j \leq J$, *are given by*

$$\hat{\boldsymbol{\Theta}}^{\boldsymbol{inh}} = A\boldsymbol{B} + (I - A)\mathbf{1}\theta_0,$$
$$\widehat{\boldsymbol{\Theta}}^{\boldsymbol{hom}} = A\boldsymbol{B} + (I - A)\mathbf{1}\hat{\theta}_0,$$

where $\boldsymbol{B} = (B_0, \ldots, B_I)'$, *with*

$$B_i = \sum_{j=0}^{\iota(i)} \frac{\gamma_j/\delta_j^2}{w_{[\iota(i)]}} Y_{i,j}, \quad w_j = \frac{\gamma_j^2}{\delta_j^2},$$

and $\mathbf{1} = (1, \ldots, 1)'$, I *is the identity matrix and* $A = T(T + S)^{-1}$ *where* $S = E[\mathrm{Cov}(\boldsymbol{B}|\boldsymbol{\Theta})]$ *is the diagonal matrix with entries*

$$(S)_{i,i} = E[\mathrm{Var}(B_i|\boldsymbol{\Theta})] = \frac{\sigma^2\theta_0}{p_i w_{[\iota(i)]}}.$$

Moreover,

$$\hat{\theta}_0 = \left(\mathbf{1}'(S + T)^{-1}\mathbf{1}\right)^{-1}\mathbf{1}'(S + T)^{-1}\boldsymbol{B}.$$

A proof of Theorem 3.3 is given in Paper B. In the case where the Θ_i's are uncorrelated, we obtain simpler formulas:

Corollary 3.4 *Under Model Assumptions 3.1 and if the* Θ_i *are uncorrelated, then the credibility estimators of* Θ_i *based on all observations* $Y_{i,j}$, $i + j \leq I$, $0 \leq j \leq J$, *are given by*

$$\hat{\Theta}_i^{inh} = \alpha_i B_i + (1 - \alpha_i)\theta_0,$$
$$\hat{\Theta}_i^{hom} = \alpha_i B_i + (1 - \alpha_i)\hat{\theta}_0,$$

where B_i and w_j are given in Theorem 3.3 and

$$\alpha_i = \frac{p_i w_{[\iota(i)]}}{p_i w_{[\iota(i)]} + \frac{\sigma^2\theta_0}{\tau^2}}, \quad \hat{\theta}_0 = \sum_{i=0}^{I} \frac{\alpha_i}{\alpha_{[I]}} B_i.$$

If all $\gamma_j \neq 0$, Corollary 3.4 can also be derived from the Bühlmann-Straub model for $Z_{i,j} = Y_{i,j}/\gamma_j$. In the case where the Θ_i's are uncorrelated we have the following 'balance property' for the homogeneous credibility estimator (see Bühlmann-Gisler, 2005).

Theorem 3.5 (Balance Property) *Under Model Assumptions 3.1 and if the Θ_i's are uncorrelated, then the homogeneous credibility estimator of Θ_i satisfies the 'balance property'*

$$\sum_{i+j\leq I} \frac{\gamma_j}{\delta_j^2} \hat{X}_{i,j}^{hom} = \sum_{i+j\leq I} \frac{\gamma_j}{\delta_j^2} X_{i,j},$$

where $\hat{X}_{i,j}^{hom} = p_i \hat{\Theta}_i^{hom} \gamma_j$ with $\hat{\Theta}_i^{hom}$ given in Corollary 3.4.

Remark 3.6

- In the ODP case, we obtain the balance property

  $$\sum_{i+j\leq I} \hat{X}_{i,j}^{hom} = \sum_{i+j\leq I} X_{i,j},$$

  which means that the sum of the predicted past increments is equal to the sum of the observed claims.

- In the case $\tau^2 \to \infty$ we have credibility weights $\alpha_i = 1$ and we additionally get the following 'balance property' in each row

  $$\sum_{j=0}^{\iota(i)} \frac{\gamma_j}{\delta_j^2} \hat{X}_{i,j}^{hom} = \sum_{j=0}^{\iota(i)} \frac{\gamma_j}{\delta_j^2} X_{i,j}.$$

  In the ODP case we then get

  $$p_i \hat{\Theta}_i^{hom} = \frac{C_{i,\iota(i)}}{\beta_{\iota(i)}}.$$

3.3 Estimation of the Development Pattern

For the credibility estimators in Section 3.2 the development pattern $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_J)$ was assumed to be known. However, the second main question in the BF method is how this development pattern should be estimated. The best linear unbiased estimate of γ_j given $\boldsymbol{\Theta} = (\Theta_0, \ldots, \Theta_I)$ is given by

$$\gamma_j^{raw} = \frac{X_{[I-j],j}}{\sum_{i=0}^{I-j} \mu_i(\Theta_i)}. \tag{3.2}$$

Since the $\mu_i(\Theta_i)$ are not exactly known they are replaced by estimates $\hat{\mu}_i$ of $\mu_i(\Theta_i)$. In order to normalise the resulting estimates of γ_j we consider the simplest normalisation given by

$$\hat{\gamma}_j^s = \hat{\gamma}_j^{raw} \left(\sum_{j=0}^{J} \hat{\gamma}_j^{raw} \right)^{-1}, \tag{3.3}$$

where $\hat{\gamma}_j^{raw}$ is obtained from (3.2) by replacing the $\mu_i(\Theta_i)$ by the estimates $\hat{\mu}_i = p_i \hat{\Theta}_i^{inh}$ or $\hat{\mu}_i = p_i \hat{\Theta}_i^{hom}$.

In Paper B we also consider the normalisations derived from ML considerations in Paper A.

Note that the estimate of the development pattern will depend on $\hat{\boldsymbol{\mu}}$ that is, $\hat{\gamma}_j = \hat{\gamma}_j(\hat{\boldsymbol{\mu}})$ and the $\hat{\mu}_i$ depend themselves on $\boldsymbol{\gamma}$ that is, $\hat{\mu}_i = \hat{\mu}_i(\hat{\boldsymbol{\gamma}})$. Therefore an iterative procedure is necessary. In Paper B we give an iterative algorithm for the simultaneous estimation of the development pattern and the a posteriori adjusted $\hat{\mu}_i$. Further, Paper B provides estimates for the variance parameters. If the iterative procedure converges we call the resulting estimates a compatible pair.

If the Θ_i's are uncorrelated and if θ_0 is unknown, then the compatible pair $(\tilde{\boldsymbol{\mu}}^{hom}, \tilde{\boldsymbol{\gamma}}^{hom})$ is automatically normalised that is,

$$\tilde{\gamma}_j^{hom} = \frac{X_{[I-j],j}}{\tilde{\mu}_{[I-j]}^{hom}},$$

and we therefore have the balance property in each column j. A proof of this result is given in Paper B.

3.3.1 Relations to Other Methods

In this subsection we have a closer look at the ODP case ($\gamma_j = \delta_j^2$) under the assumption that the Θ_i's are uncorrelated. If θ_0 is known, we have

$$B_i = \frac{C_{i,\iota(i)}}{p_i \beta_{\iota(i)}} \quad \text{and} \quad \alpha_i = \frac{p_i \beta_{\iota(i)}}{p_i \beta_{\iota(i)} + \frac{\sigma^2 \theta_0}{\tau^2}},$$

and we obtain the following formula for the reserve which is also given in Wüthrich-Merz (2008)

$$\hat{R}_i = (1 - \beta_{\iota(i)}) \left(\alpha_i \frac{C_{i,\iota(i)}}{\beta_{\iota(i)}} + (1 - \alpha_i) p_i \theta_0 \right).$$

We therefore have a mixture between a data based projection method (CL type), where

$$\hat{C}^d_{i,J} = C_{i,\iota(i)} + (1 - \beta_{\iota(i)}) \frac{C_{i,\iota(i)}}{\beta_{\iota(i)}} = \frac{C_{i,\iota(i)}}{\beta_{\iota(i)}},$$

and a method based on external information (BF type), where

$$\hat{C}^e_{i,J} = C_{i,\iota(i)} + (1 - \beta_{\iota(i)}) p_i \theta_0.$$

In these formulas the development pattern is still assumed to be known. If the development pattern is estimated by the CL development pattern, we have $\hat{C}^d_{i,J} = \hat{C}^{CL}_{i,J}$. In practice it is not unusual to combine different reserving methods in such an additive way.

If $\tau^2 \to 0$ then $\alpha_i = 0$ and we get

$$\lim_{\tau^2 \to 0} \frac{\alpha_i}{\alpha_{[I]}} = \frac{p_i \beta_{\iota(i)}}{\sum_{l=0}^{I} p_l \beta_{\iota(l)}},$$

and therefore $\hat{\theta}_0$ is then the Cape Cod estimate of the expected loss ratio (see (1.8))

$$\hat{\theta}_0 = \frac{\sum_{i=0}^{I} C_{i,\iota(i)}}{\sum_{l=0}^{I} p_l \beta_{\iota(l)}}.$$

In the case $\tau^2 \to \infty$ the iterative algorithm for the estimation of the development pattern and the a posteriori adjusted $\hat{\mu}_i$'s yields the CL reserve and the CL development pattern, which is in line with the well-known result of the BF reserve in the ODP case, where the $\hat{\mu}_i$ and $\hat{\gamma}_j$ are the ML estimators (see Hachemeister-Stanard, 1975; Kremer 1985; Mack, 1991 and Renshaw-Verrall, 1998). An alternative proof, which also shows that the CL reserve is the unique solution, is given in Paper B. Hence in this special case we have a unique solution for the compatible pair.

If the premium coincides with the CL ultimate that is, if $p_i = \hat{C}^{CL}_{i,J}$, $0 \leq i \leq I$, then the algorithm delivers the CL development pattern and reserves.

3.4 Conditional Mean Square Error of Prediction (MSEP)

The final estimates resulting from the iterative estimation are denoted by $\tilde{\mu}_i$, $\tilde{\Theta}_i$, $\tilde{\beta}_{\iota(i)}$, $\tilde{\tau}^2$, $\widetilde{\sigma^2\theta_0}$ and $\tilde{\theta}_0$. The corresponding reserve is denoted by $\tilde{R}_i$ and we recall that $R_i = \sum_{j=I-i+1}^{J} X_{i,j}$. In Paper B we derive an estimate for the conditional MSEP in order to assess the prediction uncertainty. The conditional MSEP for accident year $i > I - J$, given the data $\mathcal{D}_I$, is given by

$$\operatorname{msep}_{R_i|\mathcal{D}_I}\left(\tilde{R}_i\right) = E\left[\left(R_i - \tilde{R}_i\right)^2 \middle| \mathcal{D}_I\right].$$

In Paper B the conditional MSEP is split into different additive terms. Some of these terms are approximated by unconditional means which allows to derive distribution-free estimators for these terms. One remaining term is estimated by parametric simulation. One could also estimate the conditional MSEP with a simulation, however, without closed formulas, interpretations are more difficult. Moreover, the conditional MSEP would then depend heavily on the distributional assumptions. In the approach presented in Paper B distributional assumptions are only needed for one term.

We define

$$R_i(\Theta_i) = p_i\Theta_i(1-\beta_{\iota(i)}) = E[R_i|\Theta_i], \quad R_i^{Bayes} = p_iE[\Theta_i|\mathcal{D}_I](1-\beta_{\iota(i)}),$$
$$R_i^{cred} = p_i\Theta_i^{inh}(1-\beta_{\iota(i)}),$$

where Θ_i^{inh} is the inhomogeneous credibility estimator as given in Theorem 3.3. The conditional MSEP, given all available data $\mathcal{D}_I$, is given by

$$\begin{aligned}\operatorname{msep}_{R_i|\mathcal{D}_I}\left(\tilde{R}_i\right) = {} & \underbrace{\sum_{j=I-i+1}^{J} p_i\sigma^2\delta_j^2 E\left[\Theta_i|\mathcal{D}_I\right]}_{=\mathrm{PV}_i} + \underbrace{E\left[\left(R_i(\Theta_i) - R_i^{cred}\right)^2 \middle| \mathcal{D}_I\right]}_{=\mathrm{EEI}_i} \\ & + \underbrace{\left(R_i^{cred} - \tilde{R}_i\right)^2}_{=\mathrm{EEII}_i} + 2\left(R_i^{cred} - \tilde{R}_i\right)\left(R_i^{Bayes} - R_i^{cred}\right).\end{aligned} \tag{3.4}$$

In the process variance PV_i the term $E[\Theta_i|\mathcal{D}_I]$ is approximated by $\tilde{\Theta}_i$ and after inserting the parameter estimates we obtain the following estimate

$$\widehat{\mathrm{PV}}_i = p_i\tilde{\Theta}_i\widetilde{\sigma^2\theta_0}/\tilde{\theta}_0 \sum_{j=I-i+1}^{J} \tilde{\delta}_j^2.$$

The estimation error term EEI_i in (3.4) is approximated by the unconditional expected value and by using the quadratic loss matrix

$$E\left[\left(\boldsymbol{\Theta}-\boldsymbol{\Theta}^{\boldsymbol{inh}}\right)\left(\boldsymbol{\Theta}-\boldsymbol{\Theta}^{\boldsymbol{inh}}\right)'\right]=(I-A)T.$$

Inserting the parameter estimates in the resulting approximation yields the following estimate for EEI_i

$$\widehat{\mathrm{EEI}}_i=p_i^2(1-\tilde{\beta}_{\iota(i)})^2((I-\tilde{A})\tilde{T})_{i,i}.$$

Given $\mathcal{D}_I$, the second estimation error term EEII_i in (3.4) is a constant, but since R_i^{cred} is not known, we have to estimate it. This has to be done by simulation. Finally, we assume that R_i^{cred} is close to R_i^{Bayes} so that the last mixed term in (3.4) is negligible.

Note that the closed formula for $\mathrm{PV}_i+\mathrm{EEI}_i$ is a lower bound for the conditional MSEP of accident year i, which is obtained without any distributional assumptions.

For the parametric bootstrap simulation in Paper B the following distributional assumptions are used: $\boldsymbol{\Theta}=|\bar{\boldsymbol{\Theta}}|:=(|\bar{\Theta}_0|,\ldots,|\bar{\Theta}_I|)$, where

$$\bar{\boldsymbol{\Theta}}\sim \mathrm{N}_{I+1}(\mathbf{1}\bar{\theta}_0,\bar{T}),\quad X_{i,j}|\Theta_i\sim \mathrm{N}(p_i\Theta_i\gamma_j,p_i\Theta_i\sigma^2\delta_j^2),$$

and where $\bar{\theta}_0$ and $\bar{T}$ are chosen such that $E[\boldsymbol{\Theta}]=\mathbf{1}\theta_0$ and $\mathrm{Cov}(\boldsymbol{\Theta})=T$. For our numerical examples, the differences $\bar{\theta}_0-\theta_0$ and $\bar{T}-T$ are negligible. Note that $\bar{\boldsymbol{\Theta}}$ has a multivariate normal distribution and for $\boldsymbol{\Theta}$ we consider componentwise the absolute values to ensure positivity. Of course, these assumptions can be changed easily and one can switch to different distributional assumptions. We have for instance tried a lognormal distribution for the $X_{i,j}$ with the assumptions on $\boldsymbol{\Theta}$ as above. Moreover, for the case of independent $(\Theta_i)_i$ we have also used a lognormal distribution for each Θ_i. We have compared the results for the simulated terms $\widehat{\mathrm{EEII}}$ under these different distributional assumptions. In all our examples the relative differences in the corresponding terms $\widehat{\mathrm{EEII}}$ were less than 0.5%. The details of the simulation procedure are explained in Paper B. Moreover, Paper B also provides an estimate for the conditional MSEP of aggregated accident years.

4. Best Estimate Reserves and Claims Development Results

Claims reserves are usually the most important item on the liability side of the balance sheet of a non-life insurance company. A change of the reserves by a small percentage can change the yearly result from quite positive to rather negative and vice versa. Therefore, the management wants to be confident that the actuarial reserves are best estimates that is, reserves that are neither optimistic nor pessimistic. As long as the CDRs in different accounting years fluctuate around zero, there is no reason to question that the reserves are best estimates because they are centred around zero. But if the CDRs in a given line of business (lob) are negative over several consecutive years, it often happens that actuaries are blamed for causing the resulting losses in the profit & loss account. On the other hand, if the CDRs are positive it is often complained that the company has missed profitable opportunities because of too high and not competitive premium calculation due to too pessimistic reserves.

The question whether CDRs with the same sign over several consecutive accounting years necessarily mean that the reserves are not best estimates is investigated in Paper C.

4.1 Basic Bayesian Framework

In Paper C two Bayesian models with diagonal effects are introduced. Diagonal effects are effects that act on accounting years and, therefore, such effects influence different accident years simultaneously. This leads to dependencies between accident years. Since diagonal effects affect accounting years they are also called accounting year effects or calendar year effects in the literature. A typical example of a diagonal effect is claims inflation. Diagonal effects received considerable attention in recent literature, see for instance Wüthrich (2010), Shi et al. (2012), Salzmann-Wüthrich (2012) and Merz et al. (2013). For further references we refer to the paper of Merz et al. (2013).

Apart from the data in the claims triangle there is often expert knowledge available in a company. This should be taken into account when calculating best estimate reserves. Let $\mathcal{D}_t$ denote the observed data until time t and in this chapter $\mathcal{I}_t$ denotes additional external information independent of the data, such as observations and expert opinion with regard to legal and economic developments.

A Bayesian framework allows to consider models, which incorporate a priori knowledge on the one hand and observed claims data on the other hand. In Paper C the diagonal effect for accounting year t is modelled by a random variable Ψ_t. The basic Bayesian framework considered in Paper C is a multiplicative model with the Bayesian structure given as follows:

Model Assumptions 4.1 (Bayes Model with Diagonal Effects) *Conditionally, given* $\mathbf{\Psi} = \{\Psi_0, \Psi_1, \ldots\}$ *and* $\mathcal{I}_t$ *for any* $t \in \mathbb{N}$*, the increments* $(X_{i,j})_{i,j}$ *are independent and there exist positive parameters* $\mu_0, \ldots, \mu_t$*, and parameters* $\gamma_0, \ldots, \gamma_J$*, with* $\sum_j \gamma_j = 1$ *such that*

$$E\left[X_{i,j}|\mathbf{\Psi}, \mathcal{I}_t\right] = E\left[X_{i,j}|\Psi_{i+j}\right] = \mu_i \gamma_j \Psi_{i+j}.$$

Further we assume that $\mathrm{Var}(\Psi_t) < \infty$ *and* $\mathrm{Var}(X_{i,j}|\mathbf{\Psi}, \mathcal{I}_t) < \infty$ *for all* $t \in \mathbb{N}$.

The variables $\mathbf{\Psi}$ are latent (non-observable) factors that drive accounting year effects in the $X_{i,j}$. Based on the information $\mathcal{D}_t$ and $\mathcal{I}_t$ these latent factors need to be filtered.

Best estimate reserves $\hat{R}_i^{(t)}$ of accident year i at time $t \geq i$ are defined as the reserves minimizing the conditional MSEP

$$\mathrm{msep}_{R_i^{(t)}|\mathcal{I}_t,\mathcal{D}_t}\left(\widehat{R}_i^{(t)}\right) = E\left[\left(\widehat{R}_i^{(t)} - R_i^{(t)}\right)^2 \middle| \mathcal{I}_t, \mathcal{D}_t\right],$$

where $R_i^{(t)} = \sum_{j=t-i+1}^{J} X_{i,j}$. Under Model Assumptions 4.1 the best estimate reserve of accident year i at the end of accounting year $t \geq J$ is given by

$$\widehat{R}_i^{(t)} = E\left[R_i^{(t)} \middle| \mathcal{I}_t, \mathcal{D}_t\right] = E\left[E\left[R_i^{(t)} \middle| \mathcal{I}_t, \mathcal{D}_t, \mathbf{\Psi}\right] \middle| \mathcal{I}_t, \mathcal{D}_t\right] \tag{4.1}$$

$$= \mu_i \sum_{j=t-i+1}^{J} \gamma_j E\left[\Psi_{i+j} \middle| \mathcal{I}_t, \mathcal{D}_t\right], \tag{4.2}$$

and the best estimate reserve at the end of accounting year t is given by the sum over all accident years:

$$\widehat{R}^{(t)} = \sum_{i=t-J+1}^{t} \widehat{R}_i^{(t)}.$$

Recall that the CDR of accident year $t - J \leq i \leq t - 1$ in the financial statement of accounting year t is defined by

$$\mathrm{CDR}_i^{(t)} = \widehat{C}_{i,J}^{(t-1)} - \widehat{C}_{i,J}^{(t)},$$

where $\widehat{C}_{i,J}^{(t)} = C_{i,t-i} + \widehat{R}_i^{(t)}$. Since

$$\widehat{C}_{i,J}^{(t)} = E\left[C_{i,J} | \mathcal{I}_t, \mathcal{D}_t\right],$$

the sequence $(\widehat{C}_{i,J}^{(t)})_{t \geq i}$ is a martingale with respect to the filtration generated by $\sigma(\mathcal{I}_t, \mathcal{D}_t)$, $t \in \mathbb{N}$. Hence the sequence $(\mathrm{CDR}_i^{(t)})_{t \geq i+1}$ of CDRs in consecutive years consists of increments of a martingale. Therefore, these CDRs are uncorrelated and have expected value zero. From this fact it is often concluded that CDRs having the same sign over several consecutive accounting years indicate that the reserves are not best estimates.

However, claim payments and the claims development depend on 'hidden' not directly observable (latent) characteristics. Under Model Assumptions 4.1 these latent characteristics are the diagonal effects and the martingale argument only holds in average over these characteristics. But in a claims development triangle a specific but unknown realisation of the latent variables is given. The observed CDRs resulting from best estimates based on such a conditional latent factor are no longer increments of a martingale with expected value zero. In the following we illustrate this fact with an example.

4.2 Illustrative Example

We consider the simple situation where there was a stable period in the years $t \leq s_0 - 1$ with no diagonal effects observable in the claims triangle and with no indication that this situation will change in the future. However, in accounting year s_0, there is a sudden change in the environment (e.g. legislation), which leads to a change of the expected value of the total claim payments in each 'cell' $\{(i, j) : i + j \geq s_0\}$ by some unknown factor Ψ_{s_0}. After accounting year s_0 the situation is again stable. For simplicity and illustrative purposes we assume that at time s_0 we know that there are no further changes in future accounting years. Of course, this is a simplification but it is just a special case of the Jump Model in Paper C, where several jumps can occur. In the case we consider here, the external information $\mathcal{I}_{s_0}$ contains the information that a change happens in year s_0 and that no further changes occur. For simplicity, we formulate the model for the time after s_0, so we can insert realised values from the information $\mathcal{I}_{s_0}$ and we then do not need to take the external information into account.

Model Assumptions 4.2

- *Conditionally, given* Ψ_{s_0}*, the* $(X_{i,j})_{i,j}$ *are independent and normally distributed with*

$$E[X_{i,j}|\Psi_{s_0}] = \begin{cases} \mu\gamma_j, & i+j \leq s_0-1, \\ \mu\gamma_j\Psi_{s_0}, & i+j \geq s_0, \end{cases}$$
$$\mathrm{Var}(X_{i,j}|\Psi_{s_0}) = \mu\gamma_j\sigma^2,$$

  where $\sum_{j=0}^{J}\gamma_j = 1$*, and* μ *and* $\gamma_0,\ldots,\gamma_J$ *are positive parameters.*

- $\Psi_{s_0} \sim N(m,\tau^2)$ *for some constants* $m \in \mathbb{R}$ *and* $\tau^2 > 0$*.*

Due to the assumption on the variance of the $X_{i,j}$, all γ_j need to be positive under Model Assumptions 4.2. The model in Paper C is more general and uses different parameters for the variances. In order to simplify the formulas we assume in the following that $s_0 \geq J$. Under Model Assumptions 4.2 we have for $t \geq s_0$

$$\Psi_{s_0}^{(t)} := E[\Psi_{s_0}|\mathcal{D}_t] = \alpha_t \frac{1}{\mu(t-s_0+1)}\sum_{l=s_0}^{t}\sum_{j=0}^{J} X_{l-j,j} + (1-\alpha_t)m, \tag{4.3}$$

where

$$\alpha_t = \frac{\mu(t-s_0+1)}{\mu(t-s_0+1)+\frac{\sigma^2}{\tau^2}}.$$

The predictors $\Psi_{s_0}^{(t)}$ and the variances $q_{s_0}^{(t)}$ can also be calculated recursively (see Proposition 3.4 in Paper C) and for $t > s_0$ we have

$$\Psi_{s_0}^{(t)} = \tilde{\alpha}_t \frac{\sum_{j=0}^{J} X_{t-j,j}}{\mu} + (1-\tilde{\alpha}_t)\Psi_{s_0}^{(t-1)}, \quad q_{s_0}^{(t)} = (1-\tilde{\alpha}_t)q_{s_0}^{(t-1)}, \tag{4.4}$$

where

$$\tilde{\alpha}_t = \frac{\mu}{\mu+\frac{\sigma^2}{q_{s_0}^{(t-1)}}}.$$

For early contributions on recursive credibility estimation we refer to Jones-Gerber (1975) and Sundt (1981).

From (4.3) an explicit formula for the CDR can be derived and for $t+1 > s_0$ we have

$$\mathrm{CDR}_{t+1} = \underbrace{\Psi_{s_0}^{(t)}\mu(1-\beta_0) - \sum_{j=1}^{J} X_{t+1-j,j}}_{\text{difference between forecast and observations}} + \underbrace{\left(\Psi_{s_0}^{(t)} - \Psi_{s_0}^{(t+1)}\right)\mu\sum_{k=2}^{J}(1-\beta_{k-1})}_{\text{change future}}.$$

From the recursive representation (4.4) we observe that in the case $t+1 > s_0$ the second term 'change future' converges to zero as $t \to \infty$ since the weights $\tilde{\alpha}_{t+1} \to 0$. On the other hand, the difference between forecast and observations depends more on the volatility of the data.

Next we want to consider the sign of the CDR over several consecutive accounting years in the period starting at year $s_0 + 1$. We recall that Ψ_{s_0} is not directly observable and hence the actuary needs to base his predictions on the available information $\mathcal{D}_t$. However, in one triangle there is one fixed realisation of Ψ_{s_0} which affects the data. Hence, if we consider one triangle, we are interested in the CDR conditional on Ψ_{s_0} and not on the average over all possible Ψ_{s_0}. Therefore we calculate the conditional expectation of the CDR, given Ψ_{s_0}. From Paper C we have for $t+1 > s_0$

$$\begin{aligned} &E\left[\mathrm{CDR}_{t+1}|\mathcal{D}_{s_0-1}, \Psi_{s_0}\right] \\ &\quad = (m - \Psi_{s_0})\mu \left((1-\beta_0)(1-\alpha_t) + (\alpha_{t+1} - \alpha_t) \sum_{k=2}^{J} (1-\beta_{k-1}) \right). \end{aligned} \tag{4.5}$$

Note that the second factor in (4.5) is positive since α_t is increasing in t. The first factor $m - \Psi_{s_0}$ is unknown but fixed for one triangle. From this we deduce the following properties for the conditional expected value of the CDR:

- The conditional expected value of the CDR has the same sign for all $t+1 \geq s_0 + 1$.
- The absolute value of $E\left[\mathrm{CDR}_{t+1}|\mathcal{D}_{s_0-1}, \Psi_{s_0}\right]$ increases with the absolute value $|m - \Psi_{s_0}|$.
- The absolute value of $E\left[\mathrm{CDR}_{t+1}|\mathcal{D}_{s_0-1}, \Psi_{s_0}\right]$ is bigger, if the development is longer. In other words, the slower the development pattern $(\beta_j)_{0 \leq j \leq J}$ increases to $\beta_J = 1$, the larger is the absolute value of $E\left[\mathrm{CDR}_{t+1}|\mathcal{D}_{s_0-1}, \Psi_{s_0}\right]$.
- The absolute value of $E\left[\mathrm{CDR}_{t+1}|\mathcal{D}_{s_0-1}, \Psi_{s_0}\right]$ is decreasing in t and in the limit for $t \to \infty$ the CDR is conditionally unbiased.

In Paper C we provide a simulation example to illustrate the results. For the simulation it is assumed that in accounting year $s_0 = 10$ there is an increase of 5% in the expected values of the future claims in each cell. Further it is assumed that the expected increase is only 2%. This means that the true diagonal effect is given by $\psi_{s_0} = 1.05$ and the expected value is $m = 1.02$. Further we choose $\mu = 10'000$, $\sigma = 4$ and $\tau = 0.02$ and a development pattern

γ_j as in Paper C. We then simulate 10′000 triangles with $I = 21$ accident years and $J = 9$ and calculate the CDR for time $s_0 = 10$ to time $t = 21$.

The conditional distribution of the CDR is illustrated with the boxplots of the realised values of the CDR in the simulations in Figure 4.1. Note that Model Assumptions 4.2 are only assumed for times $t \geq s_0$. For the calculation of CDR_{s_0} we assume that $\Psi_{s_0}^{(s_0-1)} = 1$ since before time s_0 there is no indication that a change will occur (see Paper C for a more detailed explanation). Observe that the conditional expected value is negative and it gets closer to zero in later accounting years. In the first few accounting years after the jump, it is likely to observe negative CDRs. The intensity of this effect strongly depends on the difference between the realised diagonal effect ψ_{s_0} and its expected value m. The larger this difference is, the more likely it is to observe a long period with CDRs of the same sign.

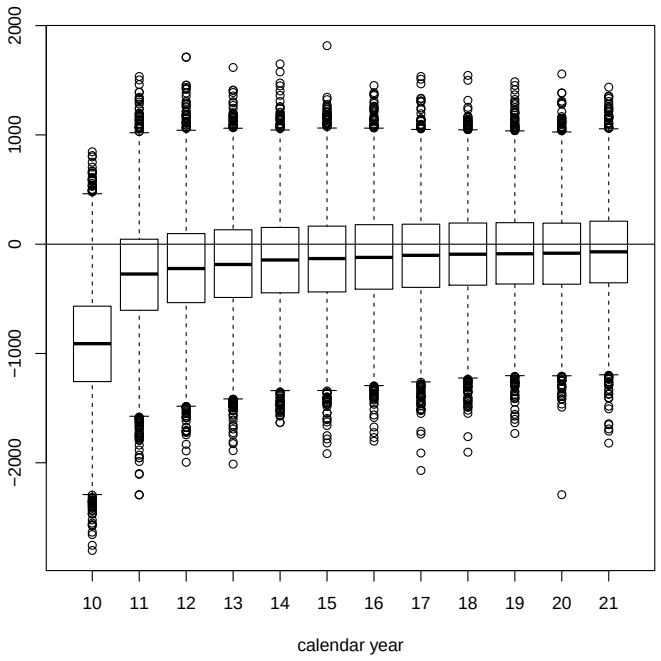

Figure 4.1: Boxplot of CDR_t, with 10'000 simulations, $\psi_{s_0} = 1.05$.

In practice, there are often several consecutive jumps going into the same direction over a period of years. In Paper C we give a more general framework in order to model such cases more adequately.

4.3 Further Remarks and Outlook

In Paper C it is shown in a simple model that the same sign of the CDR over several consecutive accounting years is not a contradiction to best estimate reserves. A more general model illustrates how situations of change can mathematically be modelled.

In our models it is assumed that there is the same diagonal factor for all elements of a diagonal. There can be situations, where this not the case, for instance, if a change of legislation only has an impact on bodily injury claims. Future research could extend the modelling framework to such situations by assuming that the diagonal effects are random vectors rather than random variables. For illustrative purposes we assumed the parameters to be known. For applications in practice one also needs to find estimators for these parameters.

5. Cape Cod Method

In the introduction we have seen that the Cape Cod method is a more robust method than the CL method. Due to its simplicity the Cape Cod method is well-established in practice. However, in the original article of Bühlmann-Straub (1983), there is no formula for the prediction uncertainty of the ultimate claim in the Cape Cod method and it is mentioned that the estimation of the development pattern in the Cape Cod method is still unsolved. In Paper D we introduce a stochastic model for the Cape Cod method and we derive estimates for the development pattern within this model. Moreover, we derive estimates for the conditional MSEP of the ultimate claim and for the uncertainty in the CDR.

5.1 Cape Cod Method in a Distribution-free Model

In Paper D we introduce the following distribution-free model for the Cape Cod method.

Model Assumptions 5.1 (Cape Cod Model)
Incremental claims $X_{i,j}$ are independent and there exist positive parameters q, $\nu_0, \nu_1, \ldots, \nu_I$, $\sigma_0^2, \sigma_1^2, \ldots, \sigma_J^2$, and a development pattern $\gamma_0, \gamma_1, \ldots, \gamma_J$, with $\sum_{j=0}^{J} \gamma_j = 1$ such that

$$E[X_{i,j}] = \nu_i q \gamma_j \quad \text{and} \quad \operatorname{Var}(X_{i,j}) = \nu_i q \sigma_j^2,$$

for $0 \leq i \leq I$ and $0 \leq j \leq J$.

The parameters $\nu_0, \ldots, \nu_I$ in Model Assumptions 5.1 denote the premiums and are assumed to be known. Note that for each $0 \leq i \leq I$, the premium ν_i is an on-level premium as described in Chapter 3.

Under Model Assumptions 5.1 the best linear unbiased estimate for γ_j is given by

$$\tilde{\gamma}_j^{raw} = \frac{X_{[I-j],j}}{q\nu_{[I-j]}}.$$

In general the $\tilde{\gamma}_j^{raw}$'s do not sum up to 1 and hence we consider the normalised pattern

$$\hat{\gamma}_j = \frac{\tilde{\gamma}_j^{raw}}{\sum_{j=0}^{J} \tilde{\gamma}_j^{raw}} = \frac{\hat{\gamma}_j^{raw}}{\hat{\beta}_J^{raw}}, \tag{5.1}$$

where

$$\hat{\gamma}_j^{raw} = \frac{X_{[I-j],j}}{\nu_{[I-j]}}, \quad \hat{\beta}_j^{raw} = \sum_{k=0}^{j} \hat{\gamma}_k^{raw}.$$

If we insert the estimates $\hat{\gamma}_j$ in the best linear unbiased estimator for q based on the observations $X_{i,j}$, $i+j \leq I$, $j \leq J$, we obtain

$$\hat{q} = \frac{\sum_{j=0}^{J} \sum_{i=0}^{I-j} X_{i,j} \frac{X_{[I-j],j}/\nu_{[I-j]}}{\hat{\beta}_J^{raw}\sigma_j^2}}{\sum_{j=0}^{J} \nu_{[I-j]} \frac{X_{[I-j],j}^2/(\nu_{[I-j]})^2}{\left(\hat{\beta}_J^{raw}\right)^2\sigma_j^2}} = \hat{\beta}_J^{raw}.$$

In Paper D it is shown that the same formula is obtained if we insert the estimates $\hat{\gamma}_j$ in the Cape Cod estimator

$$\hat{q} = \frac{\sum_{i=0}^{I} C_{i,\iota(i)}}{\sum_{i=0}^{I} \nu_i \hat{\beta}_{\iota(i)}} = \hat{\beta}_J^{raw},$$

where we recall that $\iota(i) = \min(I-i, J)$. We therefore use the following estimates under Model Assumptions 5.1

$$\hat{\gamma}_j = \frac{\frac{X_{[I-j],j}}{\nu_{[I-j]}}}{\sum_{l=0}^{J} \frac{X_{[I-l],l}}{\nu_{[I-l]}}}, \qquad \hat{q} = \frac{\sum_{i=0}^{I} C_{i,\iota(i)}}{\sum_{i=0}^{I} \nu_i \hat{\beta}_{\iota(i)}} = \sum_{l=0}^{J} \frac{X_{[I-l],l}}{\nu_{[I-l]}}. \tag{5.2}$$

The Cape Cod reserve and ultimate claim prediction for accident years $i \geq I-J+1$ are then given by

$$\hat{R}_i^{CC} = \nu_i \hat{q} \left(1 - \hat{\beta}_{I-i}\right) = \nu_i \hat{\beta}_J^{raw} \left(1 - \frac{\hat{\beta}_{I-i}^{raw}}{\hat{\beta}_J^{raw}}\right) = \nu_i \left(\hat{\beta}_J^{raw} - \hat{\beta}_{I-i}^{raw}\right), \tag{5.3}$$

$$\hat{C}_{i,J}^{CC} = C_{i,I-i} + \hat{R}_i^{CC}. \tag{5.4}$$

The variance parameters $q\sigma_j^2$ are estimated by

$$\widehat{q\sigma_j^2} = \frac{1}{I-j}\sum_{i=0}^{I-j}\frac{1}{\nu_i}\left(X_{i,j}-\nu_i\hat{\gamma}_j^{raw}\right)^2, \quad 0\leq j\leq J,\ j<I, \tag{5.5}$$

and if $I=J$ we use the extrapolation from Mack (1993) (see also (1.6))

$$\widehat{q\sigma_J^2} = \min\left(\widehat{q\sigma^2}_{J-2}, \frac{\widehat{q\sigma^2}^2_{J-1}}{\widehat{q\sigma^2}_{J-2}}\right).$$

In Paper D we derive the following estimate for the conditional MSEP of the Cape Cod predictor $\hat{C}_{i,J}^{CC}$ given in (5.4) under Model Assumptions 5.1

$$\widehat{\text{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\hat{C}_{i,J}^{CC}\right) = \widehat{\text{PV}}_i + \widehat{\text{PEE}}_i, \tag{5.6}$$

where $\widehat{\text{PV}}_i$ and $\widehat{\text{PEE}}_i$ are given by

$$\widehat{\text{PV}}_i = \sum_{j=I-i+1}^{J}\nu_i\widehat{q\sigma_j^2}, \quad \widehat{\text{PEE}}_i = \nu_i^2\sum_{j=I-i+1}^{J}\frac{\widehat{q\sigma_j^2}}{\nu_{[I-j]}},$$

and where $\widehat{q\sigma_j^2}$ is given in (5.5). For aggregated accident years the corresponding conditional MSEP is estimated by

$$\begin{aligned}&\widehat{\text{msep}}_{\sum_{i=I-J+1}^{I}C_{i,J}|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I}\hat{C}_{i,J}^{CC}\right)\\ &\quad=\sum_{i=I-J+1}^{I}\widehat{\text{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\hat{C}_{i,J}^{CC}\right)+2\sum_{I-J+1\leq i<k\leq I}\widehat{\text{PEE}}_{i,k},\end{aligned}$$

where $\widehat{\text{PEE}}_{i,k}$ is given by

$$\widehat{\text{PEE}}_{i,k} = \nu_i\nu_k\sum_{j=I-i+1}^{J}\frac{\widehat{q\sigma_j^2}}{\nu_{[I-j]}}, \quad i<k.$$

5.1.1 Conditional MSEP of the One-year Claims Development Result

From Section 1.2 we recall that the one-year CDR is the difference between two successive predictors of the ultimate claim. The CDR for accident year i and accounting year $I+1$ is given by

$$\text{CDR}_i^{(I+1)} = \hat{C}_{i,J}^{(I)} - \hat{C}_{i,J}^{(I+1)},$$

where $\hat{C}_{i,J}^{(k)}$, $k = I, I+1$, denotes the prediction of $C_{i,J}$ at time k. Under Model Assumptions 5.1 we have for $I - J + 1 \leq i \leq I$ and $k = I, I+1$,

$$\hat{C}_{i,J}^{(k)} = C_{i,k-i} + \nu_i \hat{q}^{(k)} \left(1 - \hat{\beta}_{k-i}^{(k)}\right) = C_{i,k-i} + \nu_i \left(\hat{\beta}_J^{(k)raw} - \hat{\beta}_{k-i}^{(k)raw}\right),$$

where $\hat{q}^{(k)} = \hat{\beta}_J^{(k)raw}$ and

$$\hat{\beta}_j^{(k)raw} = \sum_{l=0}^{j} \hat{\gamma}_l^{(k)raw}, \quad \text{with} \quad \hat{\gamma}_j^{(k)raw} = \frac{X_{[k-j],j}}{\nu_{[k-j]}}.$$

At time I, we consider $\hat{C}_{i,J}^{(I)}$ as best estimate based on the available information at time I. Therefore we predict $\mathrm{CDR}_i^{(I+1)}$ at time I by zero. The uncertainties in the predictions of the CDRs are studied by the conditional MSEP:

$$\mathrm{msep}_{\mathrm{CDR}_i^{(I+1)}|\mathcal{D}_I}(0) = E\left[\left(\mathrm{CDR}_i^{(I+1)}\right)^2 \middle| \mathcal{D}_I\right].$$

In Paper D we derive the following estimates for the conditional MSEP of the CDR of accident year i at time I:

$$\widehat{\mathrm{msep}}_{\mathrm{CDR}_i^{(I+1)}|\mathcal{D}_I}(0) = \nu_i \widehat{q\sigma^2}_{I+1-i} \frac{\nu_{[i]}}{\nu_{[i-1]}} + \nu_i^2 \sum_{j=I+2-i}^{J} \frac{\widehat{q\sigma^2}_j}{\nu_{[I-j]}} \frac{\nu_{I+1-j}}{\nu_{[I+1-j]}}, \tag{5.7}$$

where the $\widehat{q\sigma^2}_j$'s are given in (5.5). In order to compare this formula with the estimate (5.6) for the conditional MSEP of the predictor for the ultimate claim, we rewrite (5.7) as follows

$$\widehat{\mathrm{msep}}_{\mathrm{CDR}_i^{(I+1)}|\mathcal{D}_I}(0) = \nu_i \widehat{q\sigma^2}_{I+1-i} + \nu_i^2 \frac{\widehat{q\sigma^2}_{I+1-i}}{\nu_{[i-1]}} + \nu_i^2 \sum_{j=I+2-i}^{J} \frac{\widehat{q\sigma^2}_j}{\nu_{[I-j]}} \frac{\nu_{I+1-j}}{\nu_{[I+1-j]}}.$$

Note that the conditional MSEP of the CDR considers only the first term of $\widehat{\mathrm{PV}}_i$, which corresponds to the variance in the next diagonal. For the parameter estimation error the first term is fully incorporated and all other terms in $\widehat{\mathrm{PEE}}_i$ are scaled down by the factors $\nu_{I+1-j}/\nu_{[I+1-j]} < 1$. Similar relations are observed in the case of the CL method, see Merz-Wüthrich (2008) and Bühlmann et al. (2009).

The conditional MSEP of $\mathrm{CDR}^{(I+1)} = \sum_{i=I-J+1}^{I} \mathrm{CDR}_i^{(I+1)}$ is estimated by

$$\begin{aligned}\widehat{\mathrm{msep}}_{\mathrm{CDR}^{(I+1)}|\mathcal{D}_I}(0) = &\sum_{i=I-J+1}^{I} \widehat{\mathrm{msep}}_{\mathrm{CDR}_i^{(I+1)}|\mathcal{D}_I}(0) \\ &+ 2 \sum_{I-J+1 \leq i < k \leq I} \nu_i \nu_k \left(\frac{\widehat{q\sigma^2}_{I+1-i}}{\nu_{[i-1]}} + \sum_{j=I+2-i}^{J} \frac{\widehat{q\sigma^2}_j \nu_{I+1-j}}{\nu_{[I-j]} \nu_{[I+1-j]}} \right).\end{aligned}$$

In the derivation of these conditional MSEPs we use approximations by the unconditional expected value. Therefore these formulas can also be interpreted as unconditional MSEPs.

5.1.2 Over-dispersed Poisson (ODP) Model

A special case of Model Assumptions 5.1 is obtained by assuming that the incremental claims are ODP distributed with constant dispersion parameter ϕ. Note that incremental claims are in this case assumed to be positive.

The ODP assumption brings us back to the ODP Model 1.3, which was for instance used in England-Verrall (2002).

Under these additional distributional assumptions the estimates for q and γ_j can be calculated with the ML method. In Paper D it is shown that the ML estimates coincide with the estimates given in (5.2).

Under the ODP assumption the estimates for $q\sigma_j^2$, $0 \leq j \leq J$, need to be replaced by some estimates for $\phi q\gamma_j$. As in Wüthrich-Merz (2008) we suggest to estimate ϕ with the help of Pearson residuals

$$\hat{\phi} = \frac{1}{|\mathcal{D}_I| - (J+1)} \sum_{\substack{i+j \leq I \\ j \leq J}} \frac{(X_{i,j} - \nu_i \hat{\gamma}_j^{raw})^2}{\nu_i \hat{\gamma}_j^{raw}}, \tag{5.8}$$

and $\hat{\phi}\hat{\gamma}_j^{raw}$ is then an estimate for $\phi q\gamma_j$.

Remark 5.2 Under Model Assumptions 5.1 and if the $X_{i,j}$ are ODP distributed with constant dispersion parameter ϕ, we can derive the CL development pattern from ML estimation: if $\mu_i = \nu_i q$, $0 \leq i \leq I$, are considered as unknown parameters and if the μ_i's and γ_j's are both estimated by ML estimation then we know from Section 1.3.1 that the resulting pattern is the CL development pattern. For the estimation of the correlations in the CL development pattern one can therefore use the Fisher information matrix. This approach is also used by Alai et al. (2010) as described in Section 2.2. However, in the Cape Cod method the premium is assumed to be known and hence ignoring this information for the estimation of the development pattern is not consistent with the Cape Cod assumptions.

5.2 Example and Comparison

We use a data set from Wüthrich-Merz (2008) in order to illustrate the results. The data is also provided in Paper D. We calculate the results obtained in the Cape Cod Model 5.1 and the results obtained under the additional ODP

assumption. Moreover, for comparison we also list the results from the CL method and from the BF method. For the BF method we give the results from the ODP model of Paper A and from the model of Alai et al. (2010). Further we add the results obtained in the BF method with repricing under Model Assumptions 3.1. The a priori estimates for the BF method and the BF method with repricing are given in Table 5.2. For the coefficient of variation of the $\hat{\mu}_i$ the estimate described in Paper A is used with $n = 10$, which yields in this example a value of $\text{CoVa}(\hat{\mu}_i) = 12.3\%$. In this example the a priori estimates seem to be rather conservative and the BF method gives higher reserves than the CL and the Cape Cod method. In the BF method with repricing the conservative a priori estimates are adjusted and the resulting reserves lie between those from the CL method and from the Cape Cod method.

	Reserves	$\text{PV}^{1/2}$	$\text{PEE}^{1/2}$	$\text{MSEP}^{1/2}$	CoVa
Cape Cod Model 5.1	6'617'622	436'215	201'730	480'610	7.3%
Cape Cod ODP	6'617'622	378'170	290'414	476'815	7.2%
Mack CL	6'047'061	424'379	185'024	462'960	7.7%
BF ODP Alai et al. (2010)	7'356'580	329'007	228'252	733'129	10.0%
BF ODP Paper A	8'987'300	363'648	1'080'965	1'140'493	12.7%
BF Repricing, homogeneous	6'055'498	358'998	306'909	472'306	7.8%
BF Repricing, inhomogeneous	6'136'080	338'579	304'217	455'174	7.4%

Table 5.1: Results for aggregated accident years: Reserves, Process Standard Deviation, Parameter Estimation Error, Conditional MSEP and Coefficient of Variation.

i	$\hat{\mu}_i$
0	11'653'101
1	11'367'306
2	10'962'965
3	10'616'762
4	11'044'881
5	11'480'700
6	11'413'572
7	11'126'527
8	10'986'548
9	11'618'437

Table 5.2: A priori estimates $\hat{\mu}_i$ for $E[C_{i,J}]$ in the BF method.

6. Concluding Remarks

In this thesis we considered problems arising in claims reserving. On the one hand we have studied estimation and prediction uncertainties for well-established methods: for the BF method and the Cape Cod method we derived explicit formulas for the prediction errors. For the Cape Cod method we additionally gave estimates for the uncertainties in the one-year CDR. The resulting formulas are explicit and simple to be implemented in a spreadsheet. Moreover, we showed how the BF method can be modified in order to adjust the original a priori estimates based on the claims experience from the triangle/trapezoid. We have also shown relations between this new BF method with repricing and the CL method as well as the Cape Cod method. Additionally, the BF method with repricing allowed to include dependencies between different accident years.

In our models we have assumed that development patterns are the same for all accident years. In certain situations in practice it would be beneficial to model development patterns that are allowed to vary over the accident years. Future research on such varying development patterns would be interesting for practitioners.

On the other hand we have studied the CDR in a general Bayesian framework. We have illustrated that despite of the martingale property, CDRs with the same sign over several consecutive accounting years are not a contradiction to best estimate reserves. The illustration should solve this common misunderstanding between management and actuaries.

Paper A

Annina Saluz, Alois Gisler, Mario V. Wüthrich.

Development Pattern and Prediction Error for the Stochastic Bornhuetter-Ferguson Claims Reserving Method.

ASTIN Bulletin **41**(2), 2011, 279–313.

Development Pattern and Prediction Error for the Stochastic Bornhuetter-Ferguson Claims Reserving Method

Annina Saluz Alois Gisler Mario V. Wüthrich

ETH Zurich, RiskLab Switzerland, Department of Mathematics,
8092 Zurich, Switzerland

Abstract

We investigate the question how the development pattern in the Bornhuetter-Ferguson method should be estimated and derive the corresponding conditional mean square error of prediction (MSEP) of the ultimate claim prediction. An estimator of this conditional MSEP in a distribution-free model was given by Mack (2008), whereas in Alai et al. (2010) this conditional MSEP was studied in an over-dispersed Poisson model using the chain ladder development pattern. In this paper we consider distributional models and derive estimators (maximum likelihood) for the development pattern taking all relevant information into account. Moreover, we suggest new estimators of the correlation matrix of these estimators and new estimators of the conditional MSEP. Our findings supplement some of Mack's results. The methodology is illustrated at two numerical examples.

Keywords. Claims Reserving, Bornhuetter-Ferguson Method, Mean Square Error of Prediction, Claims Development Pattern.

1 Introduction

The chain ladder (CL) and the Bornhuetter-Ferguson (BF) method (1972) are still the most frequently used claims reserving methods in practice. For both methods it is assumed that there exists a development pattern γ_j, which is the same for all accident years. For the BF method there are, in addition

compared to the CL method, given estimates $\hat{\mu}_i$ of the expected ultimate claim of each accident year i, and the BF reserve of accident year i is obtained by multiplying $\hat{\mu}_i$ with the 'still to come percentage' of accident year i according to the development pattern γ_j. Thus, one essential difference between the CL and the BF philosophy is that the BF method incorporates $\hat{\mu}_i$ in the reserve estimate. Often, the CL development pattern is used for the development pattern in the BF method. However, this is not consistent with the BF philosophy, as the CL development pattern disregards any information contained in the $\hat{\mu}_i$'s.

Originally, both methods were established as purely pragmatic methods without an underlying stochastic model. Such a pragmatic approach allows to determine a point estimate of the claims reserve. However, in order to assess the uncertainties in this point estimate one needs an underlying stochastic model.

For the CL method Mack (1993) presented such a stochastic model and a corresponding estimator for the conditional mean square error of prediction (MSEP) in 1993. For the BF method such models and estimators are comparatively new. In 2008, Mack (2008) introduced a distribution-free model underlying the BF method and suggested estimators for the corresponding MSEP. Estimators of the MSEP for models related to BF were already derived earlier as, for instance, in England-Verrall (2002) or in Neuhaus (1992). However, Mack's model is probably the most general distribution-free model underlying the BF method. But this generality has also a drawback, namely, that many results in Mack (2008) cannot be obtained in a stringent mathematical way and that one has to resort to pragmatic considerations and approximations.

Another approach was used by Wüthrich-Merz (2008) and by Alai et al. (2010). Both studied the conditional MSEP in an over-dispersed Poisson (ODP) model where the development pattern is estimated by the CL development pattern.

In this paper we investigate the question how the development pattern in the BF method should be estimated considering all relevant information and how the corresponding correlation matrix looks like. For this purpose we consider stochastic BF models with parametric model assumptions for incremental claims. In these models we are able to derive the maximum likelihood estimators (MLEs) as well as the correlation matrix of the MLEs and derive estimators of the corresponding conditional MSEP. In contrast to the usual approach in which the CL development pattern is used, we implement the BF philosophy in the estimates of the development pattern by incorporating the a priori knowledge in this estimation. In our opinion the new estimates for the development pattern, the correlation matrix and the conditional MSEP

should generally be used for the BF method.

Moreover, we show how the variance of the a priori estimates can be estimated from the data. Finally, we apply our methods to real data from practice.

Organisation of the paper. In the next section we introduce the notation and data structure. In Section 3 we give the basic assumptions underlying the BF method and recall the model of Mack (2008). In Section 4 we introduce three distributional models underlying the BF method and derive estimates for the conditional MSEP in these models. Conclusions and numerical examples are given in Sections 5 and 6. Technical proofs of the statements are provided in Appendix A.

2 Notation and Data Structure

We denote the cumulative claims (cumulative payments or incurred losses) in accident year $i \in \{0, \dots, I\}$ at the end of development year $j \in \{0, \dots, J\}$ by $C_{i,j} > 0$ and we assume that $J \leq I$. Let $X_{i,j} = C_{i,j} - C_{i,j-1}$ denote the incremental claims, where we set $C_{i,-1} = 0$. In the sequel it is also useful to define

$$X_{[k],j} = \sum_{i=0}^{k} X_{i,j}, \quad 0 \leq k \leq I,\ 0 \leq j \leq J.$$

In general we denote the summation over an index starting from 0 with a square bracket. We assume that all claims are settled after development year J and therefore the total ultimate claim of accident year i is given by $C_{i,J}$. At time I we have information

$$\mathcal{D}_I = \{C_{i,j} : i + j \leq I,\ j \leq J\},$$

and our goal is to predict $\mathcal{D}_I^c = \{C_{i,j} : i + j > I, i \leq I,\ j \leq J\}$. The outstanding loss liabilities for accident year i at time I are given by

$$R_i = C_{i,J} - C_{i,I-i}, \quad I - J + 1 \leq i \leq I, \tag{2.1}$$

and the total outstanding loss liabilities are given by $R = \sum_{i=I-J+1}^{I} R_i$.

Remark 2.1. The 'true' outstanding loss liabilities are given by formula (2.1) only if $C_{i,j}$ denote cumulative payments. For incurred losses $C_{i,j}$ the outstanding loss liabilities are given by

$$R_i = C_{i,J} - C_{i,I-i} + C_{i,I-i} - C_{i,I-i}^{paid}, \tag{2.2}$$

with $C_{i,j}^{paid}$ denoting the cumulative payments of accident year i up to development year j. Note that the additional term $C_{i,I-i} - C_{i,I-i}^{paid}$ is observable at time I and has no impact on the claims prediction problem and uncertainty. Therefore we only consider the outstanding loss liabilities as defined in (2.1). For incurred losses, (2.1) is often referred to as IBNR.

3 BF Reserving Method

3.1 BF Reserving Method

The BF method goes back to Bornhuetter-Ferguson (1972). There are a priori estimates of the expected ultimate claim $C_{i,J}$ given by

$$\hat{\mu}_i = \nu_i \hat{q}_i, \tag{3.1}$$

where ν_i denotes the premium of accident year i and where $\hat{q}_i$ is an estimate of the expected ultimate loss ratio. The BF reserve is then given by

$$\widehat{R}_i = \hat{\mu}_i (1 - \hat{\beta}_{I-i}), \tag{3.2}$$

where $1 - \hat{\beta}_{I-i}$ is the estimated still to come factor at the end of development year $I-i$. In this paper we assume that the estimates $\hat{\mu}_i$ are independent of the observations $\mathcal{D}_I$, which is for instance the case, if these estimates represent an external expert opinion or if they come from a pricing which is mainly based on common statistics of pooled industry-wide data. Note that $\hat{\mu}_i$ is identical to an a priori estimate of the ultimate claim $C_{i,J}$, but that it is not its current predictor, which at time I is given by

$$\widehat{C}_{i,J} = C_{i,I-i} + \widehat{R}_i = C_{i,I-i} + \hat{\mu}_i (1 - \hat{\beta}_{I-i}).$$

3.2 Basic Assumptions Underlying the BF Method

From the fundamental properties of the BF method we derive some basic assumptions, which an underlying model should satisfy. These assumptions are motivated in Mack (2008). The independence assumption between the current claim amount $C_{i,I-i}$ and the reserve estimate suggests that incremental claims within the same accident year i are independent. The independence between different accident years is a standard assumption which we also adopt

here. Moreover, from the BF reserve estimate (3.2) it follows that a stochastic model for the BF method has to be cross-classified of the type

$$E[C_{i,j}] = \mu_i \beta_j \quad \text{or equivalently} \quad E[X_{i,j}] = \mu_i \gamma_j,$$

with $\beta_j = \sum_{k=0}^{j} \gamma_k$. Since $\mu_i \gamma_j = (\mu_i c)(\gamma_j / c)$ for any constant $c > 0$, the parameters μ_i and γ_j are only unique up to a constant factor. Without loss of generality one can therefore assume that $\gamma_0 + \ldots + \gamma_J = 1$. The sequence $(\beta_j)_j$ then denotes the (cumulative) development pattern and $(\gamma_j)_j$ is the (incremental) development pattern. Further, we assume that the a priori estimates $\hat{\mu}_i$ are unbiased for $E[C_{i,J}]$. Hence we make the following basic assumptions for a stochastic model underlying the BF method.

Model Assumptions 3.1 (Basic Underlying Assumptions)**.**

BF1 Incremental claims $X_{i,j}$ are independent and there exist parameters $\mu_0, \ldots, \mu_I$, $\gamma_0, \ldots, \gamma_J$, with $\sum_{j=0}^{J} \gamma_j = 1$ such that

$$E[X_{i,j}] = \mu_i \gamma_j, \quad 0 \leq i \leq I \text{ and } 0 \leq j \leq J.$$

BF2 The random variables $\hat{\mu}_i$ are unbiased estimates for $\mu_i = E[C_{i,J}]$.

Mack's Model: In addition to Model Assumptions 3.1 Mack (2008) makes the following variance assumption

$$\mathrm{Var}(X_{i,j}) = \mu_i \sigma_j^2. \tag{3.3}$$

Remark 3.2. Note that Mack's model is a distribution-free model for the BF method. Since it is a very general model several approximations are used to estimate the prediction uncertainties.

Alai et al. Model: In addition to Model Assumptions 3.1 Alai et al. (2010) assume that incremental claims $X_{i,j}$ are ODP distributed. Prediction uncertainty is then studied within a MLE and generalized linear model framework, but by using the CL development pattern.

3.3 Estimation of the Development Pattern

In order to estimate the BF reserve and its conditional MSEP we need to estimate the development pattern γ_j, $0 \leq j \leq J$. In practice one often uses the CL development pattern

$$\hat{\gamma}_j^{CL} = \prod_{k=j}^{J-1} \hat{f}_k^{-1} - \prod_{k=j-1}^{J-1} \hat{f}_k^{-1}, \quad \text{where} \quad \hat{f}_k = \frac{C_{[I-k-1],k+1}}{C_{[I-k-1],k}}. \tag{3.4}$$

However, the use of the CL development pattern is not consistent with the BF philosophy, as it disregards the information in the a priori estimates $\hat{\mu}_i$. If the μ_i were known then the best linear estimate of γ_j under Model Assumptions 3.1 and under the variance assumption (3.3) would be

$$\gamma_j^* = \frac{X_{[I-j],j}}{\mu_{[I-j]}},$$

with $\mu_{[k]} = \sum_{i=0}^{k} \mu_i$, $0 \leq k \leq I$. Therefore Mack (2008) suggests initial estimates

$$\hat{\gamma}_j^* = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}},$$

and then applies manual smoothing such that the final estimators $\hat{\gamma}_j$ sum up to 1.

The above estimators are not the only ones consistent with the BF philosophy, and the restriction to linear estimators is possibly not optimal. An alternative estimator is obtained with the following reasoning: assume first that we are given a full rectangle. Then an obvious estimator is

$$\hat{\gamma}_j = \frac{X_{[I],j}}{C_{[I],J}}.$$

Note that the $\hat{\gamma}_j$ automatically sum up to 1. As only the upper trapezoid $\mathcal{D}_I$ is given, a natural idea is to replace the unknown $X_{i,j}$ in the lower right triangle of $\mathcal{D}_I$ by the BF predictors $\hat{\mu}_i\hat{\gamma}_j$ resulting in the following system of equations

$$\hat{\gamma}_j = \frac{\sum_{i=0}^{I-j} X_{i,j} + \sum_{i=I-j+1}^{I} \hat{\mu}_i\hat{\gamma}_j}{\sum_{i=0}^{I-J} C_{i,J} + \sum_{i=I-J+1}^{I} \left(C_{i,I-i} + \sum_{l=I-i+1}^{J} \hat{\mu}_i\hat{\gamma}_l \right)}, \quad 0 \leq j \leq J. \quad (3.5)$$

In order to investigate different estimators and the corresponding conditional MSEPs more precisely we consider explicit BF models with parametric assumptions for incremental claims $X_{i,j}$. All the models considered in the following will satisfy Model Assumptions 3.1 and the variance assumption (3.3). Hence they are all special cases of the model considered in Mack (2008). In these distributional models we are able to derive estimates (MLEs) of the development pattern in a stringent mathematical way. Moreover, we find explicit solutions for the correlation matrices of these estimators, which are needed to calculate the MSEP of the BF reserves. As we will see, in one of the models, the above estimator (3.5) turns out to be the estimator corresponding to the MLE, if we replace the unknown μ_i by $\hat{\mu}_i$.

4 Stochastic BF Models based on Parametric Distributions

4.1 Over-dispersed Poisson Model with Constant Dispersion Parameter

4.1.1 Model

In Model Assumptions 3.1 we have specified the basic assumptions of the BF model. In the following model we assume that the $X_{i,j}$ follow an ODP distribution with constant dispersion parameter.

Model Assumptions 4.1 (Over-dispersed Poisson Model)**.**

P1 Incremental claims $X_{i,j}$ are independent and there exist positive parameters ϕ, $\mu_0, \ldots, \mu_I$ and $\gamma_0, \ldots, \gamma_J$ with $\sum_{j=0}^{J} \gamma_j = 1$ such that $X_{i,j}/\phi \sim$ Poisson$(\mu_i \gamma_j / \phi)$. In particular, we have

$$E[X_{i,j}] = \mu_i \gamma_j,$$
$$\text{Var}(X_{i,j}) = \phi \mu_i \gamma_j,$$

where ϕ is called dispersion parameter.

P2 The a priori estimates $\hat{\mu}_i$ for $\mu_i = E[C_{i,J}]$ are unbiased and independent from $X_{l,j}$ for $0 \leq l \leq I$, $0 \leq j \leq J$.

Remarks 4.2.

- In the ODP Model the incremental claims $X_{i,j}$ are required to be positive.
- Disregarding the μ_i, that is considering data $\mathcal{D}_I$ only, the MLEs for the claims reserves and the development pattern are identical to the CL reserves and the CL development pattern $\hat{\gamma}_j^{CL}$, see for instance Hachemeister-Stanard (1975), Mack (1991) or Schmidt-Wünsche (1998).
- Alai et al. (2010) used a similar model to derive an estimate for the conditional MSEP in the BF method.
- Note that we have not made any assumption on the dependence structure between the $\hat{\mu}_i$'s.

4.1.2 Estimation of the Development Pattern

We estimate the development pattern with the MLE method. In order to calculate the MLEs we first assume that the μ_i are known and then plug in the estimates $\hat{\mu}_i$. The MLEs are best calculated using Lagrange multipliers. Note that in the ODP case the log-likelihood function for $X_{i,j}$ is given by

$$\frac{1}{\phi}\left(X_{i,j}\left(\log\mu_i + \log\gamma_j\right) - \mu_i\gamma_j\right) - g(X_{i,j},\phi),$$

where $g(\cdot,\phi)$ is a normalising function that does not depend on the parameters μ_i and γ_j.

Remark 4.3. Multiplying the log-likelihood function by ϕ, we observe that the MLEs for the γ_j's do not depend on the dispersion parameter.

Neglecting the normalising function g we obtain the Lagrange function with Lagrange multiplier κ given by

$$\mathcal{L}_{\mathcal{D}_I}\left(\gamma_0,\ldots,\gamma_J,\kappa\right) = \sum_{i+j\le I}\left(X_{i,j}(\log\mu_i + \log\gamma_j) - \mu_i\gamma_j\right) + \kappa\left(1 - \sum_{j=0}^{J}\gamma_j\right).$$

Note that the last term comes from the normalisation requirement $\sum_{j=0}^{J}\gamma_j = 1$. For the MLEs $\tilde{\gamma}_j$ we obtain the following equations

$$\frac{\partial\mathcal{L}_{\mathcal{D}_I}}{\partial\gamma_j} = \sum_{i=0}^{I-j}\left(\frac{X_{i,j}}{\gamma_j} - \mu_i\right) - \kappa = 0,$$

and therefore we find the solutions

$$\tilde{\gamma}_j = \frac{X_{[I-j],j}}{\mu_{[I-j]} + \kappa}. \tag{4.1}$$

From the side constraint it follows that

$$\frac{\partial\mathcal{L}_{\mathcal{D}_I}}{\partial\kappa} = 1 - \sum_{j=0}^{J}\gamma_j = 0,$$

which yields in view of (4.1) an implicit equation for κ

$$1 - \sum_{j=0}^{J}\frac{X_{[I-j],j}}{\mu_{[I-j]} + \kappa} = 0. \tag{4.2}$$

From equation (4.2) it follows that κ is the root of a polynomial of degree J. Since the γ_j need to be positive in the ODP Model and since $\mu_{[I-j]}$ is decreasing in j we have the constraint $\kappa \in (-\mu_{[I-J]},\infty)$. In this interval there is a unique solution κ by monotonicity (cf. Remarks 4.6).

Remark 4.4. We do not use the partial derivatives with respect to μ_i. This is not necessary since we already have the a priori estimates $\hat{\mu}_i$ and therefore do not want to use MLEs for the μ_i.

Then, we replace the true (unknown) μ_i by the estimates $\hat{\mu}_i$ and obtain the estimators

$$\hat{\gamma}_j = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]} + \hat{\kappa}}, \tag{4.3}$$

where $\hat{\kappa}$ fulfils the implicit equation

$$1 - \sum_{j=0}^{J} \hat{\gamma}_j = 0. \tag{4.4}$$

Theorem 4.5. *The estimates* (4.3) *satisfy equations* (3.5).

Proof. Observe that the estimates $\hat{\gamma}_j$ defined in (3.5) satisfy the constraint $\sum_{j=0}^{J} \hat{\gamma}_j = 1$. Therefore it is sufficient to prove that the $\hat{\gamma}_j$'s defined in (4.3) satisfy

$$\frac{\hat{\gamma}_0}{\hat{\gamma}_j} = \frac{X_{[I],0}}{X_{[I-j],j} + \sum_{i=I-j+1}^{I} \hat{\mu}_i \hat{\gamma}_j}, \quad 1 \leq j \leq J.$$

From equation (4.3) we have $\hat{\kappa} = \frac{X_{[I-j],j}}{\hat{\gamma}_j} - \hat{\mu}_{[I-j]}$ for $0 \leq j \leq J$ and conclude

$$\begin{aligned} \frac{\hat{\gamma}_0}{\hat{\gamma}_j} &= \left(\frac{X_{[I],0}}{\hat{\mu}_{[I]} + \hat{\kappa}} \right) \frac{1}{\hat{\gamma}_j} = \left(\frac{X_{[I],0}}{\hat{\mu}_{[I]} + \frac{X_{[I-j],j}}{\hat{\gamma}_j} - \hat{\mu}_{[I-j]}} \right) \frac{1}{\hat{\gamma}_j} \\ &= \frac{X_{[I],0}}{X_{[I-j],j} + \sum_{i=I-j+1}^{I} \hat{\mu}_i \hat{\gamma}_j}. \end{aligned}$$

This proves the claim.

□

Remarks 4.6.

- If instead of $\hat{\mu}_i$ the data of the fully developed rectangular was given, then the MLE of γ_j would be

  $$\hat{\gamma}_j^{MLE} = \frac{X_{[I],j}}{C_{[I],J}},$$

  and would not depend on the μ_i at all. This shows that in the ODP Model the MLEs are not linear in the observations $X_{i,j}$. Given $\mathcal{D}_I$, the μ_i's appear in the MLEs only to the extent that the entries $X_{i,j}$ in the lower right triangle are not known and filled up with the BF predictors $\hat{\mu}_i \hat{\gamma}_j$.

- From Theorem 4.5 it follows that the estimators $\hat{\gamma}_j$ obtained by replacing the unknown μ_i in the MLEs $\tilde{\gamma}_j$ by $\hat{\mu}_i$ satisfy equations (3.5). However the easiest way of finding the solution of equations (3.5) is to calculate the estimators $\hat{\gamma}_j$ by means of equation (4.3). There we only need to calculate $\hat{\kappa}$, which can be done quite easily by starting with $\hat{\kappa} = 0$, that is with the raw estimators

$$\hat{\gamma}_j^{(0)} = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}},$$

and then by increasing or decreasing $\hat{\kappa}$ until the side constraint

$$f\left(\hat{\kappa}\right) = \sum_{j=0}^{J} \hat{\gamma}_j = 1$$

is fulfilled. Note hereby that in the interval $(-\hat{\mu}_{[I-J]}, \infty)$ the function $f\left(\hat{\kappa}\right)$ is strictly decreasing and consequently the side constraint $f(\hat{\kappa}) = 1$ has a unique solution. Of course β_j is then estimated by $\hat{\beta}_j = \sum_{k=0}^{j} \hat{\gamma}_k,\ 0 \leq j \leq J$, and satisfies $\hat{\beta}_J = 1$.

- From the estimators (4.3) we also see how a mathematically founded smoothing from the raw estimators $\hat{\gamma}_j^{(0)}$ to the adjusted estimators $\hat{\gamma}_j$ looks like in the ODP Model. If we write

$$\hat{\gamma}_j = c_j \hat{\gamma}_j^{(0)}$$

then the 'correction factors' c_j are given by

$$c_j = \frac{\hat{\mu}_{[I-j]}}{\hat{\mu}_{[I-j]} + \hat{\kappa}}.$$

Since $\hat{\mu}_{[I-j]}$ is decreasing in j, c_j is decreasing in j for $\hat{\kappa} > 0$ and increasing for $\hat{\kappa} < 0$. For instance, if $\hat{\kappa} < 0$ then the correction factor is greater for the late development years than for the newer ones.

- Consider the effect of conservative a priori estimation. If we increase some of the a priori estimates $\hat{\mu}_i$ then $\hat{\kappa}$ is decreasing because of the side constraint (cf. (4.3)-(4.4)). We denote the new estimates by $\hat{\mu}_i^c$, $\hat{\gamma}_j^c$ and $\hat{\kappa}^c$. For the change in the denominator of $\hat{\gamma}_j$ we then have

$$\hat{\mu}_{[I-k]}^c - \hat{\mu}_{[I-k]} + \hat{\kappa}^c - \hat{\kappa} \leq \hat{\mu}_{[I-j]}^c - \hat{\mu}_{[I-j]} + \hat{\kappa}^c - \hat{\kappa}, \quad 0 \leq j < k \leq J,$$

and because of the side constraint there must be an index $0 \leq l < J$ such that for all $j \leq l < k$ we have

$$\hat{\mu}_{[I-k]}^c - \hat{\mu}_{[I-k]} + \hat{\kappa}^c - \hat{\kappa} \leq 0 \leq \hat{\mu}_{[I-j]}^c - \hat{\mu}_{[I-j]} + \hat{\kappa}^c - \hat{\kappa}.$$

It follows that $\hat{\gamma}_k^c \geq \hat{\gamma}_k$ for all $k > l$ (in the tail) and for $j \leq l$ we have $\hat{\gamma}_j^c \leq \hat{\gamma}_j$. This means that more weight is given to the tail and the effect of conservative a priori estimation is therefore reinforced by the estimates of the development pattern. Note that in the case where we only increase $\hat{\mu}_i$ with $i \leq I - J$ the estimates of the development pattern remain unchanged (cf. (3.5)).

- The quality of the BF estimators strongly depends on the quality of the a priori estimates $\hat{\mu}_i$. If $\sum_{j=0}^{J} \hat{\gamma}_j^{(0)}$ deviates much from 1 and thus $\hat{\kappa}$ deviates much from 0, then this could be an indicator that the a priori estimates might be biased. As a further check one could estimate the development pattern and the reserves by disregarding the a priori estimates $\hat{\mu}_i$. In the OPD Model with a full development triangle or trapezoid the resulting MLEs are then identical to the CL estimates. If for most accident years the CL predictors for the ultimate claims were either higher or lower than the BF predictors, this would be a strong indicator that the a priori estimates $\hat{\mu}_i$ might be biased and that one of the basic assumptions of BF is violated. Sometimes one then adjusts the a priori estimates $\hat{\mu}_i$ (see e.g. the Cape Cod method in Wüthrich-Merz, 2008 or Radtke-Schmidt, 2004, which goes back to Bühlmann-Straub, 1983). There is no problem to calculate the BF reserves with 'a posteriori' adjusted $\hat{\mu}_i$. But this means that the $\hat{\mu}_i$ are no longer independent of the data $\mathcal{D}_I$, which should be taken into account for the estimation of the MSEP and is beyond the scope of this paper.

- A tail development $\hat{\gamma}_{J+1}$ can be incorporated by replacing the side constraint in the Lagrange function by $\sum_{j=0}^{J} \gamma_j = 1 - \hat{\gamma}_{J+1}$. The estimate $\hat{\gamma}_{J+1}$ needs to come from outside together with an estimated covariance structure $\widehat{\mathrm{Cov}}(\hat{\gamma}_{J+1}, \hat{\gamma}_j)$, $0 \leq j \leq J+1$, because there is no data available for the tail.

Finally, we need to estimate the dispersion parameter ϕ. As in Wüthrich-Merz (2008) we use Pearson residuals to estimate ϕ. The Pearson residuals are given by

$$R_{i,j} = \frac{X_{i,j} - \mu_i \gamma_j}{\sqrt{\mu_i \gamma_j}},$$

with $E[R_{i,j}^2] = \phi$. To estimate $R_{i,j}$ we replace μ_i and γ_j by estimates, but we need to be cautious here. We want to estimate the variance of the $X_{i,j}$ and therefore we are only interested in the process variance term and not in the parameter estimation error. The uncertainty of the external estimate $\hat{\mu}_i$ is considered in a different step. As a consequence we have to estimate μ_i and γ_j

from the data in order to obtain an estimate for ϕ. We use the MLE method to gain data based estimates for μ_i and γ_j. As mentioned in Remarks 4.6 in the case of ODP data in a full triangle or trapezoid the MLEs of μ_i and γ_j coincide with the CL estimates. Therefore we estimate $R_{i,j}$ with

$$\hat{R}_{i,j} = \frac{X_{i,j} - \hat{C}_{i,J}^{CL}\hat{\gamma}_j^{CL}}{\sqrt{\hat{C}_{i,J}^{CL}\hat{\gamma}_j^{CL}}},$$

where $\hat{C}_{i,J}^{CL} = C_{i,I-i}\prod_{j=I-i}^{J-1}\hat{f}_j$ and where the CL development pattern $\hat{\gamma}_j^{CL}$ was defined in (3.4). An estimate for ϕ is then given by

$$\hat{\phi} = \frac{\sum_{i+j\leq I}\hat{R}_{i,j}^2}{|\mathcal{D}_I| - p}, \tag{4.5}$$

where $|\mathcal{D}_I|$ is the number of observations in $\mathcal{D}_I$ and where $p = I + J + 1$ is the number of estimated parameters.

In order to quantify the uncertainties in the BF predictors $\widehat{C}_{i,J}$ and $\sum_{i=I-J+1}^{I}\widehat{C}_{i,J}$ we consider the conditional MSEP. Given information $\mathcal{I}_I$, the conditional MSEP of a predictor $\widehat{X}$ of a random variable X is defined by

$$\mathrm{msep}_{X|\mathcal{I}_I}(\widehat{X}) = E[(X - \widehat{X})^2|\mathcal{I}_I].$$

If in addition the predictor $\widehat{X}$ is $\mathcal{I}_I$-measurable we obtain

$$\mathrm{msep}_{X|\mathcal{I}_I}(\hat{X}) = \mathrm{Var}(X|\mathcal{I}_I) + (E[X|\mathcal{I}_I] - \hat{X})^2.$$

For the calculation of the conditional MSEP we therefore need to consider second moments and covariances of the estimates.

4.1.3 Covariance Matrix of the Estimated Development Pattern

In this section we consider the estimation of the covariance matrix of the $\hat{\gamma}_j$'s. A well-known result from statistics is that MLEs are asymptotically unbiased and multivariate normally distributed. Moreover, the asymptotic covariance matrix is given by the inverse of the Fisher information matrix. By definition, the entries of the Fisher information matrix $H(\boldsymbol{\gamma})$ are given by

$$H(\boldsymbol{\gamma})_{j,k} = E\left[\frac{\partial l_{\mathcal{D}_I}}{\partial\gamma_j}\frac{\partial l_{\mathcal{D}_I}}{\partial\gamma_k}\right], \quad 0 \leq j,k \leq J-1,$$

where $l_{\mathcal{D}_I}$ is the log-likelihood function. In the ODP case the log-likelihood function is given by

$$l_{\mathcal{D}_I}(\gamma_0,\dots,\gamma_{J-1}) = \sum_{\substack{i+j\le I\\ j\le J-1}} \frac{1}{\phi}\left(X_{i,j}\left(\log\mu_i+\log\gamma_j\right)-\mu_i\gamma_j\right)$$
$$+\sum_{i=0}^{I-J}\frac{1}{\phi}\left(X_{i,J}\left(\log\mu_i+\log\left(1-\sum_{j=0}^{J-1}\gamma_j\right)\right)-\mu_i\left(1-\sum_{j=0}^{J-1}\gamma_j\right)\right)+r,$$

where r contains all remaining terms, which do not depend on the parameters $\boldsymbol{\gamma}=(\gamma_0,\dots,\gamma_{J-1})$. Note that the last line of the above equality comes from the normalising condition $\sum_{j=0}^{J}\gamma_j=1$. For $0\le j\le J-1$ we obtain the diagonal elements

$$\begin{aligned}
H(\boldsymbol{\gamma})_{j,j} &= E\left[\left(\frac{\partial l_{\mathcal{D}_I}}{\partial\gamma_j}\right)^2\right]\\
&= E\left[\left(\sum_{i=0}^{I-j}\frac{1}{\phi}\left(\frac{X_{i,j}}{\gamma_j}-\mu_i\right)-\sum_{i=0}^{I-J}\frac{1}{\phi}\left(\frac{X_{i,J}}{\gamma_J}-\mu_i\right)\right)^2\right]\\
&= \sum_{i=0}^{I-j}\frac{1}{\phi^2}\mathrm{Var}\left(\frac{X_{i,j}}{\gamma_j}\right)+\sum_{i=0}^{I-J}\frac{1}{\phi^2}\mathrm{Var}\left(\frac{X_{i,J}}{\gamma_J}\right)\\
&= \sum_{i=0}^{I-j}\frac{1}{\phi}\frac{\mu_i}{\gamma_j}+\sum_{i=0}^{I-J}\frac{1}{\phi}\frac{\mu_i}{\gamma_J} \;=\; \frac{\mu_{[I-j]}}{\phi\gamma_j}+\frac{\mu_{[I-J]}}{\phi\gamma_J}.
\end{aligned}\tag{4.6}$$

Analogously, for $0\le j<k\le J-1$, we have the off-diagonal terms

$$H(\boldsymbol{\gamma})_{k,j}=H(\boldsymbol{\gamma})_{j,k}=E\left[\frac{\partial l_{\mathcal{D}_I}}{\partial\gamma_j}\frac{\partial l_{\mathcal{D}_I}}{\partial\gamma_k}\right]=\frac{\mu_{[I-J]}}{\phi\gamma_J}.\tag{4.7}$$

Hence, we write the Fisher information matrix as follows

$$H(\boldsymbol{\gamma})=\frac{\mu_{[I-J]}}{\phi\gamma_J}\begin{pmatrix}
\frac{\mu_{[I]}/\gamma_0}{\mu_{[I-J]}/\gamma_J}+1 & 1 & \cdots & 1\\
1 & \frac{\mu_{[I-1]}/\gamma_1}{\mu_{[I-J]}/\gamma_J}+1 & \cdots & 1\\
\vdots & \vdots & \ddots & \vdots\\
1 & 1 & \cdots & \frac{\mu_{[I-J+1]}/\gamma_{J-1}}{\mu_{[I-J]}/\gamma_J}+1
\end{pmatrix}\in\mathbb{R}^{J\times J}.$$

For the entries of the inverse of the Fisher information matrix we obtain (see Appendix A)

$$(H(\boldsymbol{\gamma})^{-1})_{j,k}=\frac{\phi\gamma_j}{\mu_{[I-j]}}\left(\mathbb{1}_{\{j=k\}}-\frac{\gamma_k/\mu_{[I-k]}}{\sum_{l=0}^{J}\gamma_l/\mu_{[I-l]}}\right),\quad 0\le j,k\le J-1.\tag{4.8}$$

We now assume that the $\hat{\mu}_i$ are close to the true μ_i and hence the covariances $\mathrm{Cov}(\hat{\gamma}_j, \hat{\gamma}_k)$ are close to the covariances of the true MLEs $\tilde{\gamma}_j$ (cf. (4.1)). With the asymptotic MLE approximation we then obtain

$$\mathrm{Cov}(\hat{\gamma}_j, \hat{\gamma}_k) \approx (H(\boldsymbol{\gamma})^{-1})_{j,k},$$

and of course the covariances of $\hat{\gamma}_J = 1 - \sum_{j=0}^{J-1} \hat{\gamma}_j$ are obtained by linearity. Replacing the parameters γ_j, ϕ and $\mu_{[I-j]}$ by their estimates $\hat{\gamma}_j$ (see (4.3)), $\hat{\phi}$ (see (4.5)) and $\hat{\mu}_{[I-j]}$ (see Model Assumptions 4.1) we arrive at the following estimates for the covariances

$$\widehat{\mathrm{Cov}}(\hat{\gamma}_j, \hat{\gamma}_k) = \frac{\hat{\phi}\hat{\gamma}_j}{\hat{\mu}_{[I-j]}} \left(\mathbb{1}_{\{j=k\}} - \frac{\hat{\gamma}_k / \hat{\mu}_{[I-k]}}{\sum_{l=0}^{J} \hat{\gamma}_l / \hat{\mu}_{[I-l]}} \right), \quad 0 \le j,k \le J. \tag{4.9}$$

This is clear for $0 \le j,k \le J-1$ and also holds for $0 \le j,k \le J$.
Note that

$$\gamma_j^* = \frac{X_{[I-j],j}}{\mu_{[I-j]}}$$

are uncorrelated with variance

$$\mathrm{Var}\left(\gamma_j^*\right) = \frac{\phi\gamma_j}{\mu_{[I-j]}}.$$

Hence the off-diagonal elements and the second term in (4.9) come from the side constraint that the estimators have to sum up to one.

4.1.4 Conditional MSEP

In this section we derive estimates for the conditional MSEP of the BF predictors $\widehat{C}_{i,J} = C_{i,I-i} + \hat{\mu}_i(1 - \hat{\beta}_{I-i})$ and $\sum_{i=I-J+1}^{I} \widehat{C}_{i,J}$ under Model Assumptions 4.1. Let $\hat{\boldsymbol{\mu}} = (\hat{\mu}_0, \ldots, \hat{\mu}_I)$ and let $\mathcal{I}_I = \sigma(\mathcal{D}_I, \hat{\boldsymbol{\mu}})$ denote the σ-field containing the information at time I. Due to the $\mathcal{I}_I$-measurability of $\hat{\mu}_i$ and $\hat{\beta}_{I-i}$ we have

$$\begin{aligned}
\mathrm{msep}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) &= E\left[\left(\sum_{j=I-i+1}^{J} X_{i,j} - \hat{\mu}_i(1-\hat{\beta}_{I-i}) \right)^2 \middle| \mathcal{I}_I \right] \\
&= E\left[\left(\sum_{j=I-i+1}^{J} X_{i,j} - \mu_i\gamma_j \right)^2 \middle| \mathcal{I}_I \right] + \left(\sum_{j=I-i+1}^{J} \mu_i\gamma_j - \hat{\mu}_i(1-\hat{\beta}_{I-i}) \right)^2.
\end{aligned}$$

With the independence of the $X_{i,j}$ and the definition $\sum_{j=I-i+1}^{J} \gamma_j = 1 - \beta_{I-i}$ we obtain

$$\operatorname{msep}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) = \underbrace{\sum_{j=I-i+1}^{J} \operatorname{Var}(X_{i,j})}_{\text{Process Variance (PV}_i\text{)}} + \underbrace{\Big(\hat{\mu}_i(1-\hat{\beta}_{I-i}) - \mu_i(1-\beta_{I-i})\Big)^2}_{\text{Estimation Error (EE}_i\text{)}}.$$

The process variance reflects the randomness of the $X_{i,j}$ and the estimation error describes the uncertainty in the parameter estimates.

Process Variance. For the estimation of the process variance recall that $\operatorname{Var}(X_{i,j}) = \phi\mu_i\gamma_j$ according to Model Assumptions 4.1. If we plug in the parameter estimates we obtain the estimator

$$\widehat{\mathrm{PV}}_i = \sum_{j=I-i+1}^{J} \widehat{\operatorname{Var}}(X_{i,j}) = \sum_{j=I-i+1}^{J} \hat{\phi}\hat{\mu}_i\hat{\gamma}_j = \hat{\phi}\hat{\mu}_i(1-\hat{\beta}_{I-i}).$$

Estimation Error. Given $\mathcal{I}_I$, the term $\Big(\hat{\mu}_i(1-\hat{\beta}_{I-i}) - \mu_i(1-\beta_{I-i})\Big)^2$ is an unknown constant. In order to estimate this constant we rewrite it as follows

$$\begin{aligned}\Big(\hat{\mu}_i(1-\hat{\beta}_{I-i}) - \mu_i(1-\beta_{I-i})\Big)^2 &= \Big((\hat{\mu}_i-\mu_i)(1-\hat{\beta}_{I-i}) + \mu_i(\beta_{I-i}-\hat{\beta}_{I-i})\Big)^2 \\ &= (1-\hat{\beta}_{I-i})^2(\hat{\mu}_i-\mu_i)^2 + \mu_i^2(\beta_{I-i}-\hat{\beta}_{I-i})^2 \\ &\quad + 2(1-\hat{\beta}_{I-i})\mu_i(\hat{\mu}_i-\mu_i)(\beta_{I-i}-\hat{\beta}_{I-i}),\end{aligned}$$

and estimate the unknown terms $(\hat{\mu}_i-\mu_i)^2$, $(\beta_{I-i}-\hat{\beta}_{I-i})^2$ and $(\hat{\mu}_i-\mu_i)(\beta_{I-i}-\hat{\beta}_{I-i})$ by taking the mean over all possible $\mathcal{I}_I$, that is, we study the fluctuation of $\hat{\mu}_i$ and $\hat{\beta}_{I-i}$ around the true values μ_i and β_{I-i}, respectively. If we neglect the dependencies of the $\hat{\mu}_i$ and the $\hat{\beta}_j$ we obtain with the unbiasedness of the $\hat{\mu}_i$ and the (approximate) asymptotic unbiasedness of the $\hat{\beta}_j$ the following estimator for the estimation error

$$\widehat{\mathrm{EE}}_i = (1-\hat{\beta}_{I-i})^2\widehat{\operatorname{Var}}(\hat{\mu}_i) + \hat{\mu}_i^2\widehat{\operatorname{Var}}(\hat{\beta}_{I-i}),$$

where $\widehat{\operatorname{Var}}(\hat{\mu}_i)$ and $\widehat{\operatorname{Var}}(\hat{\beta}_{I-i})$ are estimates for the variances $\operatorname{Var}(\hat{\mu}_i)$ and $\operatorname{Var}(\hat{\beta}_{I-i})$, respectively. The variance of the $\hat{\beta}_{I-i}$ is estimated using the Fisher information matrix

$$\widehat{\operatorname{Var}}(\hat{\beta}_{I-i}) = \sum_{0\le j,k\le I-i} \widehat{\operatorname{Cov}}(\hat{\gamma}_j, \hat{\gamma}_k), \tag{4.10}$$

where the covariances $\widehat{\operatorname{Cov}}(\hat{\gamma}_j, \hat{\gamma}_k)$ are given in (4.9).

It remains to estimate the variances of the a priori estimates $\hat{\mu}_i$. An actuary being able to deliver a priori estimates $\hat{\mu}_i$ is often also able to make a statement about the precision of these estimates. In this case beside the $\hat{\mu}_i$'s there are also given a priori estimates of $\mathrm{Var}(\hat{\mu}_i)$. If this is not the case we show how $\mathrm{Var}(\hat{\mu}_i)$ can be estimated from the data. To this end we assume that the $\hat{\mu}_i$ have a constant coefficient of variation $\mathrm{CoVa}(\hat{\mu}_i) \equiv c$ and that the correlation $\mathrm{Corr}(\hat{\mu}_i, \hat{\mu}_k)$ can be estimated by

$$\widehat{\mathrm{Corr}}(\hat{\mu}_i, \hat{\mu}_k) = \begin{cases} \frac{n-|i-k|}{n}, & 0 \le i,k \le I \text{ and } |i-k| < n, \\ 0, & \text{else}, \end{cases} \tag{4.11}$$

where $0 \le n \le I$.

Remark 4.7. In the case where the premiums ν_i and the variances of the ultimate loss ratios are constant, we obtain the above formula for the correlation $\widehat{\mathrm{Corr}}(\hat{\mu}_i, \hat{\mu}_k)$ if the a priori loss ratios are estimated with a moving average of the ultimate loss ratios over n years (cf. Appendix A). Observe that we do not assume that the a priori estimates $\hat{\mu}_i$ are independent.

In the numerical examples we will use formula (4.11) with $n = 10$, assuming that information more than 10 years ago does not contribute to the new forecast. We define the aggregated claim amount at time I by

$$C_I = \sum_{i=0}^{I-J} C_{i,J} + \sum_{i=I-J+1}^{I} C_{i,I-i}.$$

The corresponding pure risk premium is given by

$$\Pi_I = E[C_I] = \sum_{i=0}^{I-J} \mu_i + \sum_{i=I-J+1}^{I} \beta_{I-i}\mu_i. \tag{4.12}$$

Moreover, we define the observed loss ratio at time I by

$$Q_I = \frac{C_I}{\hat{\Pi}_I} = \frac{C_I}{\Pi_I}\frac{\Pi_I}{\hat{\Pi}_I},$$

where $\hat{\Pi}_I$ is obtained from (4.12) by replacing all μ_i by the a priori estimates $\hat{\mu}_i$. With a first order Taylor approximation around Π_I we obtain

$$E[\Pi_I/\hat{\Pi}_I] \approx 1 \quad \text{and} \quad \mathrm{Var}(\Pi_I/\hat{\Pi}_I) \approx \mathrm{CoVa}^2(\hat{\Pi}_I),$$

and therefore with the independence of $\hat{\Pi}_I$ and C_I we conclude that $E[Q_I] \approx 1$. For the variance $\mathrm{Var}(Q_I)$ we use the following general decomposition for the product of independent random variables X and Y

$$\mathrm{Var}(XY) = E[X]^2\mathrm{Var}(Y) + E[Y]^2\mathrm{Var}(X) + \mathrm{Var}(X)\mathrm{Var}(Y).$$

Henceforth we have with the Taylor approximation

$$\mathrm{Var}(Q_I) \approx \mathrm{Var}(C_I)/\Pi_I^2 + \mathrm{CoVa}^2(\hat{\Pi}_I) + \mathrm{Var}(C_I/\Pi_I)\mathrm{Var}(\Pi_I/\hat{\Pi}_I).$$

Neglecting the last term and asserting positivity we arrive at the following estimator

$$\widehat{\mathrm{CoVa}}^2(\hat{\Pi}_I) = \max\{0, \widehat{\mathrm{Var}}(Q_I) - \widehat{\mathrm{Var}}(C_I)/\widehat{\hat{\Pi}}_I^2\},$$

where

$$\begin{aligned}
\widehat{\hat{\Pi}}_I &= \sum_{i=0}^{I-J} \hat{\mu}_i + \sum_{i=I-J+1}^{I} \hat{\beta}_{I-i}\hat{\mu}_i, \\
\widehat{\mathrm{Var}}(Q_I) &= (\hat{Q}_I - 1)^2, \quad \text{with} \quad \hat{Q}_I = C_I/\widehat{\hat{\Pi}}_I, \\
\widehat{\mathrm{Var}}(C_I) &= \sum_{i=0}^{I-J} \hat{\mu}_i\, \widehat{\sigma}^2_{[J]} + \sum_{i=I-J+1}^{I} \hat{\mu}_i\, \widehat{\sigma}^2_{[I-i]},
\end{aligned}$$

with $\widehat{\sigma}^2_{[j]} = \hat{\phi}\hat{\beta}_j$ in the ODP Model. Note that if the $\hat{\mu}_i$ were fully correlated then $\mathrm{CoVa}(\hat{\mu}_i) = \mathrm{CoVa}(\hat{\Pi}_I)$. In general we obtain with the calculation provided in Appendix A

$$\begin{aligned}
c^2 &= \mathrm{CoVa}^2(\hat{\mu}_i) = \max\left\{0, \tilde{c}^2\right\}, \quad \text{where} \\
\tilde{c}^2 &= \mathrm{CoVa}^2(\hat{\Pi}_I)\left(1 - \frac{2\sum_{0\le i<k\le I} \beta_{I-i}\beta_{I-k}\mu_i\mu_k\,(1-\mathrm{Corr}(\hat{\mu}_i,\hat{\mu}_k))}{\left(\sum_{i=0}^{I}\mu_i\beta_{I-i}\right)^2}\right)^{-1}
\end{aligned} \tag{4.13}$$

where we set $\beta_j = 1$ if $j > J$. The variance is then estimated by

$$\widehat{\mathrm{Var}}(\hat{\mu}_i) = \hat{c}^2 \cdot \hat{\mu}_i^2, \tag{4.14}$$

where $\hat{c}^2$ is obtained from (4.13) by replacing the unknown parameters and $\mathrm{CoVa}^2(\hat{\Pi}_I)$, $\mathrm{Corr}(\hat{\mu}_i, \hat{\mu}_k)$ by their estimates.
Thus, we have derived the following estimate for the conditional MSEP of $\widehat{C}_{i,J}$:

Estimate 4.8 (MSEP single accident year)**.** *Under Model Assumptions 4.1 the conditional MSEP of the BF predictor $\widehat{C}_{i,J}$ is estimated by*

$$\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) = \sum_{j=I-i+1}^{J} \hat{\phi}\hat{\mu}_i\hat{\gamma}_j + \widehat{\mathrm{Var}}(\hat{\mu}_i)(1-\hat{\beta}_{I-i})^2 + \hat{\mu}_i^2\widehat{\mathrm{Var}}(\hat{\beta}_{I-i}),$$

where $\widehat{\mathrm{Var}}(\hat\beta_{I-i})$ *is given in* (4.10) *and* $\widehat{\mathrm{Var}}(\hat\mu_i)$ *is given in* (4.14). *For the estimates* $\hat\mu_i$, $\hat\gamma_j$ *and* $\hat\phi$ *see Model Assumptions 4.1, equation* (4.3) *and* (4.5), *respectively.*

Our next goal is the derivation of an estimate of the conditional MSEP for aggregated accident years. We have

$$\begin{aligned}\mathrm{msep}_{\sum_{i=I-J+1}^I C_{i,J}|\mathcal{I}_I}\left(\sum_{i=I-J+1}^I \widehat{C}_{i,J}\right) &= E\left[\left(\sum_{i=I-J+1}^I \left(C_{i,J}-\widehat{C}_{i,J}\right)\right)^2\middle|\mathcal{I}_I\right]\\ &= \sum_{i=I-J+1}^I \mathrm{msep}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) + 2\sum_{i<k} E\left[\left(C_{i,J}-\widehat{C}_{i,J}\right)\left(C_{k,J}-\widehat{C}_{k,J}\right)\middle|\mathcal{I}_I\right].\end{aligned}$$

As above using the $\mathcal{I}_I$ measurability of $\hat\beta_j$ and $\hat\mu_i$ we obtain for $i<k$

$$\begin{aligned}E\left[\left(C_{i,J}-\widehat{C}_{i,J}\right)\left(C_{k,J}-\widehat{C}_{k,J}\right)\middle|\mathcal{I}_I\right] &= \mathrm{Cov}\left(\sum_{j=I-i+1}^J X_{i,j}, \sum_{j=I-k+1}^J X_{k,j}\right)\\ &+ \left(\hat\mu_i\left(1-\hat\beta_{I-i}\right)-\mu_i\left(1-\beta_{I-i}\right)\right)\left(\hat\mu_k\left(1-\hat\beta_{I-k}\right)-\mu_k\left(1-\beta_{I-k}\right)\right).\end{aligned}$$

With the independence of the incremental claims $X_{i,j}$ we see that the covariance term is equal to zero and given $\mathcal{I}_I$ the last expression is an unknown constant. In order to estimate this unknown term we proceed as above and define

$$\begin{aligned}\mathrm{EE}_{i,k} &= \left(\hat\mu_i\left(1-\hat\beta_{I-i}\right)-\mu_i\left(1-\beta_{I-i}\right)\right)\left(\hat\mu_k\left(1-\hat\beta_{I-k}\right)-\mu_k\left(1-\beta_{I-k}\right)\right)\\ &= \left((1-\hat\beta_{I-i})(\hat\mu_i-\mu_i)-\mu_i(\hat\beta_{I-i}-\beta_{I-i})\right)\\ &\quad\times\left((1-\hat\beta_{I-k})(\hat\mu_k-\mu_k)-\mu_k(\hat\beta_{I-k}-\beta_{I-k})\right),\end{aligned}$$

which is estimated by

$$\widehat{\mathrm{EE}}_{i,k} = (1-\hat\beta_{I-i})(1-\hat\beta_{I-k})\widehat{\mathrm{Cov}}(\hat\mu_i,\hat\mu_k)+\hat\mu_i\hat\mu_k\widehat{\mathrm{Cov}}(\hat\beta_{I-i},\hat\beta_{I-k}),$$

where

$$\widehat{\mathrm{Cov}}(\hat\mu_i,\hat\mu_k) = \widehat{\mathrm{Corr}}(\hat\mu_i,\hat\mu_k)\sqrt{\widehat{\mathrm{Var}}\left(\hat\mu_i\right)\widehat{\mathrm{Var}}\left(\hat\mu_k\right)}, \tag{4.15}$$

and

$$\widehat{\mathrm{Cov}}\left(\hat\beta_{I-i},\hat\beta_{I-k}\right) = \sum_{j=0}^{I-i}\sum_{l=0}^{I-k}\widehat{\mathrm{Cov}}(\hat\gamma_j,\hat\gamma_l), \quad I-J+1\le i,k\le I, \tag{4.16}$$

(see (4.14) and (4.9)).
Putting everything together we obtain the following estimate for the conditional MSEP:

Estimate 4.9 (MSEP aggregated accident years)**.**
Under Model Assumptions 4.1 the conditional MSEP for aggregated accident years is estimated by

$$\widehat{\operatorname{msep}}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{I}_I}\left(\sum_{i=I-J+1}^{I}\widehat{C}_{i,J}\right) = \sum_{i=I-J+1}^{I}\widehat{\operatorname{msep}}_{C_{i,J}|\mathcal{I}_I}\left(\widehat{C}_{i,J}\right)$$
$$+2\sum_{I-J+1\le i<k\le I}\left((1-\hat\beta_{I-i})(1-\hat\beta_{I-k})\widehat{\operatorname{Cov}}(\hat\mu_i,\hat\mu_k)+\hat\mu_i\hat\mu_k\widehat{\operatorname{Cov}}(\hat\beta_{I-i},\hat\beta_{I-k})\right).$$

Remark 4.10. In the literature one often calculates the conditional MSEP given $\mathcal{D}_I$ instead of $\mathcal{I}_I$. This corresponds to a different interpretation. If the $\hat\mu_i$ are considered as observations one should condition on $\mathcal{I}_I$ otherwise if the $\hat\mu_i$ are considered as a priori information one should condition on $\mathcal{D}_I$. That is, in the latter case we are in a Bayesian framework, where μ_i are random variables with an a priori distribution and $\hat\mu_i$ is the expected value of μ_i. In the first case $\hat\mu_i$ is treated as an 'expert observation'. In the case where we condition on $\mathcal{D}_I$ we obtain for single accident years

$$\operatorname{msep}_{C_{i,J}|\mathcal{D}_I}(\widehat{C}_{i,J})$$
$$=\underbrace{\sum_{j=I-i+1}^{J}\operatorname{Var}(X_{i,j})}_{\mathrm{PV}_i}+\underbrace{E\left[\left(\hat\mu_i(1-\hat\beta_{I-i})-\mu_i(1-\beta_{I-i})\right)^2\middle|\mathcal{D}_I\right]}_{\mathrm{EE}_i}.$$

In order to estimate the last term we consider the mean over all possible $\mathcal{D}_I$ (we could also use this procedure when we condition on $\mathcal{I}_I$ and the resulting estimates differ slightly).
With the approximation $E[\hat\mu_i(1-\hat\beta_{I-i})]\approx\mu_i(1-\beta_{I-i})$ we obtain

$$E\left[\left(\hat\mu_i(1-\hat\beta_{I-i})-\mu_i(1-\beta_{I-i})\right)^2\right]\approx\operatorname{Var}\left(\hat\mu_i(1-\hat\beta_{I-i})\right)$$
$$=E\left[\operatorname{Var}(\hat\mu_i(1-\hat\beta_{I-i})|\hat\mu_i)\right]+\operatorname{Var}\left(E\left[\hat\mu_i(1-\hat\beta_{I-i})\middle|\hat\mu_i\right]\right),$$

and estimating the last term we obtain the following estimate for the estimation error

$$\widehat{\mathrm{EE}}_i=\left(\hat\mu_i^2+\widehat{\operatorname{Var}}(\hat\mu_i)\right)\widehat{\operatorname{Var}}(1-\hat\beta_{I-i})+(1-\hat\beta_{I-i})^2\widehat{\operatorname{Var}}(\hat\mu_i).$$

We therefore have the additional term $\widehat{\operatorname{Var}}(\hat\mu_i)\widehat{\operatorname{Var}}(1-\hat\beta_{I-i})$ in the estimation error. Similarly we have for aggregated accident years

$$\mathrm{EE}_{i,k}$$
$$=E\left[\left(\hat\mu_i(1-\hat\beta_{I-i})-\mu_i\left(1-\beta_{I-i}\right)\right)\left(\hat\mu_k(1-\hat\beta_{I-k})-\mu_k\left(1-\beta_{I-k}\right)\right)\middle|\mathcal{D}_I\right],$$

which is estimated by

$$\begin{aligned}\widehat{\mathrm{EE}}_{i,k} = &\left(1-\hat{\beta}_{I-i}\right)\left(1-\hat{\beta}_{I-k}\right)\widehat{\mathrm{Cov}}(\hat{\mu}_i,\hat{\mu}_k)\\ &+\left(\widehat{\mathrm{Cov}}(\hat{\mu}_i,\hat{\mu}_k)+\hat{\mu}_i\hat{\mu}_k\right)\widehat{\mathrm{Cov}}(\hat{\beta}_{I-i},\hat{\beta}_{I-k}).\end{aligned}$$

The resulting formulas for the conditional MSEP have the same form as the estimates derived in Mack (2008) (but the estimates of the parameters and covariances are, of course, different).

4.2 General Over-dispersed Poisson Model

4.2.1 Model

In Model Assumptions 4.1 we have assumed that the dispersion parameter ϕ is constant over all development years. The assumption of a constant dispersion is quite restrictive and often not appropriate in practical applications. Therefore we consider in this subsection the general case, where the dispersion parameter varies between development years.

Model Assumptions 4.11 (General Over-dispersed Poisson Model)**.**

GP1 Incremental claims $X_{i,j}$ are independent and there exist positive parameters $\phi_0,\ldots,\phi_J$, $\mu_0,\ldots,\mu_I$ and $\gamma_0,\ldots,\gamma_J$ with $\sum_{j=0}^{J}\gamma_j=1$ such that $X_{i,j}/\phi_j \sim \text{Poisson}(\mu_i\gamma_j/\phi_j)$. In particular, we have

$$\begin{aligned}E[X_{i,j}] &= \mu_i\gamma_j,\\ \mathrm{Var}(X_{i,j}) &= \phi_j\mu_i\gamma_j,\end{aligned}$$

where ϕ_j is the dispersion parameter for development year j.

GP2 The a priori estimates $\hat{\mu}_i$ for $\mu_i = E[C_{i,J}]$ are unbiased and independent from $X_{l,j}$ for $0\le l\le I$, $0\le j\le J$.

4.2.2 Estimation of the Development Pattern

As in Section 4.1 we estimate γ_j with the MLE method using Lagrange multipliers. The Lagrange function is given by

$$\mathcal{L}_{\mathcal{D}_I}(\gamma_0,\ldots,\gamma_J,\kappa) = \sum_{i+j\le I}\frac{1}{\phi_j}\left(X_{i,j}\left(\log\mu_i+\log\gamma_j\right)-\mu_i\gamma_j\right)+\kappa\left(1-\sum_{j=0}^{J}\gamma_j\right), \tag{4.17}$$

where κ is the Lagrange multiplier. For $0 \leq j \leq J$ the MLE $\tilde{\gamma}_j$ is the solution of

$$\frac{\partial \mathcal{L}_{\mathcal{D}_I}}{\partial \gamma_j} = \sum_{i=0}^{I-j} \frac{1}{\phi_j} \left(\frac{X_{i,j}}{\gamma_j} - \mu_i \right) - \kappa = 0, \tag{4.18}$$

and hence

$$\tilde{\gamma}_j = \frac{\sum_{i=0}^{I-j} \frac{1}{\phi_j} X_{i,j}}{\sum_{i=0}^{I-j} \frac{1}{\phi_j} \mu_i + \kappa} = \frac{X_{[I-j],j}/\phi_j}{\mu_{[I-j]}/\phi_j + \kappa}. \tag{4.19}$$

Remark 4.12. Note that (4.19) has the same structure as (4.1) with respect to the normalised random variables $Z_{i,j} = X_{i,j}/\phi_j$.

With the side constraint we obtain the following implicit equation for κ

$$\frac{\partial \mathcal{L}_{\mathcal{D}_I}}{\partial \kappa} = 1 - \sum_{j=0}^{J} \frac{X_{[I-j],j}/\phi_j}{\mu_{[I-j]}/\phi_j + \kappa} = 0.$$

It remains to estimate μ_i and ϕ_j.

For μ_i we use the a priori estimates $\hat{\mu}_i$. In order to estimate the dispersion parameters ϕ_j we consider $Y_{i,j} = X_{i,j}/\mu_i$ with

$$E[Y_{i,j}] = \gamma_j \quad \text{and} \quad \operatorname{Var}(Y_{i,j}) = \phi_j \frac{\gamma_j}{\mu_i} = \frac{\sigma_j^2}{\mu_i},$$

where $\sigma_j^2 = \phi_j \gamma_j$. An unbiased estimate for σ_j^2 is given by

$$\tilde{\sigma}_j^2 = \frac{1}{I-j} \sum_{i=0}^{I-j} \mu_i (Y_{i,j} - \bar{Y}_j)^2, \quad 0 \leq j \leq J,\ j \neq I, \tag{4.20}$$

with $\bar{Y}_j = \sum_{i=0}^{I-j} \frac{\mu_i}{\mu_{[I-j]}} Y_{i,j}$. If $J = I$ we use an extrapolation to obtain an estimate for σ_J^2, see Mack (1993). The dispersion parameter ϕ_j can then be estimated by

$$\hat{\phi}_j = \frac{\hat{\sigma}_j^2}{\hat{\gamma}_j}, \tag{4.21}$$

where $\hat{\sigma}_j^2$ is obtained from $\tilde{\sigma}_j^2$ by replacing the unknown μ_i's by the a priori estimates $\hat{\mu}_i$ and where

$$\hat{\gamma}_j = \frac{X_{[I-j],j}/\hat{\phi}_j}{\hat{\mu}_{[I-j]}/\hat{\phi}_j + \hat{\kappa}}, \tag{4.22}$$

with $\hat{\kappa} \in \left(- \min_{0 \le j \le J} \left\{ \frac{\hat{\mu}_{[I-j]}}{\hat{\phi}_j} \right\}, \infty \right)$ given by the implicit equation

$$1 - \sum_{j=0}^{J} \frac{X_{[I-j],j}/\hat{\phi}_j}{\hat{\mu}_{[I-j]}/\hat{\phi}_j + \hat{\kappa}} = 0.$$

Note that (4.21) is an implicit equation since $\hat{\phi}_j$ also appears in the equation for $\hat{\gamma}_j$. If we plug in $\hat{\phi}_j = \hat{\sigma}_j^2/\hat{\gamma}_j$ in (4.22) we obtain

$$\hat{\gamma}_j = \frac{\hat{\gamma}_j X_{[I-j],j}/\hat{\sigma}_j^2}{\hat{\gamma}_j \hat{\mu}_{[I-j]}/\hat{\sigma}_j^2 + \hat{\kappa}},$$

and solving for $\hat{\gamma}_j$ we get

$$\hat{\gamma}_j = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}} - \hat{\kappa} \frac{\hat{\sigma}_j^2}{\hat{\mu}_{[I-j]}}.$$

With the side constraint $\sum_{j=0}^{J} \hat{\gamma}_j = 1$ we arrive at the following explicit solution

$$\hat{\gamma}_j = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}} + \frac{\hat{\sigma}_j^2/\hat{\mu}_{[I-j]}}{\sum_{l=0}^{J} \hat{\sigma}_l^2/\hat{\mu}_{[I-l]}} \left(1 - \sum_{l=0}^{J} \frac{X_{[I-l],l}}{\hat{\mu}_{[I-l]}} \right). \tag{4.23}$$

The cumulative development pattern is estimated by $\hat{\beta}_j = \sum_{k=0}^{j} \hat{\gamma}_k$, $0 \le j \le J$.

Remarks 4.13.

- For the case of a constant dispersion parameter ϕ we cannot use the above estimation procedure for ϕ. In that case $\sigma_j^2/\gamma_j \equiv \phi$ but $\hat{\sigma}_j^2/\hat{\gamma}_j$ might not be constant over all j. Moreover, if ϕ is constant over all development years, it does not make sense to consider the development years j separately for the estimation of ϕ.

- The estimators (4.23) do not fulfil equations (3.5) anymore.

- Contrary to the ODP Model, the MLEs based only on the observed data $\mathcal{D}_I$ and disregarding the μ_i, are no longer identical to the CL forecasts. Instead, the MLEs $\hat{\mu}_i^{MLE}$ and the modified 'development pattern' $\hat{\beta}_j^{mod} = \sum_{k=0}^{j} \hat{\gamma}_k/\phi_k$ are proportional to the CL forecasts and the CL development pattern obtained from the modified data $X_{i,j}/\phi_j$.

- Denote again by
$$\hat{\gamma}_j^{(0)} = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}}$$
the raw estimates. Then we again see how the smoothing from these raw estimates to the adjusted estimates $\hat{\gamma}_j$ looks like. Here the smoothing corrections are additive.

- A big deviation of $\sum_{j=0}^{J} \hat{\gamma}_j^{(0)}$ from 1 could be an indicator that the a priori estimates might be biased.

- A tail development $\hat{\gamma}_{J+1}$ can be incorporated in the same way as described in Remarks 4.6.

4.2.3 Covariance of the Estimated Development Pattern

As in Section 4.1 the covariance matrix of the $\hat{\gamma}_j$'s can be estimated using the Fisher information matrix. The log-likelihood function $l_{\mathcal{D}_I}$ is given by

$$\begin{aligned} l_{\mathcal{D}_I}(\gamma_0, \ldots, \gamma_{J-1}) &= \sum_{\substack{i+j \le I \\ j \le J-1}} \frac{1}{\phi_j} \left(X_{i,j} \left(\log \mu_i + \log \gamma_j \right) - \mu_i \gamma_j \right) \\ &+ \sum_{i=0}^{I-J} \frac{1}{\phi_J} \left(X_{i,J} \left(\log \mu_i + \log \left(1 - \sum_{j=0}^{J-1} \gamma_j \right) \right) - \mu_i \left(1 - \sum_{j=0}^{J-1} \gamma_j \right) \right) + r, \end{aligned} \tag{4.24}$$

where r contains all remaining terms, which do not depend on the parameters $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_{J-1})$. Analogously to (4.6) we obtain for $0 \le j \le J-1$

$$\begin{aligned} H(\boldsymbol{\gamma})_{j,j} &= E\left[\left(\frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_j} \right)^2 \right] \\ &= E\left[\left(\sum_{i=0}^{I-j} \frac{1}{\phi_j} \left(\frac{X_{i,j}}{\gamma_j} - \mu_i \right) - \sum_{i=0}^{I-J} \frac{1}{\phi_J} \left(\frac{X_{i,J}}{\gamma_J} - \mu_i \right) \right)^2 \right] \\ &= \frac{\mu_{[I-j]}}{\phi_j \gamma_j} + \frac{\mu_{[I-J]}}{\phi_J \gamma_J}, \end{aligned}$$

and for $0 \le j < k \le J-1$ we get analogously to (4.7)

$$H(\boldsymbol{\gamma})_{k,j} = H(\boldsymbol{\gamma})_{j,k} = E\left[\frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_j} \frac{\partial l_{\mathcal{D}_I}}{\partial \gamma_k} \right] = \frac{\mu_{[I-J]}}{\phi_J \gamma_J}.$$

The inversion of the Fisher information matrix is provided in Appendix A and we obtain for $0 \leq j,k \leq J-1$ the approximation

$$\mathrm{Cov}(\hat{\gamma}_j, \hat{\gamma}_k) \approx (H(\boldsymbol{\gamma})^{-1})_{j,k} = \frac{\phi_j \gamma_j}{\mu_{[I-j]}} \left(\mathbb{1}_{\{j=k\}} - \frac{\phi_k \gamma_k / \mu_{[I-k]}}{\sum_{l=0}^{J} \phi_l \gamma_l / \mu_{[I-l]}} \right). \quad (4.25)$$

By linearity we obtain analogous formulas for $\hat{\gamma}_J = 1 - \sum_{j=0}^{J-1} \hat{\gamma}_j$. Replacing the unknown parameters $\mu_{[I-j]}$, γ_j and ϕ_j by their estimates $\hat{\mu}_{[I-j]}$ (see Model Assumptions 4.11), $\hat{\gamma}_j$ (see (4.22)) and $\hat{\phi}_j$ (see (4.21)) we obtain the following estimates for the covariances

$$\widehat{\mathrm{Cov}}(\hat{\gamma}_j, \hat{\gamma}_k) = \frac{\hat{\phi}_j \hat{\gamma}_j}{\hat{\mu}_{[I-j]}} \left(\mathbb{1}_{\{j=k\}} - \frac{\hat{\phi}_k \hat{\gamma}_k / \hat{\mu}_{[I-k]}}{\sum_{l=0}^{J} \hat{\phi}_l \hat{\gamma}_l / \hat{\mu}_{[I-l]}} \right), \quad 0 \leq j,k \leq J. \quad (4.26)$$

For $\hat{\beta}_{I-i}$ we get

$$\widehat{\mathrm{Cov}}(\hat{\beta}_{I-i}, \hat{\beta}_{I-k}) = \sum_{j=0}^{I-i} \sum_{l=0}^{I-k} \widehat{\mathrm{Cov}}(\hat{\gamma}_j, \hat{\gamma}_l), \quad I-J+1 \leq i,k \leq I. \quad (4.27)$$

In the next section we give estimates for the conditional MSEP of the BF predictors $\widehat{C}_{i,J} = C_{i,I-i} + \hat{\mu}_i(1-\hat{\beta}_{I-i})$ and $\sum_{i=I-J+1}^{I} \widehat{C}_{i,J}$ under Model Assumptions 4.11.

4.2.4 Conditional MSEP

The derivation of an estimate of the conditional MSEP for the BF predictor $\widehat{C}_{i,J}$ in the General ODP Model is analogous to Section 4.1.4. We obtain the following estimate for single accident years:

Estimate 4.14 (MSEP single accident year)**.**
Under Model Assumptions 4.11 the conditional MSEP of the BF predictor $\widehat{C}_{i,J}$ is estimated by

$$\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) = \sum_{j=I-i+1}^{J} \hat{\phi}_j \hat{\mu}_i \hat{\gamma}_j + \widehat{\mathrm{Var}}(\hat{\mu}_i)(1-\hat{\beta}_{I-i})^2 + \hat{\mu}_i^2 \widehat{\mathrm{Var}}(\hat{\beta}_{I-i}),$$

where $\widehat{\mathrm{Var}}(\hat{\beta}_{I-i})$ is given in (4.27) *and $\widehat{\mathrm{Var}}(\hat{\mu}_i)$ is given in* (4.14)*.*

Using the same ideas as in the derivation of Estimate 4.9 we obtain the following estimate for aggregated accident years:

Estimate 4.15 (MSEP aggregated accident years)**.**
Under Model Assumptions 4.11 the conditional MSEP for aggregated accident years is estimated by

$$\widehat{\operatorname{msep}}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{I}_I}\left(\sum_{i=I-J+1}^{I}\widehat{C}_{i,J}\right) = \sum_{i=I-J+1}^{I}\widehat{\operatorname{msep}}_{C_{i,J}|\mathcal{I}_I}\left(\widehat{C}_{i,J}\right) + 2\sum_{I-J+1\le i<k\le I}\left((1-\hat{\beta}_{I-i})(1-\hat{\beta}_{I-k})\widehat{\operatorname{Cov}}(\hat{\mu}_i,\hat{\mu}_k) + \hat{\mu}_i\hat{\mu}_k\widehat{\operatorname{Cov}}(\hat{\beta}_{I-i},\hat{\beta}_{I-k})\right),$$

where $\widehat{\operatorname{Cov}}(\hat{\beta}_{I-i},\hat{\beta}_{I-k})$ *is given in* (4.27) *and* $\widehat{\operatorname{Cov}}(\hat{\mu}_i,\hat{\mu}_k)$ *is given in* (4.15)*.*

4.3 Normal Model

4.3.1 Model

As mentioned in Remarks 4.2 the ODP Model can only be used if the incremental claims $X_{i,j}$ are positive, which is often appropriate for claims payments. In contrast, incurred losses increments are sometimes negative and therefore an ODP model is not suitable in this case. The following model does not have these restrictions.

Model Assumptions 4.16 (Normal Model)**.**

N1 Incremental claims $X_{i,j}$ are independent and normally distributed and there exist parameters $\mu_0,\dots,\mu_I,\ \gamma_0,\dots,\gamma_J$ with $\sum_{j=0}^{J}\gamma_j = 1$ and $\sigma_0^2,\dots,\sigma_J^2$ such that

$$E[X_{i,j}] = \mu_i\gamma_j,$$
$$\operatorname{Var}(X_{i,j}) = \mu_i\sigma_j^2,$$

where σ_j^2 is strictly positive.

N2 The a priori estimates $\hat{\mu}_i$ of $\mu_i = E[C_{i,J}]$ are unbiased and independent from $X_{l,j}$ for $0 \le l \le I$, $0 \le j \le J$.

In the following we derive the MLEs and the corresponding conditional MSEP under Model Assumptions 4.16.

4.3.2 Estimation of the Development Pattern

As in the ODP Model we first assume that the μ_i are known and calculate the MLEs for the γ_j using Lagrange multipliers. In the Normal Model this

procedure allows us to find explicit solutions for the estimates. The Lagrange function is given by

$$\mathcal{L}_{\mathcal{D}_I}(\gamma_0,\dots,\gamma_J,\kappa) = \sum_{i+j\le I} -\frac{(X_{i,j}-\mu_i\gamma_j)^2}{2\mu_i\sigma_j^2} + \kappa\left(1-\sum_{j=0}^{J}\gamma_j\right),$$

where κ is the Lagrange multiplier. For $0 \le j \le J$ we have

$$\frac{\partial\mathcal{L}_{\mathcal{D}_I}}{\partial\gamma_j} = \sum_{i=0}^{I-j}\frac{X_{i,j}-\mu_i\gamma_j}{\sigma_j^2} - \kappa = 0,$$

and therefore

$$\tilde{\gamma}_j = \frac{X_{[I-j],j}-\kappa\sigma_j^2}{\mu_{[I-j]}}. \tag{4.28}$$

With the side constraint

$$1 = \sum_{j=0}^{J}\tilde{\gamma}_j = \sum_{j=0}^{J}\frac{X_{[I-j],j}}{\mu_{[I-j]}} - \sum_{j=0}^{J}\kappa\frac{\sigma_j^2}{\mu_{[I-j]}},$$

it follows that

$$\kappa = \frac{\sum_{j=0}^{J}\frac{X_{[I-j],j}}{\mu_{[I-j]}} - 1}{\sum_{j=0}^{J}\frac{\sigma_j^2}{\mu_{[I-j]}}}.$$

We insert κ in equation (4.28) and obtain the MLE for γ_j

$$\tilde{\gamma}_j = \frac{X_{[I-j],j}}{\mu_{[I-j]}} + \frac{\sigma_j^2/\mu_{[I-j]}}{\sum_{l=0}^{J}\sigma_l^2/\mu_{[I-l]}}\left(1-\sum_{l=0}^{J}\frac{X_{[I-l],l}}{\mu_{[I-l]}}\right). \tag{4.29}$$

The μ_i and σ_j^2 in formula (4.29) are unknown and have to be replaced by estimators. For μ_i we insert $\hat{\mu}_i$ and σ_j^2 can be estimated analogously as in the General ODP Model, that is, by replacing the μ_i in formula (4.20) by $\hat{\mu}_i$.

Remark 4.17. Using MLEs for σ_j^2 results in a system of equations for γ_j and σ_j that is only iteratively solvable.

We arrive at the final estimate

$$\hat{\gamma}_j = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}} + \frac{\hat{\sigma}_j^2/\hat{\mu}_{[I-j]}}{\sum_{l=0}^{J}\hat{\sigma}_l^2/\hat{\mu}_{[I-l]}}\left(1-\sum_{l=0}^{J}\frac{X_{[I-l],l}}{\hat{\mu}_{[I-l]}}\right), \tag{4.30}$$

which is exactly the same estimator as in the General ODP Model (see (4.23)). The cumulative development pattern is estimated by $\hat{\beta}_j = \sum_{k=0}^{j}\hat{\gamma}_k$, $0 \le j \le J$.

Remark 4.18. A tail development $\hat{\gamma}_{J+1}$ can be incorporated as described in Remarks 4.6.

4.3.3 Covariance Matrix of the Estimated Development Pattern

Since the estimated development pattern is the same as in the General ODP Model we could use the estimated covariance matrix of the $\hat{\gamma}_j$ derived in Section 4.2.3. However, the $\hat{\gamma}_j$ are estimated with the MLE method and for consistency with the derivations in the other models we use the Fisher information matrix to estimate the covariance matrix of the $\hat{\gamma}_j$. The likelihood function for the Normal Model is given by

$$\begin{aligned} &l_{\mathcal{D}_I}(\gamma_0, \ldots, \gamma_{J-1}) \\ &= \sum_{\substack{i+j \leq I \\ j \leq J-1}} -\frac{(X_{i,j} - \mu_i \gamma_j)^2}{2\mu_i \sigma_j^2} - \sum_{i=0}^{I-J} \frac{\left(X_{i,J} - \mu_i \left(1 - \sum_{j=0}^{J-1} \gamma_j\right)\right)^2}{2\mu_i \sigma_j^2} + r, \end{aligned}$$

where r contains all remaining terms, which do not depend on the parameters $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_{J-1})$. For the entries of the Fisher information matrix we get analogously to Section 4.1.2

$$H(\boldsymbol{\gamma})_{j,k} = \frac{\mu_{[I-j]}}{\sigma_j^2} \left(\mathbb{1}_{\{j=k\}} + \frac{\mu_{[I-J]}/\sigma_J^2}{\mu_{[I-j]}/\sigma_j^2} \right),$$

where $0 \leq j, k \leq J-1$. With the calculation provided in Appendix A we obtain the entries of the inverse of the Fisher information matrix and for $0 \leq j, k \leq J-1$ we get the approximations

$$\operatorname{Cov}(\hat{\gamma}_j, \hat{\gamma}_k) \approx \left(H(\boldsymbol{\gamma})^{-1}\right)_{j,k} = \frac{\sigma_j^2}{\mu_{[I-j]}} \left(\mathbb{1}_{\{j=k\}} - \frac{\sigma_k^2/\mu_{[I-k]}}{\sum_{l=0}^{J} \sigma_l^2/\mu_{[I-l]}} \right). \tag{4.31}$$

By linearity we obtain the same formulas for the variance and covariances of $\hat{\gamma}_J$. By replacing the unknown parameters $\mu_{[I-j]}$ and σ_j^2 by their estimates we obtain exactly the same formulas as in the General ODP Model, that is,

$$\widehat{\operatorname{Cov}}(\hat{\gamma}_j, \hat{\gamma}_k) = \frac{\hat{\sigma}_j^2}{\hat{\mu}_{[I-j]}} \left(\mathbb{1}_{\{j=k\}} - \frac{\hat{\sigma}_k^2/\hat{\mu}_{[I-k]}}{\sum_{l=0}^{J} \hat{\sigma}_l^2/\hat{\mu}_{[I-l]}} \right), \quad 0 \leq j, k \leq J. \tag{4.32}$$

The covariances of the $\hat{\beta}_{I-i}$ are estimated by

$$\widehat{\operatorname{Cov}}\left(\hat{\beta}_{I-i}, \hat{\beta}_{I-k}\right) = \sum_{j=0}^{I-i} \sum_{l=0}^{I-k} \widehat{\operatorname{Cov}}\left(\hat{\gamma}_j, \hat{\gamma}_l\right), \quad I-J+1 \leq i, k \leq I. \tag{4.33}$$

Remark 4.19. The coincidence of the estimated covariance matrices in the General ODP Model and the Normal Model is of course meaningful, however not obvious because of the approximations used.

In the next section we give an estimate for the conditional MSEP of the BF predictor in the Normal Model.

4.3.4 Conditional MSEP

Since the estimates in the Normal Model coincide with the estimates in the General ODP Model we also obtain the same estimates for the conditional MSEP:

Estimate 4.20 (MSEP single accident year)**.**
Under Model Assumptions 4.16 the conditional MSEP of the BF predictor $\widehat{C}_{i,J}$ *is estimated by*

$$\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{I}_I}(\widehat{C}_{i,J}) = \sum_{j=I-i+1}^{J} \hat{\mu}_i \hat{\sigma}_j^2 + \widehat{\mathrm{Var}}(\hat{\mu}_i)(1-\hat{\beta}_{I-i})^2 + \hat{\mu}_i^2 \widehat{\mathrm{Var}}(\hat{\beta}_{I-i}),$$

where $\widehat{\mathrm{Var}}(\hat{\beta}_{I-i})$ *is given in* (4.33) *and* $\widehat{\mathrm{Var}}(\hat{\mu}_i)$ *is given in* (4.14)*.*

For aggregated accident years we obtain in the Normal Model:

Estimate 4.21 (MSEP aggregated accident years)**.**
Under Model Assumptions 4.16 the conditional MSEP for aggregated accident years is estimated by

$$\begin{aligned}\widehat{\mathrm{msep}}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{I}_I}\left(\sum_{i=I-J+1}^{I} \widehat{C}_{i,J}\right) &= \sum_{i=I-J+1}^{I} \widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{I}_I}\left(\widehat{C}_{i,J}\right) \\ &+ 2 \sum_{I-J+1 \le i < k \le I} \left((1-\hat{\beta}_{I-i})(1-\hat{\beta}_{I-k})\widehat{\mathrm{Cov}}(\hat{\mu}_i, \hat{\mu}_k) + \hat{\mu}_i \hat{\mu}_k \widehat{\mathrm{Cov}}(\hat{\beta}_{I-i}, \hat{\beta}_{I-k})\right),\end{aligned}$$

where $\widehat{\mathrm{Cov}}(\hat{\beta}_{I-i}, \hat{\beta}_{I-k})$ *is given in* (4.33) *and* $\widehat{\mathrm{Cov}}(\hat{\mu}_i, \hat{\mu}_k)$ *is given in* (4.15)*,* $I-J+1 \le i < k \le I$*.*

Remark 4.22. Analogous estimators can be derived in the more general Tweedie's exponential dispersion family models, similar to Alai-Wüthrich (2009).

5 Conclusions and Remarks

In this paper we have investigated the question how the development pattern should be estimated in the BF method and we have derived the corresponding MSEP of the ultimate claim prediction. For this purpose we have considered three distributional models. For these models we have been able to find estimators for the development pattern γ_j which are consistent with the BF philosophy. Moreover, we have found formulas for smoothing from the raw

estimates $\hat{\gamma}_j^{(0)}$ to the final estimates $\hat{\gamma}_j$ and we have been able to find explicit formulas for the correlation matrix of these estimates in terms of the inverse Fisher information matrix.

The ODP Model is presumably not an adequate model for most practical cases. The General ODP Model is a reasonable model for claims payments modelling and the Normal Model can be used for incurred claims studies. For the latter two models we have found the same estimators. But also in the case where the distributional assumptions are not fully satisfied, we suggest applying these estimators because currently there are no estimators available from which we know that they perform better. Therefore we suggest

- to estimate the development pattern by means of formula (4.30)
- to estimate the correlation matrix of these estimates by means of formula (4.32)
- to estimate the conditional MSEP of the ultimate claim by means of Estimate 4.20 (single accident year) and Estimate 4.21 (aggregated accident years).

6 Numerical Examples

The data for the numerical examples are from a Swiss insurance company and for confidentiality purposes the figures are scaled with a constant. The a priori estimates are obtained from pricing. To be more precise $\hat{\mu}_i$ corresponds to the initial forecast of the expected ultimate claim at the end of year $i-1$ and there is not done any repricing afterwards. We consider claims payments data from industrial property insurance and incurred losses data from motor liability insurance. Industrial property insurance is a short tailed line of business meaning that the development is usually finished after short time. On the contrary, motor liability is a long tailed line of business, that is, we have longer settlement periods.

For comparison we also give the results obtained with the CL method. The conditional MSEP for the CL method is calculated according to the distribution-free model by Mack (1993).

6.1 Industrial Property Insurance

Let us first consider data from industrial property, which is given in Table 6 in Appendix B. We apply the estimators suggested in Section 5 and refer to them as BF in the following tables. In this example we additionally give

the results obtained in the ODP Model (BF ODP) and the results obtained using Mack's model (2008) (BF Mack). With the trapezoid of claims payments we obtain estimated coefficients of variation of the a priori estimates of $\widehat{\mathrm{CoVa}}(\hat{\mu}_i) = 5.25\%$ in the ODP Model and $\widehat{\mathrm{CoVa}}(\hat{\mu}_i) = 4.56\%$ for BF. The resulting development patterns and the reserves are given in Tables 1 and 2. The development patterns are obtained from formulas (4.3) and (4.30) without manual smoothing. Note that all three BF development patterns are close to the CL development pattern. Moreover, we calculated the 'usual' BF reserves, which are obtained with the BF method using the CL development pattern. Due to the rather small value on the diagonal in the newest accident year the BF reserves are higher than the CL reserves. In Table 3 the conditional MSEP and the corresponding coefficient of variation are given. In BF Mack we applied the coefficient of variation $\widehat{\mathrm{CoVa}}(\hat{\mu}_i) = 4.56\%$ from the Normal Model. The process standard deviation and estimation error are given in Table 4. Note that the difference between the conditional MSEP in the ODP Model and in BF in the newest accident year comes from the process variance. More precisely, it is due to the rather small value $\hat{\phi}\hat{\gamma}_1 = 187$ compared to $\hat{\sigma}_1^2 = 323$ appearing in the corresponding process variances. Similarly we have $\hat{s}_j^2 = 347$ for the corresponding term in Mack's model. The assumption of a constant dispersion parameter seems therefore questionable.

The results of BF are very close to the results of BF Mack, especially for older accident years. For newer accident years the process variance term contains more parameter estimates $\hat{\sigma}_j^2$ and $\hat{s}_j^2$, respectively. The differences in the estimation error are mainly due to the different estimation of the variances $\widehat{\mathrm{Var}}(\hat{\beta}_{I-i})$. The additional term $\widehat{\mathrm{Var}}(\hat{\mu}_i)\widehat{\mathrm{Var}}(\hat{\beta}_{I-i})$ in Mack's formula for the estimation error is negligible compared to the other terms.

	dev. pattern $\hat{\beta}_{I-i}$			
i	BF ODP	BF	BF Mack	CL
9	99.77%	99.78%	99.77%	99.78%
10	99.55%	99.57%	99.56%	99.59%
11	99.25%	99.29%	99.27%	99.29%
12	98.45%	98.48%	98.47%	98.50%
13	94.08%	94.24%	94.13%	94.14%
14	60.21%	60.59%	60.31%	60.40%

Table 1: Industrial property, claims payments: estimated development pattern.

	estimated reserves				
i	BF ODP	BF	BF Mack	CL	'usual' BF
9	268	257	261	230	246
10	505	481	492	290	467
11	766	731	751	636	725
12	1'501	1'468	1'479	1'313	1'454
13	5'830	5'677	5'786	5'946	5'774
14	38'611	38'240	38'520	34'502	38'426
total	**47'481**	**46'854**	**47'288**	**42'916**	**47'091**

Table 2: Industrial property, claims payments: estimated reserves.

	$\text{msep}^{1/2}$				CoVa			
i	BF ODP	BF	BF Mack	CL	BF ODP	BF	BF Mack	CL
9	410	373	373	341	152.6%	145.1%	143.3%	148.3%
10	560	435	435	325	110.9%	90.3%	88.4%	112.1%
11	685	508	509	457	89.3%	69.5%	67.8%	72.0%
12	953	1'097	1'099	1'064	63.5%	74.7%	74.3%	81.0%
13	1'886	1'861	1'876	1'946	32.3%	32.8%	32.4%	32.7%
14	5'133	6'257	6'516	6'073	13.3%	16.4%	16.9%	17.6%
total	**5'875**	**6'829**	**6'988**	**6'587**	**12.4%**	**14.6%**	**14.8%**	**15.3%**

Table 3: Industrial property, claims payments: estimated conditional $\text{MSEP}^{1/2}$ and coefficient of variation.

	process std. dev.				$(\text{estimation error})^{1/2}$			
i	BF ODP	BF	BF Mack	CL	BF ODP	BF	BF Mack	CL
9	385	351	351	323	139	126	127	111
10	529	410	410	313	185	146	146	86
11	651	483	483	438	211	160	160	133
12	911	1'053	1'053	1'024	279	310	312	286
13	1'796	1'777	1'788	1'869	575	554	567	542
14	4'622	5'874	6'067	5'885	2'232	2'156	2'377	1'501
total	**5'126**	**6'268**	**6'453**	**6'291**	**2'871**	**2'710**	**2'682**	**1'952**

Table 4: Industrial property, claims payments: estimated process standard deviation and square root of the estimation error.

6.2 Motor Liability Insurance

The data from motor liability are given in Table 7 and Table 8 in Appendix B. Note that the observations $C_{0,0}$, $C_{1,0}$ and $C_{0,1}$ are missing. But Table 8 is not really an incomplete triangle because $C_{0,j}$, $j \geq 2$ and $C_{1,j}$, $j \geq 1$ contain all claims of accident years 0 and 1, respectively, that is, also the closed one's. The results obtained with the data from Table 8 are given in Table 5. We apply the estimators suggested in Section 5 and obtain the estimate $\widehat{\text{CoVa}}(\hat{\mu}_i) = 9.47\%$. The reserves are calculated using formula (2.2) with the diagonal claims payments given in Table 7. The BF reserves are

rather high compared to the CL and to the 'usual' BF reserves. Interestingly, the reserves obtained by the 'usual' BF method using the CL development pattern are closer to the CL reserves and even slightly lower. This means that the reason for the higher reserves obtained with BF cannot be conservative a priori estimates $\hat{\mu}_i$, but is rather the new BF consistent estimate of the development pattern. This example shows, that the way how the development pattern is estimated in the BF method can have a big impact on the resulting reserves. As already pointed out previously, the usual way of simply using the CL development pattern is not consistent with the BF philosophy.

	dev. pattern $\hat{\beta}_{I-i}$		estimated reserves			$\text{msep}^{1/2}$		CoVa	
i	BF	CL	BF	CL	'usual' BF	BF	CL	BF	CL
5	99.98%	100.01%	7'236	7'130	7'130	948	926	13.1%	13.0%
6	99.90%	99.93%	11'504	11'337	11'358	1'159	1'100	10.1%	9.7%
7	100.08%	100.12%	14'520	14'341	14'326	1'322	1'283	9.1%	8.9%
8	100.05%	100.12%	9'141	8'837	8'804	1'952	1'914	21.4%	21.7%
9	99.80%	99.93%	22'902	22'274	22'297	2'931	2'847	12.8%	12.8%
10	99.80%	99.94%	26'574	25'870	25'888	3'166	3'079	11.9%	11.9%
11	99.66%	99.83%	39'181	38'354	38'357	3'638	3'690	9.3%	9.6%
12	99.71%	99.90%	42'358	41'487	41'519	3'687	3'599	8.7%	8.7%
13	99.54%	99.78%	43'398	42'296	42'303	4'500	4'553	10.4%	10.8%
14	99.49%	99.79%	64'401	63'116	63'092	5'126	5'275	8.0%	8.4%
15	99.57%	99.93%	71'833	70'258	70'253	5'765	5'897	8.0%	8.4%
16	99.29%	99.69%	90'656	88'897	88'864	6'373	6'549	7.0%	7.4%
17	98.91%	99.44%	102'764	100'412	100'425	7'707	7'824	7.5%	7.8%
18	98.72%	99.41%	90'343	87'150	87'332	9'092	8'906	10.1%	10.2%
19	98.67%	99.53%	142'365	138'381	138'445	10'693	10'738	7.5%	7.8%
20	98.36%	99.60%	142'585	136'460	136'744	13'654	12'774	9.6%	9.4%
21	98.40%	100.04%	126'460	118'925	118'880	15'938	14'291	12.6%	12.0%
22	98.48%	100.48%	125'213	116'434	115'704	18'159	15'146	14.5%	13.0%
23	99.69%	102.21%	153'111	144'103	141'922	19'740	17'725	12.9%	12.3%
24	100.66%	104.49%	205'725	191'687	188'956	24'655	23'581	12.0%	12.3%
25	102.59%	107.98%	262'721	243'520	240'382	28'535	28'107	10.9%	11.5%
total			**1'794'990**	**1'711'269**	**1'702'980**	**61'926**	**60'628**	**3.4%**	**3.5%**

Table 5: Motor liability, incurred losses: estimated development pattern, reserves, conditional $\text{MSEP}^{1/2}$ and coefficient of variation.

Appendix

A Proofs

Inversion of the Fisher Information Matrix. In order to calculate the inverse of the Fisher information matrix in (4.8), (4.25) and (4.31) we need to invert a matrix of the form

$$A = \begin{pmatrix} a_0+1 & 1 & 1 & \cdots & 1 \\ 1 & a_1+1 & 1 & \cdots & 1 \\ 1 & 1 & a_2+1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & a_{J-1}+1 \end{pmatrix},$$

where $a_0, a_1, \ldots, a_{J-1} \in \mathbb{R} \setminus \{0\}$. We number the rows and columns of A starting from 0 and denote the entry in row j and column k by $A_{j,k}$, $0 \leq j,k \leq J-1$, that is, $A_{j,k} = 1 + a_j \mathbb{1}_{\{j=k\}}$. We claim that

$$\left(A^{-1}\right)_{j,k} = \frac{1}{a_j}\left(\mathbb{1}_{\{j=k\}} - \frac{a_0}{d a_k}\right), \quad 0 \leq j,k \leq J-1,$$

where $d = a_0\left(1 + \sum_{j=0}^{J-1} \frac{1}{a_j}\right)$.

Proof. For $0 \leq j,k \leq J-1$ we have

$$\begin{aligned}
(A \cdot A^{-1})_{j,k} &= \sum_{l=0}^{J-1} \left(1 + a_j \cdot \mathbb{1}_{\{j=l\}}\right) \frac{1}{a_l}\left(\mathbb{1}_{\{l=k\}} - \frac{a_0}{d a_k}\right) \\
&= \frac{1}{a_k} - \sum_{l=0}^{J-1} \frac{a_0}{a_l d a_k} + \mathbb{1}_{\{j=k\}} - \frac{a_0}{d a_k} \\
&= \mathbb{1}_{\{j=k\}} + \frac{1}{a_k}\left(1 - \frac{a_0}{d}\left(\sum_{l=0}^{J-1} \frac{1}{a_l} + 1\right)\right) \\
&= \mathbb{1}_{\{j=k\}},
\end{aligned}$$

where we used the definition of d in the last equation.

□

Proof of Equation (4.13)**.** With $\beta_j = 1$ for $j > J$ we have

$$\begin{aligned}
\mathrm{CoVa}^2(\hat{\Pi}_I) &= \mathrm{CoVa}^2\left(\sum_{i=0}^{I}\hat{\mu}_i\beta_{I-i}\right) = \frac{\mathrm{Var}\left(\sum_{i=0}^{I}\hat{\mu}_i\beta_{I-i}\right)}{\left(\sum_{i=0}^{I}\mu_i\beta_{I-i}\right)^2} \\
&= \frac{\sum_{i=0}^{I}\beta_{I-i}^2\mathrm{Var}(\hat{\mu}_i) + 2\sum_{i<k}\beta_{I-i}\beta_{I-k}\mathrm{Corr}(\hat{\mu}_i,\hat{\mu}_k)\mu_i\mathrm{CoVa}(\hat{\mu}_i)\mu_k\mathrm{CoVa}(\hat{\mu}_k)}{\left(\sum_{i=0}^{I}\mu_i\beta_{I-i}\right)^2} \\
&= \frac{c^2\left(\sum_{i=0}^{I}\beta_{I-i}^2\mu_i^2 + 2\sum_{i<k}\beta_{I-i}\beta_{I-k}\mu_i\mu_k\mathrm{Corr}(\hat{\mu}_i,\hat{\mu}_k)\right)}{\left(\sum_{i=0}^{I}\mu_i\beta_{I-i}\right)^2} \\
&= c^2\frac{\left(\left(\sum_{i=0}^{I}\beta_{I-i}\mu_i\right)^2 - 2\sum_{i<k}\beta_{I-i}\beta_{I-k}\mu_i\mu_k\left(1-\mathrm{Corr}(\hat{\mu}_i,\hat{\mu}_k)\right)\right)}{\left(\sum_{i=0}^{I}\mu_i\beta_{I-i}\right)^2}.
\end{aligned}$$

□

Proof of Remark 4.7. Estimating the a priori loss ratios with a moving average of the ultimate loss ratios over n years and assuming that $\nu_i \equiv \nu$ and $\mathrm{Var}(U_i/\nu) = \sigma^2/\nu^2$ yields the following formulas

$$\hat{\mu}_i = \nu\hat{q}_i = \frac{\sum_{l=i}^{n+i-1}U_l}{n}, \quad 0 \le i \le I,$$

where U_i denotes the ultimate claim amount for accident year i. For $i \le k$ it follows with the independence of accident years

$$\mathrm{Cov}(\hat{\mu}_i,\hat{\mu}_k) = \frac{1}{n^2}\mathrm{Cov}\left(\sum_{l=i}^{n+i-1}U_l, \sum_{l=k}^{n+k-1}U_l\right) = \frac{1}{n^2}\sum_{l=k}^{n+i-1}\sigma^2 = \frac{n+i-k}{n^2}\sigma^2,$$

and since $\mathrm{Var}(\hat{\mu}_i) = \sigma^2/n$ we obtain

$$\mathrm{Corr}(\hat{\mu}_i,\hat{\mu}_k) = \frac{n-|i-k|}{n^2\sigma^2/n}\sigma^2 = \frac{n-|i-k|}{n}.$$

□

B Data

i/j	0	1	2	3	4	5	ultimate	$\hat{\mu}_i$
0	52'572	76'651	80'044	80'524	80'870	81'459	81'587	81'552
1	58'623	89'190	94'040	95'592	95'637	95'765	95'898	87'138
2	71'086	108'235	110'410	110'917	110'883	111'092	111'049	100'276
3	58'236	86'079	91'586	90'303	90'490	90'507	90'372	99'319
4	66'661	108'829	113'347	114'785	115'656	115'756	116'481	102'035
5	56'059	90'688	96'389	96'661	97'015	97'160	97'542	100'963
6	52'443	87'856	91'063	91'846	92'414	92'855	92'920	101'178
7	67'307	102'881	107'783	108'279	108'644	108'844	109'599	102'764
8	67'829	98'815	102'008	102'374	102'775	102'868	102'792	111'570
9	69'259	100'684	104'879	106'717	106'602	106'668		114'284
10	41'714	66'880	69'390	69'697	69'869			113'055
11	54'717	82'924	86'781	89'270				102'519
12	46'429	79'564	86'174					96'879
13	55'001	95'511						98'517
14	52'630							97'041

Table 6: Industrial property, cumulative payments and a priori estimates $\hat{\mu}_i$.

i	$C^{paid}_{i,I-i}$	$\hat{\mu}_i$
0	268'392	242'359
1	286'310	270'229
2	272'781	309'063
3	341'679	337'683
4	337'137	348'026
5	381'388	398'224
6	399'724	439'606
7	424'117	450'296
8	419'528	456'387
9	411'082	463'645
10	410'387	468'488
11	435'980	476'198
12	372'513	445'631
13	411'770	457'050
14	380'622	432'273
15	384'000	447'010
16	364'883	443'103
17	336'143	438'830
18	317'801	435'961
19	300'874	453'019
20	262'034	469'713
21	242'768	461'548
22	206'808	475'731
23	199'872	442'506
24	185'856	438'441
25	132'116	414'974

Table 7: Motor liability, diagonal payments and a priori estimates $\hat{\mu}_i$.

i/j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	ultimate
0			284'451	277'284	268'633	265'993	263'702	263'260	262'612	262'596	264'525	264'165	265'054	266'590	267'035	267'291	267'217	270'073	269'350	269'308	268'740	268'392
1		314'004	307'747	289'276	283'281	281'033	277'636	276'586	277'460	277'042	279'303	281'596	284'693	284'264	284'787	284'472	285'261	286'117	286'307	285'422	285'683	286'310
2	321'578	309'394	289'928	282'004	275'502	270'268	271'434	269'822	266'568	267'388	268'193	268'081	267'289	266'985	266'706	266'820	267'891	270'210	272'464	272'152	272'473	272'781
3	371'156	367'120	346'871	339'593	342'688	339'325	332'822	333'296	329'576	333'859	337'626	338'856	338'417	341'972	341'159	342'548	342'949	343'067	343'512	342'166	342'871	341'679
4	370'661	369'491	358'572	344'401	346'665	342'370	344'807	344'529	346'520	344'774	345'346	344'592	339'747	340'214	338'824	338'687	338'119	337'285	336'834	336'271	336'647	337'137
5	419'289	416'218	393'048	384'661	383'508	376'001	382'602	383'968	384'388	383'803	387'951	389'023	386'576	389'510	388'477	388'833	389'687	390'600	389'190	388'147	388'548	
6	437'568	427'975	415'222	407'791	406'460	418'175	411'091	406'724	410'959	413'780	414'044	410'973	410'142	409'120	410'870	414'860	412'483	410'468	410'964	410'768		
7	450'610	448'418	431'294	424'361	432'825	427'773	436'947	434'592	437'282	437'315	438'579	439'037	440'049	439'942	439'561	437'789	437'218	439'781	438'979			
8	468'979	450'942	432'591	435'162	423'583	420'876	424'588	424'632	429'180	431'788	429'673	433'284	430'848	432'002	431'379	429'732	429'539	428'875				
9	443'189	434'315	433'325	425'332	424'921	422'432	421'233	422'904	424'484	426'661	431'462	431'676	434'368	433'599	431'820	432'994	433'037					
10	435'307	422'149	412'660	405'190	402'367	402'597	411'096	412'873	420'973	432'325	437'067	433'610	435'824	435'656	434'796	436'011						
11	475'948	438'817	432'413	436'710	436'904	443'156	448'347	463'171	468'412	476'070	474'593	473'324	477'058	472'283	473'524							
12	447'021	422'678	405'919	399'462	400'047	398'297	406'939	404'834	409'056	411'421	412'002	410'739	409'744	413'587								
13	457'229	444'054	436'390	436'853	442'292	453'494	456'363	460'272	459'591	456'975	455'336	454'500	453'068									
14	462'989	464'776	447'833	432'893	432'124	442'743	451'994	451'534	450'528	450'845	448'398	442'810										
15	484'915	468'800	454'958	447'601	461'106	470'358	465'346	468'879	461'537	456'753	453'919											
16	462'028	429'610	438'929	454'797	468'116	468'721	469'907	463'823	459'524	452'385												
17	450'908	456'030	476'259	483'129	476'952	464'941	453'391	445'089	434'103													
18	426'385	428'504	456'796	449'886	445'397	432'021	412'353	402'565														
19	461'078	477'458	480'960	471'869	462'978	444'670	437'203															
20	444'123	430'684	433'664	419'422	403'126	396'903																
21	433'830	407'931	393'723	371'800	361'853																	
22	418'202	374'855	338'598	324'790																		
23	426'853	373'282	351'590																			
24	410'810	394'477																				
25	405'597																					

Table 8: Motor liability, incurred losses.

Paper

B

Annina Saluz, Hans Bühlmann, Alois Gisler, Franco Moriconi.

The Bornhuetter-Ferguson Method with Repricing.

Submitted.

Bornhuetter-Ferguson Reserving Method with Repricing

Annina Saluz* Hans Bühlmann* Alois Gisler*
Franco Moriconi†

*ETH Zurich, RiskLab Switzerland, Department of Mathematics,
8092 Zurich, Switzerland

†Università di Perugia, Dipartimento di Economia,
06123 Perugia, Italy

Abstract

The Bornhuetter-Ferguson (BF) reserve of an accident year is the product of an a priori estimate of the expected ultimate claim and the estimated 'still to come percentage'. In practice, these a priori estimates are often adjusted over time. We call this process 'repricing'. In this paper we study the incorporation of claims experience for the repricing process by using credibility theory. We simultaneously estimate the still to come percentages and we quantify the prediction uncertainty of the resulting reserves.

In a special case the method can be interpreted as a weighted average between chain ladder (CL) and BF.

Keywords. Claims Reserving, Bornhuetter-Ferguson Method, Credibility Theory, Repricing.

1 Introduction

The Bornhuetter-Ferguson (BF) method (Bornhuetter-Ferguson, 1972) is one of the most frequently used claims reserving methods in practice. In this method the reserves of accident year i at time I are estimated by

$$\widehat{R}_i^{BF} = \hat{\mu}_i \left(1 - \hat{\beta}_{I-i}\right), \tag{1.1}$$

where

$$\begin{aligned} \hat{\mu}_i &= \text{estimate of the ultimate claim } \mu_i = E\left[C_{i,J}\right] \text{ of accident year } i, \\ \hat{\beta}_{I-i} &= \text{estimate of the cumulative claims development pattern } \beta_{I-i}. \end{aligned}$$

The essential questions in the BF method are

i) Where do the estimates $\hat{\mu}_i$ come from?
ii) How should the development pattern β_j be estimated?

Regarding question i):
The basic idea behind the BF method is to incorporate the information contained in the premium for claims reserving. The premium is used to calculate an a priori estimate $\hat{\mu}_i^a$ of the ultimate claim $E\left[C_{i,J}\right]$ to be used for $\hat{\mu}_i$ in (1.1). If p_i denotes the earned premium of accident year i, μ_i is usually estimated by $\hat{\mu}_i^a = p_i\hat{\theta}_i$, where $\hat{\theta}_i$ is an a priori estimate of the expected loss ratio of accident year i. In practice such a priori estimates mostly come from pricing and/or a budgeting and planning process that is, $\hat{\mu}_i^a$ is an estimate of the expected claim amount at the beginning of year i based on all information available at that time: on internal data such as observed claims from previous years as well as on external information such as industry-wide data and expert opinions. Note that the premiums p_i can be considered as a volume measure.

A priori estimates should not change during the claims development in later years, otherwise they are no longer a priori estimates. However, in practice the $\hat{\mu}_i$'s used in the BF reserving formula (1.1) are often changed in later years. We call this process 'repricing'. Usually, such a repricing or change of the estimates $\hat{\mu}_i^a$ is not done in a systematic way. As a consequence there is quite a lot of subjective judgement in the resulting BF reserves. In this paper we show how the repricing can be done, based on a sound actuarial basis which eliminates the subjective element in the BF reserve estimates.

Indeed, a repricing or an update of the original a priori estimates in later years makes sense. At the beginning of year i there is given some a priori estimate of the expected claim amount of accident year i. During the development process more claims information becomes available and one can do an a posteriori update of this estimate in later years. The whole situation perfectly fits into the framework of Bayesian statistics and credibility. A classical example of a method that includes a repricing is the Benktander-Hovinen method (Benktander 1976; Hovinen 1981). In this method the updated estimate $\hat{\mu}_i$ is an average of the chain ladder (CL) ultimate and an a priori estimate $\hat{\mu}_i^a$. In a special case our repricing method can be interpreted as a weighted average between the CL and the BF method. Such additive mixtures of different

reserving methods are not unusual in practice. Our methodology gives a mathematical framework for such a mixture.

Papers using credibility for claims reserving include De Vylder (1982), Mack (1990), Witting (1987), Hesselager-Witting (1988) and Neuhaus (1992). Witting (1987) used a credibility model for claim numbers. In Hesselager-Witting (1988) a credibility model with random fluctuations in the development pattern is introduced. Neuhaus (1992) studies the efficiency of an approximate credibility method for loss reserving. A comparison between the method in Neuhaus (1992) and the Benktander-Hovinen method can be found in Mack (2000). A more recent paper by Jessen-Rietdorf (2011) considers a credibility model including diagonal effects within an over-dispersed Poisson framework that is, the variance is assumed to be proportional to the mean. Mack (1990) reconsidered the model of De Vylder (1982) with a more general assumption on the variance of incremental claims. The models in De Vylder (1982) and Mack (1990) are closest to our model, however, the assumptions with regard to the variance as well as with regard to the dependence structure are much more restrictive in their models (see Remarks 3.3). Moreover De Vylder (1982) and Mack (1990) do not provide estimates for the (conditional) mean square error of prediction (MSEP).

Regarding question ii):
In practice the development pattern in the BF method is often estimated by the CL pattern. However, as shown in Mack (2008) the use of the CL pattern is not compatible with the assumptions underlying the BF method. As stated above, the incorporation of the premiums is crucial in the BF method. The availability of the a priori estimates in the BF method is a fundamental difference to the CL method. Hence it is argued in Saluz et al. (2011) that estimators of the development pattern in the BF method should also take this additional information into account. Estimators for the development pattern which take the a priori estimates into account are for instance discussed in Mack (2008) and Saluz et al. (2011). In our case the a priori estimates are a posteriori updated and hence the estimates for the development pattern have to be updated, too. This leads to an iterative algorithm for the simultaneous estimation of the development pattern and the a posteriori adjusted $\hat{\mu}_i$'s. This iterative procedure is an innovation compared to De Vylder (1982) and Mack (1990). In De Vylder (1982) and Mack (1990) the estimation of the development pattern is based on the a priori estimates and the development pattern is therefore not updated over time.

Organisation of the paper. In Section 2 we introduce the notation and data structure. In Section 3 we introduce a credibility model for a BF method

with repricing. The credibility estimates are given in Section 4 and the corresponding estimates for the development pattern are given in Section 6. In Section 5 we show how the variance parameters can be estimated. An estimation procedure for the MSEP is provided in Section 7 and an example is provided in Section 8. Whereas the component of the MSEP deriving from the uncertainty of the updated $\hat{\mu}_i$ can be expressed in a closed formula, we need a simulation step to cope with the uncertainty of the development pattern.

2 Notation and Data Structure

We denote the cumulative claims (cumulative payments or incurred losses) in accident year $i \in \{0, \dots, I\}$ at the end of development year $j \in \{0, \dots, J\}$ by $C_{i,j} \geq 0$ and we assume $J \leq I$. The earned premium of accident year i is denoted by p_i, $0 \leq i \leq I$. Let $X_{i,j} = C_{i,j} - C_{i,j-1}$ denote the incremental claims, where we set $C_{i,-1} = 0$. The summation over an index starting from 0 is denoted with a square bracket, for example,

$$X_{[k],j} = \sum_{i=0}^{k} X_{i,j}, \quad 0 \leq k \leq I, \quad 0 \leq j \leq J.$$

We assume that all claims are settled after development year J and therefore the total ultimate claim of accident year i is given by $C_{i,J}$. The observations of accident year i are given by $\boldsymbol{X}_i = (X_{i,0}, \dots, X_{i,(I-i)\wedge J})'$, where $\boldsymbol{X}'$ denotes the transpose of a vector (or a matrix) $\boldsymbol{X}$. At time I we have information

$$\mathcal{D}_I = \{C_{i,j} : i + j \leq I,\ j \leq J\},$$

and our goal is to predict $\mathcal{D}_I^c = \{C_{i,j} : i + j > I,\ i \leq I,\ j \leq J\}$. The outstanding loss liabilities for accident year i at time I are given by

$$R_i = C_{i,J} - C_{i,I-i}, \quad I - J + 1 \leq i \leq I, \tag{2.1}$$

and the total outstanding loss liabilities are given by $R = \sum_{i=I-J+1}^{I} R_i$. In order to simplify notation we define $\iota(i) = (I - i) \wedge J$ to be the last observed development year in accident year i.

Remark 2.1. The 'true' outstanding loss liabilities are given by formula (2.1) only if $C_{i,j}$ denote cumulative payments. For incurred losses $C_{i,j}$ the outstanding loss liabilities are given by

$$R_i = (C_{i,J} - C_{i,\iota(i)}) + (C_{i,\iota(i)} - C_{i,\iota(i)}^{paid}), \tag{2.2}$$

with $C_{i,j}^{paid}$ denoting the cumulative payments of accident year i up to development year j. Note that the additional term $C_{i,\iota(i)} - C_{i,\iota(i)}^{paid}$ is observable at time I and has no impact on the uncertainties in the claims prediction problem. Therefore we only consider the outstanding loss liabilities as defined in (2.1). For incurred losses, R_i in (2.1) is often referred to as IBNR.

3 Stochastic Models behind the BF Method

3.1 Basic Assumptions Underlying the BF Method

The basic assumptions underlying the BF method are independent increments $X_{i,j}$, $0 \leq i \leq I$, $0 \leq j \leq J$ and a cross-classified model

$$E[X_{i,j}] = \mu_i \gamma_j, \quad 0 \leq i \leq I \quad \text{and} \quad 0 \leq j \leq J,$$

with $\sum_{j=0}^{J} \gamma_j = 1$ and consequently $\mu_i = E[C_{i,J}]$. The cumulative development pattern is given by $\beta_j = \sum_{k=0}^{j} \gamma_k$. In contrast to the CL method, the BF method takes the information contained in the earned premiums p_i into account. More precisely, in the BF method, there are a priori estimates $\hat{\mu}_i^a$ of μ_i available. Often these a priori estimates $\hat{\mu}_i^a$ are assumed to be unbiased and given by

$$\hat{\mu}_i^a = p_i \hat{\theta}_i,$$

where $\hat{\theta}_i$ is an estimate of the expected loss ratio of accident year i. Additionally, Mack (2008) assumes that

$$\mathrm{Var}(X_{i,j}) = \mu_i \sigma_j^2, \quad 0 \leq i \leq I \quad \text{and} \quad 0 \leq j \leq J.$$

We formalise these basic assumptions in the following model formulated for the $Y_{i,j} = X_{i,j}/p_i$. We denote the modified observations of accident year i by $\boldsymbol{Y}_i = (Y_{i,0}, \ldots, Y_{i,\iota(i)})'$. Moreover, we assume in the following that the premiums p_i are 'on-level' premiums that is, earned premiums adjusted for premium-cycles such that the expected loss ratio is the same for all accident years.

Model Assumptions 3.1 (Basic BF Model)**.**

BF1 *All $Y_{i,j}$ are independent and there exist parameters $\gamma_0, \ldots, \gamma_J$ and positive parameters θ_0, σ^2 and $\delta_0^2, \ldots, \delta_J^2$ with $\sum_{j=0}^{J} \gamma_j = \sum_{j=0}^{J} \delta_j^2 = 1$ such*

that for $0 \leq i \leq I$ *and* $0 \leq j \leq J$

$$E[Y_{i,j}] = \theta_0 \gamma_j, \quad \mathrm{Var}(Y_{i,j}) = \theta_0 \frac{\delta_j^2 \sigma^2}{p_i}.$$

BF2 There are given unbiased a priori estimates $\hat{\mu}_i^a$ *of* $p_i\theta_0$ *and* $\hat{\mu}_i^a$ *is independent of* $\boldsymbol{Y}_l$, $l \geq i$.

3.2 Stochastic Model for the BF Method with Repricing

In order to include a repricing we modify the assumptions of the Basic BF Model and switch to a Bayesian framework. By this we mean that the unknown loss ratio parameter is modelled as a random variable. More precisely, we assume that the true expected ultimate claim amount is given by $\mu_i(\Theta_i) = p_i\Theta_i$, where p_i are assumed to be known (premiums) and Θ_i is a random variable describing the expected loss ratio of accident year i. These considerations bring us to the following model.

Model Assumptions 3.2 (BF Model with Repricing)**.**
Each accident year i *is characterised by its true expected loss ratio* Θ_i.

BF1' Given $\boldsymbol{\Theta} = (\Theta_0, \ldots, \Theta_I)$, *all* $Y_{i,j}$ *are conditionally independent and there exist parameters* $\gamma_1, \ldots, \gamma_J$ *and positive parameters* γ_0, σ^2 *and* $\delta_0^2, \ldots, \delta_J^2$ *with* $\sum_{j=0}^J \gamma_j = \sum_{j=0}^J \delta_j^2 = 1$ *such that for* $0 \leq i \leq I$ *and* $0 \leq j \leq J$

$$E[Y_{i,j}|\boldsymbol{\Theta}] = \Theta_i \gamma_j, \quad \mathrm{Var}(Y_{i,j}|\,\boldsymbol{\Theta}) = \Theta_i \frac{\delta_j^2 \sigma^2}{p_i}.$$

BF2' We assume

$$E[\Theta_i] = \theta_0 \quad \textit{and} \quad \mathrm{Cov}(\boldsymbol{\Theta}) = T,$$

where T *is symmetric and positive definite with* $(T)_{i,k} = \tau_{i,k}$ *and* $\tau_{i,i} = \tau^2$ *for all* i.

We distinguish two cases in the sequel:

i) Homogeneous case: θ_0 is unknown and to be estimated from the data.

ii) Inhomogeneous case: θ_0 is known.

Remarks 3.3.

- Note that $\mu_i(\Theta_i) = p_i\Theta_i = E[C_{i,J}|\,\Theta_i]$ is the 'true' individual expected ultimate claim amount of accident year i given the specific characteristic Θ_i of year i, whereas $p_i\theta_0 = E[C_{i,J}] = E[\mu_i(\Theta_i)]$ is the a priori expected ultimate claim amount. If $\mu_i(\Theta_i)$ was known, we would use it in the BF formula. Since it is not known, we have to estimate it.

- The 'true' individual expected ultimate loss ratio Θ_i is unknown and can deviate from the a priori expected loss ratio θ_0. The distribution of Θ_i thus reflects the uncertainty of our a priori knowledge with regard to the 'true' individual expected loss ratio.

- We assume that $\gamma_0 > 0$ in order to simplify the formulas below. The case $\gamma_0 \leq 0$ is not meaningful for practical situations and therefore this is not a restriction.

- In the special case $\delta_j^2 = \gamma_j$ we have $\mathrm{Var}(X_{i,j}|\,\Theta_i) = \sigma^2 E[X_{i,j}|\Theta_i]$, which is referred to as the over-dispersed Poisson (ODP) case. Note that the ODP case is not appropriate if negative incremental claims occur.

- If we assume that $\Theta_i \equiv \theta_0$ we distinguish the following cases. In the inhomogeneous case the expected values $E[C_{i,J}]$ are then exactly known and we are in the situation of the original BF framework (Section 3.1) with $\mu_i(\Theta_i) \equiv p_i\theta_0$. In this case a repricing makes no sense and one should use $p_i\theta_0$ in the BF reserving formula. In the homogeneous case the expected loss ratio θ_0 needs to be estimated from the data which is similar to the Cape Cod method of Bühlmann-Straub (1983).

- The model in De Vylder (1982) is formulated in a regression framework using the assumption

$$\boldsymbol{X}_i = \boldsymbol{\Gamma}_i\beta(\Theta_i),$$

where $\boldsymbol{\Gamma}_i = (\Gamma_{i,0}, \ldots, \Gamma_{i,\iota(i)})'$ is a random vector independent of Θ_i with

$$E[\boldsymbol{\Gamma}_i] = (\gamma_0, \ldots, \gamma_{\iota(i)})', \quad 0 \leq i \leq I.$$

The development pattern γ_j is not normalised. The variance of the $X_{i,j}$ is assumed to be constant over all development years that is, De Vylder (1982) assumes

$$\mathrm{Var}(Y_{i,j}|\Theta_i) = \frac{r^2\beta^2(\Theta_i)}{p_i}, \quad 0 \leq j \leq J.$$

Mack (1990) generalised the model of De Vylder and embedded the model in the Bühlmann-Straub model. However, the latter can only be done in the case where $\gamma_j \neq 0$ for all $0 \leq j \leq J$. For the variance of the $Y_{i,j}$ Mack (1990) assumes

$$\mathrm{Var}(Y_{i,j}|\Theta_i) = \frac{\Psi(\Theta_i)}{p_i}\gamma_j^{1-\alpha}E[Y_{i,j}|\Theta_i],$$

where $\alpha \in \mathbb{R}$. For $\alpha \in \{0, 1, 2\}$ a micro model illustrates the variance assumption. Finally, in De Vylder (1982) and Mack (1990) it is assumed that the Θ_i's are independent.

If we make additional assumptions on the dependencies of the Θ_i's we obtain simpler formulas and additional properties for the estimators of $\mu_i(\Theta_i)$. In order to refer to this special case, we introduce the following model, which is a special case of Model Assumptions 3.2.

Model Assumptions 3.4 (Independence Model)**.**
Each accident year i is characterised by its true expected loss ratio Θ_i.

BF1" Assumption BF1' from Model Assumptions 3.2 is satisfied.

BF2" The pairs $\{(\boldsymbol{Y}_i, \Theta_i), i = 0, 1, \ldots, I\}$ are independent and $\Theta_0, \Theta_1, \ldots$ are iid with

$$E[\Theta_i] = \theta_0 \quad \textit{and} \quad \mathrm{Var}(\Theta_i) = \tau^2.$$

4 Estimation of the $\mu_i(\Theta_i)$

The $\hat{\mu}_i$ in the BF reserve formula (1.1) are estimates of the true expected value of the ultimate claim $C_{i,J}$. In our model, this expected value depends on the risk characteristic Θ_i of year i, which means that the $\hat{\mu}_i$ in (1.1) should be estimates of $\mu_i(\Theta_i) = p_i\Theta_i$. We estimate the Θ_i's by credibility estimators.

Theorem 4.1. *Under Model Assumptions 3.2 the inhomogeneous credibility estimator of $\boldsymbol{\Theta} = (\Theta_0, \ldots, \Theta_I)'$ based on all observations $Y_{i,j}$, $i + j \leq I$, $0 \leq j \leq J$, is given by*

$$\hat{\boldsymbol{\Theta}}^{\boldsymbol{inh}} = A\boldsymbol{B} + (I - A)\mathbf{1}\theta_0,$$

where $\boldsymbol{B} = (B_0, \ldots, B_I)'$, with

$$B_i = \sum_{j=0}^{\iota(i)} \frac{\gamma_j/\delta_j^2}{w_{[\iota(i)]}} Y_{i,j}, \quad w_j = \frac{\gamma_j^2}{\delta_j^2},$$

and $\mathbf{1} = (1, \ldots, 1)'$*,* I *is the identity matrix and* $A = T(T+S)^{-1}$ *with* $T = \mathrm{Cov}(\boldsymbol{\Theta})$ *and* $S = E\left[\mathrm{Cov}\left(\boldsymbol{B}|\,\boldsymbol{\Theta}\right)\right]$ *is the diagonal matrix with entries*

$$(S)_{i,i} = E\left[\mathrm{Var}\left(B_i|\,\boldsymbol{\Theta}\right)\right] = \frac{\sigma^2\theta_0}{p_i w_{[\iota(i)]}}.$$

The proof of Theorem 4.1 is given in Appendix A.

Note that $w_{[\iota(i)]} > 0$ for all i, because $\gamma_0 > 0$. However some of the γ_j can be zero or even negative, which is often the case for incurred claims and later development years.

Corollary 4.2. *Under Model Assumptions 3.4 (Independence Model) the inhomogeneous credibility estimator of* Θ_i *based on all observations* $Y_{i,j}$*,* $i+j \leq I$*,* $0 \leq j \leq J$*, is given by*

$$\hat{\Theta}_i^{inh} = \alpha_i B_i + (1-\alpha_i)\theta_0,$$

where B_i *and* w_j *are given in Theorem 4.1 and*

$$\alpha_i = \frac{p_i w_{[\iota(i)]}}{p_i w_{[\iota(i)]} + \frac{\sigma^2\theta_0}{\tau^2}}.$$

The proof of Corollary 4.2 is given in Appendix A.

Remarks 4.3.

- In the ODP case ($\gamma_j = \delta_j^2$) we have, under Model Assumptions 3.4 (Independence Model),

$$B_i = \frac{C_{i,\iota(i)}}{p_i\beta_{\iota(i)}} \quad \text{and} \quad \alpha_i = \frac{p_i\beta_{\iota(i)}}{p_i\beta_{\iota(i)} + \frac{\sigma^2\theta_0}{\tau^2}}. \tag{4.1}$$

- In the general case with dependent Θ_i the estimator in Corollary 4.2 is the credibility estimator for Θ_i based only on the observations $\boldsymbol{Y}_i$ of accident year i. This formula follows from the regression credibility model (see Bühlmann-Gisler, 2005).

If θ_0 is unknown we use the following homogeneous credibility estimator.

Theorem 4.4. *Under Model Assumptions 3.2 the homogeneous credibility estimator of* $\boldsymbol{\Theta} = (\Theta_0, \ldots, \Theta_I)$ *based on all observations* $Y_{i,j}$*,* $i+j \leq I$*,* $0 \leq j \leq J$*, is given by*

$$\widehat{\boldsymbol{\Theta}}^{\boldsymbol{hom}} = A\boldsymbol{B} + (I-A)\mathbf{1}\hat{\theta}_0,$$

where $\boldsymbol{B}$, $\mathbf{1}$ and A are given in Theorem 4.1 and where

$$\hat{\theta}_0 = \left(\mathbf{1}'(S+T)^{-1}\mathbf{1}\right)^{-1}\mathbf{1}'(S+T)^{-1}\boldsymbol{B},$$

with S and T as in Theorem 4.1.

The proof of Theorem 4.4 is given in Appendix A.

Corollary 4.5. *Under Model Assumptions 3.4 (Independence Model) the homogeneous credibility estimator based on all observations $Y_{i,j}$, $i+j \leq I$, $0 \leq j \leq J$, is given by*

$$\hat{\Theta}_i^{hom} = \alpha_i B_i + (1-\alpha_i)\hat{\theta}_0,$$

where B_i and α_i are as in Corollary 4.2 and

$$\hat{\theta}_0 = \sum_{i=0}^{I} \frac{\alpha_i}{\alpha_{[I]}} B_i.$$

The proof of Corollary 4.5 is given in Appendix A.

Remark 4.6. If $\tau^2 \to 0$ then $\alpha_i = 0$ and we get

$$\lim_{\tau^2\to 0}\frac{\alpha_i}{\alpha_{[I]}} = \lim_{\tau^2\to 0}\frac{p_i w_{[\iota(i)]}}{\left(\tau^2 p_i w_{[\iota(i)]} + \sigma^2\theta_0\right)\sum_{l=0}^{I}\frac{p_l w_{[\iota(l)]}}{\tau^2 p_l w_{[\iota(l)]}+\sigma^2\theta_0}} = \frac{p_i w_{[\iota(i)]}}{\sum_{l=0}^{I} p_l w_{[\iota(l)]}},$$

and therefore

$$\hat{\theta}_0 = \sum_{i=0}^{I}\frac{p_i w_{[\iota(i)]} B_i}{\sum_{l=0}^{I} p_l w_{[\iota(l)]}}.$$

In the ODP case $\hat{\theta}_0$ is then the Cape Cod estimate of the expected loss ratio (see Bühlmann-Straub, 1983, or Wüthrich-Merz, 2008)

$$\hat{\theta}_0 = \frac{\sum_{i=0}^{I} C_{i,\iota(i)}}{\sum_{l=0}^{I} p_l \beta_{\iota(l)}}.$$

In the case of independence we have the following 'balance property' for the homogeneous credibility estimator (see Bühlmann-Gisler, 2005).

Theorem 4.7 (Balance Property)**.** *Under Model Assumptions 3.4 (Independence Model) the homogeneous credibility estimator of Θ_i satisfies the 'balance property'*

$$\sum_{i+j\leq I}\frac{\gamma_j}{\delta_j^2}\hat{X}_{i,j}^{hom} = \sum_{i+j\leq I}\frac{\gamma_j}{\delta_j^2}X_{i,j},$$

where $\hat{X}_{i,j}^{hom} = p_i\hat{\Theta}_i^{hom}\gamma_j$ with $\hat{\Theta}_i^{hom}$ given in Corollary 4.5.

Remarks 4.8.

- In the ODP case, we obtain the balance property

$$\sum_{i+j\leq I} \hat{X}_{i,j}^{hom} = \sum_{i+j\leq I} X_{i,j},$$

  which means that the sum of the predicted past increments is equal to the sum of the observed claims.

- In the case $\tau^2 \to \infty$ we have the credibility weights $\alpha_i = 1$ and we additionally get the following 'balance property' in each row

$$\sum_{j=0}^{\iota(i)} \frac{\gamma_j}{\delta_j^2} \hat{X}_{i,j}^{hom} = \sum_{j=0}^{\iota(i)} \frac{\gamma_j}{\delta_j^2} X_{i,j}.$$

  In the ODP case we then get

$$p_i \hat{\Theta}_i^{hom} = \frac{C_{i,\iota(i)}}{\beta_{\iota(i)}}.$$

4.1 Different Interpretations

So far we derived the formulas under the idea of introducing a repricing in the BF method. However, a closer look at the formulas allows for different interpretations. In the ODP case, under Model Assumptions 3.4 (Independence Model) and assuming that θ_0 is known (inhomogeneous case), we obtain the following formula for the reserve (see Remarks 4.3) which is also given in Wüthrich-Merz (2008)

$$\hat{R}_i = (1 - \beta_{\iota(i)}) \left(\alpha_i \frac{C_{i,\iota(i)}}{\beta_{\iota(i)}} + (1 - \alpha_i) p_i \theta_0 \right).$$

We therefore have a mixture between a data based projection method (CL type), where

$$\hat{C}_{i,J}^{d} = C_{i,\iota(i)} + (1 - \beta_{\iota(i)}) \frac{C_{i,\iota(i)}}{\beta_{\iota(i)}} = \frac{C_{i,\iota(i)}}{\beta_{\iota(i)}},$$

and a method based on external information (BF type), where

$$\hat{C}_{i,J}^{e} = C_{i,\iota(i)} + (1 - \beta_{\iota(i)}) p_i \theta_0.$$

In these formulas the pattern is still assumed to be known. If the pattern is estimated by the CL pattern, we have $\hat{C}_{i,J}^{d} = \hat{C}_{i,J}^{CL}$. In practice it is not

unusual to combine different reserving methods in such an additive way. Our methodology gives a mathematical framework for such mixtures.

If $\gamma_j \neq \delta_j^2$ we have an analogous expression

$$\hat{R}_i = (1-\beta_{\iota(i)}) \left(\alpha_i \frac{C^*_{i,\iota(i)}}{\beta_{\iota(i)}} + (1-\alpha_i) p_i \theta_0 \right),$$

where

$$C^*_{i,\iota(i)} = B_i = \sum_{j=0}^{\iota(i)} \frac{w_j \beta_{\iota(i)}}{\gamma_j w_{[\iota(i)]}} X_{i,j}$$

is interpreted as an 'adjusted diagonal' of accident year i.

5 Estimation of the Variance Components

Before we study the estimation of the development pattern we also need the estimates for the variance components $\sigma^2\theta_0$, δ_j^2 and τ^2.

We suggest to estimate δ_j^2 by

$$\tilde{\delta}_j^2 = \frac{\hat{\delta}_j^{2^{raw}}}{\hat{\delta}_{[J]}^{2^{raw}}}, \tag{5.1}$$

where

$$\hat{\delta}_j^{2^{raw}} = \frac{1}{I-j} \sum_{i=0}^{I-j} p_i \left(Y_{i,j} - \sum_{k=0}^{I-j} \frac{p_k}{p_{[I-j]}} Y_{k,j} \right)^2, \quad j < I, \tag{5.2}$$

and where we use the extrapolation $\hat{\delta}_J^{2^{raw}} = \min(\hat{\delta}_{J-1}^{4^{raw}} / \hat{\delta}_{J-2}^{2^{raw}}, \hat{\delta}_{J-2}^{2^{raw}})$ if $J = I$, see Mack (1993). The estimate $\hat{\delta}_j^{2^{raw}}$ for $\delta_j^2\sigma^2\theta_0$ is not unbiased, but the bias should be comparatively small. More precisely, we have for $j < I$

$$E\left[\hat{\delta}_j^{2^{raw}}\right]$$
$$= \delta_j^2\sigma^2\theta_0 + \frac{p_{[I-j]}}{I-j}\tau^2\gamma_j^2 \left(\sum_{i=0}^{I-j} \frac{p_i}{p_{[I-j]}} \left(1 - \frac{p_i}{p_{[I-j]}} \right) \right) - \frac{2\gamma_j^2}{p_{[I-j]}(I-j)} \sum_{i<k} p_i p_k \tau_{i,k}.$$

For estimating $\sigma^2\theta_0$ and τ^2 we adjust the standard estimators of the Bühlmann-Straub model (see Bühlmann-Gisler, 2005) to our model with correlated Θ_i.

For $\sigma^2\theta_0$ we use the standard estimator, where the number of observations differs between years,

$$\widehat{\sigma^2\theta_0} = \frac{1}{N} \sum_{i=0}^{I-1} \sum_{j=0}^{\iota(i)} \frac{p_i}{\tilde{\delta}_j^2} \left(Y_{i,j} - \hat{\gamma}_j B_i \right)^2,$$

and

$$N = \sum_{i=0}^{I-1} \iota(i) = IJ - \frac{1}{2}J(J-1);$$

here $\hat{\gamma}_j$ is an estimate for γ_j.

For τ^2 we distinguish the inhomogeneous and the homogeneous case. If θ_0 is known we consider the following estimate for τ^2

$$\widehat{\widehat{\tau}}^2 = \sum_{i=0}^{I} \frac{p_i \widehat{w}_{[\iota(i)]}}{\sum_{l=0}^{I} p_l \widehat{w}_{[\iota(l)]}} (B_i - \theta_0)^2 - \frac{(I+1)\widehat{\sigma^2\theta_0}}{\sum_{i=0}^{I} p_i \widehat{w}_{[\iota(i)]}}, \tag{5.3}$$

where $\widehat{w}_j$ is an estimate for w_j. Since $\widehat{\widehat{\tau}}^2$ in (5.3) can become negative we use

$$\hat{\tau}^2 = \max\left(\widehat{\widehat{\tau}}^2, 0\right). \tag{5.4}$$

Moreover, due to the uncertainties in the loss ratios Θ_i, an estimate $\hat{\tau}^2 = 0$ is not realistic, since it would mean that the expected values $E[C_{i,J}]$ are exactly specified by the experts. Therefore a lower bound $\mathrm{Var}(\Theta_i) = \theta_0^2 \mathrm{CoVa}(\Theta_i)^2 > 0$ for $\hat{\tau}^2$ could be meaningful. The standard values for the parameter risk delivered by the Swiss supervising authority FINMA range between 3.5% and 5.75% (see the Technical Document on the Swiss Solvency Test, FINMA, 2006).

In the case where θ_0 is unknown we use

$$\begin{aligned}
\hat{\tau}^2 &= \max\left(\widehat{\widehat{\tau}}^2, 0\right), \quad \text{where} \qquad (5.5)\\
\widehat{\widehat{\tau}}^2 &= \frac{1}{c}\left(t - \frac{I\widehat{\sigma^2\theta_0}}{\sum_{i=0}^{I} p_i \widehat{w}_{[\iota(i)]}}\right), \text{with}\\
c &= \sum_{i=0}^{I} \frac{p_i \widehat{w}_{[\iota(i)]}}{\sum_{l=0}^{I} p_l \widehat{w}_{[\iota(l)]}} \left(1 - \sum_{k=0}^{I} \rho_{i,k} \frac{p_k \widehat{w}_{[\iota(k)]}}{\sum_{l=0}^{I} p_l \widehat{w}_{[\iota(l)]}}\right),\\
t &= \sum_{i=0}^{I} \frac{p_i \widehat{w}_{[\iota(i)]}}{\sum_{l=0}^{I} p_l \widehat{w}_{[\iota(l)]}} \left(B_i - \bar{B}\right)^2,\\
\bar{B} &= \sum_{i=0}^{I} \frac{p_i \widehat{w}_{[\iota(i)]}}{\sum_{l=0}^{I} p_l \widehat{w}_{[\iota(l)]}} B_i,
\end{aligned}$$

and where the correlations $\rho_{l,k} = \tau_{l,k}/\tau^2$ are externally given. An example how these correlations can be specified is given in Saluz et al. (2011), where it is assumed that the correlation is linearly decreasing in the difference of the accident years and it is zero after m years, that is,

$$\hat{\rho}_{i,k} = \max\left(0, \frac{m - |i-k|}{m}\right), \tag{5.6}$$

with e.g. $m = 10$. Of course, a lower bound $\mathrm{Var}(\Theta_i) = \theta_0^2 \mathrm{CoVa}(\Theta_i)^2 > 0$ can be used for $\hat{\tau}^2$ also in this case.

6 Estimation of the Development Pattern and Iterative Algorithm

For the credibility estimators in Section 4 the development pattern $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_J)$ was assumed to be known. However the second main question in the BF method is how this development pattern should be estimated. If the 'true' expected ultimate claims $\mu_i(\Theta_i)$ were known then the best linear unbiased estimate of γ_j given $\boldsymbol{\Theta} = (\Theta_0, \ldots, \Theta_I)$ would be

$$\gamma_j^{raw} = \frac{X_{[I-j],j}}{\sum_{i=0}^{I-j} \mu_i(\Theta_i)}. \tag{6.1}$$

Since the $\mu_i(\Theta_i)$ are not exactly known they are replaced by estimates $\hat{\mu}_i$ of $\mu_i(\Theta_i)$. However these raw estimates do not sum up to one and are therefore not a claims development pattern. To get a claims development pattern one has to normalise them. The simplest normalisation is given by

$$\hat{\gamma}_j^s = \hat{\gamma}_j^{raw} \left(\sum_{j=0}^{J} \hat{\gamma}_j^{raw} \right)^{-1}, \tag{6.2}$$

where $\hat{\gamma}_j^{raw}$ is obtained from (6.1) by replacing the $\mu_i(\Theta_i)$ by some estimates $\hat{\mu}_i$.

Remark 6.1. Examples for different normalisations are given in Saluz et al. (2011). Based on maximum likelihood (ML) considerations Saluz et al. (2011) suggest to use the following additive normalisation

$$\hat{\gamma}_j^N = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}} + \frac{\hat{\delta}_j^2 / \hat{\mu}_{[I-j]}}{\sum_{l=0}^{J} \hat{\delta}_l^2 / \hat{\mu}_{[I-l]}} \left(1 - \sum_{l=0}^{J} \frac{X_{[I-l],l}}{\hat{\mu}_{[I-l]}} \right) = \hat{\gamma}_j^{raw} + \kappa_j. \tag{6.3}$$

In the ODP case ($\delta_j^2 = \gamma_j$) Saluz et al. (2011) suggest to use the normalisation

$$\hat{\gamma}_j^{ODP} = \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]} + \kappa}, \tag{6.4}$$

instead of (6.3), where κ fulfils the implicit equation

$$\sum_{j=0}^{J} \hat{\gamma}_j^{ODP} = 1,$$

under the constraints $\hat{\gamma}_j^{ODP} > 0$, $0 \leq j \leq J$, that is, $\kappa > -\hat{\mu}_{[I-J]}$. Although (6.3) and (6.4) were derived under distributional assumptions, one can use these normalisations also in a distribution-free framework.

In any case the estimate of the pattern will depend on $\hat{\boldsymbol{\mu}}$ that is, $\hat{\gamma}_j = \hat{\gamma}_j(\hat{\boldsymbol{\mu}})$. In our case we use as estimators $\hat{\mu}_i$ either $\hat{\mu}_i^{hom} = p_i\hat{\Theta}_i^{hom}$ or $\hat{\mu}_i^{inh} = p_i\hat{\Theta}_i^{inh}$. The point is that $\hat{\mu}_i^{hom}$ as well as $\hat{\mu}_i^{inh}$ depend themselves on $\boldsymbol{\gamma}$ that is, $\hat{\mu}_i = \hat{\mu}_i(\boldsymbol{\gamma})$.

Definition 6.2. *For any estimators $\widehat{\mu}_i(\boldsymbol{\gamma})$ and $\widehat{\gamma}_j(\boldsymbol{\mu})$ we call $(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\gamma}})$ a compatible pair if $\widehat{\mu}_i(\widetilde{\boldsymbol{\gamma}}) = \widetilde{\mu}_i$ and $\widehat{\gamma}_j(\widetilde{\boldsymbol{\mu}}) = \widetilde{\gamma}_j$ for all i and j.*

In order to calculate such a compatible pair, an iterative procedure will be necessary. We will start with a pattern $\hat{\boldsymbol{\gamma}}^{(0)}$, then calculate the corresponding credibility estimates $\hat{\boldsymbol{\mu}}^{(1)}$, which are then used to calculate a new pattern $\hat{\boldsymbol{\gamma}}^{(1)}$ and so on. If this iterative procedure converges we have a compatible pair that is, the equations for the credibility estimates and for the pattern are all simultaneously fulfilled.

6.1 Iterative Algorithm

We state the formulas for the homogeneous case (unknown θ_0). In the inhomogeneous case (known θ_0), the formulas for $\hat{\mu}_i^{(n)}$ and $\hat{\tau}^{2^{(n)}}$ need to be adjusted accordingly. The correlations $\rho_{l,k}$ are assumed as in (5.6). If information on the correlation structure is available this should be used and the covariance matrix T can be estimated differently.

Iterative Algorithm 6.3.

i) We start with $\hat{\boldsymbol{\mu}}^{(0)} = \boldsymbol{p} = (p_0, \ldots, p_I)'$ and the corresponding estimate of the development pattern that is,

$$\begin{aligned}
\hat{\mu}_i^{(0)} &= p_i, \\
\hat{\gamma}_j^{(0)} &= \frac{X_{[I-j],j}}{p_{[I-j]}} \left(\sum_{l=0}^{J} \frac{X_{[I-l],l}}{p_{[I-l]}} \right)^{-1}, \\
\hat{\beta}_j^{(0)} &= \sum_{k=0}^{j} \hat{\gamma}_k^{(0)}, \\
\widehat{w}_j^{(0)} &= \frac{\left(\hat{\gamma}_j^{(0)}\right)^2}{\tilde{\delta}_j^2}.
\end{aligned}$$

ii) Iteration step from $n-1$ to n

$$
\begin{aligned}
B_i^{(n)} &= \sum_{j=0}^{\iota(i)} \frac{\hat{\gamma}_j^{(n-1)}/\tilde{\delta}_j^2}{\widehat{w}_{[\iota(i)]}^{(n-1)}} Y_{i,j}, \\
\widehat{\sigma^2\theta_0}^{(n)} &= \frac{1}{N} \sum_{i=0}^{I-1} \sum_{j=0}^{\iota(i)} \frac{p_i}{\tilde{\delta}_j^2} \left(Y_{i,j} - \hat{\gamma}_j^{(n-1)} B_i^{(n)} \right)^2, \quad N = IJ - \frac{1}{2} J(J-1), \\
S_{i,i}^{(n)} &= \frac{\widehat{\sigma^2\theta_0}^{(n)}}{p_i w_{[\iota(i)]}^{(n-1)}}, \quad S_{i,k}^{(n)} = 0, \quad i \neq k, \\
c^{(n)} &= \sum_{i=0}^{I} \frac{p_i \widehat{w}_{[\iota(i)]}^{(n-1)}}{\sum_{l=0}^{I} p_l \widehat{w}_{[\iota(l)]}^{(n-1)}} \left(1 - \sum_{k=0}^{I} \hat{\rho}_{i,k} \frac{p_k \widehat{w}_{[\iota(k)]}^{(n-1)}}{\sum_{l=0}^{I} p_l \widehat{w}_{[\iota(l)]}^{(n-1)}} \right), \\
\bar{B}^{(n)} &= \sum_{i=0}^{I} \frac{p_i \widehat{w}_{[\iota(i)]}^{(n-1)}}{\sum_{l=0}^{I} p_l \widehat{w}_{[\iota(l)]}^{(n-1)}} B_i^{(n)}, \\
t^{(n)} &= \sum_{i=0}^{I} \frac{p_i \widehat{w}_{[\iota(i)]}^{(n-1)}}{\sum_{l=0}^{I} p_l \widehat{w}_{[\iota(l)]}^{(n-1)}} \left(B_i^{(n)} - \bar{B}^{(n)} \right)^2, \\
T_{i,i}^{(n)} &= \hat{\tau}^{2^{(n)}} = \max \left(0, \frac{1}{c^{(n)}} \left(t^{(n)} - \frac{I \widehat{\sigma^2\theta_0}^{(n)}}{\sum_{i=0}^{I} p_i \widehat{w}_{[\iota(i)]}^{(n-1)}} \right) \right),
\end{aligned}
$$

$$
\begin{aligned}
T_{i,k}^{(n)} &= \hat{\tau}_{i,k}^{(n)} = \hat{\tau}^{2^{(n)}} \hat{\rho}_{i,k}, \quad i \neq k, \\
A^{(n)} &= T^{(n)} \left(S^{(n)} + T^{(n)} \right)^{-1}, \\
\hat{\theta}_0^{(n)} &= \left(\mathbf{1}' \left(S^{(n)} + T^{(n)} \right)^{-1} \mathbf{1} \right)^{-1} \mathbf{1}' \left(S^{(n)} + T^{(n)} \right)^{-1} \boldsymbol{B}^{(n)} \\
\widehat{\boldsymbol{\Theta}}^{(n)} &= A^{(n)} \boldsymbol{B}^{(n)} + \left(I - A^{(n)} \right) \mathbf{1} \hat{\theta}_0^{(n)}, \\
\hat{\mu}_i^{(n)} &= p_i \hat{\Theta}_i^{(n)}, \\
\hat{\gamma}_j^{(n)} &= \frac{X_{[I-j],j}}{\hat{\mu}_{[I-j]}^{(n)}} \left(\sum_{l=0}^{J} \frac{X_{[I-l],l}}{\hat{\mu}_{[I-l]}^{(n)}} \right)^{-1}, \\
\hat{\beta}_j^{(n)} &= \sum_{k=0}^{j} \hat{\gamma}_k^{(n)}, \\
\widehat{w}_j^{(n)} &= \frac{\left(\hat{\gamma}_j^{(n)} \right)^2}{\tilde{\delta}_j^2}.
\end{aligned}
$$

iii) If a predefined convergence criterion is fulfilled then the iteration is stopped and the resulting (approximately) compatible pair is denoted by $\left(\widetilde{\boldsymbol{\mu}}^{hom}, \widetilde{\boldsymbol{\gamma}}^{hom}\right)$.

iv) Resulting reserves:

$$\widetilde{R}_i^{hom} = \widetilde{\mu}_i^{hom}\left(1 - \widetilde{\beta}_{\iota(i)}^{hom}\right), \quad i = I - J + 1, \ldots, I.$$

Remark 6.4. In our examples we used the following convergence criterion: The iteration is stopped after n iterations if for some given values ϵ_γ and ϵ_μ we have

$$\sqrt{\sum_{j=0}^{J}\left(\hat{\gamma}_j^{(n)} - \hat{\gamma}_j^{(n-1)}\right)^2} < \epsilon_\gamma \qquad \text{or} \qquad \sqrt{\sum_{i=0}^{I}\left(\widehat{\mu}_i^{(n)} - \widehat{\mu}_i^{(n-1)}\right)^2} < \epsilon_\mu. \quad (6.5)$$

6.2 Compatible Parameter Pairs

The iterative estimation procedure raises the question of existence and uniqueness of the compatible pair. In the special case $\tau^2 \to 0$ no iteration is necessary and we have a unique solution. In the inhomogeneous case we then have $\tilde{\mu}_i = p_i\theta_0$ and

$$\tilde{\gamma}_j = \hat{\gamma}_j^{(0)} = \frac{X_{[I-j],j}}{p_{[I-j]}}\left(\sum_{l=0}^{J}\frac{X_{[I-l],l}}{p_{[I-l]}}\right)^{-1}.$$

In the homogeneous case we obtain (cf. Remark 4.6)

$$\begin{aligned}
\tilde{\mu}_i = p_i\tilde{\theta}_0 = p_i\hat{\theta}_0^{(1)} &= p_i\frac{\sum_{i=0}^{I}\sum_{j=0}^{\iota(i)}\frac{X_{[I-j],j}}{p_{[I-j]}}\left(\sum_{l=0}^{J}\frac{X_{[I-l],l}}{p_{[I-l]}}\right)^{-1}\frac{X_{i,j}}{\tilde{\delta}_j^2}}{\sum_{i=0}^{I}p_i\sum_{j=0}^{\iota(i)}\frac{(X_{[I-j],j})^2}{(p_{[I-j]})^2\tilde{\delta}_j^2}\left(\sum_{l=0}^{J}\frac{X_{[I-l],l}}{p_{[I-l]}}\right)^{-2}} \\
&= p_i\left(\sum_{l=0}^{J}\frac{X_{[I-l],l}}{p_{[I-l]}}\right)\frac{\sum_{j=0}^{J}\sum_{i=0}^{I-j}\frac{X_{[I-j],j}}{p_{[I-j]}}\frac{X_{i,j}}{\tilde{\delta}_j^2}}{\sum_{j=0}^{J}\sum_{i=0}^{I-j}p_i\frac{(X_{[I-j],j})^2}{(p_{[I-j]})^2\tilde{\delta}_j^2}} \\
&= p_i\sum_{l=0}^{J}\frac{X_{[I-l],l}}{p_{[I-l]}},
\end{aligned}$$

and $\tilde{\gamma}_j = \hat{\gamma}_j^{(0)}$ as above. The same formulas are obtained in the ODP case. In general the convergence of the algorithm is not as simple to show.

However in our numerical examples the algorithm always converged if we used the simple normalisation (6.2). Moreover, the solutions did not depend on the starting values of the pattern that is, we received the same results, if we started with some arbitrary initial pattern $\hat{\boldsymbol{\gamma}}^{(0)}$. In our numerical calculations we have also tried the normalisation given in (6.3), but then we did not always obtain convergence.

Under Model Assumptions 3.4 (Independence Model) the compatible pair in the homogeneous case is automatically normalised. In other words, the raw estimates coincide with the pattern obtained as the solution from the compatible pair $(\tilde{\boldsymbol{\mu}}^{hom}, \tilde{\boldsymbol{\gamma}}^{hom})$ that is,

$$\tilde{\gamma}_j^{hom} \tilde{\mu}_{[I-j]}^{hom} = X_{[I-j],j}.$$

Consequently we have the balance property in each column j. This is shown by the following theorem.

Theorem 6.5. *Let $\hat{\gamma}_j$ denote an estimate for the claims development pattern, which is normalised using* (6.2)*,* (6.3) *or* (6.4)*. Under Model Assumptions 3.4 (Independence Model) the compatible pair $(\tilde{\boldsymbol{\mu}}^{hom}, \tilde{\boldsymbol{\gamma}}^{hom})$ satisfies for $0 \leq j \leq J$*

$$\tilde{\gamma}_j^{hom} = \frac{X_{[I-j],j}}{\tilde{\mu}_{[I-j]}^{hom}}.$$

The proof of Theorem 6.5 is given in Appendix C.

Theorem 6.5 states that if there are compatible estimates $\tilde{\gamma}_j^{hom}$ and $\tilde{\mu}_i^{hom}$, where $\tilde{\gamma}_j$ are normalised using (6.2), (6.3) or (6.4), then

$$\tilde{\gamma}_j^{hom} = \frac{X_{[I-j],j}}{\tilde{\mu}_{[I-j]}^{hom}}, \tag{6.6}$$

and hence no normalisation is necessary. Moreover, presupposed that for each normalisation there is a unique solution for the compatible pair, these solutions coincide for all the three normalisations. In all our numerical calculations the algorithm converged to the same values independent of the normalisations used. Moreover there are as many variables as equations for the compatible pair. Hence there is some plausibility for the uniqueness of the solution (however the system of equations is not linear).

Let us have a closer look at the ODP case under Model Assumptions 3.4 (Independence Model). In this case we have

$$B_i^{(n)} = \frac{C_{i,\iota(i)}}{p_i \hat{\beta}_{\iota(i)}^{(n-1)}}, \quad \widehat{w}_j^{(n)} = \hat{\gamma}_j^{(n)} \quad \text{and} \quad \alpha_i^{(n)} = \frac{p_i \hat{\beta}_{\iota(i)}^{(n-1)}}{p_i \hat{\beta}_{\iota(i)}^{(n-1)} + \frac{\widehat{\sigma^2 \theta_0}^{(n)}}{\hat{\tau}^{2(n)}}}.$$

Moreover, if $p_i = \hat{C}_{i,J}^{CL}$, then the algorithm delivers the CL pattern and reserves (see Appendix B).
In the case $\tau^2 \to \infty$ the algorithm yields the CL reserve and the CL pattern, which is in line with the well-known result of the BF reserve in the ODP case, where the $\hat{\mu}_i$ and $\hat{\gamma}_j$ are the ML estimators (see Hachemeister-Stanard, 1975, Mack, 1991 and Renshaw-Verrall, 1998). An alternative proof, which also shows that the CL reserve is the unique solution, is given in Appendix B. Hence in this special case we have a unique solution for the compatible pair.

7 Mean Square Error of Prediction (MSEP)

7.1 Introductory Remarks

There are different ways to look at the MSEP: the conditional view and the unconditional view. The *conditional MSEP* for all accident years is the MSEP given the data $\mathcal{D}_I$ at time I and it is denoted by

$$\operatorname{msep}_{\sum_{i=I-J+1}^{I} R_i|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I}\tilde{R}_i\right) = E\left[\left(\sum_{i=I-J+1}^{I}\left(R_i-\tilde{R}_i\right)\right)^2\middle|\mathcal{D}_I\right],$$

whereas the *unconditional MSEP* is given by

$$\operatorname{msep}_{\sum_{i=I-J+1}^{I} R_i}\left(\sum_{i=I-J+1}^{I}\tilde{R}_i\right) = E\left[\left(\sum_{i=I-J+1}^{I}\left(R_i-\tilde{R}_i\right)\right)^2\right].$$

In claims reserving we are interested in the prediction of outstanding liabilities given the specific data $\mathcal{D}_I$ and in assessing the accuracy of this prediction. Hence it is the conditional MSEP we are interested in and which we want to estimate. However the conditional MSEP depends on the conditional distribution which is usually unknown. In the following we split the conditional MSEP into different additive terms. If we approximate some of these terms by the unconditional means we are able to derive distribution-free estimators for these terms. There will remain only one term which we have to estimate by parametric simulation and where we have to specify some underlying distributional assumptions.

7.2 Decomposition and Closed Formula

For any estimator $\hat{\mu}_i = p_i\hat{\Theta}_i^{hom}$ or $\hat{\mu}_i = p_i\hat{\Theta}_i^{inh}$ we denote the final estimates resulting from the iteration by $\tilde{\mu}_i$, $\tilde{\Theta}_i$, $\tilde{\beta}_{\iota(i)}$, $\tilde{\tau}^2$, $\widetilde{\sigma^2\theta_0}$ and $\tilde{\theta}_0$. The corresponding

reserve is denoted by $\tilde{R}_i$ and we recall that $R_i = \sum_{j=I-i+1}^{J} X_{i,j}$. We define

$$R_i(\Theta_i) = p_i\Theta_i(1-\beta_{\iota(i)}) = E[R_i|\Theta_i], \quad R_i^{Bayes} = p_i E[\Theta_i|\mathcal{D}_I](1-\beta_{\iota(i)}),$$
$$R_i^{cred} = p_i\Theta_i^{inh}(1-\beta_{\iota(i)}),$$

where Θ_i^{inh} is the inhomogeneous credibility estimator as given in Theorem 4.1. The conditional MSEP, given all available data $\mathcal{D}_I$, is given by

$$\begin{aligned}
&\mathrm{msep}_{\sum_{i=I-J+1}^{I} R_i|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I}\tilde{R}_i\right) = E\left[\left(\sum_{i=I-J+1}^{I}\left(R_i-\tilde{R}_i\right)\right)^2\middle|\mathcal{D}_I\right]\\
&= E\left[\left(\sum_{i=I-J+1}^{I}\left(R_i-R_i^{cred}\right)\right)^2\middle|\mathcal{D}_I\right] + \left(\sum_{i=I-J+1}^{I}\left(R_i^{cred}-\tilde{R}_i\right)\right)^2\\
&\quad + 2\left(\sum_{i=I-J+1}^{I}\left(R_i^{cred}-\tilde{R}_i\right)\right)\left(\sum_{i=I-J+1}^{I}\left(R_i^{Bayes}-R_i^{cred}\right)\right).
\end{aligned}$$

Since incremental claims from different accident years are conditionally independent, given $\boldsymbol{\Theta}$, we obtain

$$\begin{aligned}
&E\left[\left(\sum_{i=I-J+1}^{I}\left(R_i-R_i^{cred}\right)\right)^2\middle|\mathcal{D}_I\right]\\
&= \sum_{i=I-J+1}^{I} E\left[\mathrm{Var}\left(R_i|\,\Theta_i\right)|\,\mathcal{D}_I\right] + E\left[\left(\sum_{i=I-J+1}^{I}\left(R_i(\Theta_i)-R_i^{cred}\right)\right)^2\middle|\mathcal{D}_I\right]\\
&= \sum_{i=I-J+1}^{I}\sum_{j=I-i+1}^{J} p_i\sigma^2\delta_j^2 E\left[\Theta_i|\mathcal{D}_I\right] + E\left[\left(\sum_{i=I-J+1}^{I}\left(R_i(\Theta_i)-R_i^{cred}\right)\right)^2\middle|\mathcal{D}_I\right],
\end{aligned}$$

and therefore

$$\begin{aligned}
\mathrm{msep}_{\sum_{i=I-J+1}^{I} R_i|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I}\tilde{R}_i\right) &= \underbrace{\sum_{i=I-J+1}^{I}\sum_{j=I-i+1}^{J} p_i\sigma^2\delta_j^2 E\left[\Theta_i|\mathcal{D}_I\right]}_{\sum_{i=I-J+1}^{I}\mathrm{PV}_i}\\
&\quad + \underbrace{E\left[\left(\sum_{i=I-J+1}^{I}\left(R_i(\Theta_i)-R_i^{cred}\right)\right)^2\middle|\mathcal{D}_I\right]}_{\mathrm{EEI}} + \underbrace{\left(\sum_{i=I-J+1}^{I}\left(R_i^{cred}-\tilde{R}_i\right)\right)^2}_{\mathrm{EEII}}\\
&\quad + \underbrace{2\left(\sum_{i=I-J+1}^{I}\left(R_i^{cred}-\tilde{R}_i\right)\right)\left(\sum_{i=I-J+1}^{I}\left(R_i^{Bayes}-R_i^{cred}\right)\right)}_{IV}. \qquad (7.1)
\end{aligned}$$

In the first term $\sum_{i=I-J+1}^{I} \mathrm{PV}_i$, which describes the process variance, we approximate $E[\Theta_i|\mathcal{D}_I]$ by $\tilde{\Theta}_i$. Defining a matrix Q and a vector $\boldsymbol{v} = (v_0, \ldots, v_I)'$ by

$$\begin{aligned}(Q)_{i,k} &= \left(\Theta_i - \Theta_i^{inh}\right)\left(\Theta_k - \Theta_k^{inh}\right), \quad 0 \le i,k \le I,\\ v_i &= p_i(1-\beta_{\iota(i)}),\end{aligned}$$

the term EEI in (7.1) can be written as

$$\mathrm{EEI} = E\left[\left(\sum_{i=I-J+1}^{I}\left(R_i(\Theta_i) - R_i^{cred}\right)\right)^2 \middle| \mathcal{D}_I\right] = \boldsymbol{v}' E[Q|\mathcal{D}_I]\boldsymbol{v}.$$

In general there is no explicit formula for $E[Q|\mathcal{D}_I]$ and if there is any, it will heavily depend on the distributional assumptions with regard to the Θ_i's as well as with regard to the conditional distributions of the $X_{i,j}$ given Θ_i. These distributions are usually not known in practice. Therefore we approximate $E[Q|\mathcal{D}_I]$ by the unconditional expected value which is given by the quadratic loss matrix

$$\begin{aligned}&E\left[\left(\boldsymbol{\Theta} - \boldsymbol{\Theta}^{\boldsymbol{inh}}\right)\left(\boldsymbol{\Theta} - \boldsymbol{\Theta}^{\boldsymbol{inh}}\right)'\right]\\ &= E[(A(\boldsymbol{\Theta} - \boldsymbol{B}) + (I-A)(\boldsymbol{\Theta} - \mathbf{1}\theta_0))(A(\boldsymbol{\Theta} - \boldsymbol{B}) + (I-A)(\boldsymbol{\Theta} - \mathbf{1}\theta_0))']\\ &= ASA' + (I-A)T(I-A)'\\ &= T(T+S)^{-1}S(T+S)^{-1}T' + S(T+S)^{-1}T(T+S)^{-1}S\\ &= S(T+S)^{-1}T\\ &= (I-A)T,\end{aligned}$$

where we used that $T(T+S)^{-1}S = S^{-1} + T^{-1}$ and hence we have the symmetry $T(T+S)^{-1}S = S(T+S)^{-1}T$. Given $\mathcal{D}_I$ the term EEII in (7.1) is a constant, but since R_i^{cred} is not known, we have to estimate it. This has to be done by simulation. Finally, we assume that R_i^{cred} is close to R_i^{Bayes} so that the last mixed term IV in (7.1) is negligible. We then arrive at the following approximation for the conditional MSEP of accident year i:

$$\mathrm{msep}_{R_i|\mathcal{D}_I}\left(\tilde{R}_i\right) \approx \mathrm{PV}_i + \mathrm{EEI}_i + \mathrm{EEII}_i,$$

where

$$\mathrm{PV}_i = \text{Process Variance} = E[\mathrm{Var}(R_i|\Theta_i)|\mathcal{D}_I],$$

which we approximate by $p_i\tilde{\Theta}_i\sigma^2\sum_{j=I-i+1}^{J}\delta_j^2$,

$$\mathrm{EEI}_i = \text{estimation error I} = E\left[\left(R_i(\Theta_i) - R_i^{cred}\right)^2 \middle| \mathcal{D}_I\right],$$

which we approximate by $p_i^2(1-\beta_{\iota(i)})^2((I-A)T)_{i,i}$ and

$$\text{EEII}_i = \text{estimation error II} = \left(R_i^{cred} - \tilde{R}_i\right)^2.$$

Similarly, we approximate $\text{msep}_{\sum_{i=I-J+1}^{I} R_i|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I}\tilde{R}_i\right)$ by

$$\sum_{i=I-J+1}^{I} \text{PV}_i + \text{EEI} + \text{EEII},$$

where we approximate EEI by

$$\boldsymbol{v}'(I-A)T\boldsymbol{v}. \tag{7.2}$$

Inserting the parameter estimates we obtain the estimates

$$\widehat{\text{PV}}_i = p_i\tilde{\Theta}_i\widetilde{\sigma^2\theta_0}/\tilde{\theta}_0 \sum_{j=I-i+1}^{J}\tilde{\delta}_j^2, \tag{7.3}$$

$$\widehat{\text{EEI}}_i = p_i^2\left(1-\tilde{\beta}_{\iota(i)}\right)^2\left(\left(I-\tilde{A}\right)\tilde{T}\right)_{i,i}. \tag{7.4}$$

Observe that the closed formula for PV_i+EEI_i is a lower bound for the conditional MSEP of accident year i, which is obtained without any distributional assumptions.

7.3 Parametric Bootstrap Simulation

The terms EEII_i and EEII have to be estimated by a parametric bootstrap simulation. For this purpose we make the following distributional assumptions

$$\begin{aligned}\bar{\boldsymbol{\Theta}} &\sim \text{N}_{I+1}(\mathbf{1}\bar{\theta}_0, \bar{T}), \quad \Theta_i = |\bar{\Theta}_i|,\\ X_{i,j}|\Theta_i &\sim \text{N}(p_i\Theta_i\gamma_j, p_i\Theta_i\sigma^2\delta_j^2),\end{aligned}$$

where $\bar{\theta}_0$ and $\bar{T}$ are chosen such that $E[\boldsymbol{\Theta}] = \mathbf{1}\theta_0$ and $\text{Cov}(\boldsymbol{\Theta}) = T$. We introduce $\bar{\boldsymbol{\Theta}} = (\bar{\Theta}_0, \ldots, \bar{\Theta}_I)$ to ensure positivity of the Θ_i. For the values we consider in applications, the differences $\bar{\theta}_0 - \theta_0$ and $\bar{T} - T$ are negligible and will be set to zero. The covariance matrix T is estimated using the correlations in (5.6) and the variance estimate $\tilde{\tau}^2$ resulting from the chosen iterative algorithm.

Simulation Steps:
With the compatible estimates $(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\gamma}})$ we have the following steps in the s-th simulation $(s = 1, \ldots, S)$:

1. For $0 \leq i \leq I$ and $0 \leq j \leq I$ generate

$$\bar{\boldsymbol{\Theta}}^{(s)} \sim \mathrm{N}_{I+1}\left(\mathbf{1}\tilde{\theta}_0, \tilde{T}\right), \quad \Theta_i^{(s)} = \left|\bar{\Theta}_i^{(s)}\right|$$
$$X_{i,j}^{(s)} \sim \mathrm{N}\left(p_i \Theta_i^{(s)} \tilde{\gamma}_j, p_i \Theta_i^{(s)} \frac{\widetilde{\sigma^2\theta_0}}{\tilde{\theta}_0} \tilde{\delta}_j^2\right), \quad i + j \leq I.$$

2. For $0 \leq i \leq I$ and $0 \leq j \leq J$ calculate the compatible pair $\tilde{\mu}_i^{(s)}$, $\tilde{\gamma}_j^{(s)}$ with Iterative Algorithm 6.3 (adjusted to the chosen framework) using the new data $\mathcal{D}_I^{(s)} = \{X_{i,j}^{(s)}, i + j \leq I, j \leq J\}$.

3. For $0 \leq i \leq I$ calculate

$$\boldsymbol{\Theta}^{\boldsymbol{inh(s)}} = \tilde{A}\boldsymbol{B} + \left(I - \tilde{A}\right)\mathbf{1}\tilde{\theta}_0,$$

where $\tilde{A}$ is obtained from A in Theorem 4.1 by replacing the parameters by the estimates from the original data $\mathcal{D}_I$. These parameters are the true parameters for the simulated data $\mathcal{D}_I^{(s)}$.

4. Calculate for $I - J + 1 \leq i \leq I$

$$\Delta_i^{(s)} = \left(p_i \Theta_i^{inh(s)} \left(1 - \tilde{\beta}_{\iota(i)}\right) - \tilde{\mu}_i^{(s)} \left(1 - \tilde{\beta}_{\iota(i)}^{(s)}\right)\right)^2,$$

and calculate

$$\Delta^{(s)} = \left(\sum_{i=I-J+1}^{I} \left(p_i \Theta_i^{inh(s)} \left(1 - \tilde{\beta}_{\iota(i)}\right) - \tilde{\mu}_i^{(s)} \left(1 - \tilde{\beta}_{\iota(i)}^{(s)}\right)\right)\right)^2.$$

With S simulations the estimation errors EEII_i and EEII are estimated as

$$\widehat{\mathrm{EEII}}_i = \frac{1}{S}\sum_{s=1}^{S} \Delta_i^{(s)} \quad \text{and} \quad \widehat{\mathrm{EEII}} = \frac{1}{S}\sum_{s=1}^{S} \Delta^{(s)}.$$

The total conditional MSEP is estimated by

$$\widehat{\mathrm{msep}}_{\sum_{i=I-J+1}^{I} R_i | \mathcal{D}_I}\left(\sum_{i=I-J+1}^{I} \tilde{R}_i\right) = \sum_{i=I-J+1}^{I} \widehat{\mathrm{PV}}_i + \widehat{\mathrm{EEI}} + \widehat{\mathrm{EEII}},$$

where $\widehat{\mathrm{PV}}_i$ is given in (7.3) and $\widehat{\mathrm{EEI}}$ is obtained from (7.2) by inserting the parameter estimates. Similarly, the conditional MSEP of a single accident year $I - J + 1 \leq i \leq I$ is estimated by

$$\widehat{\mathrm{msep}}_{R_i | \mathcal{D}_I}\left(\tilde{R}_i\right) = \widehat{\mathrm{PV}}_i + \widehat{\mathrm{EEI}}_i + \widehat{\mathrm{EEII}}_i,$$

with $\widehat{\mathrm{PV}}_i$ and $\widehat{\mathrm{EEI}}_i$ given in (7.3) and (7.4).

Remark 7.1. In a full Bayesian model with distributional assumptions, one could also calculate the total conditional MSEP by simulation. However, the results will then heavily depend on the distributional assumptions. With our approach a distributional assumption is only needed for EEII_i and EEII.

8 Example

We consider a data set of payments from liability insurance, which is given in Table 6. For the iterative algorithm we use in all cases the convergence criterion (6.5) with $\epsilon_\gamma = \epsilon_\mu = 10^{-10}$ and additionally set an upper bound of 1000 iterations to find the compatible pairs in each step. The premiums are given in Table 1 together with the initial pattern $\hat{\gamma}_j^{(0)}$, the 'initial reserves' $\hat{R}_i = p_i(1-\hat{\beta}_{\iota(i)}^{(0)})$ and the estimates for the δ_j^2's. For reasons of confidentiality the figures are multiplied by a constant.

i, j	$\tilde{\delta}_j^2$	p_i	$\hat{\gamma}_j^{(0)}$	$p_i(1-\hat{\beta}_{\iota(i)}^{(0)})$
0	69.58%	21'472	60.03%	0
1	19.74%	22'219	29.14%	301
2	1.73%	22'873	3.85%	406
3	2.96%	23'833	2.30%	578
4	1.58%	22'794	1.23%	786
5	2.37%	23'265	1.02%	1'088
6	1.56%	23'518	0.65%	1'642
7	0.38%	28'128	0.42%	3'047
8	0.09%	25'628	1.35%	10'244
total				**18'093**

Table 1: estimated parameters and reserves without iteration

Independent Θ_i

Homogeneous Case (Unknown θ_0)

We first consider the case where θ_0 is unknown and where the Θ_i's are independent. Under Model Assumptions 3.4 (Independence Model) we obtain the following results with Iterative Algorithm 6.3. The compatible pair is found after 6 iterations and we have the following estimates

$$\tilde{\theta}_0 = 0.884, \quad \tilde{\tau} = 1.62\% \quad \text{and} \quad \widetilde{\sigma\sqrt{\theta_0}} = 6.05.$$

The $\tilde{\delta}_j^2$ and the starting pattern $\hat{\gamma}_j^{(0)}$ are as in Table 1. The compatible estimates $\tilde{\mu}_i^{hom}$ and $\tilde{\gamma}_j^{hom}$ and the corresponding reserves $\tilde{R}_i^{hom}$ are given in Table 2. The MSEP results are also given in Table 2.

i,j	$\tilde{\mu}_i^{hom}$	$\tilde{\gamma}_j^{hom}$	$\tilde{R}_i^{hom}$	$\sqrt{\widehat{\text{msep}}}$	$\widehat{\text{CoVa}}$	$\sqrt{\widehat{\text{PV}}_i}$	$\sqrt{\widehat{\text{EEI}}_i}$	$\sqrt{\widehat{\text{EEII}}_i}$
0	18'932	60.03%	0	0	0		0	0
1	19'567	29.14%	266	40	15%	28	5	29
2	20'228	3.85%	360	82	23%	63	6	51
3	21'126	2.30%	514	160	31%	133	9	88
4	20'357	1.23%	704	222	32%	193	12	110
5	20'806	1.03%	974	260	27%	227	16	125
6	20'665	0.65%	1'442	312	22%	276	25	142
7	24'618	0.42%	2'665	380	14%	330	45	182
8	22'661	1.36%	9'058	614	7%	534	158	259
total			**15'983**	**1'120**	**7.0%**	**762**	**168**	**803**

Table 2: Unknown θ_0: compatible pair $(\tilde{\boldsymbol{\mu}}^{hom}, \tilde{\boldsymbol{\gamma}}^{hom})$ with corresponding reserves, MSEP, coefficient of variation, process standard deviation and root of the estimation error terms.

Note that $\tilde{\theta}_0 = 0.884 < 1$ indicates that the losses are systematically lower than the premiums. Consequently, the 'initial reserves' in Table 1 are larger than the reserves in Table 2. The effect of the iterative algorithm on the pattern is negligible in this example.

Inhomogeneous Case (Assuming $\theta_0 = 1$)

Assuming $\theta_0 = 1$ that is, assuming that the p_i's are the pure risk premiums, we have $\hat{\theta}_0^{(n)} \equiv \theta_0 = 1$. The estimation of τ^2 in Iterative Algorithm 6.3 is replaced by the formulas resulting from (5.4). We need 10 iterations to find the compatible pair. For the structural parameters we obtain $\tilde{\tau} = 11.62\%$ and $\tilde{\sigma} = 6.04$. The results are summarised in Table 3. The compatible estimates are denoted by $\tilde{\mu}_i^{inh}$, $\tilde{\gamma}_j^{inh}$ and let $\tilde{R}_i^{inh}$ be the corresponding reserves.

i,j	$\tilde{\mu}_i^{inh}$	$\tilde{\gamma}_j^{inh}$	$\tilde{R}_i^{inh}$	$\sqrt{\widehat{\text{msep}}}$	$\widehat{\text{CoVa}}$	$\sqrt{\widehat{\text{PV}}_i}$	$\sqrt{\widehat{\text{EEI}}_i}$	$\sqrt{\widehat{\text{EEII}}_i}$
0	18'864	60.02%	0	0	0		0	0
1	19'423	29.17%	268	50	18%	26	11	41
2	20'507	3.82%	372	87	23%	60	15	61
3	21'648	2.26%	535	161	30%	127	21	96
4	21'648	1.22%	758	223	29%	187	29	119
5	22'256	1.03%	1'051	264	25%	221	40	139
6	20'268	0.66%	1'416	310	22%	257	60	163
7	23'720	0.43%	2'563	396	15%	304	107	231
8	23'153	1.38%	9'256	946	10%	507	490	630
total			**16'218**	**1'363**	**8.4%**	**721**	**508**	**1'039**

Table 3: Assumption $\theta_0 = 1$: compatible $(\tilde{\boldsymbol{\mu}}^{inh}, \tilde{\boldsymbol{\gamma}}^{inh})$ and corresponding reserves, MSEP, coefficient of variation, process standard deviation and root of the estimation error terms.

Despite the difference of the assumption $\theta_0 = 1$ to the estimate $\tilde{\theta}_0 = 0.884$ from the homogeneous case, the difference between the reserves in Table 2 and Table 3 is surprisingly small. This is due to the estimates for $\tilde{\tau}$: in the inhomogeneous case $\tilde{\tau}$ is much larger than in the homogeneous case and hence the data get more weight in the inhomogeneous case. The differences in the pattern are negligible.

In the case of independent Θ_i, the first estimation error terms are given by

$$\begin{aligned}\mathrm{EEI}_i &= \big(p_i(1-\beta_{\iota(i)})\big)^2 (1-\alpha_i)\tau^2 \\ &= \big(p_i(1-\beta_{\iota(i)})\big)^2 \frac{\sigma^2\theta_0}{p_i w_{[\iota(i)]} + \frac{\sigma^2\theta_0}{\tau^2}}, \\ \mathrm{EEI} &= \sum_{i=I-J+1}^{I} \mathrm{EEI}_i.\end{aligned}$$

Since in the case of unknown θ_0 the estimate for τ^2 is much smaller than in the case where we assume $\theta_0 = 1$, we have larger values for EEI_i and EEI in Table 3 compared to the numbers in Table 2.

Dependent Θ_i

In the case of dependent Θ_i, the covariance matrix T is estimated as in Iterative Algorithm 6.3 with $m = 10$ in the estimates $\hat{\rho}_{i,k}$.

Homogeneous Case (Unknown θ_0)

Using Iterative Algorithm 6.3, the compatible pair is found after 7 iterations and for the structural parameters we obtain

$$\tilde{\theta}_0 = 0.879, \quad \tilde{\tau} = 2.86\% \quad \text{and} \quad \widetilde{\sigma\sqrt{\theta_0}} = 6.04.$$

The results are summarised in Table 4. Comparing the results for the conditional MSEP with the results in Table 2, we see that in this example, the dependencies of the Θ_i's have hardly an impact on the conditional MSEP.

Inhomogeneous Case (Assuming $\theta_0 = 1$)

In the case where θ_0 is known, we replace the estimate for $\hat{\mu}_i^{(n)}$ in Iterative Algorithm 6.3 by the corresponding formula from Theorem 4.1. The estimate

i,j	$\tilde{\mu}_i^{hom}$	$\tilde{\gamma}_j^{hom}$	$\tilde{R}_i^{hom}$	$\sqrt{\widehat{\text{msep}}}$	$\widehat{\text{CoVa}}$	$\sqrt{\widehat{\text{PV}}_i}$	$\sqrt{\widehat{\text{EEI}}_i}$	$\sqrt{\widehat{\text{EEII}}_i}$
0	18'908	60.06%	0	0	0		0	0
1	19'598	29.12%	266	40	15%	28	5	28
2	20'264	3.84%	361	82	23%	63	7	51
3	21'209	2.30%	516	160	31%	134	9	87
4	20'354	1.22%	704	223	32%	193	13	110
5	20'717	1.02%	970	260	27%	227	17	125
6	20'728	0.65%	1'446	313	22%	278	27	143
7	24'641	0.42%	2'666	382	14%	331	53	183
8	22'461	1.36%	8'972	627	7%	533	200	262
total			**15'901**	**1'152**	**7.2%**	**763**	**281**	**816**

Table 4: Unknown θ_0: compatible pair $(\tilde{\boldsymbol{\mu}}^{hom}, \tilde{\boldsymbol{\gamma}}^{hom})$ with corresponding reserves, MSEP, coefficient of variation, process standard deviation and root of the estimation error terms.

of $\hat{\tau}^{2^{(n)}}$ is adjusted according to formula (5.4). The compatible pair is found after 9 iterations and for the structural parameters we obtain

$$\tilde{\tau} = 11.60\% \quad \text{and} \quad \tilde{\sigma} = 6.04.$$

The results are summarised in Table 5.

i,j	$\tilde{\mu}_i^{inh}$	$\tilde{\gamma}_j^{inh}$	$\tilde{R}_i^{inh}$	$\sqrt{\widehat{\text{msep}}}$	$\widehat{\text{CoVa}}$	$\sqrt{\widehat{\text{PV}}_i}$	$\sqrt{\widehat{\text{EEI}}_i}$	$\sqrt{\widehat{\text{EEII}}_i}$
0	18'788	60.09%	0	0	0		0	0
1	19'359	29.11%	265	41	15%	26	9	30
2	20'280	3.82%	365	81	22%	59	12	53
3	21'561	2.27%	529	157	30%	127	17	90
4	21'111	1.22%	735	218	30%	184	23	115
5	21'435	1.03%	1'007	255	25%	217	32	131
6	20'263	0.66%	1'413	303	21%	257	47	154
7	24'077	0.43%	2'600	384	15%	306	88	215
8	22'347	1.37%	8'918	771	9%	498	397	435
total			**15'833**	**1'286**	**8.1%**	**714**	**452**	**969**

Table 5: Assumption $\theta_0 = 1$: compatible $(\tilde{\boldsymbol{\mu}}^{inh}, \tilde{\boldsymbol{\gamma}}^{inh})$ and corresponding reserves, MSEP, coefficient of variation, process standard deviation and root of the estimation error terms.

Note that the reserves in this example are even slightly smaller than the reserves in Table 4, which is again due to a smaller value of the estimate for $\tilde{\tau}$ in the homogeneous case.

Appendices

A Proof of the Credibility Formulas

Proof of Theorem 4.1: We first derive the inhomogeneous credibility estimators of the Θ_i's based on the B_l's. We denote these credibility estimators by the vector $\hat{\boldsymbol{\Theta}}^{\boldsymbol{inh}} = A\boldsymbol{B} + \boldsymbol{a}$, where $\boldsymbol{a}$ is a constant vector of dimension $I+1$. The normal equations (see Bühlmann-Gisler, 2005) are given by

$$\begin{aligned} E\left[\hat{\boldsymbol{\Theta}}^{\boldsymbol{inh}}\right] &= AE\left[\boldsymbol{B}\right] + \boldsymbol{a} = \theta_0 \mathbf{1}, \\ \mathrm{Cov}\left(\boldsymbol{\Theta}, \boldsymbol{B}\right) &= A\mathrm{Cov}\left(\boldsymbol{B}\right). \end{aligned}$$

From the first equation we obtain $\boldsymbol{a} = (I-A)\theta_0\mathbf{1}$ and for the second equation we have

$$T = A(T+S),$$

which is satisfied by the credibility matrix $A = T(T+S)^{-1}$. It remains to prove that the credibility estimator based on all $Y_{i,j}$ depends on the observations only through the B_i's. Let

$$\hat{\Theta}_i = b_i\theta_0 + \sum_{l=0}^{I}\sum_{k=0}^{\iota(l)} a_{l,k}^{(i)} Y_{l,k}$$

denote the credibility estimator of Θ_i based on all $Y_{i,j}$. From the corresponding normal equations we obtain for all $0 \le r \le I$ and $0 \le j \le J$

$$b_i + \sum_{l=0}^{I}\sum_{k=0}^{\iota(l)} a_{l,k}^{(i)}\gamma_k = 1, \tag{A.1}$$

$$\sum_{l=0}^{I}\sum_{k=0}^{\iota(l)} a_{l,k}^{(i)}\gamma_k\gamma_j\tau_{l,r} + a_{r,j}^{(i)}\frac{\sigma^2\theta_0\delta_j^2}{p_r} = \gamma_j\tau_{i,r}. \tag{A.2}$$

We show that the coefficients of the $Y_{l,k}$'s in

$$\begin{aligned} \hat{\Theta}_i^{inh} &= \sum_{l=0}^{I}(A)_{i,l}B_l + \left(1 - \sum_{l=0}^{I}(A)_{i,l}\right)\theta_0 \\ &= \sum_{l=0}^{I}(A)_{i,l}\sum_{k=0}^{\iota(l)}\frac{\gamma_k/\delta_k^2}{w_{[\iota(l)]}}Y_{l,k} + \left(1 - \sum_{l=0}^{I}(A)_{i,l}\right)\theta_0, \end{aligned}$$

satisfy the normal equations (A.1) and (A.2). Inserting the coefficients $a_{l,k}^{(i)} = (A)_{i,l}\frac{\gamma_k/\delta_k^2}{w_{[\iota(l)]}}$ and $b_i = \left(1 - \sum_{l=0}^{I}(A)_{i,l}\right)$ we obtain (A.1), for (A.2) we have

$$\begin{aligned}
&\gamma_j \sum_{l=0}^{I}\sum_{k=0}^{\iota(l)}(A)_{i,l}\frac{\gamma_k^2/\delta_k^2}{w_{[\iota(l)]}}\tau_{l,r} + \gamma_j(A)_{i,r}\frac{\sigma^2\theta_0}{p_r w_{[I-r]}} \\
&= \gamma_j \sum_{l=0}^{I}(A)_{i,l}(T)_{l,r} + \gamma_j(A)_{i,r}(S)_{r,r} \\
&= \gamma_j(AT)_{i,r} + \gamma_j \sum_{k=0}^{I}(A)_{i,k}(S)_{k,r} \\
&= \gamma_j(A(T+S))_{i,r} \\
&= \gamma_j\left(T(T+S)^{-1}(T+S)\right)_{i,r} \\
&= \gamma_j\tau_{i,r}.
\end{aligned}$$

□

Proof of Corollary 4.2: If the Θ_i are iid the covariance matrix T is diagonal with $T = \tau^2 I$ and therefore the credibility matrix A is diagonal with

$$(A)_{i,i} = \frac{\tau^2}{\tau^2 + \frac{\sigma^2\theta_0}{p_i w_{[\iota(i)]}}} = \alpha_i.$$

Hence the claim follows from Theorem 4.1.

□

Proof of Theorem 4.4: Due to Theorem 4.1 the homogeneous credibility estimator of $\boldsymbol{\Theta}$ is of the form

$$\widehat{\boldsymbol{\Theta}}^{\boldsymbol{hom}} = A\boldsymbol{B} + (I - A)\mathbf{1}\mathrm{Pro}(\theta_0|L_e^{ind}(\boldsymbol{Y_0}, \ldots, \boldsymbol{Y_I})),$$

where

$$\begin{aligned}
&L_e^{ind}(\boldsymbol{Y_0}, \ldots, \boldsymbol{Y_I}) \\
&= \{\hat{\theta}_0 : \hat{\theta}_0 = \sum_{i=0}^{I}\sum_{j=0}^{\iota(i)} a_{i,j}Y_{i,j},\ a_{i,j} \in \mathbb{R},\ 0 \le i \le I,\ 0 \le j \le \iota(i),\ E[\hat{\theta}_0] = \theta_0\}.
\end{aligned}$$

Here $\mathrm{Pro}(\theta_0|L_e^{ind}(\boldsymbol{Y_0}, \ldots, \boldsymbol{Y_I}))$ is the orthogonal projection of θ_0 on the affine subspace $L_e^{ind}(\boldsymbol{Y_0}, \ldots, \boldsymbol{Y_I})$. The normal equations for

$$\hat{\theta}_0 = \mathrm{Pro}(\theta_0|L_e^{ind}(\boldsymbol{Y_0}, \ldots, \boldsymbol{Y_I}))$$

are given by

$$\mathbb{E}\left[\hat{\theta}_0\right] = \theta_0, \quad \mathrm{Cov}\left(\hat{\theta}_0, \hat{\hat{\theta}}_0\right) = \mathrm{Var}\left(\hat{\theta}_0\right),$$

for all $\hat{\hat{\theta}}_0 \in L_e^{ind}(\boldsymbol{Y_0}, \ldots, \boldsymbol{Y_I})$. If $\hat{\theta}_0 = \sum_{i=0}^{I} \sum_{j=0}^{\iota(i)} a_{i,j} Y_{i,j}$ it follows that

$$\sum_{i=0}^{I} \sum_{j=0}^{\iota(i)} a_{i,j}\gamma_j = 1,$$

$$\sum_{i=0}^{I} \sum_{j=0}^{\iota(i)} (a_{i,j} - b_{i,j})\gamma_j \left(\sum_{l=0}^{I} \sum_{k=0}^{\iota(l)} a_{l,k}\tau_{i,l}\gamma_k + a_{i,j} \frac{\sigma^2\theta_0}{p_i\gamma_j/\delta_j^2} \right) = 0, \tag{A.3}$$

for all $b_{i,j} \in \mathbb{R}$ with $\sum_{i=0}^{I} \sum_{j=0}^{\iota(i)} b_{i,j}\gamma_j = 1$. We show that the normal equations are satisfied by the coefficients of $Y_{i,j}$ in $\left(\mathbf{1}'(S+T)^{-1}\mathbf{1}\right)^{-1} \mathbf{1}'(S+T)^{-1}\boldsymbol{B}$, which are given by

$$a_{i,j} = \left(\mathbf{1}'(S+T)^{-1}\mathbf{1}\right)^{-1} \mathbf{1}'(S+T)^{-1}\boldsymbol{e}_i \frac{\gamma_j/\delta_j^2}{w_{[\iota(i)]}},$$

with $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0 \ldots, 0)'$, where the 1 is at position $i+1$. With these coefficients we obtain

$$\sum_{i=0}^{I} \sum_{j=0}^{\iota(i)} a_{i,j}\gamma_j = \left(\mathbf{1}'(S+T)^{-1}\mathbf{1}\right)^{-1} \mathbf{1}'(S+T)^{-1} \sum_{i=0}^{I} \boldsymbol{e}_i \sum_{j=0}^{\iota(i)} \frac{\gamma_j^2/\delta_j^2}{w_{[\iota(i)]}} = 1.$$

Moreover, we get

$$\begin{aligned}
&\sum_{l,k} a_{l,k}\tau_{i,l}\gamma_k + a_{i,j} \frac{\sigma^2\theta_0}{p_i\gamma_j/\delta_j^2} \\
&= \left(\mathbf{1}'(S+T)^{-1}\mathbf{1}\right)^{-1} \mathbf{1}'(S+T)^{-1} \left(\sum_{l=0}^{I} \boldsymbol{e}_l \tau_{i,l} \sum_{k=0}^{\iota(i)} \frac{\gamma_k^2/\delta_k^2}{w_{[\iota(k)]}} + \boldsymbol{e}_i \frac{\sigma^2\theta_0}{p_i w_{[\iota(i)]}} \right) \\
&= \left(\mathbf{1}'(S+T)^{-1}\mathbf{1}\right)^{-1} \mathbf{1}'(S+T)^{-1} \left(\sum_{l=0}^{I} \boldsymbol{e}_l \left((T)_{i,l} + (S)_{i,l} \right) \right) \\
&= \left(\mathbf{1}'(S+T)^{-1}\mathbf{1}\right)^{-1} \mathbf{1}'(S+T)^{-1} (T+S)\, \boldsymbol{e}_i \\
&= \left(\mathbf{1}'(S+T)^{-1}\mathbf{1}\right)^{-1}.
\end{aligned}$$

Since the last expression does not depend on i and j, the normal equation (A.3) is satisfied. This proves the claim.

□

Proof of Corollary 4.5: As in the proof of Corollary 4.2 the credibility matrix A is diagonal with

$$(A)_{i,i} = \alpha_i.$$

Similarly $(T+S)^{-1}$ is diagonal with

$$\left((T+S)^{-1}\right)_{i,i} = \frac{\alpha_i}{\tau^2},$$

and therefore

$$\hat{\theta}_0 = \left(\mathbf{1}'(S+T)^{-1}\mathbf{1}\right)^{-1}\mathbf{1}'(S+T)^{-1}\boldsymbol{B} = \sum_{i=0}^{I}\frac{\alpha_i}{\alpha_{[I]}}B_i.$$

Hence the claim follows from Theorem 4.4.

□

B Results in the ODP Case under Model Assumptions 3.4 (Independence Model)

In the following section we assume that Model Assumptions 3.4 (Independence Model) are satisfied and we consider the ODP case, that is we assume that $\gamma_j = \delta_j^2$.

If $\tau^2 \to \infty$ we obtain in the ODP case

$$\begin{aligned}\tilde{\mu}_i^{hom} &= \frac{C_{i,\iota(i)}}{\tilde{\beta}_{\iota(i)}^{hom}}, \\ \tilde{\gamma}_j^{hom} &= \frac{X_{[I-j],j}}{\tilde{\mu}_{[I-j]}^{hom}}.\end{aligned} \tag{B.1}$$

The first equality holds because $\alpha_i = 1$ in this case and the second equality is due to Theorem 6.5.

Theorem B.1. *In the case of a non-informative prior ($\tau^2 \to \infty$) we obtain for the compatible pair in the ODP case the CL results*

$$\tilde{\mu}_i^{hom} = \hat{C}_{i,J}^{CL} \quad \text{and} \quad \tilde{\gamma}_j^{hom} = \hat{\gamma}_j^{CL}.$$

Proof of Theorem B.1: First note that from the CL algorithm it follows that

$$C_{[I-j],j}\prod_{l=j}^{J-1}\hat{f}_l = \hat{C}^{CL}_{[I-j],J}, \tag{B.2}$$

$$C_{[I-j],j-1}\prod_{l=j-1}^{J-1}\hat{f}_l = \hat{C}^{CL}_{[I-j],J}, \tag{B.3}$$

where $\hat{f}_j = C_{[I-j-1],j+1}/C_{[I-j-1],j}$ are the CL factors. Moreover we have from (B.1) that $\tilde{\mu}^{hom}_i = C_{i,J} = \hat{C}^{CL}_{i,J}$ for $0 \le i \le I-J$. Therefore

$$\begin{aligned}\tilde{\gamma}^{hom}_J &= \frac{X_{[I-J],J}}{\tilde{\mu}^{hom}_{[I-J]}} = \frac{C_{[I-J],J} - C_{[I-J],J-1}}{C_{[I-J],J}} \\ &= 1 - \hat{f}^{-1}_{J-1} = \hat{\gamma}^{CL}_J.\end{aligned}$$

It follows that

$$\tilde{\beta}^{hom}_{J-1} = 1 - \tilde{\gamma}^{hom}_J = 1 - \hat{\gamma}^{CL}_J = \hat{\beta}^{CL}_{J-1} = \frac{1}{\hat{f}_{J-1}}.$$

From (B.1) we then obtain

$$\tilde{\mu}^{hom}_{I-J+1} = \frac{C_{I-J+1,J-1}}{\tilde{\beta}_{J-1}} = C_{I-J+1,J-1}\hat{f}_{J-1} = \hat{C}^{CL}_{I-J+1,J},$$

and therefore $\tilde{\mu}^{hom}_{[I-J+1]} = \hat{C}^{CL}_{[I-J+1],J}$. With (B.2) and (B.3) we get

$$\begin{aligned}\tilde{\gamma}^{hom}_{J-1} &= \frac{X_{[I-J+1],J-1}}{\tilde{\mu}^{hom}_{[I-J+1]}} = \frac{C_{[I-J+1],J-1} - C_{[I-J+1],J-2}}{\tilde{\mu}^{hom}_{[I-J+1]}} \\ &= \left(\hat{f}^{-1}_{J-1} - \prod_{l=J-2}^{J-1}\hat{f}^{-1}_l\right)\frac{\hat{C}^{CL}_{[I-J+1],J}}{\hat{C}^{CL}_{[I-J+1],J}} = \hat{\gamma}^{CL}_{J-1}.\end{aligned}$$

Iteratively we conclude $\tilde{\mu}^{hom}_i = \hat{C}^{CL}_{i,J}$ for $0 \le i \le I$ and $\tilde{\gamma}^{hom}_j = \hat{\gamma}^{CL}_j$ for $0 \le j \le J$.

□

Iteration with $p_i = \hat{C}^{CL}_{i,J}$:
If $p_i = \hat{C}^{CL}_{i,J}$ we derive, as in the proof of Theorem B.1, that

$$\hat{\gamma}^{(0)}_j = \frac{X_{[I-j],j}}{p_{[I-j]}} = \hat{\gamma}^{CL}_j,$$

and hence we obtain in the first iteration

$$B_i^{(1)} = 1 \quad \text{and} \quad \left(\hat{\theta}_0\right)^{(1)} = 1.$$

It follows that for any estimator $\hat{\mu}_i = \hat{\mu}_i^{inh}$ or $\hat{\mu}_i = \hat{\mu}_i^{hom}$ that $\hat{\mu}_i^{(1)} = \hat{\mu}_i^{(0)} = \hat{C}_{i,J}^{CL}$. We therefore have immediate convergence and obtain the CL algorithm.

C Proof of Theorem 6.5

We first consider the simple normalisation (6.2). It follows from the definition of $\tilde{\gamma}_j^{hom} = \hat{\gamma}_j^{raw}/c$ with $c = \sum_{j=0}^{J} \hat{\gamma}_j^{raw}$ that

$$\tilde{\mu}_{[I-j]}^{hom} \tilde{\gamma}_j^{hom} c = X_{[I-j],j}.$$

Multiplying by $\tilde{\gamma}_j^{hom}/\delta_j^2$ and taking the sum over all j we obtain

$$\sum_{j=0}^{J} c\tilde{\mu}_{[I-j]}^{hom} \frac{\left(\tilde{\gamma}_j^{hom}\right)^2}{\delta_j^2} = \sum_{j=0}^{J} X_{[I-j],j} \frac{\tilde{\gamma}_j^{hom}}{\delta_j^2} = \sum_{j=0}^{J} \tilde{\mu}_{[I-j]}^{hom} \frac{\left(\tilde{\gamma}_j^{hom}\right)^2}{\delta_j^2},$$

where the last equality is due to the balance property given in Theorem 4.7. It follows that $c = 1$ that is,

$$c = \sum_{j=0}^{J} \hat{\gamma}_j^{raw} = 1,$$

and hence $\tilde{\gamma}_j^{hom} = \hat{\gamma}_j^{raw}$ as claimed.

For the normalisation (6.3) it follows from the definition of $\tilde{\gamma}_j^{hom} = \hat{\gamma}_j^{raw} + \kappa_j$ that

$$\tilde{\mu}_{[I-j]}^{hom} \tilde{\gamma}_j^{hom} - \tilde{\mu}_{[I-j]}^{hom} \kappa_j = X_{[I-j],j}.$$

As above we multiply by $\tilde{\gamma}_j^{hom}/\delta_j^2$ and take the sum over all j to obtain

$$\sum_{j=0}^{J} \tilde{\mu}_{[I-j]}^{hom} \frac{\left(\tilde{\gamma}_j^{hom}\right)^2}{\delta_j^2} - \tilde{\mu}_{[I-j]}^{hom} \kappa_j \frac{\tilde{\gamma}_j^{hom}}{\delta_j^2} = \sum_{j=0}^{J} X_{[I-j],j} \frac{\tilde{\gamma}_j^{hom}}{\delta_j^2} = \sum_{j=0}^{J} \tilde{\mu}_{[I-j]}^{hom} \frac{\left(\tilde{\gamma}_j^{hom}\right)^2}{\delta_j^2}.$$

Therefore

$$\sum_{j=0}^{J} \kappa_j \frac{\tilde{\mu}_{[I-j]}^{hom} \tilde{\gamma}_j^{hom}}{\delta_j^2} = 0,$$

and with

$$\kappa_j = \frac{\delta_j^2/\tilde{\mu}_{[I-j]}^{hom}}{\sum_{l=0}^{J} \delta_l^2/\tilde{\mu}_{[I-l]}^{hom}} \left(1 - \sum_{j=0}^{J} \hat{\gamma}_j^{raw}\right),$$

we obtain

$$\frac{1}{\sum_{l=0}^{J}\delta_l^2/\tilde{\mu}_{[I-l]}^{hom}}\left(1-\sum_{j=0}^{J}\hat{\gamma}_j^{raw}\right)\sum_{j=0}^{J}\tilde{\gamma}_j^{hom}=\frac{1}{\sum_{l=0}^{J}\delta_l^2/\tilde{\mu}_{[I-l]}^{hom}}\left(1-\sum_{j=0}^{J}\hat{\gamma}_j^{raw}\right)$$
$$=0.$$

We conclude

$$\sum_{j=0}^{J}\hat{\gamma}_j^{raw}=1,$$

and hence $\tilde{\gamma}_j^{hom}=\hat{\gamma}_j^{raw}$ as claimed.
In case of the ODP normalisation (6.4) we obtain

$$\sum_{j=0}^{J}\tilde{\mu}_{[I-j]}^{hom}\frac{\left(\tilde{\gamma}_j^{hom}\right)^2}{\delta_j^2}+\kappa\frac{\left(\tilde{\gamma}_j^{hom}\right)^2}{\delta_j^2}=\sum_{j=0}^{J}X_{[I-j],j}\frac{\tilde{\gamma}_j^{hom}}{\delta_j^2}=\sum_{j=0}^{J}\tilde{\mu}_{[I-j]}^{hom}\frac{\left(\tilde{\gamma}_j^{hom}\right)^2}{\delta_j^2},$$

and therefore $\kappa=0$, as claimed.
In the ODP case we set $\tilde{\gamma}_j^{hom}=\delta_j^2$ and the proof is as above.

□

D Data

i	p_i	$X_{i,0}$	$X_{i,1}$	$X_{i,2}$	$X_{i,3}$	$X_{i,4}$	$X_{i,5}$	$X_{i,6}$	$X_{i,7}$	$X_{i,8}$
0	21'472	10'236	5'695	776	571	277	197	261	124	257
1	22'219	11'599	6'047	483	376	96	58	98	39	
2	22'873	13'079	5'369	872	232	119	426	23		
3	23'833	12'845	6'303	773	456	368	138			
4	22'794	13'549	5'819	760	731	369				
5	23'265	12'742	6'813	855	413					
6	23'518	11'896	5'585	930						
7	28'128	13'894	6'834							
8	25'628	13'595								

Table 6: Incremental payments and premiums p_i from liability insurance.

Paper

C

Annina Saluz, Alois Gisler.

Best Estimate Reserves and the Claims Development Result in Consecutive Calendar Years.

Annals of Actuarial Science, **8**(2), 2014, 351–373.

Best Estimate Reserves and the Claims Development Results in Consecutive Calendar Years

Annina Saluz Alois Gisler

ETH Zurich, RiskLab Switzerland, Department of Mathematics,
8092 Zurich, Switzerland

Abstract

The claims development result (CDR) is the difference between the best estimate predictions of the ultimate claim in two successive years. With best estimate reserves it is often argued that CDRs in consecutive years should fluctuate randomly around zero. However, in practice one frequently observes that CDRs in a given line of business have the same sign over several consecutive years. We show that this is a phenomenon which is not unusual and to be expected in situations of change. Moreover, we show how situations of change can adequately be described by a model taking into account the evolving external information.

Keywords. Claims Reserving, Best Estimate Reserves, Claims Development Result, Diagonal Effects.

1 Introduction

1.1 A Fundamental Question

Accurate claims reserves are essential for an insurance company. If the reserves are wrong, the premiums will be wrong, which would have serious consequences on the profitability. Claims reserves are by far the most important item on the liability side of the balance sheet. They have a big impact on the profit & loss account of a given calendar year and a change of the

reserves by a small percentage of say 3% might well change the yearly result from positive to negative and vice versa.

For these reasons the management wants to be confident that the actuarial reserves are best estimates that is, reserves, which are neither optimistic nor pessimistic, neither on the prudent nor on the aggressive side. In order to calculate best estimate reserves at time I the available information needs to be taken into account. One year later more information is available and the best estimate reserves need to be updated accordingly. The difference between two successive best estimate predictions is the claims development result (CDR). An easy and obvious first check for management is to look at the CDRs over several consecutive years: As long as the CDRs in different years fluctuate around zero, there is no reason to question that the reserves are best estimates. But as soon as the CDR in a certain line of business (lob) has the same sign over several consecutive years, be it negative or positive, the management and/or the executive directors responsible for the profitability of the lob might become suspicious. If the CDRs are negative, the actuaries will be blamed to be responsible for the resulting losses in the profit & loss account. If the CDRs are positive they will be blamed that the company has missed profitable opportunities because of too high and not competitive premiums due to pessimistic reserves.

However, do CDRs with the same sign over several consecutive calendar years necessarily mean that the reserves are not best estimates? Are CDRs with the same sign over several consecutive calendar years a contradiction to best estimate reserves? This is the fundamental question we are going to discuss in this paper.

In order to answer this question we have to model situations of change in a mathematical way. A suitable modelling of realistic situations of change is a second aim of the paper. For this purpose we will consider Bayesian models with calendar year effects. For related structural ideas on the behaviour of best estimates and the updating process we refer to Arjas (1989). In a recent paper on dependence modelling in claims run-off triangles Merz et al. (2013) define a lognormal model that allows to model calendar year effects. Moreover their paper gives a summary on existing literature on calendar year effects modelling. An example for a Bayesian claims reserving model is the Bayesian model for the chain ladder (CL) method introduced by Gisler-Wüthrich (2008). Bühlmann et al. (2009) derive a recursive formula for the estimation of the CL factors in the model of Gisler-Wüthrich (2008). The updating of the CL factors in this recursive procedure is based on new data. This is the usual way of Bayesian modelling: One has one prior distribution and one considers the posterior distribution, given the data. More literature on Bayesian claims reserving is for instance given in de Alba (2002)

or Wüthrich-Merz (2008).

In contrast to many other reserving models we suggest a model that allows to update a priori assumptions due to new external information. We provide a recursive procedure with explicit formulas for the updating of the parameters. In our recursive procedure the parameter updates are not only based on new data, but also based on new external information. In other words, we consider models, in which the prior distribution can be modified dynamically. As stated in De Jong-Zehnwirth (1983) external (or subjective) information is of crucial importance in times of rapid change. De Jong-Zehnwirth (1983) use a state-space model where the updating process is done via Kalman filter equations.

Organisation of the paper. In the remainder of this section we introduce the problem from a mathematical point of view. In Section 2 we look at an example from pricing, where the basic properties to be discussed can most simply be explained. Then we will introduce in Section 3.2 a simple model for reserving, where we can explicitly calculate the best estimate reserves. We provide a numerical example to illustrate the results. In the model for reserving the a priori assumptions can be updated due to new external information.

1.2 Best Estimate Reserves

In claims reserving one usually considers a claims development triangle or trapezoid

$$\mathcal{D}_t = \{X_{i,j} : i = 0, \dots, t;\ j = 0, \dots, J;\ i + j \leq t\}, \tag{1.1}$$

where $X_{i,j}$ denotes the incremental claim (incurred or paid claims) of accident year i at the end of development year j. Let $C_{i,j}$ be the corresponding cumulative claim and denote by U_i the ultimate claim amount of accident year i. If all claims are settled at the end of development year $J \leq t$ then $U_i = C_{i,J}$.

The outstanding liabilities of accident year $i \geq t-J$ at the end of calendar year t are defined by

$$R_i^{(t)} := \sum_{j=t-i+1}^{J} X_{i,j}^{paid} = U_i - C_{i,t-i}^{paid},$$

where $X_{i,j}^{paid}$ are the incremental claim payments of accident year i in development year j and where $C_{i,t-i}^{paid}$ are the cumulative claim payments of accident year i until the end of calendar year t.

For claims reserving one has to predict these outstanding liabilities, which is equivalent to predicting the ultimate claim, that is

$$\widehat{R}_i^{(t)} = \widehat{U}_i^{(t)} - C_{i,t-i}^{paid},$$

where $\widehat{U}_i^{(t)}$ is a predictor of the ultimate claim U_i at the end of calendar year t.

Best estimate reserves are reserves where the expected value of the CDR is zero and which are as accurate as possible. In mathematical terms one has to be a bit more precise. Denote by $\mathcal{A}_t$ the σ-algebra containing all information available at time t. Best estimate reserves of accident year i at time t are then defined as the reserves minimizing the conditional mean square error of prediction (MSEP)

$$\operatorname{msep}_{R_i^{(t)}|\mathcal{A}_t}\left(\widehat{R}_i^{(t)}\right) = E\left[\left(\widehat{R}_i^{(t)} - R_i^{(t)}\right)^2 \middle| \mathcal{A}_t\right]$$

and are given by

$$\widehat{R}_i^{(t)} = E\left[R_i^{(t)} \middle| \mathcal{A}_t\right] = E\left[U_i \middle| \mathcal{A}_t\right] - C_{i,t-i}^{paid}. \tag{1.2}$$

The best estimate reserve at the end of calendar year t is then the sum over all accident years

$$\widehat{R}^{(t)} = \sum_{i=t-J+1}^{t} \widehat{R}_i^{(t)}.$$

Remarks:

- The available information $\mathcal{A}_t$ might consist of two components $\mathcal{A}_t = \sigma\left(\mathcal{D}_t, \mathcal{I}_t\right)$, where $\mathcal{D}_t$ are the observed data until time t and where $\mathcal{I}_t$ denotes some additional external information independent of the data such as observations and expert opinions with regard to legal and economic developments. In the actuarial literature often only the information contained in the data $\mathcal{D}_t$ is considered. However we believe that additional available information should also be taken into account. In mathematical terms the sequences $\{\mathcal{A}_t : t = 0, 1, \ldots\}$ as well as $\{\mathcal{D}_t : t = 0, 1, \ldots\}$ and $\{\mathcal{I}_t : t = 0, 1, \ldots\}$ are filtrations.

- $E\left[U_i \middle| \mathcal{A}_t\right]$ is not exactly known at time t, as it depends itself on unknown parameters. For instance in the CL model of Mack (1993) $E\left[U_i \middle| \mathcal{A}_t\right]$ depends on the unknown CL factors f_j, which have to be estimated from the data. Thus, to be strict, the best estimate reserve is only an estimate of the conditional expected value of the outstanding liabilities.

However, we would like to study a model where best estimate reserves can explicitly be calculated in order to show that CDRs with the same sign over several consecutive calendar years are not a contradiction to best estimate reserves. This means that even if we know the parameters exactly, CDRs with the same sign over several consecutive calendar years are a phenomenon which is to be expected in certain situations. Therefore it will be assumed that the best estimate reserves as defined in (1.2) are exactly known.

1.3 The Martingale Argument

The CDR of accident year $t - J \leq i \leq t - 1$ in the financial statement of calendar year t is defined by

$$\mathrm{CDR}_i^{(t)} := \widehat{R}_i^{(t-1)} - X_{i,t-i}^{paid} - \widehat{R}_i^{(t)}, \tag{1.3}$$

where $X_{i,t-i}^{paid}$ denotes the incremental claim payments of accident year i in calendar year t. This can also be written as

$$\mathrm{CDR}_i^{(t)} = \widehat{U}_i^{(t-1)} - \widehat{U}_i^{(t)}.$$

Since

$$\widehat{U}_i^{(t)} = E\left[U_i \middle| \mathcal{A}_t\right],$$

the sequence

$$\left\{\widehat{U}_i^{(t)} : t = i, i+1, \ldots\right\} \tag{1.4}$$

is a martingale with respect to the filtration $(\mathcal{A}_t)_{t\in\mathbb{N}}$. Hence the sequence

$$\left\{\mathrm{CDR}_i^{(t)} : \ t = i+1, i+2, \ldots\right\},$$

representing CDRs in consecutive years, consists of increments of a martingale. Therefore these CDRs are uncorrelated and have expected value zero. From this fact it is often concluded that CDRs having the same sign over several consecutive calendar years are a strong indication that the reserves are not best estimates.

Remark:
Also with a martingale it is not necessarily unlikely that the increments have the same sign over several consecutive years. This is in particular the case if the distribution of the CDRs is strongly skewed to the left and if the probability of a positive CDR is much higher than 0.5. This can be the case in special lob such as product liability with potential large IBNR claims occurring with

a small probability. But even if the distribution is symmetric around zero and if we assume that CDRs of different calendar years are independent, within a long time horizon of say 50 years there is for instance a probability of 54.4% to have a consecutive period of 6 or more years with the same sign of the CDR. However such long time horizons are not considered in insurance, and for a reasonable time horizon of say 10 years this probability reduces to 9.4%. Thus the martingale argument is not a very strict argument from a mathematical point of view but it reinforces the general belief, that CDRs with the same sign over several consecutive years are a contradiction to best estimate reserves.

1.4 Observations in Practice

In practice however, one often observes the same sign of the CDR over several consecutive years. A striking example was the claims development after the mid-nineties in the line motor liability in Switzerland. Most of the Swiss insurance companies experienced a negative CDR over a longer period of about eight consecutive years. The reason was the emergence of a new phenomenon called 'whiplash'. It is mainly observed after rear-end collisions. The victims complain about permanent pain, chronic fatigue, loss of concentration and inability to work. However, as a peculiarity of this phenomenon, there is no objective medical injury which could be diagnosed by a medical doctor and which would explain these symptoms. Because of this lack of an observable medical indication the courts did not acknowledge a disability for such cases for a long time. But after a decision at the Swiss Federal court, the court practice began gradually to change. This was the beginning of a long process leading to a permanently increasing claims load, which was impossible to predict and which manifested itself only with time. For instance, there are many rear-end collisions, but fortunately most of them turned out to be harmless and only a small percentage developed to a case with long-term injury. However, at the beginning, there was no claims experience allowing to estimate neither this percentage nor the expected claim amount of the corresponding claims. There was also a change in the whole environment. For instance more and more lawyers and law offices began to specialise on whiplash cases which itself increased the number of persons claiming for a disability. Or the advocates of the victims were not prepared to settle a claim in the hope that time works in favour of the victims and that they can get a higher amount at a later settlement date. It was also a cultural issue, as the whiplash phenomenon was mainly observed in the German speaking part of Switzerland and was practically non-existent in the French and Italian speaking part. The question was whether and when it would spill over to these parts of Switzerland,

too. Fortunately it did not.

What are best reserve estimates in such periods of change and what should a reserving actuary do? We claim that it is not sufficient just to look at the claims triangle. In the paid claims triangle there is usually hardly seen any change at the early years of such a period and the new situation is reflected in the claims data only gradually and with a delay. In the incurred triangles one might observe an increase in the newer diagonals. However such a new situation is also a challenge for the claims adjusters and there is a big uncertainty in their case estimates and as a consequence in the information observed in the newer diagonals of the incurred triangle. One of the authors has experienced other situations where the claims adjusters overreacted at the beginning. We claim that best practice is to communicate closely with the claims department, to take their judgement and evaluation as an expert opinion, which is permanently updated based on the development at the claims front (verdicts, court-practice, etc.) and on the observed data in the claims triangles. In the motor liability example of the mid-nineties it took several years until the situation stabilized and the new court practice was reflected in the claims data, and hence the resulting CDRs were negative over several years. In 2008 there was another decision at the Swiss Federal court with the effect that courts have become much more reluctant to attribute a permanent disability after an accident with whiplash indications. After that there was observed a similar development of the CDRs in the opposite direction.

The Swiss motor liability example shows that also with best reserving practice the CDRs can have the same sign over a longer period. The question arises whether we have overseen something in the martingale argument? The answer is yes, we have overseen something.

There is nothing wrong in the mathematics and the sequence $\{\widehat{U}_i^{(t)} : t = i, i+1, \ldots\}$ as defined in (1.4) is a martingale. However one has to be careful in the interpretation what this martingale property exactly means and what conclusions one can draw from the martingale argument. The point is that the claim payments and the claims development depend on 'hidden' not directly observable characteristics describing the 'state space' at different times such as legal and economic environment. These 'hidden' characteristics are best modelled in a Bayesian way as random variables. The martingale argument only holds in the average over these state space characteristics. But what we observe in a claims development triangle are conditional observations that is, claims data generated on the condition that a specific but unknown realisation of the state space variables is given. The observed CDRs resulting from best estimates based on such conditional observations are no longer increments of a martingale with expected value zero. This fact will become more clear,

when looking at some concrete examples.

2 An Example from Pricing

The Poisson-Gamma model is a well-known model for experience rating based on observed claim frequencies and was for instance the basis for constructing the Bonus-Malus scheme in motor liability in Switzerland (see Bichsel, 1964).

Assume we have a portfolio of risks numbered $i = 1, 2, \ldots, I$ and denote by $N_{i,j}$ the claim number of risk i in year j. Each risk is characterised by its risk characteristic ϑ_i which is a realisation of a random variable Θ_i.

Model Assumptions 2.1 (Poisson-Gamma)**.**

PG1: Conditionally, given $\Theta_i = \vartheta_i$, the random variables $N_{i,j}$ $(j = 1, 2, \ldots)$ are independent and Poisson distributed with Poisson parameter $\vartheta_i \lambda_0$.

PG2: The random variables Θ_i are independent and Gamma distributed with $E[\Theta_i] = 1$ and $\mathrm{Var}(\Theta_i) = \gamma^{-1}$.

PG3: $N_{i,j}$ and $N_{k,l}$ are independent for $i \neq k$.

Denote by $\mathcal{D}_t = \{N_{i,j} : j = 1, 2, \ldots, t;\ i = 1, 2, \ldots, I\}$ the observations up to year t. Based on these observed data we want to predict for each risk i next years claim number. The following theorem is well-known from credibility literature (see for instance Bühlmann-Gisler (2005), Chapter 2).

Theorem 2.2. *The best prediction of $N_{i,t+1}$ given $\mathcal{D}_t$ under Model Assumptions 2.1 is given by*

$$\widehat{N}_{i,t+1} = \alpha_t \bar{N}_{i,t} + (1 - \alpha_t)\lambda_0,$$

where

$$\bar{N}_{i,t} = \frac{1}{t}\sum_{j=1}^{t} N_{i,j} \quad \textit{and} \quad \alpha_t = \frac{t\lambda_0}{t\lambda_0 + \gamma}.$$

Remarks:

- The individual risk characteristics ϑ_i are drawn at the very beginning from a first 'urn' as realisations of independent and identically distributed Gamma random variables.

- The sequence $\{\widehat{N}_{i,t} : t = 1, 2, \ldots\}$ is a martingale with respect to $(\mathcal{D}_t)_{t\in\mathbb{N}}$ and it holds that

$$E\left[\widehat{N}_{i,t+1} - N_{i,t+1}\middle| N_{i,1}, \ldots, N_{i,t}\right] = 0. \tag{2.1}$$

  This means that for an infinitely large portfolio there is a balance between forecasted and future claim frequency in all subportfolios of drivers having the same claim experience $N_{i,1}, \ldots, N_{i,t}$. This is the best we can achieve for a risk adjusted pricing. However the left hand side of (2.1) is an expected value over the whole Θ-space.

- But the martingale property does not hold for the observations of a specific driver i. A specific driver i has its specific risk characteristic $\Theta_i = \vartheta_i$ drawn from a first urn, and the observed claim numbers $N_{i,j}$ of this particular driver are independent and Poisson distributed with Poisson parameter $\vartheta_i\lambda_0$. We then have

$$\begin{aligned} E\left[\widehat{N}_{i,t+1} - N_{i,t+1}\middle| N_{i,1}, \ldots, N_{i,t}, \Theta_i\right] &= \widehat{N}_{i,t+1} - E\left[N_{i,t+1}\middle|\Theta_i\right] \\ &= \alpha_t\left(\bar{N}_{i,t} - \lambda_0\right) + \lambda_0\left(1 - \Theta_i\right), \end{aligned}$$

  which is in general different from zero. Moreover,

$$\begin{aligned} E\left[\widehat{N}_{i,t+1} - N_{i,t+1}\middle|\Theta_i\right] &= \alpha_t\left(\lambda_0\Theta_i - \lambda_0\right) + \lambda_0\left(1 - \Theta_i\right) \\ &= \lambda_0\left(1 - \alpha_t\right)\left(1 - \Theta_i\right). \end{aligned}$$

  Thus our best estimate predictor has conditional on Θ_i an upward ($\Theta_i < 1$) or downward ($\Theta_i > 1$) bias for all years t. This bias decreases when the observed number t of years increases.

In reserving we do not have a portfolio of different similar claims triangles and we are not interested in the mean over these triangles. We are rather looking at one specific triangle with given hidden state space characteristics comparable to a specific driver in the pricing example. Conditional on this hidden risk characteristics the best estimate reserves are no longer unbiased, and this bias can have the same sign over several consecutive calendar years. If the one-year process variance is small compared to this bias, we will observe

a CDR with the same sign over several consecutive years. In Section 3 we will consider examples from reserving.

3 Examples from Reserving

3.1 Basic Bayesian Framework

It is our personal belief, based on many years of practical experience of one of the authors, that reserving is as much art as science. By this we mean that a good reserving actuary in practice should not only analyse claims data from a claims triangle. There is much more information available which one could and should take into account. The lawyers and claims adjusters in the company will have some knowledge about changes in legislation, about claims handling, claims processes and the assessment of case estimates by the claims adjusters. Economists will have some opinion about the future development of economic factors. In short, there is a lot of expert knowledge available in a company from which one can learn something about the expected future claim payments. Beside the data in the triangle this expert knowledge should also be taken into account when calculating best estimate claims reserves.

In mathematical terms such a situation can best be modelled in a Bayesian framework, which is predestinated to deal with a priori knowledge on the one hand and observed claims data on the other hand. In such a framework the best estimate reserve is given by formula (1.2). Often there is no explicit formula for (1.2) and in this case a good approximation could be to use credibility theory instead. There is one difference to the credibility models usually considered in the literature: the a priori information does also change in time. In suitable models this can still be handled by recursive estimators.

The basic Bayesian framework considered in this paper is a multiplicative model with the Bayesian structure given below in Model Assumptions 3.1. Recall that $\mathcal{I}_t$ denotes the external information at time t and $\mathcal{D}_t$ is defined as in (1.1). By calendar year effect we mean that

$$E\left[X_{i,j}\middle|\,\mathbf{\Psi},\mathcal{I}_t\right] = E\left[X_{i,j}\middle|\,\Psi_{i+j}\right] = \mu_i\gamma_j\Psi_{i+j}, \quad t \in \mathbb{N}.$$

Since the Ψ_t are multiplicative factors to be applied on the conditional expected value of all elements of a diagonal, they are also called diagonal effects or accounting year effects. We denote by $\mathbf{\Psi} = \{\Psi_0, \Psi_1, \ldots\}$ the set of diagonal effects in all calendar years $t \in \mathbb{N}$.

Model Assumptions 3.1 (Bayes Model with Diagonal Effects)**.**
Conditionally, given $\mathbf{\Psi}$ and $\mathcal{I}_t$ for any $t \in \mathbb{N}$, the increments $(X_{i,j})_{i,j}$ are independent and there exist positive parameters $\mu_0, \ldots, \mu_t$, and parameters

$\gamma_0, \ldots, \gamma_J$, *with* $\sum_j \gamma_j = 1$ *such that*

$$E\left[X_{i,j}|\mathbf{\Psi}, \mathcal{I}_t\right] = E\left[X_{i,j}|\Psi_{i+j}\right] = \mu_i \gamma_j \Psi_{i+j}.$$

Further we assume that $\mathrm{Var}(\Psi_t) < \infty$ *and* $\mathrm{Var}(X_{i,j}|\mathbf{\Psi}, \mathcal{I}_t) < \infty$ *for all* $t \in \mathbb{N}$.

Here $(\gamma_j)_{0 \le j \le J}$ is the incremental payment pattern and $(\beta_j)_{0 \le j \le J}$ with $\beta_j = \sum_{k=0}^{j} \gamma_k$ is the cumulative payment pattern.

Remark:
If $\Psi_t = 1$ for all $t \ge 0$, then there are no diagonal effects and if we omit the external information $\mathcal{I}_t$ in Model Assumptions 3.1 we obtain the basic assumptions of the Bornhuetter-Ferguson (BF) model (see for instance Wüthrich-Merz, 2008). The best estimate reserve for accident year i at time t is then given by

$$\widehat{R}_i = (1 - \beta_{t-i})\, \mu_i,$$

which is the BF reserve formula.

Under Model Assumptions 3.1 the best estimate reserve of accident year i at the end of calendar year t is given by

$$\begin{aligned} \widehat{R}_i^{(t)} &= E\left[R_i|\, \mathcal{I}_t, \mathcal{D}_t\right] = E\left[E\left[R_i|\, \mathcal{I}_t, \mathcal{D}_t, \mathbf{\Psi}\right]|\, \mathcal{I}_t, \mathcal{D}_t\right] \\ &= \mu_i \sum_{j=t-i+1}^{J} \gamma_j E\left[\Psi_{i+j}|\, \mathcal{I}_t, \mathcal{D}_t\right], \end{aligned} \tag{3.1}$$

and the best estimate of the total reserve at the end of calendar year t is

$$\begin{aligned} \widehat{R}^{(t)} &= \sum_{i=t-J+1}^{t} \widehat{R}_i^{(t)} \\ &= \sum_{i=t-J+1}^{t} \mu_i \sum_{j=t-i+1}^{J} \gamma_j E\left[\Psi_{i+j}|\, \mathcal{I}_t, \mathcal{D}_t\right] \\ &= \sum_{k=1}^{J} \left(E\left[\Psi_{t+k}|\, \mathcal{I}_t, \mathcal{D}_t\right] \sum_{j=k}^{J} \gamma_j \mu_{t+k-j} \right). \end{aligned} \tag{3.2}$$

Hence to find the best estimate reserve $\widehat{R}^{(t)}$ we have to determine

$$\Psi_{t+k}^{(t)} = E\left[\Psi_{t+k}|\, \mathcal{I}_t, \mathcal{D}_t\right],$$

for $k = 1, \ldots, J$. Note that $\Psi_{t+k}^{(t)}$ is the best k-year prediction of Ψ_{t+k} at time t.

One year later there are the observations of the diagonal of the next calendar year $t+1$ and possibly updated external information $\mathcal{I}_{t+1}$ available. With these new informations we get the updated forecasts of $\Psi_{t+k}^{(t+1)}$.

From (3.2) it follows that the total CDR in calendar year $t+1$ is given by

$$\begin{aligned}
\mathrm{CDR}_{t+1} &= E\left[\sum_{i=0}^{t} U_i \middle| \mathcal{I}_t, \mathcal{D}_t\right] - E\left[\sum_{i=0}^{t} U_i \middle| \mathcal{I}_{t+1}, \mathcal{D}_{t+1}\right] \\
&= \sum_{i=t-J+1}^{t} \mu_i \sum_{j=t-i+1}^{J} \gamma_j \Psi_{i+j}^{(t)} - \sum_{i=t-J+1}^{t} X_{i,t-i+1} - \sum_{i=t-J+2}^{t} \mu_i \sum_{j=t-i+2}^{J} \gamma_j \Psi_{i+j}^{(t+1)} \\
&= \sum_{k=t+1}^{t+J} \Psi_k^{(t)} \sum_{j=k-t}^{J} \mu_{k-j}\gamma_j - \sum_{j=1}^{J} X_{t+1-j,j} - \sum_{k=t+2}^{t+J} \Psi_k^{(t+1)} \sum_{j=k-t}^{J} \mu_{k-j}\gamma_j \\
&= \sum_{l=1}^{J} \Psi_{t+l}^{(t)} \sum_{j=l}^{J} \mu_{t+l-j}\gamma_j - \sum_{j=1}^{J} X_{t+1-j,j} - \sum_{l=2}^{J} \Psi_{t+l}^{(t+1)} \sum_{j=l}^{J} \mu_{t+l-j}\gamma_j. \qquad (3.3)
\end{aligned}$$

In the special case where $\mu_i = \mu$ for all i, formulas (3.2) and (3.3) simplify to

$$\widehat{R}^{(t)} = \mu \sum_{k=1}^{J} \Psi_{t+k}^{(t)} (1-\beta_{k-1}),$$

$$\begin{aligned}
&\mathrm{CDR}_{t+1} \\
&= \underbrace{\Psi_{t+1}^{(t+1)} \mu(1-\beta_0) - \sum_{j=1}^{J} X_{t+1-j,j}}_{\text{difference between updated forecast and observations}} + \underbrace{\sum_{k=1}^{J} \left(\Psi_{t+k}^{(t)} - \Psi_{t+k}^{(t+1)}\right) \mu(1-\beta_{k-1})}_{\text{change in update}} \qquad (3.4) \\
&= \underbrace{\Psi_{t+1}^{(t)} \mu(1-\beta_0) - \sum_{j=1}^{J} X_{t+1-j,j}}_{\text{difference between forecast and observations}} + \underbrace{\sum_{k=2}^{J} \left(\Psi_{t+k}^{(t)} - \Psi_{t+k}^{(t+1)}\right) \mu(1-\beta_{k-1})}_{\text{change future}}. \qquad (3.5)
\end{aligned}$$

In the following we consider two specific models within this basic Bayesian framework in which we can explicitly calculate best estimate reserves.

3.2 Jump Model

We consider the situation where in each accounting year a change in the environment (e.g. legislation) can occur which leads to a 'jump' in the claim

size. Based on the external information $\mathcal{I}_t$ it is known whether a change occurs in accounting year t. The size of the possible jump at time t is modelled with a random variable whose distribution depends on the information $\mathcal{I}_t, \mathcal{D}_{t-1}$. This situation is reflected by the following model.

Model Assumptions 3.2 (Jump Model)**.**

i) Conditionally, given $\mathbf{\Psi}$ and $\mathcal{I}_t$ for any $t \in \mathbb{N}$, the $(X_{i,j})_{i,j}$ are independent and normally distributed with

$$E[X_{i,j}|\mathbf{\Psi}, \mathcal{I}_t] = E[X_{i,j}|\mathbf{\Psi}] = \mu_i \gamma_j \Psi_{i+j},$$
$$\mathrm{Var}(X_{i,j}|\mathbf{\Psi}, \mathcal{I}_t) = \mathrm{Var}(X_{i,j}|\mathbf{\Psi}) = \mu_i \eta_j^2 \sigma^2,$$

where $\sum_{j=0}^{J} \gamma_j = \sum_{j=0}^{J} \eta_j^2 = 1$, and $\mu_0, \ldots, \mu_t$, are positive parameters.

ii) The diagonal effects are given by

$$\Psi_0 = 1, \quad \Psi_t = \Psi_{t-1} + C_t J_t, \quad t \geq 1,$$

where J_t, $t \geq 1$, are $\mathcal{I}_t$-measurable Bernoulli random variables which explain whether a change (jump) occurs in calendar year t, and where C_t, $t \geq 1$, describe the jump size. We assume that $C_t|\mathcal{I}_t, \mathcal{D}_{t-1} \sim N(m_t, \tau_t^2)$ and conditionally on $\mathcal{I}_t, \mathcal{D}_{t-1}$, C_t is independent of $\Psi_0, \ldots, \Psi_{t-1}$.

iii) For each $t \geq 1$ and $l \geq 0$

$$\Psi_t|(\mathcal{I}_{t+l}, \mathcal{D}_t) \overset{d}{=} \Psi_t|(\mathcal{I}_t, \mathcal{D}_t),$$
$$E[\Psi_{t+l}|\mathcal{I}_t, \mathcal{D}_t] = E[\Psi_t|\mathcal{I}_t, \mathcal{D}_t] =: \Psi_t^{(t)}.$$

From Model Assumptions 3.2 we conclude inductively that for each $t \geq 1$, $\Psi_t|\mathcal{I}_t, \mathcal{D}_{t-1}$ is either normally distributed or equal to 1 and we have

$$\begin{aligned}
\Psi_t^{(t,t-1)} &:= E[\Psi_t|\mathcal{I}_t, \mathcal{D}_{t-1}] \\
&= E[\Psi_{t-1}|\mathcal{I}_t, \mathcal{D}_{t-1}] + m_t J_t \\
&= \begin{cases} \Psi_{t-1}^{(t-1)}, & J_t = 0, \\ \Psi_{t-1}^{(t-1)} + m_t, & J_t = 1, \end{cases}
\end{aligned}$$

where $\Psi_t^{(t)} = E[\Psi_t|\mathcal{I}_t, \mathcal{D}_t]$. Moreover,

$$\begin{aligned}
q_t^{(t,t-1)} &:= \mathrm{Var}(\Psi_t|\mathcal{I}_t, \mathcal{D}_{t-1}) \\
&= \mathrm{Var}(\Psi_{t-1}|\mathcal{I}_t, \mathcal{D}_{t-1}) + \mathrm{Var}(C_t|\mathcal{I}_t, \mathcal{D}_{t-1}) J_t \\
&= q_{t-1}^{(t-1)} + \tau_t^2 J_t \\
&= \begin{cases} q_{t-1}^{(t-1)}, & J_t = 0, \\ q_{t-1}^{(t-1)} + \tau_t^2, & J_t = 1, \end{cases}
\end{aligned}$$

where $q_t^{(t)} = \mathrm{Var}(\Psi_t|\mathcal{I}_t, \mathcal{D}_t)$.

Conditional on $\mathcal{I}_t$, $\mathcal{D}_t$ we can calculate the posterior density of Ψ_t. If $J_l = 0$ for all $l \leq t$, then $\Psi_t = 1$. Otherwise Ψ_t is normally distributed with density

$$\begin{aligned} f_{\Psi_t|\mathcal{I}_t,\mathcal{D}_t} &\propto \exp\left(-\frac{1}{2} \sum_{i+j=t} \frac{(X_{i,j} - \mu_i\gamma_j\Psi_t)^2}{\mu_i\eta_j^2\sigma^2} - \frac{\left(\Psi_t - \Psi_t^{(t,t-1)}\right)^2}{2q_t^{(t,t-1)}} \right) \\ &\propto \exp\left(-\frac{1}{2}\left(\Psi_t^2 \left(\sum_{j=0}^{J\wedge t} \frac{\mu_{t-j}\gamma_j^2}{\eta_j^2\sigma^2} + \frac{1}{q_t^{(t,t-1)}} \right) \right.\right. \\ &\qquad \left.\left. - 2\Psi_t \left(\sum_{j=0}^{J\wedge t} \frac{X_{t-j,j}\gamma_j}{\eta_j^2\sigma^2} + \frac{\Psi_t^{(t,t-1)}}{q_t^{(t,t-1)}} \right) \right) \right), \quad \text{if} \quad \exists l \leq t: \quad J_l = 1. \end{aligned}$$

Hence

$$\Psi_t^{(t)} = \begin{cases} 1, & \text{if} \quad J_l = 0, \ \forall l \leq t, \\ \tilde{\alpha}_t \frac{\sum_{j=0}^{J\wedge t} X_{t-j,j} \frac{\gamma_j}{\eta_j^2}}{w_t} + (1-\tilde{\alpha}_t)\Psi_t^{(t,t-1)}, & \text{if} \quad \exists l \leq t \text{ such that } J_l = 1, \end{cases}$$

where

$$\tilde{\alpha}_t = \frac{w_t}{w_t + \frac{\sigma^2}{q_t^{(t,t-1)}}} \quad \text{and} \quad w_t = \sum_{j=0}^{J\wedge t} \mu_{t-j} \frac{\gamma_j^2}{\eta_j^2},$$

and

$$q_t^{(t)} = \begin{cases} 0, & \text{if} \quad J_l = 0, \ \forall l \leq t, \\ (1-\tilde{\alpha}_t) q_t^{(t,t-1)}, & \text{if} \quad \exists l \leq t \text{ such that } J_l = 1. \end{cases}$$

In the following we consider the situation where only one jump can occur. This situation can be modelled if in the Jump Model 3.2 we additionally assume that

$$J_t = B_t \mathbb{1}\{J_1 = 0, \ldots, J_{t-1} = 0\}, \quad t > 1,$$

where B_t are $\mathcal{I}_t$-measurable Bernoulli random variables. Note that if the jump occurs in calendar year s_0, then we have the realisation $j_{s_0} = 1$ and at time s_0 we also know that $j_t = 0$ for all $t > s_0$. At times $t \leq s_0 - 1$ we then have the realisations $\psi_l = 1$ for all $l \leq t$ and the predictions $\Psi_{t+k}^{(t)} = 1$ for all $k \geq 0$. Therefore the model is only interesting for times $t \geq s_0$. Conditionally on $\mathcal{I}_{s_0}$ we know in this case that $\Psi_t = \Psi_{s_0}$, $t \geq s_0$. Hence, after time $t = s_0$, only the variable Ψ_{s_0} needs to be considered. If we insert the realised information available at time s_0, Model Assumptions 3.2 can be simplified. We assume that the realised values for $\Psi_{s_0-1}^{(s_0,s_0-1)}$ and $q_{s_0-1}^{(s_0,s_0-1)}$ are given by m and $\tau > 0$, respectively. At time $t \geq s_0$ we then have the following model:

Model Assumptions 3.3 (One Jump Model)**.**

- *Conditionally, given* Ψ_{s_0}*, the* $(X_{i,j})_{i,j}$ *are independent and normally distributed with*

$$E[X_{i,j}|\Psi_{s_0}] = \begin{cases} \mu_i\gamma_j, & i+j \leq s_0-1, \\ \mu_i\gamma_j\Psi_{s_0}, & i+j \geq s_0, \end{cases}$$
$$\mathrm{Var}(X_{i,j}|\Psi_{s_0}) = \mu_i\eta_j^2\sigma^2,$$

  where $\sum_{j=0}^{J}\gamma_j = \sum_{j=0}^{J}\eta_j^2 = 1$*, and* $(\mu_i)_i$ *are positive parameters.*

- $\Psi_{s_0} \sim N(m,\tau^2)$ *for some constants* $m \in \mathbb{R}$ *and* $\tau^2 > 0$*.*

The predictions for Ψ_{s_0} under Model Assumptions 3.3 can be obtained from the derivations in the Jump Model 3.2 above by inserting the realised values $j_t = \mathbb{1}_{t=s_0}$, m and τ^2. We summarise these results in the following proposition (see also Theorem 9.6 in Bühlmann-Gisler, 2005):

Proposition 3.4. *Under Model Assumptions 3.3 we have the following recursive formulas for the prediction of* Ψ_{s_0}*:*

i) For $t = s_0$

$$\Psi_{s_0}^{(s_0)} = E[\Psi_{s_0}|\mathcal{D}_{s_0}] = \tilde{\alpha}_{s_0}\frac{\sum_{j=0}^{J}X_{s_0-j,j}\frac{\gamma_j}{\eta_j^2}}{w_{s_0}} + (1-\tilde{\alpha}_{s_0})m,$$
$$q_{s_0}^{(s_0)} = \mathrm{Var}\left(\Psi_{s_0}|\mathcal{D}_{s_0}\right) = (1-\tilde{\alpha}_{s_0})\tau^2,$$

where

$$\tilde{\alpha}_{s_0} = \frac{w_{s_0}}{w_{s_0}+\frac{\sigma^2}{\tau^2}} \quad \textit{and} \quad w_{s_0} = \sum_{j=0}^{J}\mu_{s_0-j}\frac{\gamma_j^2}{\eta_j^2}.$$

ii) For $t > s_0$

$$\Psi_{s_0}^{(t)} = E[\Psi_{s_0}|\mathcal{D}_t] = \tilde{\alpha}_t\frac{\sum_{j=0}^{J}X_{t-j,j}\frac{\gamma_j}{\eta_j^2}}{w_t} + (1-\tilde{\alpha}_t)\Psi_{s_0}^{(t-1)},$$
$$q_{s_0}^{(t)} = \mathrm{Var}(\Psi_{s_0}|\mathcal{D}_t) = (1-\tilde{\alpha}_t)q_{s_0}^{(t-1)},$$

where

$$\tilde{\alpha}_t = \frac{w_t}{w_t+\frac{\sigma^2}{q_{s_0}^{(t-1)}}} \quad \textit{and} \quad w_t = \sum_{j=0}^{J}\mu_{t-j}\frac{\gamma_j^2}{\eta_j^2}.$$

The predictions given in Proposition 3.4 can also be calculated directly instead of using a recursion:

Proposition 3.5. *Under Model Assumptions 3.3, for each $l \geq 0$, Ψ_{s_0} is conditioned on $\mathcal{D}_{s_0+l}$ normally distributed with*

$$\begin{aligned}\Psi_{s_0}^{(s_0+l)} &= E[\Psi_{s_0}|\mathcal{D}_{s_0+l}] = \alpha_{s_0+l}\bar{D}_{s_0+l} + (1-\alpha_{s_0+l})m,\\ q_{s_0}^{(s_0+l)} &= \mathrm{Var}(\Psi_{s_0}|\mathcal{D}_{s_0+l}) = (1-\alpha_{s_0+l})\tau^2,\end{aligned}$$

where

$$\bar{D}_{s_0+l} = \frac{\sum_{t=s_0}^{s_0+l}\sum_{j=0}^{J} X_{t-j,j}\frac{\gamma_j}{\eta_j^2}}{\sum_{t=s_0}^{s_0+l}\sum_{j=0}^{J}\frac{\mu_{t-j}\gamma_j^2}{\eta_j^2}} \quad \textit{and} \quad \alpha_{s_0+l} = \frac{\sum_{t=s_0}^{s_0+l}\sum_{j=0}^{J}\frac{\mu_{t-j}\gamma_j^2}{\eta_j^2}}{\sum_{t=s_0}^{s_0+l}\sum_{j=0}^{J}\frac{\mu_{t-j}\gamma_j^2}{\eta_j^2}+\frac{\sigma^2}{\tau^2}}.$$

Remark:
Proposition 3.5 still yields the credibility estimator if we drop the normality assumption for Ψ_{s_0} and the $X_{i,j}$. In a distribution-free approach, it is the best linear approximation to the Bayes estimator and in this sense still a best estimate. However it is then no longer exact Bayesian and a best estimate as defined in (1.2).

Proof of Proposition 3.5:
In the One Jump Model 3.3 we have exact credibility that is, the credibility estimator is exact Bayesian (see for example Jewell (1974a; 1974b) or Bühlmann-Gisler (2005), Chapter 2). If $\gamma_j \neq 0$ for $j = 0, \ldots, J$, then the random variables

$$Y_{i,j} = \frac{X_{i,j}}{\mu_i\gamma_j}, \quad i+j \geq s_0,$$

fulfil the assumptions of the Bühlmann-Straub model with weights $w_{i,j} = \mu_i\gamma_j^2/(\sigma^2\eta_j^2)$. The formula then follows from the Bühlmann-Straub credibility formula. For incurred claims some of the γ_j might be zero. The formula can then be derived with the regression credibility model (see Theorem 8.4 in Bühlmann-Gisler, 2005) by using the observation vector $\boldsymbol{X}^{(s_0+l)}$ with all observations from calendar year s_0 to calendar year s_0+l as elements, where the regression equation is given by $E[X_{i,j}|\Psi_{s_0}] = \mu_i\gamma_j\Psi_{s_0}$ and where the conditional covariance matrix of $\boldsymbol{X}^{(s_0+l)}$ given Ψ_{s_0} is diagonal with diagonal elements $\mathrm{Var}(X_{i,j}|\Psi_{s_0}) = \mu_i\eta_j^2\sigma^2$.

□

With Proposition 3.4 we can also find an explicit formula for the CDR. For simplicity we consider the case with $\mu_i \equiv \mu$. In order to calculate CDR_{s_0} we recall that Model Assumptions 3.3 are only obtained after inserting realised

values from the available information at time s_0 and we have $\Psi_t^{(s_0-1)} = 1$ for all $t \geq s_0 - 1$. By using (3.4) and (3.5) we obtain:

- for $t+1 = s_0$

$$\begin{aligned}
&\mathrm{CDR}_{s_0} \\
&= \underbrace{\Psi_{s_0}^{(s_0)}\mu(1-\beta_0) - \sum_{j=1}^{J} X_{s_0-j,j}}_{\text{difference between updated forecast and observations}} + \underbrace{\left(1-\Psi_{s_0}^{(s_0)}\right)\mu\sum_{k=1}^{J}(1-\beta_{k-1})}_{\text{change in update}} \\
&= \underbrace{\mu(1-\beta_0) - \sum_{j=1}^{J} X_{s_0-j,j}}_{\text{difference between forecast and observations}} + \underbrace{\left(1-\Psi_{s_0}^{(s_0)}\right)\mu\sum_{k=2}^{J}(1-\beta_{k-1})}_{\text{change future}},
\end{aligned}$$

and

$$1 - \Psi_{s_0}^{(s_0)} = 1 - m + \tilde{\alpha}_{s_0}\left(m - \frac{\sum_{j=0}^{J} X_{s_0-j,j}\frac{\gamma_j}{\eta_j^2}}{w_{s_0}}\right),$$

- for $t+1 > s_0$

$$\begin{aligned}
&\mathrm{CDR}_{t+1} \\
&= \underbrace{\Psi_{s_0}^{(t+1)}\mu(1-\beta_0) - \sum_{j=1}^{J} X_{t+1-j,j}}_{\text{difference between updated forecast and observations}} + \underbrace{\left(\Psi_{s_0}^{(t)} - \Psi_{s_0}^{(t+1)}\right)\mu\sum_{k=1}^{J}(1-\beta_{k-1})}_{\text{change in update}} \\
&= \underbrace{\Psi_{s_0}^{(t)}\mu(1-\beta_0) - \sum_{j=1}^{J} X_{t+1-j,j}}_{\text{difference between forecast and observations}} + \underbrace{\left(\Psi_{s_0}^{(t)} - \Psi_{s_0}^{(t+1)}\right)\mu\sum_{k=2}^{J}(1-\beta_{k-1})}_{\text{change future}},
\end{aligned}$$

and

$$\Psi_{s_0}^{(t)} - \Psi_{s_0}^{(t+1)} = \tilde{\alpha}_{t+1}\left(\Psi_{s_0}^{(t)} - \frac{\sum_{j=0}^{J} X_{t+1-j,j}\frac{\gamma_j}{\eta_j^2}}{w_{t+1}}\right).$$

We observe that in the case $t+1 > s_0$ the second terms 'change in update' and 'change future' respectively, converge to zero as $t \to \infty$ since the weights $\tilde{\alpha}_{t+1} \to 0$. On the other hand, the difference between (updated) forecast and observations depends more on the volatility of the data.

Next we want to consider the sign of the CDR over several consecutive calendar years in the period starting at year $s_0 + 1$. What we will observe are conditional data generated by the model, given a specific realisation of Ψ_{s_0}. More precisely, in one triangle we have one fixed Ψ_{s_0} which has an impact on the data. The actuary cannot observe Ψ_{s_0} and hence needs to base his prediction on the available information $\mathcal{D}_t$. However, although Ψ_{s_0} is not observable, if we consider a specific triangle, then Ψ_{s_0} is fixed. For this reason we do not average over different values of Ψ_{s_0} but we rather condition on the unknown variable Ψ_{s_0} in the following. First, we investigate the behaviour of the conditional expected value

$$E\left[\Psi_{s_0}^{(t)} | \mathcal{D}_{s_0-1}, \Psi_{s_0}\right].$$

With the estimator $\Psi_{s_0}^{(t)}$ in Proposition 3.5 and from $E[\bar{D}_t | \mathcal{D}_{s_0-1}, \Psi_{s_0}] = \Psi_{s_0}$, $t \geq s_0$, it follows that

$$E\left[\Psi_{s_0}^{(t)} - \Psi_{s_0}^{(t+1)} \middle| \mathcal{D}_{s_0-1}, \Psi_{s_0}\right] = (\alpha_t - \alpha_{t+1})(\Psi_{s_0} - m).$$

We then obtain for the conditional expectation of the CDR for $t + 1 > s_0$

$$\begin{aligned}
&E\left[\mathrm{CDR}_{t+1} | \mathcal{D}_{s_0-1}, \Psi_{s_0}\right] \\
&= \mu(1-\beta_0)(1-\alpha_t)(m - \Psi_{s_0}) \qquad (3.6) \\
&\quad + E\left[\Psi_{s_0}^{(t)} - \Psi_{s_0}^{(t+1)} \middle| \mathcal{D}_{s_0-1}, \Psi_{s_0}\right] \mu \sum_{k=2}^{J} (1-\beta_{k-1}) \\
&= \mu(1-\beta_0)(1-\alpha_t)(m - \Psi_{s_0}) + (\alpha_{t+1} - \alpha_t)(m - \Psi_{s_0}) \mu \sum_{k=2}^{J} (1-\beta_{k-1}) \\
&= (m - \Psi_{s_0}) \mu \left((1-\beta_0)(1-\alpha_t) + (\alpha_{t+1} - \alpha_t) \sum_{k=2}^{J} (1-\beta_{k-1}) \right). \qquad (3.7)
\end{aligned}$$

If $\beta_j \leq 1$ for all j, or $\gamma_j \geq 0$ for all j respectively, then the second factor in (3.7) is positive since α_t is increasing in t. Moreover this second factor decreases in t since $1 - \alpha_t$ and $\alpha_{t+1} - \alpha_t$ are decreasing. In the limit for $t \to \infty$, it vanishes. Additionally, we can see from (3.7) that the absolute value of the conditional expectation of the CDR is larger if more weight is given to the tail that is, for long tailed lob the bias is higher.

Hence we have found the following result with regard to the behaviour of the conditional expected value of the CDR in the case where $\gamma_j \geq 0$ for all j:

- The sign of $E\left[\mathrm{CDR}_{t+1} | \mathcal{D}_{s_0-1}, \Psi_{s_0}\right]$ is the same for all $t + 1 \geq s_0 + 1$.
- The absolute value of $E\left[\mathrm{CDR}_{t+1} | \mathcal{D}_{s_0-1}, \Psi_{s_0}\right]$ is the bigger, the bigger $|m - \Psi_{s_0}|$ is.

- The absolute value of $E\left[\mathrm{CDR}_{t+1}|\mathcal{D}_{s_0-1}, \Psi_{s_0}\right]$ is the bigger, the longer the development is. In other words, $|E\left[\mathrm{CDR}_{t+1}|\mathcal{D}_{s_0-1}, \Psi_{s_0}\right]|$ is bigger, if the development pattern $(\beta_j)_{0\leq j\leq J}$ increases slower to $\beta_J = 1$.
- The absolute value of $E\left[\mathrm{CDR}_{t+1}|\mathcal{D}_{s_0-1}, \Psi_{s_0}\right]$ is decreasing in t.
- In the limit for $t \to \infty$ the CDR is conditionally unbiased.

3.2.1 Numerical Example

For the example we choose a special case of Model Assumptions 3.3 in which $\mu_i \equiv \mu$ and $\eta_j^2 = \gamma_j$. Moreover, we assume that $m = 1.02$ and $\tau = 0.02$. In this case the formula for $\Psi_{s_0}^{(t)}$ simplifies to

$$\Psi_{s_0}^{(t)} = \alpha_t \frac{1}{\mu(t - s_0 + 1)} \sum_{l=s_0}^{t} \sum_{j=0}^{J} X_{l-j,j} + (1 - \alpha_t)m, \quad t \geq s_0,$$

where

$$\alpha_t = \frac{\mu(t - s_0 + 1)}{\mu(t - s_0 + 1) + \frac{\sigma^2}{\tau^2}}.$$

We choose the following parameters

$$\mu = 10'000, \quad \sigma = 4,$$

and the development pattern γ_j given in Table 1.

j	0	1	2	3	4	5	6	7	8	9
γ_j	0.2	0.23	0.12	0.11	0.09	0.07	0.058	0.05	0.043	0.029

Table 1: Development pattern γ_j.

In our simulation of the specific observed triangle we start with a 10 by 10 triangle and we assume the jump occurs in calendar year $s_0 = 10$. We then add calendar years and calculate the CDR for time $s_0 = 10$ to time $t = 21$. The true diagonal effect to be considered as a specific realisation of Ψ_{s_0} is given by $\psi_{s_0} = 1.05$. Hence we consider a rather extreme situation where Ψ_{s_0} deviates substantially from the a priori expert's expectation of $m = 1.02$. Of course if the realisation ψ_{s_0} coincides with m, then there is no systematic bias in the CDR_t for $t \geq s_0 + 1$, and the bigger the deviation of ψ_{s_0} from m, the bigger is this bias. The simulated incremental claims data is given in Table 2.

	0	1	2	3	4	5	6	7	8	9
0	1930	2426	1208	1266	903	946	636	480	389	169
1	2185	2372	1099	1110	932	793	516	452	462	362
2	2406	2334	1304	959	819	696	525	466	550	294
3	2035	2548	1272	1159	1149	871	459	324	473	375
4	2199	2559	1437	1152	1038	657	681	575	290	283
5	2169	2486	1237	951	1052	660	692	537	496	307
6	2164	2277	1564	1035	1098	822	459	462	597	322
7	1918	2495	958	1289	862	825	663	648	484	332
8	2347	2186	1267	1123	954	892	526	492	534	281
9	2366	2614	1251	1052	961	952	597	596	607	384
10	2138	2581	1408	1291	990	758	541	581	532	308
11	1958	2564	1182	1119	739	762	609	503	347	412
12	2068	2540	1389	1181	989	895	605	552	527	343
13	1752	2627	1267	1378	788	847	517	560	337	
14	1674	2355	1326	1142	1064	497	543	420		
15	2245	2176	1172	994	1143	770	456			
16	1980	2425	1285	1067	854	880				
17	1992	2420	1420	1020	1010					
18	2024	2387	1229	1521						
19	2488	2687	1361							
20	1979	2498								
21	1816									

Table 2: Incremental claims

In Table 3 the CDR's are given with the splits in the terms 'change future', 'difference between forecast and observations', 'difference between updated forecast and observations' and 'change in update'. Moreover we give the conditional expected value of the CDR, which is always negative in this example. In the CDR we observe negative values in five consecutive years. More precisely, we have 10 times a loss in the 12 years from calendar year 10 to 21. This is also illustrated in the plots in Figure 1 and Figure 2. From the table and from the figures one can see that in this specific example the deviation of the forecasts from the observations in the newest diagonal contributes much more to the observed CDR than the effect of the parameter update for future years.

The predictions $\Psi_{s_0}^{(t)}$ are given in Table 3 and a corresponding plot is given in Figure 3.

t	CDR_t	Future	Forecast	Up Forecast	Update	$E[\text{CDR}_t \mid \mathcal{D}_{s_0-1}, \Psi_{s_0}]$	$\Psi_{s_0}^{(t)}$
10	-1563	-718	-845	-560	-1003	-923	1.036
11	-479	-92	-387	-351	-128	-273	1.040
12	-181	-38	-143	-128	-53	-218	1.042
13	-459	-26	-433	-423	-36	-180	1.043
14	-897	-88	-808	-773	-124	-154	1.048
15	194	8	187	184	11	-134	1.047
16	-281	-26	-256	-245	-36	-118	1.049
17	-112	-1	-111	-111	-1	-106	1.049
18	-270	-26	-244	-233	-37	-95	1.050
19	129	-32	162	175	-45	-87	1.052
20	-198	-9	-189	-186	-12	-80	1.052
21	-425	-15	-410	-404	-21	-74	1.053

Table 3: CDRs, change future (Future), difference between forecast and observations (Forecast), difference between updated forecast and observations (Up forecast), change in update (Update), conditional expected value of the CDR and predictions $\Psi_{s_0}^{(t)}$

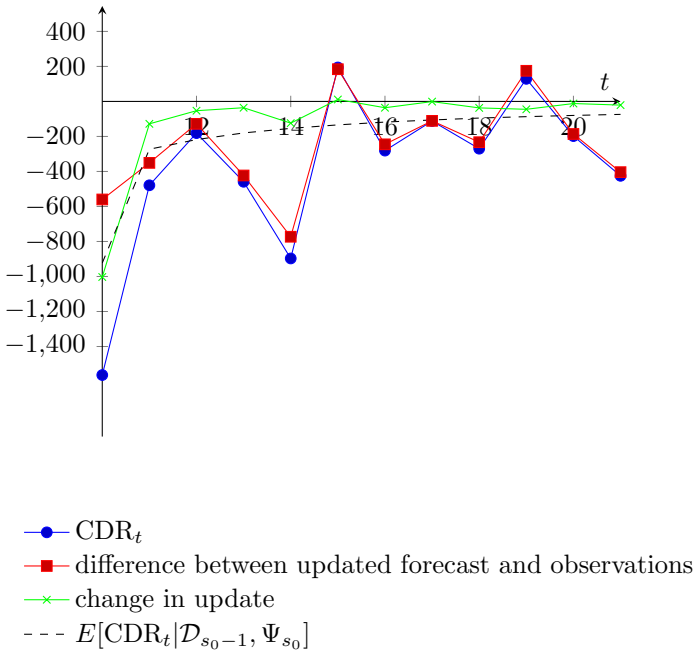

Figure 1: CDR, split in 'difference between updated forecast and observations' plus 'change in update' and $E[\text{CDR}_t \mid \mathcal{D}_{s_0-1}, \Psi_{s_0}]$.

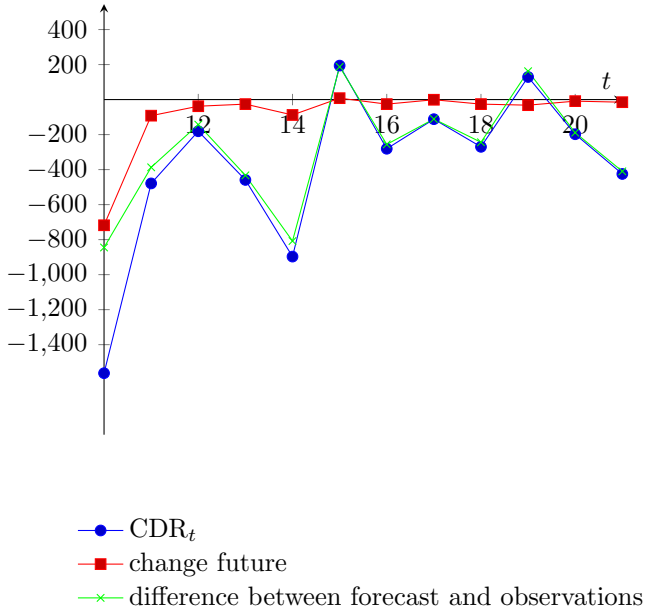

Figure 2: CDR and split in 'difference between forecast and observations' plus 'change future'.

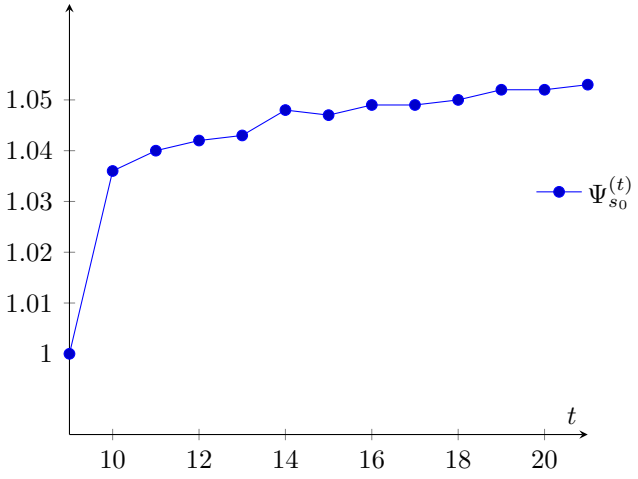

Figure 3: Predictions $\Psi_{s_0}^{(t)}$ of Ψ_{s_0}.

For illustrative purposes we have only considered one realisation of a claims triangle with the given diagonal effect ψ_{s_0} in the above example. We additionally simulate 10'000 triangles with this diagonal effect and consider the conditional distribution of the CDR. It is important to keep the diagonal effect fixed, because in the situation considered we have exactly one effect and therefore we should not average over different effects.

We give boxplots of the realised values of the CDR in the simulations in Figure 4. The conditional expected value is negative and we observe that it gets closer to zero in later calendar years. In the first few calendar years after the jump, it is likely to observe negative CDRs.

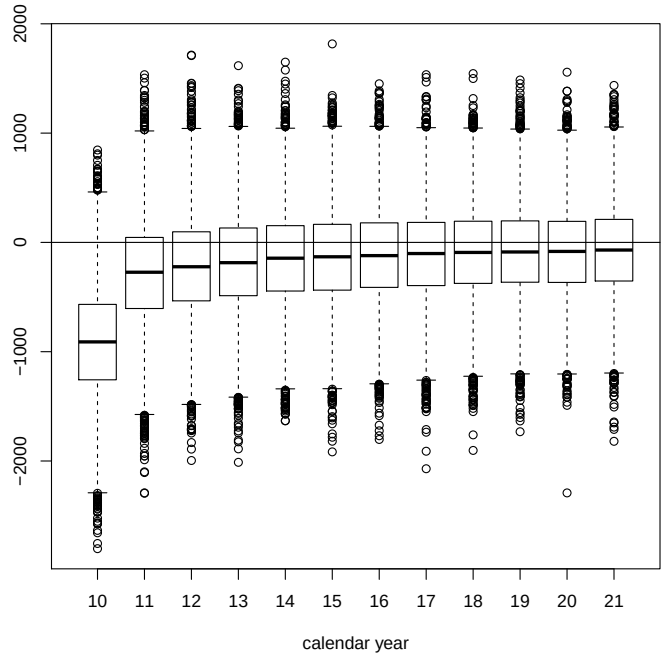

Figure 4: Boxplot of CDR_t, with 10'000 simulations, $\psi_{s_0} = 1.05$.

In practice, there are usually several consecutive jumps, often going into the same direction over a period of years. For such cases the more general Jump Model 3.2 can be used. Moreover, in the real world there can be several sources and events (change of legislation, economic factors and so on) which lead to a diagonal effect. Some of them lead to an additive change (jump) in the diagonal effects, others to a multiplicative one or to a combination of both. For instance, the impact of economic factors such as consumer price inflation

on Ψ_t could possibly be modelled by an autoregressive process meaning that

$$\Psi_t = \rho\Psi_{t-1} + (1-\rho) + \Delta_t,$$

where the Δ_t are iid with $E[\Delta_t] = 0$. Such an autoregressive process is for instance considered in Shi et al., 2012).

4 Concluding Remarks

The primary aim of this paper was to show that the same sign of the CDR over several consecutive calendar years is not a contradiction to best estimate reserves. This has already been achieved by considering the One Jump Model 3.3.

More realistic situations can be mathematically well described by the Jump Model 3.2 introduced in Section 3.2. This model covers a variety of real life situations.

In our models we have assumed that there is the same diagonal factor for all elements of a diagonal. There can be situations, where this not the case, for instance, if a change of legislation only has an impact on bodily injury claims. Our modelling framework can be extended to such situations by assuming that the diagonal effects are random vectors rather than random variables. But this is beyond the aim and the scope of this paper.

In this paper we have considered full Bayesian models in order to be able to explicitly calculate best estimate reserves defined as the posterior expected values of future claim payments. One might object that a normal distribution is not a realistic assumption for claims data but one can easily drop the distributional assumptions and switch to a corresponding credibility model and credibility estimators. But this is again out of the primary goal and scope of this paper. The same holds for μ_i and γ_j, which were assumed to be known. Of course, in practice one will have to estimate them.

We want to emphasize once more that in our model the expert opinion changes over time. This idea can also be found in De Jong-Zehnwirth (1983). However, in most of the actuarial literature on Bayes- and credibility models there is only one a priori distribution not changing over time. We believe that models that allow for changing expert opinions over time have a big potential to bring models nearer to the processes in the real world and to solve problems in practice, not only for reserving, but also for pricing and other applications.

Acknowledgements

The authors would like to thank two anonymous referees for valuable comments that helped to improve an earlier version of this paper.

Paper

D

Annina Saluz.

Prediction Uncertainties in the Cape Cod Reserving Method.

Annals of Actuarial Science, forthcoming in 2015.

Prediction Uncertainties in the Cape Cod Reserving Method

Annina Saluz
ETH Zurich, RiskLab Switzerland, Department of Mathematics,
8092 Zurich, Switzerland

Abstract

The Cape Cod method was designed by Bühlmann-Straub (1983) in order to overcome some shortcomings of the chain ladder (CL) method. Due to its simplicity and due to the advantages over the CL method, the Cape Cod method has become a well-established method in practice. In this paper we consider a distribution-free stochastic model for the Cape Cod method. Within this model we give the parameter estimates and we derive estimates for the conditional mean square error of prediction for the Cape Cod method. Additionally, we derive an estimate for the uncertainty in the claims development result (CDR).

Keywords. Cape Cod Method, Claims Reserving, Mean Square Error of Prediction, Claims Development Result.

1 Introduction

The chain ladder (CL) and the Bornhuetter-Ferguson (1972) (BF) method are probably the two most popular claims reserving methods. The CL predictor of the ultimate claim is obtained by multiplying the current claims amount by a product of development factors. Hence, the CL predictors of the ultimate claims are directly proportional to the current claims amounts and therefore, if the current claims amount is zero or an outlier, the CL prediction gives unsatisfactory results. This is in particular the case for long tailed lines of business in recent accident years. Moreover, the CL method is very sensitive

to changes in individual claim numbers. Finally, the CL method is purely based on the claims data and it disregards the information in the premiums.

The BF reserve of an accident year is the product of an estimate of the expected ultimate claim and the estimated 'still to come percentage' of the corresponding accident year. For the estimation of the latter one assumes that there is a development pattern, which is the same for all accident years. The a priori estimate of the ultimate claim incorporates the information from the earned premiums. In the past, the development pattern for the BF method was often estimated by the development pattern resulting from the CL method. However, Mack (2006) criticises the use of the CL development pattern in the BF method, because the CL method assumes a multiplicative connection between past and future loss amounts, whereas the BF method establishes an additive connection (independence). Moreover, the incorporation of the premiums in the BF method is a fundamental difference to the CL method. Hence, Saluz et al. (2011) argue that this additional information should also be incorporated in the estimates of the development pattern and that the use of the CL development pattern is inconsistent with the BF philosophy.

Bühlmann-Straub (1983) developed the Cape Cod method in order to overcome the above mentioned shortcomings of the CL method. A derivation of the Cape Cod method is published in Straub (1988). In a numerical example Bühlmann-Straub (1983) use the CL development pattern for the Cape Cod method, however, they remark that the estimation of the development pattern is still unsolved. The Cape Cod method is more robust than the CL method and, in contrast to the CL method, the Cape Cod method incorporates the information contained in the premiums. Hence, as in the BF method, we think that an estimate of the development pattern for the Cape Cod method should incorporate this information, too.

Due to its simplicity, the Cape Cod method is a well-established method in practice. However, in current literature there seems to be no formula for the estimation of the prediction uncertainty in the Cape Cod method. In this paper we consider a stochastic model for the Cape Cod method and we derive estimates for the development pattern within this model. The incorporation of the information from the premiums in the estimation of the development pattern is consistent with our model assumptions. In order to quantify prediction uncertainty we derive an estimate for the conditional mean square error of prediction (MSEP) of the Cape Cod reserve and we give an estimate for the uncertainty in the one-year claims development result (CDR).

Organisation of the paper. In the remainder of this section we introduce the notation and we review the Cape Cod and the CL method. In Sec-

tion 2 we consider a distribution-free model for the Cape Cod method. In this model we calculate a development pattern that incorporates the information from the premiums and we derive estimates for the prediction uncertainty of the ultimate claim and for the uncertainty in the CDR. In Section 2.2 we consider a distributional model and show that in this special case the parameter estimates given in the distribution-free model can be derived with maximum likelihood (ML) estimation. A numerical example is given in Section 3.

1.1 Notation and Data Structure

We denote the cumulative claims (cumulative payments or incurred losses) in accident year $i \in \{0, \dots, I\}$ at the end of development year $j \in \{0, \dots, J\}$ by $C_{i,j} > 0$ and we assume $J \leq I$. Let $X_{i,j} = C_{i,j} - C_{i,j-1}$ denote the incremental claims, where we set $C_{i,-1} = 0$. The summation over an index starting from 0 is denoted with a square bracket, for example

$$C_{[k],j} = \sum_{i=0}^{k} C_{i,j}, \quad 0 \leq k \leq I, \quad 0 \leq j \leq J.$$

We assume that all claims are settled after development year J and therefore the total ultimate claim of accident year i is given by $C_{i,J}$. At time I we have information in the upper left trapezoid/triangle

$$\mathcal{D}_I = \{C_{i,j} : i + j \leq I,\ j \leq J\},$$

and our goal is to predict the lower right triangle $\mathcal{D}_I^c = \{C_{i,j} : i + j > I, i \leq I,\ j \leq J\}$. An illustration of the data is given in Table 1.

accident year i	development year j 0	1	…	j	…	J
0						
1	observations $\mathcal{D}_I$					
⋮						
i						
⋮				$\mathcal{D}_I^c$		
I						

Table 1: Claims development triangle: The observations are given in the upper left trapezoid $\mathcal{D}_I$. The lower right triangle $\mathcal{D}_I^c$ needs to be predicted.

The outstanding loss liabilities for accident year i at time I are given by

$$R_i = C_{i,J} - C_{i,I-i}, \quad I - J + 1 \leq i \leq I, \tag{1.1}$$

and the total outstanding loss liabilities are given by $R = \sum_{i=I-J+1}^{I} R_i$. The earned premium for accident year i is denoted by ν_i.

Remark 1.1. The 'true' outstanding loss liabilities are given by formula (1.1) only if $C_{i,j}$ denote cumulative payments. For incurred losses $C_{i,j}$ the outstanding loss liabilities are given by

$$R_i = C_{i,J} - C_{i,I-i} + C_{i,I-i} - C_{i,I-i}^{paid}, \tag{1.2}$$

with $C_{i,j}^{paid}$ denoting the cumulative payments of accident year i up to development year j. Note that the additional term $C_{i,I-i} - C_{i,I-i}^{paid}$ is observable at time I and has no impact on the claims prediction problem. Therefore we only consider the outstanding loss liabilities as defined in (1.1). For incurred losses, (1.1) is often referred to as IBNR.

In order to simplify notation we denote the index of the last observed development year in accident year i by $\iota(i) = (I-i) \wedge J$.

1.2 Chain Ladder Method

The CL prediction of the ultimate claim $C_{i,J}$ of accident year $i > I - J$ is given by

$$\hat{C}_{i,J}^{CL} = C_{i,\iota(i)} \prod_{j=\iota(i)}^{J-1} \hat{f}_j, \quad \text{where} \quad \hat{f}_j = \frac{C_{[I-j-1],j+1}}{C_{[I-j-1],j}}. \tag{1.3}$$

The CL reserve is given by $(i > I - J)$

$$\begin{aligned} \hat{R}_i^{CL} &= C_{i,\iota(i)} \left(\prod_{j=\iota(i)}^{J-1} \hat{f}_j - 1 \right) = C_{i,\iota(i)} \prod_{j=\iota(i)}^{J-1} \hat{f}_j \left(1 - \prod_{j=\iota(i)}^{J-1} \hat{f}_j^{-1} \right) \\ &= \hat{C}_{i,J}^{CL} \left(1 - \hat{\beta}_{\iota(i)}^{CL} \right), \end{aligned}$$

where the CL development pattern

$$\hat{\beta}_j^{CL} = \prod_{k=j}^{J-1} \hat{f}_k^{-1}, \quad 0 \le j \le J-1, \quad \hat{\beta}_J^{CL} = 1, \tag{1.4}$$

is an estimate of the development pattern β_j, $0 \le j \le J$, which describes the percentage of claims which emerge up to development year j.

There are different stochastic models that justify the use of the CL algorithm. A distribution-free model was suggested by Mack (1993).

As mentioned above, the proportionality of the CL predictor to the current claims amount $C_{i,\iota(i)}$ is sometimes problematic in long tailed lines of business for recent accident years. Note that the $\hat{f}_j$ for late development years are based on few observations. A change of $C_{0,J}$ has a multiplicative impact on the prediction of the ultimate claims of all accident years. Therefore the CL method is very sensitive to changes of individual numbers.

1.3 Cape Cod Method

The Cape Cod predictor (see Bühlmann-Straub, 1983) for the ultimate claim is given by

$$\hat{C}_{i,J}^{CC} = C_{i,\iota(i)} + \nu_i \hat{q} \left(1 - \hat{\beta}_{\iota(i)}\right), \tag{1.5}$$

where

$$\hat{q} = \frac{\sum_{i=0}^{I} C_{i,\iota(i)}}{\sum_{i=0}^{I} \nu_i \hat{\beta}_{\iota(i)}}, \tag{1.6}$$

and where $\hat{\beta}_{\iota(i)}$ is an estimate of $\beta_{\iota(i)}$, which describes the percentage of claims emerging up to development year $\iota(i)$. The incremental development pattern $\gamma_j = \beta_j - \beta_{j-1}$ is estimated by

$$\hat{\gamma}_0 = \hat{\beta}_0 \quad \text{and} \quad \hat{\gamma}_{j+1} = \hat{\beta}_{j+1} - \hat{\beta}_j, \quad 0 \leq j \leq J-1.$$

In the original article of Bühlmann-Straub (1983) it is mentioned that the estimation of the development pattern $\hat{\beta}_j$ is an unsolved problem. In practice the development pattern is often estimated by the chain ladder (CL) development pattern given in (1.4). In this case the Cape Cod predictor of the ultimate claim can be written as

$$\hat{C}_{i,J}^{CC} = C_{i,\iota(i)} + \nu_i \hat{q} \hat{\beta}_{\iota(i)}^{CL} \left(\prod_{j=\iota(i)}^{J-1} \hat{f}_j - 1 \right),$$

that is, we replace the diagonal element $C_{i,\iota(i)}$ in the CL prediction

$$\hat{C}_{i,J}^{CL} = C_{i,\iota(i)} \prod_{j=\iota(i)}^{J-1} \hat{f}_j = C_{i,\iota(i)} + C_{i,\iota(i)} \left(\prod_{j=\iota(i)}^{J-1} \hat{f}_j - 1 \right), \tag{1.7}$$

by the more robust value

$$\nu_i \hat{q} \hat{\beta}_{\iota(i)}^{CL} = \nu_i \hat{\beta}_{\iota(i)}^{CL} \frac{\sum_{i=0}^{I} C_{i,\iota(i)}}{\sum_{i=0}^{I} \nu_i \hat{\beta}_{\iota(i)}^{CL}}.$$

Therefore one can interpret the Cape Cod method as an application of the CL algorithm to these more robust diagonal values (see Wüthrich-Merz, 2008). On the other hand, if we take formula (1.5), the Cape Cod method is close to the Bornhuetter-Ferguson (1972) (BF) method in which the predictor of the ultimate claim is given by

$$\hat{C}_{i,J}^{BF} = C_{i,\iota(i)} + \hat{\mu}_i \left(1 - \hat{\beta}_{\iota(i)}\right),$$

where $\hat{\mu}_i$ is an a priori estimate of the expected ultimate claim $E[C_{i,J}]$. In the BF method $\hat{\mu}_i$ is usually an estimate that incorporates the premiums ν_i. Mack (2006) criticises the use of the CL pattern in the BF method and suggests different estimates. Saluz et al. (2011) argue that the information from the premiums should be incorporated in the estimation of the development pattern in the BF method and derive corresponding ML estimators. For the Cape Cod method we suggest similar estimators that incorporate the premiums. The suggested estimators are derived from the best linear unbiased estimators. In Section 2.2 we will see that under the assumption that incremental claims are over-dispersed Poisson (ODP) distributed, these estimators coincide with the maximum likelihood (ML) estimators.

In order to quantify the uncertainties in the Cape Cod predictions we will consider second moments. The conditional MSEP for accident year $i > I-J$, given $\mathcal{D}_I$, is given by

$$\begin{aligned}
\operatorname{msep}_{C_{i,J}|\mathcal{D}_I}(\hat{C}_{i,J}^{CC}) &= E\left[\left(C_{i,J} - \hat{C}_{i,J}^{CC}\right)^2 \middle| \mathcal{D}_I\right] \\
&= E\left[\left(\sum_{j=I-i+1}^{J} X_{i,j} - \nu_i \hat{q}\left(1 - \hat{\beta}_{I-i}\right)\right)^2 \middle| \mathcal{D}_I\right] \\
&= \underbrace{\sum_{j=I-i+1}^{J} \operatorname{Var}(X_{i,j})}_{\mathrm{PV}_i} + \underbrace{\nu_i^2 \left(q(1-\beta_{I-i}) - \hat{q}\left(1 - \hat{\beta}_{I-i}\right)\right)^2}_{=\mathrm{PEE}_i},
\end{aligned} \tag{1.8}$$

where PV_i is the process variance of accident year i, which describes the randomness of the $X_{i,j}$, and PEE_i is the parameter estimation error, which describes the uncertainties in the parameter estimates.

Similarly, for aggregated accident years we want to estimate

$$\operatorname{msep}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I} \hat{C}_{i,J}^{CC}\right) = E\left[\left(\sum_{i=I-J+1}^{I}\left(C_{i,J} - \hat{C}_{i,J}^{CC}\right)\right)^2 \middle| \mathcal{D}_I\right].$$

In order to estimate the conditional MSEP we need to introduce a stochastic model.

2 Stochastic Model for the Cape Cod Method

2.1 Cape Cod Method in a Distribution-free Model

For the estimation of the prediction uncertainty in the Cape Cod method we assume the following underlying distribution-free model.

Model Assumptions 2.1 (Cape Cod Model)**.**
Incremental claims $X_{i,j}$ are independent and there exist positive parameters q, σ_j^2, $0 \leq j \leq J$, and a development pattern $\gamma_0, \ldots, \gamma_J$, with $\sum_{j=0}^{J} \gamma_j = 1$ such that

$$E[X_{i,j}] = \nu_i q \gamma_j \quad \text{and} \quad \operatorname{Var}(X_{i,j}) = \nu_i q \sigma_j^2,$$

for $0 \leq i \leq I$ and $0 \leq j \leq J$.

Note that the premiums ν_i are 'on-level' premiums, that is, premiums adjusted for business cycles such that the expected loss ratio q is the same for all accident years.

Parameter Estimation

Under Model Assumptions 2.1 the best linear unbiased estimate for γ_j is given by

$$\tilde{\gamma}_j^{raw} = \frac{X_{[I-j],j}}{q \nu_{[I-j]}}.$$

In general the $\tilde{\gamma}_j^{raw}$'s do not sum up to 1 and hence we consider the normalised development pattern

$$\hat{\gamma}_j = \frac{\tilde{\gamma}_j^{raw}}{\sum_{j=0}^{J} \tilde{\gamma}_j^{raw}}.$$

Note that the normalised development pattern $\hat{\gamma}_j$ does not depend on the unknown loss ratio q. Hence we can replace the raw development pattern by

$$\hat{\gamma}_j^{raw} = \frac{X_{[I-j],j}}{\nu_{[I-j]}}, \tag{2.1}$$

which will be used in the sequel. The cumulative development pattern is estimated by

$$\hat{\beta}_j = \sum_{k=0}^{j} \hat{\gamma}_k = \frac{\hat{\beta}_j^{raw}}{\hat{\beta}_J^{raw}}, \tag{2.2}$$

where $\hat{\beta}_j^{raw} = \sum_{k=0}^{j} \hat{\gamma}_k^{raw}$ and $\hat{\gamma}_j^{raw}$ is given in (2.1).

The best linear unbiased estimator for q based on the observations $X_{i,j}$, $i + j \leq I$, $j \leq J$, is given by

$$\tilde{q} = \frac{\sum_{j=0}^{J} \sum_{i=0}^{I-j} X_{i,j} \frac{\gamma_j}{\sigma_j^2}}{\sum_{j=0}^{J} \sum_{i=0}^{I-j} \nu_i \frac{\gamma_j^2}{\sigma_j^2}}.$$

If we insert the estimate

$$\hat{\gamma}_j = \frac{X_{[I-j],j}/\nu_{[I-j]}}{\sum_{l=0}^{J} X_{[I-l],l}/\nu_{[I-l]}} = \hat{\gamma}_j^{raw}/\hat{\beta}_J^{raw} \tag{2.3}$$

for γ_j, we obtain

$$\widehat{\tilde{q}} = \frac{\sum_{j=0}^{J} \sum_{i=0}^{I-j} X_{i,j} \frac{X_{[I-j],j}/\nu_{[I-j]}}{\hat{\beta}_J^{raw} \sigma_j^2}}{\sum_{j=0}^{J} \nu_{[I-j]} \frac{X_{[I-j],j}^2/(\nu_{[I-j]})^2}{\left(\hat{\beta}_J^{raw}\right)^2 \sigma_j^2}} = \hat{\beta}_J^{raw} = \hat{q}.$$

The third equation is derived as follows

$$\hat{q} = \frac{\sum_{i=0}^{I} C_{i,\iota(i)}}{\sum_{i=0}^{I} \nu_i \hat{\beta}_{\iota(i)}} = \frac{\sum_{i=0}^{I} C_{i,\iota(i)}}{\sum_{\substack{i+j \leq I,\\ j \leq J}} \nu_i \frac{X_{[I-j],j}/\nu_{[I-j]}}{\hat{\beta}_J^{raw}}} = \frac{\hat{\beta}_J^{raw} \sum_{j=0}^{J} X_{[I-j],j}}{\sum_{j=0}^{J} \nu_{[I-j]} \frac{X_{[I-j],j}}{\nu_{[I-j]}}} = \hat{\beta}_J^{raw} = \widehat{\tilde{q}}. \tag{2.4}$$

Hence, in the Cape Cod Model 2.1 the use of the Cape Cod loss ratio estimate (1.6) can be justified by a derivation from the best linear unbiased estimate of q. Under Model Assumptions 2.1 we therefore use the following estimates

$$\hat{\gamma}_j = \frac{\frac{X_{[I-j],j}}{\nu_{[I-j]}}}{\sum_{l=0}^{J} \frac{X_{[I-l],l}}{\nu_{[I-l]}}}, \qquad \hat{q} = \frac{\sum_{i=0}^{I} C_{i,\iota(i)}}{\sum_{i=0}^{I} \nu_i \hat{\beta}_{\iota(i)}} = \sum_{l=0}^{J} \frac{X_{[I-l],l}}{\nu_{[I-l]}}. \tag{2.5}$$

Note that the loss ratio $\hat{q}$ corresponds to the multiplicative normalisation factor $\hat{\beta}_J^{raw}$ of the non-normalised development pattern given in (2.1). We could therefore omit the loss ratio estimate if the development pattern is not

normalised: For the Cape Cod reserve and ultimate prediction for accident years $i \geq I - J + 1$ we have

$$\hat{R}_i^{CC} = \nu_i \hat{q} \left(1 - \hat{\beta}_{I-i}\right) = \nu_i \hat{\beta}_J^{raw} \left(1 - \frac{\hat{\beta}_{I-i}^{raw}}{\hat{\beta}_J^{raw}}\right) = \nu_i \left(\hat{\beta}_J^{raw} - \hat{\beta}_{I-i}^{raw}\right), \tag{2.6}$$

$$\hat{C}_{i,J}^{CC} = C_{i,I-i} + \hat{R}_i^{CC}. \tag{2.7}$$

For the estimation of the conditional MSEP we further need estimates for $q\sigma_j^2$. Note that

$$\widehat{q\sigma_j^2} = \frac{1}{I-j} \sum_{i=0}^{I-j} \frac{1}{\nu_i} \left(X_{i,j} - \nu_i \hat{\gamma}_j^{raw}\right)^2, \quad 0 \leq j \leq J,\ j \neq I, \tag{2.8}$$

is an unbiased estimator for $q\sigma_j^2$, $j \neq I$. If $I = J$ we use the extrapolation from Mack (1993)

$$\widehat{q\sigma_J^2} = \min\left(\widehat{q\sigma^2}_{J-1}, \widehat{q\sigma^2}_{J-2}, \frac{\widehat{q\sigma^2}^2_{J-1}}{\widehat{q\sigma^2}_{J-2}}\right) = \min\left(\widehat{q\sigma^2}_{J-2}, \frac{\widehat{q\sigma^2}^2_{J-1}}{\widehat{q\sigma^2}_{J-2}}\right).$$

Prediction Uncertainty

The conditional MSEP can be estimated as follows:

Estimate 2.2 (MSEP, Cape Cod Model)**.** *Under Model Assumptions 2.1 the conditional MSEP of the Cape Cod predictor $\hat{C}_{i,J}^{CC}$ given in* (2.7) *is estimated by*

$$\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\hat{C}_{i,J}^{CC}\right) = \widehat{\mathrm{PV}}_i + \widehat{\mathrm{PEE}}_i,$$

where $\widehat{\mathrm{PV}}_i$ and $\widehat{\mathrm{PEE}}_i$ are given by

$$\widehat{\mathrm{PV}}_i = \sum_{j=I-i+1}^{J} \nu_i \widehat{q\sigma_j^2}, \quad \widehat{\mathrm{PEE}}_i = \nu_i^2 \sum_{j=I-i+1}^{J} \frac{\widehat{q\sigma_j^2}}{\nu_{[I-j]}}, \tag{2.9}$$

and where $\widehat{q\sigma_j^2}$ is given in (2.8)*. For aggregated accident years the corresponding conditional MSEP is estimated by*

$$\widehat{\mathrm{msep}}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I} \hat{C}_{i,J}^{CC}\right)$$
$$= \sum_{i=I-J+1}^{I} \widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\hat{C}_{i,J}^{CC}\right) + 2 \sum_{I-J+1 \leq i < k \leq I} \widehat{\mathrm{PEE}}_{i,k},$$

where $\widehat{\text{PEE}}_{i,k}$ *is given by*

$$\widehat{\text{PEE}}_{i,k} = \nu_i \nu_k \sum_{j=I-i+1}^{J} \frac{\widehat{q\sigma_j^2}}{\nu_{[I-j]}}, \quad i < k.$$

Derivation of Estimate 2.2:
For the conditional MSEP of accident year $i \geq I - J$ we have

$$\begin{aligned}
&\text{msep}_{C_{i,J}|\mathcal{D}_I}\left(\hat{C}_{i,J}^{CC}\right) = E\left[\left(C_{i,J} - \hat{C}_{i,J}^{CC}\right)^2 \middle| \mathcal{D}_I\right] \\
&= E\left[\left(\sum_{j=I-i+1}^{J} X_{i,j} - \nu_i\left(\hat{\beta}_J^{raw} - \hat{\beta}_{I-i}^{raw}\right)\right)^2 \middle| \mathcal{D}_I\right] \\
&= \sum_{j=I-i+1}^{J} \text{Var}(X_{i,j}) + \nu_i^2\left(\hat{\beta}_J^{raw} - \hat{\beta}_{I-i}^{raw} - q(1-\beta_{I-i})\right)^2 \\
&= \underbrace{\sum_{j=I-i+1}^{J} \nu_i q\sigma_j^2}_{\text{PV}_i} + \underbrace{\nu_i^2\left(\sum_{j=I-i+1}^{J}\left(\hat{\gamma}_j^{raw} - q\gamma_j\right)\right)^2}_{\text{PEE}_i}.
\end{aligned}$$

We estimate the process variance PV_i by inserting the parameter estimates given in (2.8)

$$\widehat{\text{PV}}_i = \sum_{j=I-i+1}^{J} \nu_i \widehat{q\sigma_j^2}.$$

For the estimation of the parameter estimation error of accident year i we observe that $\hat{\gamma}_j^{raw}$ is an unbiased estimator for $q\gamma_j$. We approximate PEE_i by taking the average over all possible observations $\mathcal{D}_I$ that is, we consider the expected value

$$\begin{aligned}
E[\text{PEE}_i] &= \nu_i^2 \text{Var}\left(\sum_{j=I-i+1}^{J} \hat{\gamma}_j^{raw}\right) = \nu_i^2 \sum_{j=I-i+1}^{J} \frac{1}{\left(\nu_{[I-j]}\right)^2} \sum_{l=0}^{I-j} \text{Var}(X_{l,j}) \\
&= \nu_i^2 \sum_{j=I-i+1}^{J} \frac{1}{\left(\nu_{[I-j]}\right)^2} \sum_{l=0}^{I-j} \nu_l q\sigma_j^2 = \nu_i^2 \sum_{j=I-i+1}^{J} \frac{q\sigma_j^2}{\nu_{[I-j]}}.
\end{aligned}$$

Hence we estimate the parameter estimation error by

$$\widehat{\text{PEE}}_i = \nu_i^2 \sum_{j=I-i+1}^{J} \frac{\widehat{q\sigma_j^2}}{\nu_{[I-j]}},$$

which yields the estimator $\widehat{\text{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\hat{C}_{i,J}^{CC}\right)$ given in Estimate 2.2.

Next we consider the conditional MSEP of aggregated accident years given by

$$\begin{aligned}
&\text{msep}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I}\hat{C}_{i,J}^{CC}\right) = E\left[\left(\sum_{i=I-J+1}^{I}\left(C_{i,J}-\hat{C}_{i,J}^{CC}\right)\right)^2\middle|\mathcal{D}_I\right]\\
&= \sum_{i=I-J+1}^{I}\text{msep}_{C_{i,J}|\mathcal{D}_I}(\hat{C}_{i,J}^{CC}) + 2\sum_{i<k}E\left[\left(\sum_{j=I-i+1}^{J}X_{i,j}-\nu_i\left(\hat{\beta}_J^{raw}-\hat{\beta}_{I-i}^{raw}\right)\right)\right.\\
&\qquad\left.\times\left(\sum_{j=I-k+1}^{J}X_{k,j}-\nu_k\left(\hat{\beta}_J^{raw}-\hat{\beta}_{I-k}^{raw}\right)\right)\middle|\mathcal{D}_I\right]\\
&= \sum_{i=I-J+1}^{I}\text{msep}_{C_{i,J}|\mathcal{D}_I}(\hat{C}_{i,J}^{CC})\\
&+ 2\sum_{i<k}\underbrace{\nu_i\nu_k\left(q(1-\beta_{I-i})-\left(\hat{\beta}_J^{raw}-\hat{\beta}_{I-i}^{raw}\right)\right)\left(q(1-\beta_{I-k})-\left(\hat{\beta}_J^{raw}-\hat{\beta}_{I-k}^{raw}\right)\right)}_{\text{PEE}_{i,k}}.
\end{aligned}$$

Therefore, we need to estimate for $i<k$

$$\text{PEE}_{i,k} = \nu_i\nu_k\left(\hat{\beta}_J^{raw}-\hat{\beta}_{I-i}^{raw}-q(1-\beta_{I-i})\right)\left(\hat{\beta}_J^{raw}-\hat{\beta}_{I-k}^{raw}-q(1-\beta_{I-k})\right).$$

We approximate $\text{PEE}_{i,k}$ by its expected value

$$\begin{aligned}
E[\text{PEE}_{i,k}] &= \nu_i\nu_k\text{Cov}\left(\sum_{j=I-i+1}^{J}\hat{\gamma}_j^{raw},\sum_{l=I-k+1}^{J}\hat{\gamma}_l^{raw}\right)\\
&= \nu_i\nu_k\sum_{j=I-i+1}^{J}\text{Var}\left(\hat{\gamma}_j^{raw}\right) = \nu_i\nu_k\sum_{j=I-i+1}^{J}\frac{q\sigma_j^2}{\nu_{[I-j]}},\quad i<k,
\end{aligned}$$

and estimate $\text{PEE}_{i,k}$ by

$$\widehat{\text{PEE}}_{i,k} = \nu_i\nu_k\sum_{j=I-i+1}^{J}\frac{\widehat{q\sigma_j^2}}{\nu_{[I-j]}},\quad i<k.$$

Therefore we derived the formulas given in Estimate 2.2.

□

Conditional MSEP of the One-year Claims Development Result

So far we have considered the prediction of the outstanding liabilities at time I. At the end of accounting year I we have an additional diagonal of observations in the claims triangle/trapezoid and the predictions of the ultimate claims are updated according to this new information. The one-year CDR is the difference between two successive predictors of the ultimate claim (see for instance Merz-Wüthrich, 2008). More precisely, we define the CDR for accounting year $I+1$ and for accident year i by

$$\mathrm{CDR}_i^{(I+1)} = \hat{C}_{i,J}^{(I)} - \hat{C}_{i,J}^{(I+1)}, \tag{2.10}$$

where $\hat{C}_{i,J}^{(k)}$, $k = I, I+1$, denotes the prediction of $C_{i,J}$ at time k. Under the Cape Cod Model 2.1 we have for $I - J + 1 \leq i \leq I$ and $k = I, I+1$,

$$\hat{C}_{i,J}^{(k)} = C_{i,k-i} + \nu_i \hat{q}^{(k)} \left(1 - \hat{\beta}_{k-i}^{(k)}\right) = C_{i,k-i} + \nu_i \left(\hat{\beta}_J^{(k)raw} - \hat{\beta}_{k-i}^{(k)raw}\right),$$

where $\hat{q}^{(k)} = \hat{\beta}_J^{(k)raw}$ and

$$\hat{\beta}_j^{(k)raw} = \sum_{l=0}^{j} \hat{\gamma}_l^{(k)raw}, \quad \text{with} \quad \hat{\gamma}_j^{(k)raw} = \frac{X_{[k-j],j}}{\nu_{[k-j]}}.$$

Similarly, we define the CDR for aggregated accident years at time $I+1$ by

$$\mathrm{CDR}^{(I+1)} = \sum_{i=I-J+1}^{I} \mathrm{CDR}_i^{(I+1)}.$$

At time I we predict $\mathrm{CDR}_i^{(I+1)}$ by 0 since we consider $\hat{C}_{i,J}^{(I)}$ as best estimate based on the available information at time I. New solvency regulations such as the Swiss Solvency Test, FINMA (2006), require additional risk capital for protection against possible shortfalls in this one-year CDR. In order to quantify uncertainties in the CDRs we consider second moments given by

$$\mathrm{msep}_{\mathrm{CDR}_i^{(I+1)}|\mathcal{D}_I}(0) = E\left[\left(\mathrm{CDR}_i^{(I+1)}\right)^2 \middle| \mathcal{D}_I\right],$$
$$\mathrm{msep}_{\mathrm{CDR}^{(I+1)}|\mathcal{D}_I}(0) = E\left[\left(\mathrm{CDR}^{(I+1)}\right)^2 \middle| \mathcal{D}_I\right].$$

These conditional MSEPs are estimated as follows:

Estimate 2.3 (MSEP CDR, Cape Cod Model)**.**
Under Model Assumptions 2.1 the conditional MSEP of the CDR of accident year i given in (2.10) *is estimated by*

$$\widehat{\operatorname{msep}}_{\operatorname{CDR}_i^{(I+1)}|\mathcal{D}_I}(0) = \nu_i \widehat{q\sigma^2}_{I+1-i} \frac{\nu_{[i]}}{\nu_{[i-1]}} + \nu_i^2 \sum_{j=I+2-i}^{J} \widehat{q\sigma^2}_j \frac{\nu_{I+1-j}}{\nu_{[I-j]}\nu_{[I+1-j]}},$$

where the $\widehat{q\sigma^2}_j$'s are given in (2.8)*. For aggregated accident years the corresponding conditional MSEP is estimated by*

$$\begin{aligned}\widehat{\operatorname{msep}}_{\operatorname{CDR}^{(I+1)}|\mathcal{D}_I}(0) = &\sum_{i=I-J+1}^{I} \widehat{\operatorname{msep}}_{\operatorname{CDR}_i^{(I+1)}|\mathcal{D}_I}(0) \\ &+ 2\sum_{i<k} \nu_i\nu_k \left(\frac{\widehat{q\sigma^2}_{I+1-i}}{\nu_{[i-1]}} + \sum_{j=I+2-i}^{J} \frac{\widehat{q\sigma^2}_j \nu_{I+1-j}}{\nu_{[I-j]}\nu_{[I+1-j]}} \right).\end{aligned}$$

Derivation of Estimate 2.3:
We have for $I+1-i \le J$

$$\begin{aligned}
&\operatorname{CDR}_i^{(I+1)} = -X_{i,I+1-i} + \nu_i \left(\hat\beta_J^{(I)raw} - \hat\beta_{I-i}^{(I)raw} - \hat\beta_J^{(I+1)raw} + \hat\beta_{I+1-i}^{(I+1)raw} \right) \\
&= -X_{i,I+1-i} + \nu_i \left(\sum_{j=I+1-i}^{J} \hat\gamma_j^{(I)raw} - \sum_{j=I+2-i}^{J} \hat\gamma_j^{(I+1)raw} \right) \\
&= -X_{i,I+1-i} + \nu_i \left(\sum_{j=I+2-i}^{J} \left(\hat\gamma_j^{(I)raw} \left(1 - \frac{\nu_{[I-j]}}{\nu_{[I+1-j]}} \right) - \frac{X_{I+1-j,j}}{\nu_{[I+1-j]}} \right) + \hat\gamma_{I+1-i}^{(I)raw} \right) \\
&= \nu_i q\gamma_{I+1-i} - X_{i,I+1-i} \\
&\quad + \nu_i \left(\sum_{j=I+2-i}^{J} \left(\hat\gamma_j^{(I)raw} \frac{\nu_{I+1-j}}{\nu_{[I+1-j]}} - \frac{X_{I+1-j,j}}{\nu_{[I+1-j]}} \right) + \hat\gamma_{I+1-i}^{(I)raw} - q\gamma_{I+1-i} \right), \quad (2.11)
\end{aligned}$$

and therefore

$$\begin{aligned}
&\operatorname{msep}_{\operatorname{CDR}_i^{(I+1)}|\mathcal{D}_I}(0) = \nu_i q\sigma^2_{I+1-i} + \nu_i^2 \left(\hat\gamma_{I+1-i}^{(I)raw} - q\gamma_{I+1-i} \right)^2 \\
&+ 2\nu_i^2 \left(\hat\gamma_{I+1-i}^{(I)raw} - q\gamma_{I+1-i} \right) \left(\sum_{j=I+2-i}^{J} \left(\hat\gamma_j^{(I)raw} - q\gamma_j \right) \frac{\nu_{I+1-j}}{\nu_{[I+1-j]}} \right) + \nu_i^2 \\
&\times E\left[\left(\sum_{j=I+2-i}^{J} \left(\left(\hat\gamma_j^{(I)raw} - q\gamma_j \right) \frac{\nu_{I+1-j}}{\nu_{[I+1-j]}} - \frac{X_{I+1-j,j} - \nu_{I+1-j} q\gamma_j}{\nu_{[I+1-j]}} \right) \right)^2 \middle| \mathcal{D}_I \right].
\end{aligned}$$

For the last term we have

$$
\begin{aligned}
&E\left[\left(\sum_{j=I+2-i}^{J}\left(\left(\hat{\gamma}_j^{(I)raw}-q\gamma_j\right)\frac{\nu_{I+1-j}}{\nu_{[I+1-j]}}-\frac{X_{I+1-j,j}-\nu_{I+1-j}q\gamma_j}{\nu_{[I+1-j]}}\right)\right)^2\middle|\,\mathcal{D}_I\right]\\
&=\sum_{j=I+2-i}^{J}\left\{\left(\hat{\gamma}_j^{(I)raw}-q\gamma_j\right)^2\frac{\nu_{I+1-j}^2}{\left(\nu_{[I+1-j]}\right)^2}+E\left[\left(\frac{X_{I+1-j,j}-\nu_{I+1-j}q\gamma_j}{\nu_{[I+1-j]}}\right)^2\right]\right\}\\
&\quad+2\sum_{j<k}\left(\hat{\gamma}_j^{(I)raw}-q\gamma_j\right)\left(\hat{\gamma}_k^{(I)raw}-q\gamma_k\right)\frac{\nu_{I+1-j}\nu_{I+1-k}}{\nu_{[I+1-j]}\nu_{[I+1-k]}}.
\end{aligned}
$$

As in the case of the conditional MSEP of the ultimate claim, we approximate the unknown terms by their expected values. After replacing the unknown parameters by their estimates we arrive at the following estimator

$$
\begin{aligned}
&\widehat{\mathrm{msep}}_{\mathrm{CDR}_i^{(I+1)}|\mathcal{D}_I}(0)\\
&=\nu_i\widehat{q\sigma^2}_{I+1-i}+\nu_i^2\widehat{\mathrm{Var}}\left(\hat{\gamma}_{I+1-i}^{(I)raw}\right)\\
&\quad+\nu_i^2\sum_{j=I+2-i}^{J}\left(\widehat{\mathrm{Var}}\left(\hat{\gamma}_j^{(I)raw}\right)\frac{\nu_{I+1-j}^2}{\left(\nu_{[I+1-j]}\right)^2}+\frac{\widehat{\mathrm{Var}}(X_{I+1-j,j})}{\left(\nu_{[I+1-j]}\right)^2}\right)\\
&=\nu_i\widehat{q\sigma^2}_{I+1-i}+\nu_i^2\frac{\widehat{q\sigma^2}_{I+1-i}}{\nu_{[i-1]}}+\nu_i^2\sum_{j=I+2-i}^{J}\left(\frac{\widehat{q\sigma^2}_j}{\nu_{[I-j]}}\frac{\nu_{I+1-j}^2}{\left(\nu_{[I+1-j]}\right)^2}+\frac{\nu_{I+1-j}\widehat{q\sigma^2}_j}{\left(\nu_{[I+1-j]}\right)^2}\right)\\
&=\nu_i\widehat{q\sigma^2}_{I+1-i}\frac{\nu_{[i]}}{\nu_{[i-1]}}+\nu_i^2\sum_{j=I+2-i}^{J}\widehat{q\sigma^2}_j\frac{\nu_{I+1-j}}{\nu_{[I-j]}\nu_{[I+1-j]}},
\end{aligned}
$$

where the variances $\mathrm{Var}\left(\hat{\gamma}_j^{(I)raw}\right)$ and $\mathrm{Var}(X_{I+1-j,j})$ are estimated by inserting the parameter estimates for $q\sigma_j^2$, that is,

$$
\widehat{\mathrm{Var}}\left(\hat{\gamma}_j^{(I)raw}\right)=\frac{\widehat{q\sigma^2}_j}{\nu_{[I-j]}}\quad\text{and}\quad\widehat{\mathrm{Var}}(X_{I+1-j,j})=\nu_{I+1-j}\widehat{q\sigma^2}_j.
$$

Similarly, we obtain for aggregated accident years

$$
\begin{aligned}
&\mathrm{msep}_{\mathrm{CDR}^{(I+1)}|\mathcal{D}_I}(0)\\
&=\sum_{i=I-J+1}^{I}\mathrm{msep}_{\mathrm{CDR}_i^{(I+1)}|\mathcal{D}_I}+2\sum_{i<k}E\left[\mathrm{CDR}_i^{(I+1)}\mathrm{CDR}_k^{(I+1)}\middle|\,\mathcal{D}_I\right].
\end{aligned}
$$

With (2.11) we obtain

$$\begin{aligned}
E\left[\mathrm{CDR}_i^{(I+1)}\mathrm{CDR}_k^{(I+1)}\middle|\mathcal{D}_I\right] &= \nu_k\frac{\nu_i}{\nu_{[i]}}q\sigma^2_{I+1-i} \\
&+ \nu_i\nu_k\sum_{j=I+2-i}^{J}\sum_{l=I+2-k}^{J}\left(\hat{\gamma}_j^{(I)raw}-q\gamma_j\right)\left(\hat{\gamma}_l^{(I)raw}-q\gamma_l\right)\frac{\nu_{I+1-j}\nu_{I+1-l}}{\nu_{[I+1-j]}\nu_{[I+1-l]}} \\
&+ \nu_i\nu_k\sum_{j=I+2-i}^{J}\frac{\nu_{I+1-j}}{\left(\nu_{[I+1-j]}\right)^2}q\sigma_j^2 \\
&+ \nu_i\nu_k\left(\hat{\gamma}_{I+1-i}^{(I)raw}-q\gamma_{I+1-i}\right)\sum_{j=I+2-k}^{J}\left(\hat{\gamma}_j^{(I)raw}-q\gamma_j\right)\frac{\nu_{I+1-j}}{\nu_{[I+1-j]}} \\
&+ \nu_i\nu_k\left(\hat{\gamma}_{I+1-i}^{(I)raw}-q\gamma_{I+1-i}\right)\left(\hat{\gamma}_{I+1-k}^{(I)raw}-q\gamma_{I+1-k}\right).
\end{aligned}$$

This term is approximated by its unconditional expected value

$$\begin{aligned}
&\nu_k\frac{\nu_i}{\nu_{[i]}}q\sigma^2_{I+1-i}+\nu_i\nu_k\sum_{j=I+2-i}^{J}\frac{\nu_{I+1-j}q\sigma_j^2}{\left(\nu_{[I+1-j]}\right)^2}\left(\frac{\nu_{I+1-j}}{\nu_{[I-j]}}+1\right)+\nu_i\nu_k\frac{\nu_i q\sigma^2_{I+1-i}}{\nu_{[i]}\nu_{[i-1]}} \\
&=\nu_i\nu_k\left(\frac{q\sigma^2_{I+1-i}}{\nu_{[i-1]}}+\sum_{j=I+2-i}^{J}\frac{q\sigma_j^2\nu_{I+1-j}}{\nu_{[I-j]}\nu_{[I+1-j]}}\right).
\end{aligned}$$

After inserting the parameter estimates we arrive at the formulas given in Estimate 2.3.

□

In order to compare the formula given in Estimate 2.3 with the formula for the conditional MSEP of the predictor for the ultimate claim given in Estimate 2.2, we rewrite Estimate 2.3 as follows

$$\widehat{\mathrm{msep}}_{\mathrm{CDR}_i^{(I+1)}|\mathcal{D}_I}(0)=\nu_i\widehat{q\sigma^2}_{I+1-i}+\nu_i^2\frac{\widehat{q\sigma^2}_{I+1-i}}{\nu_{[i-1]}}+\nu_i^2\sum_{j=I+2-i}^{J}\frac{\widehat{q\sigma^2}_j}{\nu_{[I-j]}}\frac{\nu_{I+1-j}}{\nu_{[I+1-j]}}.$$

Note that the conditional MSEP of the CDR considers only the first term of $\widehat{\mathrm{PV}}_i$, which corresponds to the variance in the next diagonal. For the parameter estimation error the first term is fully incorporated and all other terms in $\widehat{\mathrm{PEE}}_i$ are scaled down by the factors $\nu_{I+1-j}/\nu_{[I+1-j]}<1$. Similar relations are observed in the case of the CL method, see Merz-Wüthrich (2008) and Bühlmann et al. (2009).

2.2 Over-dispersed Poisson (ODP) Model

A special case of Model Assumptions 2.1 is obtained by assuming that the incremental claims are ODP distributed with constant dispersion parameter ϕ that is, $X_{i,j}/\phi \sim \text{Poisson}(\nu_i q \gamma_j/\phi)$. Note that incremental claims are in this case assumed to be positive.

Model Assumptions 2.4 (ODP Model)**.**
Incremental claims $X_{i,j}$ are independent and ODP distributed with

$$E[X_{i,j}] = \nu_i q \gamma_j \quad \text{and} \quad \text{Var}(X_{i,j}) = \phi \nu_i q \gamma_j = \phi E[X_{i,j}],$$

for $0 \le i \le I$ and $0 \le j \le J$, where $\gamma_0, \dots, \gamma_J$, ϕ and q are positive parameters and $\sum_{j=0}^{J} \gamma_j = 1$.

The ODP model is well-known in claims reserving and was for instance used in England-Verrall (2002).

Theorem 2.5. *Under Model Assumptions 2.4 the ML estimates for q and γ_j coincide with the estimates given in* (2.5)*.*

Proof: The log-likelihood function is given by

$$\begin{aligned} l_{\mathcal{D}_I} = & \sum_{i+j \le I, j<J} \frac{1}{\phi} \left(X_{i,j} \left(\log \nu_i + \log(q\gamma_j) \right) - \nu_i q \gamma_j \right) + r \\ & + \sum_{i=0}^{I-J} \frac{1}{\phi} \left(X_{i,J} \left(\log \nu_i + \log \left(q - \sum_{j=0}^{J-1} q\gamma_j \right) \right) - \nu_i \left(q - \sum_{j=0}^{J-1} q\gamma_j \right) \right), \end{aligned}$$

where r is a term that does not depend on the parameters q and $\gamma_0, \gamma_1, \dots, \gamma_J$. Note that the ML estimate for q yields

$$\hat{q} = \frac{\sum_{i=0}^{I} C_{i,\iota(i)}}{\sum_{i=0}^{I} \nu_i \hat{\beta}_{\iota(i)}}, \tag{2.12}$$

which is the formula for the Cape Cod loss ratio estimate. The ML estimate for γ_j, $0 \le j \le J$, is given by the equation

$$\hat{\gamma}_j = \frac{X_{[I-j],j}}{\nu_{[I-j]} \hat{q} + \frac{X_{[I-J],J}}{\hat{\gamma}_J} - \nu_{[I-J]} \hat{q}},$$

where $\hat{\gamma}_J = 1 - \sum_{j=0}^{J-1} \hat{\gamma}_j$. The equation for $\hat{\gamma}_j$ can be rewritten as

$$\hat{\gamma}_j = \frac{X_{[I-j],j}}{\nu_{[I-j]} \hat{q} + \kappa}, \tag{2.13}$$

where $\kappa = \frac{X_{[I-J],J}}{\hat{\gamma}_J} - \nu_{[I-J]}\hat{q} > -\nu_{[I-J]}\hat{q}$ ensures that $\sum_{j=0}^{J}\hat{\gamma}_j = 1$. We insert the estimates $\hat{\gamma}_j$ in the ML equation for q and obtain

$$\hat{q} = \frac{\sum_{i=0}^{I} C_{i,\iota(i)}}{\sum_{i=0}^{I}\nu_i \sum_{j=0}^{\iota(i)} \frac{X_{[I-j],j}}{\nu_{[I-j]}\hat{q}+\kappa}}.$$

It follows that

$$\sum_{i=0}^{I} C_{i,\iota(i)} = \sum_{\substack{i+j\le I,\\ j\le J}} X_{i,j} = \hat{q}\sum_{\substack{i+j\le I,\\ j\le J}} \nu_i \frac{X_{[I-j],j}}{\nu_{[I-j]}\hat{q}+\kappa} = \hat{q}\sum_{j=0}^{J}\nu_{[I-j]}\frac{X_{[I-j],j}}{\nu_{[I-j]}\hat{q}+\kappa},$$

and hence

$$0 = \sum_{j=0}^{J} X_{[I-j],j}\left(1 - \frac{\hat{q}\nu_{[I-j]}}{\nu_{[I-j]}\hat{q}+\kappa}\right) = \sum_{j=0}^{J} X_{[I-j],j}\frac{\kappa}{\nu_{[I-j]}\hat{q}+\kappa} = \kappa\sum_{j=0}^{J}\hat{\gamma}_j = \kappa.$$

From $\kappa = 0$ and the constraint $\sum_{j=0}^{J}\hat{\gamma}_j = 1$ it follows with (2.13) that

$$\hat{q} = \sum_{j=0}^{J}\frac{X_{[I-j],j}}{\nu_{[I-j]}},$$

and from (2.4) we know that $\hat{q}$ also satisfies the ML equation (2.12). We conclude that the ML estimates for q and for the γ_j's under Model Assumptions 2.4 are given by

$$\hat{\gamma}_j = \frac{\frac{X_{[I-j],j}}{\nu_{[I-j]}}}{\sum_{l=0}^{J}\frac{X_{[I-l],l}}{\nu_{[I-l]}}} = \frac{\hat{\gamma}_j^{raw}}{\hat{\beta}_J^{raw}}, \qquad \hat{q} = \frac{\sum_{i=0}^{I} C_{i,\iota(i)}}{\sum_{i=0}^{I}\nu_i\hat{\beta}_{\iota(i)}} = \sum_{l=0}^{J}\frac{X_{[I-l],l}}{\nu_{[I-l]}} = \hat{\beta}_J^{raw},$$

as in (2.5), which proves the claim.

□

Since the parameter estimates are as in (2.5), we can use Estimate 2.2 for the estimation of the conditional MSEP, we just need to replace the estimates for $q\sigma_j^2$, $0 \le j \le J$, by some estimates for $\phi q\gamma_j$. As in Wüthrich-Merz (2008) we suggest to estimate ϕ with the help of Pearson residuals

$$\hat{\phi} = \frac{1}{|\mathcal{D}_I| - (J+1)}\sum_{\substack{i+j\le I\\ j\le J}} \frac{(X_{i,j} - \nu_i\hat{\gamma}_j^{raw})^2}{\nu_i\hat{\gamma}_j^{raw}}. \tag{2.14}$$

Replacing the estimates $\hat{\sigma}_j^2$ by $\hat{\phi}\hat{\gamma}_j^{raw}$ in Estimate 2.2 we obtain the following estimators:

Estimate 2.6 (MSEP, ODP Model)**.** *Under Model Assumptions 2.4 the conditional MSEP of the Cape Cod predictor $\hat{C}_{i,J}^{CC}$ given in* (2.7) *is estimated by*

$$\widehat{\operatorname{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\hat{C}_{i,J}^{CC}\right) = \widehat{\operatorname{PV}}_i + \widehat{\operatorname{PEE}}_i = \hat{\phi}\hat{R}_i^{CC} + \nu_i^2\hat{\phi}\sum_{j=I-i+1}^{J}\frac{X_{[I-j],j}}{\left(\nu_{[I-j]}\right)^2},$$

where $\hat{\phi}$ and $\hat{R}_i^{CC}$ are given in (2.14) *and* (2.6)*. For aggregated accident years the corresponding conditional MSEP is estimated by*

$$\begin{aligned}&\widehat{\operatorname{msep}}_{\sum_{i=I-J+1}^{I} C_{i,J}|\mathcal{D}_I}\left(\sum_{i=I-J+1}^{I}\hat{C}_{i,J}^{CC}\right)\\ &= \sum_{i=I-J+1}^{I}\widehat{\operatorname{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\hat{C}_{i,J}^{CC}\right) + 2\sum_{I-J+1\le i<k\le I}\nu_i\nu_k\hat{\phi}\sum_{j=I-i+1}^{J}\frac{X_{[I-j],j}}{\left(\nu_{[I-j]}\right)^2}.\end{aligned}$$

Analogously an estimator for the uncertainty in the one-year CDR is obtained from Estimate 2.3 by replacing $\widehat{q\sigma^2}_j$ by $\hat{\phi}\hat{\gamma}_j^{raw}$, with $\hat{\phi}$ given in (2.14).

Remark 2.7. Under Model Assumptions 2.4, if $\mu_i = \nu_i q$, $0 \le i \le I$, are considered as unknown parameters and if the μ_i's and γ_j's are both estimated by ML estimation then it is well-known that the resulting development pattern is the CL development pattern (see e.g. Mack, 1991 or Renshaw-Verrall, 1998). The result goes back to Hachemeister-Stanard (1975). For the estimation of the correlations in the CL development pattern one can therefore use the Fisher information matrix. This approach is also used by Alai et al. (2010). Alai et al. (2010) derive an estimate for the conditional MSEP in the BF method in the case where the CL development pattern is used. However, in the Cape Cod method the premiums are assumed to be known and hence ignoring this information for the estimation of the development pattern is not consistent with the Cape Cod assumptions and thus we do not consider this case in more detail.

3 Example

We use data from Wüthrich-Merz (2008), which is given in Table 9. We calculate the reserves, process standard deviation, parameter estimation error and conditional MSEP for the Cape Cod Model 2.1 and for the ODP Model 2.4. Results from the Cape Cod Model 2.1 are given in Table 2. The corresponding parameter estimates are given in Table 3. The loss ratio estimates are

given by $\hat{q} = \hat{\beta}_J^{raw} = 0.674$ for the Cape Cod Model 2.1. The coefficient of variation of a reserve estimate $\hat{R}_i$ is estimated as

$$\widehat{\mathrm{CoVa}}\left(\hat{R}_i\right) = \frac{\sqrt{\widehat{\mathrm{msep}}_{R_i|\mathcal{D}_I}\left(\hat{R}_i\right)}}{\hat{R}_i} = \frac{\sqrt{\widehat{\mathrm{msep}}_{C_{i,J}|\mathcal{D}_I}\left(\hat{C}_{i,J}\right)}}{\hat{R}_i}.$$

i	Reserves	$\mathrm{PV}^{1/2}$	$\mathrm{PEE}^{1/2}$	$\mathrm{MSEP}^{1/2}$	CoVa
1	15'209	175	172	245	1.6%
2	25'619	684	486	840	3.3%
3	35'874	2'596	1'481	2'989	8.3%
4	90'234	7'556	3'836	8'474	9.4%
5	166'584	29'088	13'300	31'984	19.2%
6	314'665	67'560	28'248	73'227	23.3%
7	528'055	80'011	32'234	86'260	16.3%
8	1'200'820	152'676	55'740	162'533	13.5%
9	4'240'562	393'823	135'844	416'594	9.8%
total	**6'617'622**	**436'215**	**201'730**	**480'610**	**7.3%**

Table 2: Results Cape Cod Model 2.1: Reserves, Process Standard Deviation, Parameter Estimation Error, conditional MSEP and Coefficient of Variation according to Estimate 2.2.

j	$\hat{\gamma}_j^{raw}$	$\widehat{q\sigma_j^2}$
0	39.49%	9'760
1	19.58%	8'585
2	4.67%	1'172
3	1.51%	132
4	1.01%	251
5	0.49%	52
6	0.37%	3.5
7	0.08%	0.45
8	0.08%	0.03
9	0.10%	0.002

Table 3: Cape Cod Model 2.1: Parameter estimates (2.1) and (2.8).

For the Cape Cod Model 2.1 we also calculate the uncertainty in the CDR given by Estimate 2.3 and we compare these numbers with the conditional MSEP of the ultimate claim (see Estimate 2.2). Note that most of the uncertainty is contained in the next accounting year. As stated in Merz-Wüthrich (2008), this is to be expected for a triangle with short tailed development.

The results obtained from the ODP Model 2.4 are given in Table 5. In the ODP Model 2.4 we have a constant dispersion parameter ϕ for all development years. Since $\hat{\phi}\hat{\gamma}_j^{raw} >> \widehat{q\sigma_j^2}$ for late development years (see Tables 3 and 6),

i	$\widehat{\text{msep}}^{1/2}_{\text{CDR}_i^{(I+1)}\mid\mathcal{D}_I}(0)$	% of total uncertainty
1	245	100%
2	813	97%
3	2'886	97%
4	7'943	94%
5	30'845	96%
6	66'215	90%
7	48'072	56%
8	138'472	85%
9	382'114	92%
total	**429'567**	**89.4%**

Table 4: Cape Cod Model 2.1: Uncertainty in the CDR according to Estimate 2.3 and $\widehat{\text{msep}}^{1/2}_{\text{CDR}_i^{(I+1)}\mid\mathcal{D}_I}(0)/\widehat{\text{msep}}^{1/2}_{C_{i,J}\mid\mathcal{D}_I}\left(\hat{C}_{i,J}\right)$ for comparison with Estimate 2.2.

the process standard deviation in early accident years is much higher in this model compared to the Cape Cod Model 2.1 (see Table 2 and Table 5). However, for the total over all accident years the process standard deviations in the Cape Cod Model 2.1 (Table 2) are higher than in the ODP Model 2.4 (Table 5).

i	Reserves	$\text{PV}^{1/2}$	$\text{PEE}^{1/2}$	$\text{MSEP}^{1/2}$	CoVa
1	15'209	18'129	17'780	25'393	167.0%
2	25'619	23'530	20'246	31'041	121.2%
3	35'874	27'843	21'489	35'172	98.0%
4	90'234	44'159	27'869	52'218	57.9%
5	166'584	60'000	33'988	68'958	41.4%
6	314'665	82'463	40'870	92'035	29.2%
7	528'055	106'826	47'701	116'992	22.2%
8	1'200'820	161'092	63'570	173'182	14.4%
9	4'240'562	302'725	108'952	321'734	7.6%
total	**6'617'622**	**378'170**	**290'414**	**476'815**	**7.2%**

Table 5: Results ODP Model 2.4: Reserves, Process Standard Deviation, Parameter Estimation Error, conditional MSEP and Coefficient of Variation according to Estimate 2.6.

j	$\hat{\gamma}_j^{raw}$	$\hat{\phi}\hat{\gamma}_j^{raw}$
0	39.49%	8'535
1	19.58%	4'231
2	4.67%	1'010
3	1.51%	326
4	1.01%	219
5	0.49%	105
6	0.37%	79
7	0.08%	17
8	0.08%	16
9	0.10%	22

Table 6: ODP Model 2.4: Parameter estimates (2.1) and $\hat{\phi}$ as in (2.14).

In Table 7 we compare the CL development pattern to the development pattern given in (2.5). In this example the development patterns are very similar.

j	$\hat{\gamma}_j$	$\hat{\gamma}_j^{CL}$
0	58.62%	58.96%
1	29.06%	29.04%
2	6.94%	6.84%
3	2.24%	2.17%
4	1.50%	1.44%
5	0.72%	0.69%
6	0.54%	0.51%
7	0.12%	0.11%
8	0.11%	0.10%
9	0.15%	0.14%

Table 7: Normalised development pattern (2.3) and CL development pattern calculated from (1.4) with $\hat{\gamma}_0^{CL} = \hat{\beta}_0^{CL}$ and $\hat{\gamma}_j^{CL} = \hat{\beta}_j^{CL} - \hat{\beta}_{j-1}^{CL}$, $j > 0$.

For comparison we give the results obtained with the CL method according to Mack (1993). However, due to different model assumptions underlying the CL and the Cape Cod method it is difficult to compare these results.

i	Reserves	$PV^{1/2}$	$PEE^{1/2}$	$MSEP^{1/2}$	CoVa
1	15'126	191	187	267	1.8%
2	26'257	742	535	914	3.5%
3	34'538	2'669	1'493	3'058	8.9%
4	85'302	6'832	3'392	7'628	8.9%
5	156'494	30'478	13'517	33'341	21.3%
6	286'121	68'212	27'286	73'467	25.7%
7	449'167	80'077	29'675	85'398	19.0%
8	1'043'242	126'960	43'903	134'337	12.9%
9	3'950'815	389'783	129'769	410'817	10.4%
total	**6'047'061**	**424'379**	**185'024**	**462'960**	**7.7%**

Table 8: Results CL according to Mack (1993).

i/j	0	1	2	3	4	5	6	7	8	9	ν_i
0	5'946'975	9'668'212	10'563'929	10'771'690	10'978'394	11'040'518	11'106'331	11'121'181	11'132'310	11'148'124	15'473'558
1	6'346'756	9'593'162	10'316'383	10'468'180	10'536'004	10'572'608	10'625'360	10'636'546	10'648'192		14'882'436
2	6'269'090	9'245'313	10'092'366	10'355'134	10'507'837	10'573'282	10'626'827	10'635'751			14'456'039
3	5'863'015	8'546'239	9'268'771	9'459'424	9'592'399	9'680'740	9'724'068				14'054'917
4	5'778'885	8'524'114	9'178'009	9'451'404	9'681'692	9'786'916					14'525'373
5	6'184'793	9'013'132	9'585'897	9'830'796	9'935'753						15'025'923
6	5'600'184	8'493'391	9'056'505	9'282'022							14'832'965
7	5'288'066	7'728'169	8'256'211								14'550'359
8	5'290'793	7'648'729									14'461'781
9	5'675'568										15'210'363

Table 9: Cumulative payments and premiums ν_i.

Acknowledgements

The author would like to thank Alois Gisler and Mario V. Wüthrich for valuable comments that helped to improve an earlier version of this paper.

Bibliography

Alai, D. H., Merz, M., and Wüthrich, M. V. (2009). Mean square error of prediction in the Bornhuetter-Ferguson claims reserving method. *Annals of Actuarial Science*, 4(1):7–31.

Alai, D. H., Merz, M., and Wüthrich, M. V. (2010). Prediction uncertainty in the Bornhuetter-Ferguson claims reserving method: revisited. *Annals of Actuarial Science*, 5(1):7–17.

Alai, D. H. and Wüthrich, M. V. (2009). Taylor approximations for model uncertainty within the Tweedie exponential dispersion family. *ASTIN Bulletin*, 39(2):453–478.

Arjas, E. (1989). The claims reserving problem in non-life insurance: Some structural ideas. *ASTIN Bulletin*, 19(2):139–152.

Benktander, G. (1976). An approach to credibility in calculating IBNR for casualty excess reinsurance. *The Actuarial Review*, 3(2):7.

Bichsel, F. (1964). Erfahrungstarifierung in der Motorfahrzeug-Haftpflicht-Versicherung. *Bulletin of Swiss Ass. of Act.*, 64(1):119–130.

Bornhuetter, R. L. and Ferguson, R. E. (1972). The actuary and IBNR. *Proc. CAS*, LIX:181–195.

Buchwalder, M., Bühlmann, H., Merz, M., and Wüthrich, M. V. (2006). The mean square error of prediction in the chain ladder reserving method (Mack and Murphy revisited). *ASTIN Bulletin*, 36(2):521–542.

Bühlmann, H., De Felice, M., Gisler, A., Moriconi, F., and Wüthrich, M. V. (2009). Recursive credibility formula for chain ladder factors and the claims development result. *ASTIN Bulletin*, 39(1):275–306.

Bühlmann, H. and Gisler, A. (2005). *A Course in Credibility Theory and its Applications*. Springer-Verlag, Berlin Heidelberg.

Bühlmann, H. and Straub, E. (1983). Estimation of IBNR reserves by the methods chain ladder, Cape Cod and complementary loss ratio. In *International Summer School of the Swiss Association of Actuaries 1983, Leysin, Switzerland*.

De Alba, E. (2002). Bayesian estimation of outstanding claims reserves. *North American Actuarial Journal*, 6(4):1–20.

De Felice, M. and Moriconi, F. (2003). Risk based capital in P&C loss reserving or stressing the triangle. *Research Group on "Insurance Companies and Pension Funds"*, Working paper n.1.

De Jong, P. and Zehnwirth, B. (1983). Claims reserving, state-space models and the Kalman filter. *Journal of the Institute of Actuaries*, 110:157–181.

De Vylder, F. (1982). Estimation of IBNR claims by credibility theory. *Insurance: Mathematics and Economics*, 1(1):35–40.

England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3):443–518.

England, P. D., Verrall, R. J., and Wüthrich, M. V. (2012). Bayesian overdispersed Poisson model and the Bornhuetter-Ferguson claims reserving method. *Annals of Actuarial Science*, 6(2):258–283.

European Union (2009). Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II) (Text with EEA relevance). *Official Journal of the European Union, L 335*, 52:1–155.

FINMA (2004). White paper of the Swiss Solvency Test. *Swiss Federal Office of Private Insurance*.

FINMA (2006). Technisches Dokument zum Swiss Solvency Test, Version 02/10/2006. Technical report, Bundesamt für Privatversicherungen.

Gisler, A. and Wüthrich, M. V. (2008). Credibility for the chain ladder reserving method. *ASTIN Bulletin*, 38(2):565–600.

Hachemeister, C. and Stanard, J. (1975). IBNR claims count estimation with static lag functions. In *ASTIN Colloquium, Portimao, Portugal*.

Hesselager, O. and Witting, T. (1988). A credibility model with random fluctuations in delay probabilities for the prediction of IBNR claims. *ASTIN Bulletin*, 18(1):79–90.

Hovinen, E. (1981). Additive and continuous IBNR. In *ASTIN Colloquium, Loen, Norway*.

Jessen, A. H. and Rietdorf, N. (2011). Diagonal effects in claims reserving. *Scandinavian Actuarial Journal*, 2011(1):21–37.

Jewell, W. S. (1974a). Credibility means are exact Bayesian for exponential families. *ASTIN Bulletin*, 8(1):77–90.

Jewell, W. S. (1974b). Exact multidimensional credibility. *Bulletin of Swiss Ass. of Act.*, 74(2):193–214.

Jones, D. A. and Gerber, H. U. (1975). Credibility formulas of the updating type. *Transactions of society of actuaries*, 27:31–46.

Klugman, S. A. (1992). *Bayesian statistics in actuarial science: with emphasis on credibility*. Kluwer, Boston.

Kremer, E. (1985). *Einführung in die Versicherungsmathematik*. Vandenhoeck und Ruprecht, Göttingen.

Mack, T. (1990). Improved estimation of IBNR claims by credibility theory. *Insurance: Mathematics and Economics*, 9(1):51–57.

Mack, T. (1991). A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin*, 21(1):93–109.

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, 23(2):213–225.

Mack, T. (1994). Which stochastic model is underlying the chain ladder method? *Insurance: Mathematics and Economics*, 15(2):133–138. Special Issue in Honour of Professor G. Willem de Wit.

Mack, T. (2000). Credible claims reserves: The Benktander method. *ASTIN Bulletin*, 30(2):333–347.

Mack, T. (2006). Parameter estimation for Bornhuetter/Ferguson. *CAS Forum*, Fall 2006:141–157.

Mack, T. (2008). The prediction error of Bornhuetter/Ferguson. *ASTIN Bulletin*, 38(1):87–103.

Mack, T. and Venter, G. (2000). A comparison of stochastic models that reproduce the chain ladder reserve estimates. *Insurance: Mathematics and Economics*, 26(1):101–107.

Makov, U. E., Smith, A. F. M., and Liu, Y. H. (1996). Bayesian methods in actuarial science. *The Statistician*, 45(4):503–515.

Merz, M. and Wüthrich, M. V. (2008). Modelling the claims development result for solvency purposes. *CAS Forum*, Fall 2008:542–568.

Merz, M. and Wüthrich, M. V. (2010). Paid-incurred chain claims reserving method. *Insurance: Mathematics and Economics*, 46(3):568–579.

Merz, M., Wüthrich, M. V., and Hashorva, E. (2013). Dependence modelling in multivariate claims run-off triangles. *Annals of Actuarial Science*, 7(1):3–25.

Murphy, D. M. (1994). Unbiased development factors. *Proc. CAS*, LXXXI:154–222.

Neuhaus, W. (1992). Another pragmatic loss reserving method or Bornhuetter/Ferguson revisited. *Scandinavian Actuarial Journal*, 1992(2):151–162.

Ohlsson, E. and Lauzeningks, J. (2009). The one-year non-life insurance risk. *Insurance: Mathematics and Economics*, 45(2):203–208.

Quarg, G. and Mack, T. (2004). Munich chain ladder. *Blätter DGVFM*, 26(4):597–630.

Radtke, M. and Schmidt, K. D. (2004). *Handbuch zur Schadenreservierung*. Karlsruhe: Verlag Versicherungswirtschaft.

Renshaw, A. E. (1989). Chain ladder and interactive modelling. *Journal of the Institute of Actuaries*, 116:559–587.

Renshaw, A. E. and Verrall, R. J. (1998). A stochastic model for the chain ladder method. *British Actuarial Journal*, 14:903–923.

Saluz, A., Gisler, A., and Wüthrich, M. V. (2011). Development pattern and prediction error for the stochastic Bornhuetter-Ferguson claims reserving method. *ASTIN Bulletin*, 41(2):279–313.

Salzmann, R. and Wüthrich, M. V. (2010). Cost-of-capital margin for a general insurance liability runoff. *ASTIN Bulletin*, 40(2):415–451.

Salzmann, R. and Wüthrich, M. V. (2012). Modeling accounting year dependence in runoff triangles. *European Actuarial Journal*, 2(2):227–242.

Salzmann, R., Wüthrich, M. V., and Merz, M. (2012). Higher moments of the claims development result in general insurance. *ASTIN Bulletin*, 42(1):355–384.

Schmidt, K. D. and Wünsche, A. (1998). Chain ladder, marginal sum and maximum likelihood estimation. *Blätter DGVFM*, 23(3):267–277.

Scollnik, D. P. M. (2001). Actuarial modeling with MCMC and BUGs. *North American Actuarial Journal*, 5(2):96–124.

Shi, P., Basu, S., and Meyers, G. G. (2012). A Bayesian log-normal model for multivariate loss reserving. *North American Actuarial Journal*, 16(1):29–51.

Straub, E. (1988). *Non-Life Insurance Mathematics*. Springer-Verlag, Association of Swiss Actuaries, Zürich.

Sundt, B. (1981). Recursive credibility estimation. *Scandinavian Actuarial Journal*, 1981(1):3–21.

Taylor, G. C. (2000). *Loss Reserving: An Actuarial Perspective*. Kluwer Academic Publishers, Boston.

Taylor, G. C. (2002). Written discussion on the paper "Stochastic claims reserving in general insurance" by P. D. England and R. J. Verrall. *British Actuarial Journal*, 8(3):540–542.

Verrall, R. J. (1990). Bayes and empirical Bayes estimation for the chain ladder model. *ASTIN Bulletin*, 20(2):217–243.

Verrall, R. J. (1991). On the estimation of reserves from loglinear models. *Insurance: Mathematics and Economics*, 10(1):75–80.

Verrall, R. J. (2000). An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: Mathematics and Economics*, 26(1):91–99.

Verrall, R. J. (2004). A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving. *North American Actuarial Journal*, 8(3):67–89.

Witting, T. (1987). Kredibilitätsschätzungen für die Anzahl IBNR-Schäden. *Blätter der DGVM*, 18(1):45–58.

Wüthrich, M. V. (2010). Accounting year effects modeling in the stochastic chain ladder reserving method. *North American Actuarial Journal*, 14(2):235–255.

Wüthrich, M. V., Bühlmann, H., and Furrer, H. (2010). *Market-Consistent Actuarial Valuation*. Springer-Verlag, Berlin Heidelberg, 2nd edition.

Wüthrich, M. V., Embrechts, P., and Tsanakas, A. (2011). Risk margin for a non-life insurance run-off. *Statistics & Risk Modeling with Applications in Finance and Insurance*, 28(4):299–317.

Wüthrich, M. V. and Merz, M. (2008). *Stochastic Claims Reserving Methods in Insurance*. Wiley, Chichester.

Wüthrich, M. V. and Merz, M. (2013). *Financial Modeling, Actuarial Valuation and Solvency in Insurance*. Springer-Verlag, Berlin Heidelberg.

Curriculum Vitae

PERSONAL DATA

Name	Annina Saluz
Date of Birth	March, 27, 1986
Nationality	Swiss

EDUCATION

Ph.D. Studies **03/2010 – 11/2014**
Ph.D. studies in mathematics, ETH Zurich.
Supervisor: Prof. Dr. Alois Gisler.

Undergraduate Studies **10/2005 – 02/2010**
Master of Science ETH Mathematics.

High School **08/1999 – 06/2005**
Matura, Schweizerische Alpine Mittelschule Davos.
Major: Applied Mathematics and Physics.

EMPLOYMENT

ETH Zurich **03/2010 – 11/2014**
Research and teaching assistant.

AXA Winterthur **09/2010 – 03/2011**
Internship in Actuarial Department.

ETH Zurich **09/2008 – 12/2009**
Teaching assistant.