Prediction Uncertainty in Stochastic Claims Reserving Methods

A dissertation submitted to
ETH ZURICH
for the degree of
Doctor of Sciences

presented by
DANIEL HAKIM ALAI
Bachelor of Mathematics, University of Waterloo
born August 9, 1983
citizen of The Netherlands

accepted on the recommendation of
Prof. Dr. Paul Embrechts, examiner
Prof. Dr. Richard J. Verrall, co-examiner
Dr. Mario V. Wüthrich, co-examiner

2009
voor Emilia
Acknowledgements

I would like to thank my supervisors Paul Embrechts and Mario Wüthrich for taking a chance on an undergraduate student from Waterloo with little experience in the academic world. To Paul Embrechts, I am further indebted for the numerous important conversations we had throughout the years. Your experience and advice guided me through the ups and downs of my studies at the ETH. To Mario Wüthrich, I am grateful for the meticulous attention to detail and for having the patience necessary to properly introduce me to the art and science of research. I would also like to kindly thank Michael Merz for the valuable collaboration and Richard Verrall for agreeing to act as my co-examiner.

I gratefully acknowledge the financial support from ACE Limited and Risk-Lab Switzerland. The contacts and collaboration with members of ACE Limited provided an extra dimension to my studies for which I am very thankful. The environment at the ETH I owe to many of my fellow colleagues, with whom I shared many great experiences and who leave me with countless fond memories.

To my friends, new and old, I owe many thanks. Because of you I was able to adjust to life in Switzerland as well as enjoy glimpses of life back in Canada. I thank you for the bonds of friendship we made during the time we spent together.

Finally, I would like to thank my family. My mother, for always encouraging my academic aspirations and providing me with the foundation on which all else now flourishes. My father, for always having my best interests at heart and being there for me whenever I needed extra support in any aspect of life. Lastly my sister, for always looking out for and protecting her younger, and often less experienced, brother. I share this accomplishment with you.

Zurich, December 8, 2009. Daniel H. Alai
Abstract

Often in non-life insurance, claims reserves are the largest item on the liability side of the balance sheet. Therefore, given the available information about the past, the prediction of an adequate amount to face the responsibilities assumed by the non-life insurance company as well as the quantification of the uncertainties in these reserves are major issues in actuarial practice and science.

One avenue for development in this area is to build on the capabilities of classical claims reserving methods, methods that have been on the forefront of claims reserving for decades and hence hardwired into the system. Another avenue is to embrace new, more complex methods and put forth the time and effort necessary to incorporate them into the day-to-day operations of the company.

We aim to deal with both routes as we use the tools of stochastic modelling and generalized linear models (GLM) to update old methods as well as introduce new ones. Furthermore, we aim to make these developments accessible to practice and hence focus considerable effort on ease of implementation.

Due to their simplicity, the chain ladder (CL) and Bornhuetter-Ferguson (BF) methods are the most commonly used claims reserving methods in practice. However, in contrast to the CL method, the quantification of the uncertainty that surrounds the BF method has not been thoroughly explored in the literature. We aim to shed light on this problem in the framework of GLM, which is increasingly used in insurance.

In addition to tackling the problem of uncertainty in the BF method, we use GLM to study the problem of model uncertainty in general. Demand for this topic of study is becoming more relevant due to new developments in solvency requirements. We study the so-called Tweedie exponential dispersion family, a family that includes many important and widely-used models, and test the sensitivity of the claims reserves and their uncertainty over this family. The results provide new insight and allow models to be considered that previously presented difficulties in practice.

Finally, the complexity of standard methods is changed when one considers multivariate versions with underlying dependency. In order to improve claims reserving prediction, practitioners often separate small and large claims and then apply a two-dimensional CL algorithm to the two claims classes. We give a mathematical framework for this two-dimensional analysis that allows for the quantification of the prediction uncertainty.
Kurzfassung

Der größte Posten auf der Passivseite der Bilanz einer Nicht-Lebensversicherung sind die Schadenreserven: Die Hauptaufgabe eines Schadenreservierungsaktuars ist die Schätzung der zukünftigen Schadenzahlungen, welche mit ausreichenden Schadenreserven bedeckt werden müssen. Dabei muss er mit Hilfe von Beobachtungen in der Vergangenheit und Expertenwissen sowohl die erwarteten Verpflichtungen abschätzen, als auch die Unsicherheit in dieser Abschätzung quantifizieren.

Eine Möglichkeit diesen Bereich weiterzuentwickeln ist, sich auf die Möglichkeiten der klassischen Schadenreservierungsmethoden abzustützen, die seit Jahrzehnten im Gebiet der Schadenreservierung führend und deshalb fest im ganzen System verknüpft sind. Eine weitere Möglichkeit besteht darin, neue und komplexere Methoden mit einzubeziehen und auch die nötige Zeit und den nötigen Aufwand aufzubringen, um diese in die alltäglichen Abläufe der Gesellschaft einzubeziehen.

Wir möchten beide Möglichkeiten behandeln, indem wir die Instrumente der stochastischen Modellierung und der verallgemeinerten linearen Modelle (GLM) verwenden, um die alten Methoden auf den neuesten Stand zu bringen, aber auch um neue Methoden einzuführen. Darüber hinaus möchten wir diese Entwicklungen der Praxis zugänglich machen und konzentrieren uns deshalb stark auf eine einfache Implementierung.


Neben der Bearbeitung des Problems der Unsicherheit der BF-Methode verwenden wir GLM, um das Problem der Modell-Unsicherheit im Allgemeinen zu behandeln. Der Bedarf innerhalb dieses Gebietes steigt aufgrund der neuen Entwicklungen im Bereich der Solvenzanforderungen stetig an. Wir untersuchen die so genannte Tweedie exponential dispersion Familie, die viele wichtige und weit verbreitete Modelle umfasst und testen die Sensitivität der Schadenreserven und deren Unsicherheit bezüglich dieser Familie. Die Ergebnisse liefern eine neue Sicht auf dieses Gebiet und ermöglichen Modelle zu betrachten, die bisher in der Praxis Schwierigkeiten bereitet haben.

Schließlich ändert sich die Komplexität der Standard-Methoden, wenn man multivariate Modelle mit zugrunde liegender Abhängigkeit betrachtet. Um Scha-
Kurzfassung

denreservierungsprognosen zu verbessern, trennt man in der Praxis oft kleine von
grosen Schäden, um anschliessend auf die zwei Schadensklassen einen zweidimen-
sionalen CL-Algorithmus anzuwenden. Wir definieren die zugrundeliegende mathe-
matische Struktur dieser zweidimensionalen Analyse um die Unsicherheit der Pro-
gnose quantifizieren zu können.
Contents

Acknowledgements iii
Abstract v
Kurzfassung vii

1 Introduction 1
1.1 The Claims Settlement Process 1
1.2 Claims Reserving Notation 2
1.3 Outstanding Loss Liabilities 3
1.4 Prediction Uncertainty 4

2 Basic Claims Reserving Methods 7
2.1 The Chain Ladder Method 7
2.2 The Bornhuetter-Ferguson Method 9
2.3 Simple Bayesian Methods 11

3 Stochastic Bornhuetter-Ferguson Models 13
3.1 A Bayesian Approach 13
3.2 A Distribution-Free Approach 17
3.3 A Likelihood Approach (Papers I and II) 19
3.4 Numerical Comparison 22

4 Generalized Linear Models 25
4.1 The Exponential Dispersion Family 25
4.2 The Tweedie Exponential Dispersion Family 28
4.3 Model Uncertainty (Paper III) 29

5 Multivariate Chain Ladder Models 35
5.1 Correlated Portfolios 36
5.2 Claims Separation by Size (Paper IV) 38

Conclusions and Future Research 43

Bibliography 45
Accompanying papers

I Daniel H. Alai, Michael Merz, and Mario V. Wüthrich (2009):
  Mean square error of prediction in the Bornhuetter-Ferguson claims
  reserving method.

II Daniel H. Alai, Michael Merz, and Mario V. Wüthrich (2009):
  Prediction uncertainty in the Bornhuetter-Ferguson claims
  reserving method: revisited.
  Submitted.

III Daniel H. Alai and Mario V. Wüthrich (2009):
  Taylor approximations for model uncertainty within the
  Tweedie exponential dispersion family.

IV Daniel H. Alai and Mario V. Wüthrich (2009):
  Modelling small and large claims in a chain ladder framework.
  Submitted.
1. Introduction

1.1 The Claims Settlement Process

We begin with necessary background that motivates the study of the problems presented in this thesis. Additional references for a detailed study of the claims settlement process are e.g. Taylor (2000) and Wüthrich and Merz (2008). The branch of insurance considered in this thesis is known as non-life insurance in Continental Europe, general insurance in Great Britain, and property and casualty insurance in North America. The distinction from life insurance is fundamental, since the behaviours of the two types of contracts varies considerably.

In a non-life insurance policy, well-defined payments, also called premiums, made to the insurer provide the insured with financial coverage against a loss in case of well-specified random occurrences. If such an event occurs, the insured files a claim with the insurance company. The insurer is obliged to pay the insured the amount specified under the policy, called the claim amount or loss amount. Depending on the type of policy, the determination of the proper claim amount can often be very difficult and time consuming. Factors such as reporting delay, recovery process time required for the insurer to obtain all necessary details surrounding the claim, and new developments that can reopen closed claims contribute to this difficulty.

The assets of a non-life insurance company arrive in the form of premiums. They are deterministic in nature, thereby making the valuation of a non-life insurance company’s incoming assets a relatively easy task. The only random component is the underlying volume, i.e. the number of policies sold. The arrival of liabilities, in contrast to the assets, are completely random in nature and require careful consideration. A non-life insurance company is rarely free of potential liability with respect to any of its present and past policies because of the above mentioned difficulties present in claim amount valuation.

This fundamental difference of a non-life insurance company’s incoming assets and liabilities introduces a surplus process. The surplus is defined to be the assets minus the liabilities and is available to the insurance company, amongst other things, for investment and reinsurance. If the surplus drops below a certain threshold (the minimal capital requirement), the insurance company is deemed insolvent. In theory, this means the company is ruined. In reality, the company’s liabilities are assumed by the state or the company is taken-over by another insurer.
The study of the surplus process and ruin probabilities is not discussed further, for more details see e.g. Asmussen (2000).

Inflation also plays an important role in the claims settlement process. Claim amounts are subject to specific types of inflation that depend on the type of the insurance contract as well as general price inflation. Contract dependent inflation, also called claims inflation, differs from price inflation and can continue far beyond the claim occurrence. The impact of inflation is not discussed further in this thesis; see e.g. Taylor (2000) and Hart et al. (1996) for more about claims inflation and Teugels and Sundt (2004) and the references therein for a study of the impact of price inflation on the claims settlement process.

1.2 Claims Reserving Notation

We assume loss data for a portfolio of policies is given by a claims development triangle of observations. However, all claims reserving methods discussed in this thesis can also be applied to other shapes of loss data (e.g. claims development trapezoids). The claims development triangle has indices \( i \in \{0, 1, \ldots, I\} \) and \( j \in \{0, 1, \ldots, J\} \) with \( I = J \), which refer to accident years and development years, respectively.

The incremental claim amounts (i.e. incremental payments, change of reported claim amount or number of newly reported claims) for accident year \( i \) and development year \( j \) are denoted by \( X_{i,j} \) and the cumulative claim amounts (i.e. cumulative payments, claims incurred or total number of reported claims) of accident year \( i \) up to development year \( j \) are given by

\[
C_{i,j} = \sum_{k=0}^{j} X_{i,k}.
\]

We assume that the last development year is given by \( I \) (i.e. \( X_{i,j} \equiv 0 \) for all \( j > I \)) and the last observed accident year is given by \( I \). Since development is assumed to cease after development year \( I \), we call \( C_{i,I} \) the ultimate claim amount (or total number of reported claims) of accident year \( i \).

<table>
<thead>
<tr>
<th>accident year ( i )</th>
<th>development year ( j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>( i )</td>
<td>( D_i )</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>( I )</td>
<td>( D_I )</td>
</tr>
</tbody>
</table>

Figure 1.1: Claims development triangle.
1.3 Outstanding Loss Liabilities

Usually, at time $I$ (i.e. calendar year $I$), we have observations $D_I$ in the upper claims development triangle, defined as follows,

$$D_I = \{X_{i,j} ; i + j \leq I\}.$$  

We need to predict the random variables in its complement,

$$D_f^I = \{X_{i,j} ; i + j > I\}.$$  

Figure 1.1 shows the data structure of the claims development triangle described above.

In the remainder of the thesis, we assume that the $X_{i,j}$ denote incremental claim payments, we continue to refer to them as incremental claim amounts. Certain claims reserving methods, which are described in detail later, require the assumption of non-negative $X_{i,j}$. The assumption of individual claim payments does not guarantee this due to issues such as salvage and subrogation. However, on aggregated portfolios, negative incremental claim payments are seldomly observed.

**1.3 Outstanding Loss Liabilities**

Due to the inherent random nature of a non-life insurance company’s liabilities, the task of setting aside proper provisions to limit the chance of ruin is difficult. The task can be approached on a per claim level. For each reported claim a reserve amount is estimated by claims handlers that is deemed appropriate to cover the remaining payments of the policy. This amount is called the **case reserves**.

For an aggregate portfolio, the case reserves of all reported claims are added to provide an estimate of future liabilities on reported claims, the result is called the **claims incurred**. However, this omits consideration for claims that have already occurred but have yet to be reported, or so-called incurred but not yet reported (IBNyR) claim amounts. One could independently set aside an amount for the IBNyR, but in practice the inclusion of the IBNyR is done through modelling the aggregate claim amounts. By modelling the aggregate claim amounts, one is able to capture the emerging IBNyR pattern.

Let $R_i$ and $R$ denote the outstanding loss liabilities for accident year $i$ at time $I$,

$$R_i = \sum_{j>I-i} X_{i,j} = C_{i,I} - C_{i,I-i}, \quad i \in \{1, \ldots, I\},$$

and the total outstanding loss liabilities for all accident years,

$$R = \sum_{i=1}^I R_i,$$

respectively. The prediction of the outstanding loss liabilities $R_i$ and $R$, as well as quantifying the uncertainty in this prediction, is the classical actuarial claims reserving problem studied at every non-life insurance company. We use the term
claims reserves to mean the prediction of the outstanding loss liabilities. Hence, let \( \hat{R}_i \) and \( \hat{R} \) denote the claims reserves for accident year \( i \) at time \( I \),
\[
\hat{R}_i = \sum_{j>I} \hat{X}_{i,j} = \hat{C}_{i,I} - \hat{C}_{i,I-i}, \quad i \in \{1, \ldots, I\},
\]
and the total claims reserves for aggregated accident years,
\[
\hat{R} = \sum_{i=1}^{I} \hat{R}_i,
\]
respectively, where \( \hat{X}_{i,j} \) is a predictor for \( X_{i,j} \) and \( \hat{C}_{i,j} \) for \( C_{i,j} \). In practice a provision for adverse deviation (PAD) is added to the prediction of the outstanding loss liabilities, the sum of the two is called the risk-adjusted claims reserves. The PAD is added to further ensure the adequacy of the reserves and the stability of the insurance company.

In addition to making claim payments, insurance companies must also endure expenses to ensure the correctness of the payments and to recover losses when possible. Such expenses include the salary of the claims handling department, maintenance of IT systems, etc. These expenses are typically classified as allocated loss adjustment expenses (ALAE), those that can be traced back to a particular policy, and unallocated loss adjustment expenses (ULAE). The ALAE can be thought of variable cost and the ULAE as fixed. The ALAE can be incorporated in the observed data, whereas the ULAE cannot. For this reason ULAE requires specific consideration, but we do not further discuss this here. For more information regarding ULAE reserving, see e.g. Buchwalder et al. (2006a) and Wüthrich and Merz (2008) and the references therein.

1.4 Prediction Uncertainty

As alluded to above, finding suitable claims reserves is not the end of the story, rather the beginning. There are many claims reserving methods available to predict the outstanding loss liabilities, the challenge is to quantify not only the claims reserves but also the uncertainty of the resulting predictors. This challenge was undertaken in the work of Taylor and Ashe (1983) through statistical regression analysis and de Jong and Zehnwirth (1983) using a state-space approach and the Kalman filter as a method of estimation; see e.g. Bühlmann and Gisler (2005) or Taylor (2000) for more on the Kalman filter. Here, as in the work of Taylor and Ashe (1983), we quantify the prediction uncertainty with the aid of the most popular such measure, the so-called mean square error of prediction (MSEP). For predictor \( \hat{C}_{i,I} \) of the ultimate claim amount \( C_{i,I} \) of accident year \( i \), the conditional MSEP is defined as
\[
\text{msep}_{C_{i,I}|D_I}(\hat{C}_{i,I}) = E\left[ (\hat{C}_{i,I} - C_{i,I})^2 | D_I \right] .
\]
Note that with regards to the conditional MSEP, it does not matter whether one considers the predictor \( \hat{C}_{i,I} \) of the ultimate claim amount or the predictor \( \hat{R}_i \) of
the claims reserves of accident year $i$. Both yield the same result. We adopt the
convention of using the predictor of the ultimate claim amount. If the predictor
$\hat{C}_{i,I}$ is $D_I$-measurable, the conditional MSEP decouples as follows:

$$\text{mse}_{C_{i,I}|D_I}(\hat{C}_{i,I}) = \text{Var}(C_{i,I}|D_I) + \left(\hat{C}_{i,I} - E[C_{i,I}|D_I]\right)^2,$$

The first term on the right-hand side of the above equation is called the conditional process variance (called the statistical error in Taylor and Ashe (1983)), it represents the inherent uncertainty of the underlying model chosen for the observed data. The second term on the right-hand side is called the conditional estimation error, it constitutes the uncertainty in the estimation of the accompanying model parameters.

For aggregated accident years, the conditional MSEP of the predictor of the aggregate ultimate claim amount is defined as

$$\text{mse}_{\sum_{i=1}^l C_{i,I}|D_I}(\sum_{i=1}^l \hat{C}_{i,I}) = E\left[\left(\sum_{i=1}^l \hat{C}_{i,I} - \sum_{i=1}^l C_{i,I}\right)^2\bigg| D_I\right],$$

and similarly decouples in the event that $\sum_{i=1}^l \hat{C}_{i,I}$ is $D_I$-measurable. The objective to estimate the conditional MSEP plays a paramount role in all of the accompanying papers.

Other general uncertainty measures that could be considered include value-at-risk and expected shortfall. Furthermore, one can attempt to attain the entire distribution. This is highly beneficial to industry regarding solvency matters. However, in many cases this can only be done numerically via simulations. One avenue of attaining the entire distribution is based on statistical bootstrapping; see e.g. England and Verrall (1999, 2006) and England (2002), another in Markov chain Monte Carlo simulation methods; see e.g. Gilks et al. (1996) and Wüthrich and Merz (2008). We argue that analytic solutions can provide more insight to the nature of the risk and hence advocate the use of the conditional MSEP as a first risk measure, i.e. a risk measure for which analytic solutions can often be obtained.
2. Basic Claims Reserving Methods

In this chapter we discuss some basic methods used to predict the outstanding loss liabilities of non-life insurance companies. For an excellent overview of deterministic claims reserving methods, see e.g. Taylor (2000). For stochastic claims reserving methods, England and Verrall (2002) and Wüthrich and Merz (2008) provide excellent summaries. The chain ladder (CL) and Bornhuetter-Ferguson (BF) claims reserving methods are perhaps the two most commonly used claims reserving methods in practice today. This chapter provides the details of both methods as well as others, however, we reserve an in-depth discussion of the BF claims reserving method for Chapter 3.

2.1 The Chain Ladder Method

The CL claims reserving method dates back many years. As well as being discussed in Harnek (1966), Taylor (1986) credits its name to Prof. Beard’s work at the UK Department of Trade in the early 1970’s. At the center of the method are the so-called age to age factors that develop the cumulative claim amounts one period. Linking these factors together like a chain, one develops the age to ultimate factor. This ladder of factors describes the emerging pattern of the claim amounts. This claims reserving method is arguably the most widely used in practice and also goes by the name “the loss development triangle method”.

Kremer (1982) introduced the idea of using parameterized structures as underlying stochastic models of the CL claims reserving method. Using a linear predictor in the framework of generalized linear models (GLM), Kremer modelled the incremental observations using the log-normal distribution. In short, he modelled the log response variables and regressed on two non-interactive covariables, one attributed to the accident year and the other, the development year. His work spawned interest that further explored the underlying log-normal assumption; see e.g. Renshaw (1989) and Verrall (1989, 1990, 1991a, b).

The use of GLM evolved further due to Renshaw (1994) and Renshaw and Verrall (1998) who utilized the overdispersed Poisson distribution as the underlying stochastic model of the CL claims reserving method. The fact that the overdispersed Poisson model reproduced the CL claims reserves was known at this time, see Hachemeister and Stanard (1975), but Renshaw and Verrall were first to express it in the GLM framework. Their goal was to provide a more general framework for
Basic Claims Reserving Methods

modelling claims reserves, of which the classical CL method would be but one special case. GLM claims reserving methods are discussed in further detail in Chapter 4.

The use of GLM did not come without criticism, Mack (1994) and Mack and Venter (2000) critiqued the use of the log-normal and overdispersed Poisson assumptions, respectively (with counterarguments provided by Verrall and England (2000)). They strongly advocated the distribution-free model given below as being the only model completely faithful to the classical CL method.

The classical actuarial literature often explains the CL claims reserving method as a pure computational algorithm to estimate claims reserves. A distribution-free stochastic model underlying the CL algorithm was proposed by Mack (1993). It was based on the following model assumptions:

Model Assumptions 2.1 (Chain ladder model):

- There exist deterministic development factors \( f_0, \ldots, f_{I-1} > 0 \) such that for all \( i \in \{0, \ldots, I\} \) and all \( j \in \{1, \ldots, I\} \) we have
  \[
  E[C_{i,j} | C_{i,0}, \ldots, C_{i,j-1}] = E[C_{i,j} | C_{i,j-1}] = f_{j-1} C_{i,j-1}.
  \]

- Claims \( C_{i,j} \) of different accident years \( i \) are independent.

An easy exercise in calculating conditional expectation leads to

\[
E[C_{i,I} | D_I] = f_{I-1} E[C_{i,I-1} | C_{i,I-1}] = \ldots = C_{i,I-i} \prod_{j=I-i}^{I-1} f_j,
\]

for all \( i \in \{1, \ldots, I\} \), where the factors \( f_j \) are called the CL factors. Given the observations \( D_I \) and CL factors \( f_j \), the above equation gives a recursive algorithm for predicting the ultimate claim amount \( C_{i,I} \) by \( E[C_{i,j} | D_I] \). However, in most practical applications the CL factors \( f_j \) are not known and have to be estimated from the data \( D_I \). It is well known that the \( D_I \)-measurable estimators for the CL factors \( f_j \), defined by

\[
\hat{f}_j = \frac{\sum_{i<I-j} C_{i,j+1}}{\sum_{i<J-j} C_{i,j}},
\]

for all \( j \in \{0, \ldots, I-1\} \), are unbiased and uncorrelated; see e.g. Mack (1993). However, they are not independent since the squares of two successive estimators \( \hat{f}_j \) and \( \hat{f}_{j+1} \) are negatively correlated; see e.g. Mack et al. (2006) and Wüthrich et al. (2008).

The properties of the CL factor estimates \( \hat{f}_j \) imply that, given \( C_{i,I-i} \), the CL predictor of the ultimate claim amount \( C_{i,I} \), defined by

\[
\hat{C}_{i,I}^{CL} = C_{i,I-i} \prod_{j=I-i}^{I-1} \hat{f}_j, \quad i \in \{1, \ldots, I\},
\]
is an unbiased estimator of $E[C_{i,I} | D_I]$. The predictor of the outstanding loss liabilities of accident year $i$ using the CL claims reserving method is given by

$$\hat{R}^{CL}_i = \hat{C}_{i,I}^{CL} - C_{i,I}, \quad i \in \{1, \ldots, I\}.$$ 

There has been much discussion in the literature aimed at finding appropriate underlying stochastic models that reproduce the classical CL predictor of the outstanding loss liabilities given above; see e.g. Verrall (2000) for an excellent overview.

Given the distribution-free model of Mack (1993) where an additional assumption is made for the variance (compared to Model Assumptions 2.1), the conditional MSEP for accident year $i$ can be estimated by

$$\hat{\text{mse}}_{C_{i,I} | D_I} (\hat{C}_{i,I}^{CL}) = \left( \frac{\hat{C}_{i,I}^{CL}}{I-1} \sum_{j=I-i}^{I-1} \frac{\hat{\sigma}_j^2}{2} \left( \frac{1}{\hat{\sigma}_{C_{i,j}}^{CL}} + \frac{1}{\sum_{k=0}^{I-j-1} C_{k,j}} \right) \right);$$

see Mack (1993). A different approach to estimating the conditional MSEP is given in Buchwalder et al. (2006b). The difference therein lies in the approximation of the estimation error.

### 2.2 The Bornhuetter-Ferguson Method

The BF method goes back to Bornhuetter and Ferguson (1972). Apart from its simplicity, the BF method is a very popular claims reserving method since it is rather robust against outliers in the observations. The method is deemed very useful when the data are unstable. The presence of prior knowledge of the ultimate claim amount provides this stability and is fundamental to the method. In other words, the method allows for incorporating prior knowledge from experts, premium calculations, and strategic business plans. A rough outline of the method goes as follows, see Verrall (2004):

- Exogenously obtain an initial estimate of the expected ultimate claim amount $E[C_{i,I}]$ for each accident year $i$.
- Determine the proportion of the ultimate claim amount to be distributed amongst the development years, this is typically done using the CL factors.
- Using the above proportions, apply them to the exogenous estimate of the expected ultimate claim amount in order to predict the outstanding loss liabilities and the ultimate claim amount.

In this way, the BF estimator of the expected ultimate claim amount can be thought of as a posterior estimate. The BF predictor of the outstanding loss liabilities $R_i$, for accident year $i$, using the CL factor estimates $\hat{f}_j$ to determine the claims development pattern is defined by

$$\hat{R}^{BF}_i = (1 - \hat{\beta}_{I-i}) \hat{\nu}_i, \quad i \in \{1, \ldots, I\},$$
where
\[
\hat{\beta}_j = \prod_{k=j}^{I-1} \frac{1}{f_k},
\]
and \(\hat{\nu}_i\) is an exogenous prior estimate of the expected ultimate claim amount \(E[C_{i,I}]\).

We further define \(\hat{\beta}_I = 1\) and call \((\hat{\beta}_j)_{j=0,...,I}^\prime\) an estimate of the cumulative development pattern. In this way, it is clear to interpret the BF predictor for the claims reserves to be the outstanding portion of the prior estimate of the expected ultimate claim amount. The associated BF predictor of the ultimate claim amount is given as follows:
\[
\hat{C}^{BF}_{i,I} = C_{i,I-i} + \hat{R}^{BF}_i, \quad i \in \{1, \ldots, I\}.
\]

In contrast to the CL method, the BF method has proven to be very robust, in particular, against instability in the proportion of ultimate claim amounts paid in early development years. For accident year \(i\), the CL method relies heavily on the last observed cumulative claim amount \(C_{i,I-i}\). In fact, under the CL method, the estimate of the ultimate claim amount is directly proportional to the last observed cumulative claim \(C_{i,I-i}\). Due to the BF method’s reliance on expert opinion, this pitfall is not observed. Another fundamental difference between the CL and BF methods, as examined in Mack (2008) is with respect to tail factors. The BF method inherently includes a tail factor, whereas the CL method does not; see e.g. Mack (1999) for tail factor consideration in the CL method. As stated in Chapter 1, however, we assume the data to be fully developed and hence do not encounter this difference.

Hence, the BF method is based on the simple idea of stabilizing the BF predictor \(\hat{C}^{BF}_{i,I}\) using an initial estimate \(\hat{\nu}_i\) of the expected ultimate claim amount \(E[C_{i,I}]\) based on external knowledge. It is standard practice to use the prior estimate \(\hat{\nu}_i\) with the CL factor estimates \(\hat{f}_j\) to predict the ultimate claim amount \(C_{i,I}\). In this case, the CL method and the BF method only differ in their choice of the estimate for the expected ultimate claim amount (CL estimate versus prior estimate). This property can be seen mathematically as follows:
\[
\begin{align*}
\hat{C}^{CL}_{i,I} &= C_{i,I-i} + (1 - \hat{\beta}_{I-i}) \hat{C}^{CL}_{i,I}, \\
\hat{C}^{BF}_{i,I} &= C_{i,I-i} + (1 - \hat{\beta}_{I-i}) \hat{\nu}_i.
\end{align*}
\]

Hence, in this regard the BF method is a variant of the CL method that uses external information to obtain an initial estimate for the expected ultimate claim amount.

The BF method, as it was stated in the original work of Bornhuetter and Ferguson (1972), was not formulated in a probabilistic way. The works of Verrall (2004), Mack (2000, 2008), and Wüthrich and Merz (2008) Section 2.2, as well as our own, Paper I and Paper II, puts the BF method into a probabilistic framework.

Mack (2000) studied this from a probabilistic point of view resulting in a method referred to as the Benktander-Hovinen method, which we discuss in detail in the
2.3 Simple Bayesian Methods

In this section, we present two methods that fall into the category of simple Bayesian methods. A simple Bayesian method is defined here to be a method that incorporates some prior, external knowledge, of a single value. The BF method, for example, satisfies this criteria since an exogenous estimate of the expected ultimate claim amount is incorporated to determine the case reserves. Two other methods that fall into this category, which we discuss presently, are the Benktander-Hovinen method and the Cape Cod method.

The Benktander-Hovinen Method

The Benktander-Hovinen (BH) method proposed in Benktander (1976) and studied in detail in Mack (2000), has been in existence for many years. However, since it did not receive appropriate exposure it was rediscovered twice. First in 1981 by Eso Hovinen, a Finnish actuary who presented the work in the 1981 ASTIN Colloquium, and second by Neuhaus (1992), who published his findings in the Scandinavian Actuarial Journal. Both reinventors having been unaware of the work that preceded theirs.

As explained above, when the BF method is used in association with the CL factors, the two methods are related. They are separated only by their estimate of the expected ultimate claim amount. The CL method relies on the data, in contrast to the the BF method, which relies on an exogenous prior estimate. The
method described in Benktander (1976) marries the two concepts by introducing a credibility mixture
\[ c_i \hat{C}_{i,I}^{CL} + (1 - c_i) \hat{\nu}_i, \]
of both the CL ultimate claim amount obtained using the data and the prior estimate of the ultimate claim amount used in the BF method. This credibility mixture is then used as the prior estimate of the ultimate claim amount in another application of the BF procedure. It is for this reason that it also goes by the name “iterated Bornhuetter-Ferguson method”. For accident year \( i \), the credibility factor \( c_i \) is taken to be the appropriate value of the cumulative development pattern, \( \hat{\beta}_{I-i} \). Hence if the outstanding portion of the ultimate claim amount is close to one (i.e. \( \hat{\beta}_{I-i} \) is close to zero), we associate a high credibility to the exogenous estimate \( \hat{\nu}_i \).

The predictor of the outstanding loss liabilities for individual accident years under the BH method is given by
\[ \hat{R}_{i}^{BH} = (1 - \hat{\beta}_{I-i}) \left( \hat{\beta}_{I-i} \hat{C}_{i,I}^{CL} + (1 - \hat{\beta}_{I-i}) \hat{\nu}_i \right), \quad i \in \{1, \ldots, I\}, \]
and the associated BH predictor of the ultimate claim amount is given by
\[ \hat{C}_{i,I}^{BH} = C_{i,I-i} + \hat{R}_{i}^{BH}, \quad i \in \{1, \ldots, I\}. \]
Note that further iterations of the BF procedure are possible. Theorem 1 in Mack (2000) states that the BH method converges to the CL method as the number of iterations goes to infinity.

The Cape Cod Method
The Cape Cod (CC) method was independently derived in the works of Stanard (1980) and Bühlmann (1983); see e.g. Wüthrich and Merz (2008) for a more detailed overview of this method. As with the BF method, the CC method assumes exogenous information of the expected ultimate claim amount. However, the CC method does so using the premium received for each accident year \( i \), denoted \( \hat{\pi}_i \). In order to transition from premium received to ultimate claim amount, one requires a loss ratio. The CC method calculates an average loss ratio, denoted \( \hat{\kappa} \), to make this transition.
\[ \hat{\kappa} = \frac{\sum_{i=0}^{I} C_{i,I-i}}{\sum_{i=0}^{I} \hat{\beta}_{I-i} \hat{\pi}_i}, \]
where \( \hat{\beta} \) is the estimate of the cumulative development pattern obtained from using the CL factor estimates. This average loss ratio is applied to each \( \hat{\pi}_i \) to determine an initial estimate of the expected ultimate claim amount. Hence, under the CC method, it is assumed that \( \hat{E}[C_{i,I}] = \hat{\kappa} \hat{\pi}_i \). The predictor of the claims reserves for individual accident years under the CC method is given by
\[ \hat{R}_{i}^{CC} = (1 - \hat{\beta}_{I-i}) \hat{\kappa} \hat{\pi}_i, \quad i \in \{1, \ldots, I\}, \]
with the corresponding predictor of the ultimate claim amount, by
\[ \hat{C}_{i,I}^{CC} = C_{i,I-i} + \hat{R}_{i}^{CC}, \quad i \in \{1, \ldots, I\}. \]
3. Stochastic Bornhuetter-Ferguson Models

Notably absent from the above summary of the BF method is any indication of prediction uncertainty. This is due to the fact that the method is purely mechanical and merely specifies certain deterministic parameters that govern the determination of the claims reserves. This computational algorithm requires no underlying stochastic model. Observe below the minimal stochastic BF model assumptions formulated in Paper I:

Model Assumptions 3.1 (The Bornhuetter-Ferguson Method):

- Cumulative claim amounts $C_{i,j}$ of different accident years $i$ are independent.
- There exist parameters $\mu_0, \ldots, \mu_I > 0$ and a cumulative development pattern $\beta_0, \ldots, \beta_I > 0$ with $\beta_I = 1$ such that for all $i \in \{0, \ldots, I\}$, $j \in \{0, \ldots, I - 1\}$ and $k \in \{1, \ldots, I - j\}$, we have

$$E[C_{i,0}] = \beta_0 \mu_i,$$

$$E[C_{i,j+k}|C_{i,0}, \ldots, C_{i,j}] = C_{i,j} + (\beta_{j+k} - \beta_j) \mu_i.$$ 

The above model assumptions imply the property given below, which is often used to explain the BF method; see e.g. Radtke and Schmidt (2004).

$$E[C_{i,j}] = \beta_j \mu_i.$$ 

Given estimators of $\hat{\beta}_j$ and $\hat{\mu}_i$, which were denoted $\hat{\beta}_j$ and $\hat{\mu}_i$ above, the basic model assumptions are consistent with the BF predictor of the ultimate claim amount. However, the critical question that remains is how one should obtain these parameter estimates.

In this chapter we study three proposed stochastic models that aim to answer the question of parameter estimation as well as prediction uncertainty in the BF method.

3.1 A Bayesian Approach

In this section we study the work of Verrall (2004). He proposes a Bayesian parametric model within the framework of GLM to tackle the claims reserving problem. The
works of Jewell (1989, 1990), Ntzoufras and Dellaportas (2002) and de Alba (2002) also consider a Bayesian framework to model the outstanding loss liabilities. Excellent references of Bayesian methods in actuarial science include Klugman (1992) and Makov et al. (1996). As alluded to earlier, the BF method can be thought of as a simple Bayesian claims reserving method since prior information is given about the ultimate claim amount. However, for it to be considered a true Bayesian method, a prior distribution must be considered, rather than mere deterministic values.

The starting point of the model proposed in Verrall (2004) lies in the following model assumptions:

Model Assumptions 3.2 (Bayesian Model):
The incremental claim amounts $X_{i,j}$ are (conditionally) independent overdispersed Poisson distributed and there exist positive parameters $\mu = (\mu_i)_{i=0,...,I}$, $\gamma = (\gamma_j)_{j=0,...,I}$ and $\phi > 0$ such that
\[
E[X_{i,j} | \mu, \gamma, \phi] = m_{i,j} = \mu_i \gamma_j, \\
Var(X_{i,j} | \mu, \gamma, \phi) = \phi m_{i,j},
\]
with $\sum_{j=0}^I \gamma_j = 1$.

The parameters $(\mu_i)_{i=0,...,I}$ and $(\gamma_j)_{j=0,...,I}$ have a very intuitive explanation. The parameter $\mu_i = E[C_{i,I} | \mu, \gamma, \phi]$ is the expected ultimate claim amount or exposure for accident year $i$. The parameter $\gamma_j$ is interpreted as the proportion of the ultimate claim amount attributed to the $j^{th}$ development year. The reason for this intuitive explanation is due to the choice of the constraint $\sum_{j=0}^I \gamma_j = 1$. Note that a constraint is necessary to provide unique solutions to the parameters, since the parameters $\mu_i$ and $\gamma_j$ can only be determined up to a constant factor, i.e. $\tilde{\mu}_i = c \mu_i$ and $\tilde{\gamma}_j = \gamma_j / c$ would give the same estimate for $m_{i,j}$ for $c > 0$. One notable constraint of this model is that it assumes the incremental claim amounts to be positive, an issue that was alluded to in Chapter 1.

The column parameter $(\gamma_j)_{j=0,...,I}$ is called the development pattern. It is related to the cumulative development pattern $(\beta_j)_{j=0,...,I}$ as follows:
\[
\beta_j = \sum_{k=0}^j \gamma_k, \quad \text{for } j \in \{0, \ldots, I\}.
\]

In order to make this a Bayesian model as opposed to a simple Bayesian model, Verrall (2004) introduces prior distributions that governs the row parameters:
\[
\mu_i | \alpha_i, \theta_i \sim \text{independent } \Gamma(\alpha_i, \theta_i),
\]
where $E[\mu_i | \alpha_i, \theta_i] = \alpha_i \theta_i$ for given prior values $\alpha_i$ and $\theta_i$. For a full Bayesian model, both $(\gamma_j)_{j=0,...,I}$ and $\phi$ should be provided with appropriate prior distributions. The case where this is done for $(\gamma_j)_{j=0,...,I}$ is considered below, however, a plug-in estimate $\hat{\phi}$ is always assumed to be given for the dispersion parameter $\phi$. We presently
study the models under the different assumptions of the column parameters $\gamma_j$. We begin with given column parameters; see e.g. England and Verrall (2002) for a more detailed overview of this scenario.

**Provided Estimates of $\gamma$**

We focus our attention on a single accident year $i$. Using standard Bayesian inference, given the gamma prior for the row parameter $\mu_i$ and treating the column parameters $(\gamma_j)_{j=0,\ldots,I}$ as plug-in estimates, the predictive distribution,

$$f(X_{i,j}|X_{i,0},\ldots,X_{i,j-1},\gamma,\phi),$$

follows an overdispersed negative binomial distribution with mean

$$\left(Z_{i,j} \frac{C_{i,j-i}}{\beta_j-1} + (1-Z_{i,j})\alpha_i\theta_i\right)\gamma_j,$$

where $Z_{i,j} = \frac{\theta_i\beta_{j-1}}{\phi + \theta_i\beta_{j-1}}$.

One immediately recognizes the form of the above mean as being credibility weighted. Suppose the column parameters are estimated using the standard CL factors, that is, $\hat{\beta}_j$ and $\hat{\gamma}_j$ are estimated using $\hat{\beta}_j$ and $\hat{\gamma}_j = \hat{\beta}_j - \hat{\beta}_{j-1}$. Plugging in the CL estimates implies that the term corresponding to weight $Z_{i,j}$ represents the prediction of $X_{i,j}$ under the CL method. Furthermore, the term corresponding to $(1-Z_{i,j})$ represents the prediction of $X_{i,j}$ using prior information, i.e. the CL development pattern and the gamma prior estimate of the ultimate, $\alpha_i\theta_i$, which falls under the spirit of the BF method. Hence the mean of the incremental claim amount is a credibility weighting between the CL method and the BF method; see also Wüthrich (2007).

Not surprisingly, two major factors determine where credibility is given. The first considers the cumulative development pattern: the larger the cumulative development pattern, the larger the credibility given to the data. The second considers the uncertainty of the prior distribution: the smaller the variance, manifested by a small value $\theta_i$, the larger the credibility given to the prior information. Hence a strongly informative prior replicates the BF method, and a noninformative prior, the CL method. The results of these trends are very intuitive and the resulting model provides a more general study of claims reserving that encompasses both the CL and BF methods.

However, the model, although representative of the BF method, has not provided any guidance towards the underlying stochastic assumptions of the column parameters and the estimation of prediction uncertainty. The estimation of the column parameters can be tackled in two ways, simultaneously with the row parameters, or separately before estimating the row parameters. We presently consider the latter case, where the column parameters are estimated first using a noninformative gamma prior distribution, before estimating the row parameters.
Bayesian Negative Binomial Model

To include a provision for the uncertainty of the column parameters $\gamma$, noninformative gamma prior distributions are used and the posterior distribution is obtained under standard Bayesian inference. The reason for using noninformative prior distributions is because no information about the column parameters is required. A typical noninformative gamma prior distribution is one with $\alpha$ small and $\theta$ large, producing a distribution with low mean and high variance.

Furthermore, the prior-posterior analysis is first performed on the column parameters to avoid the influence of estimating the row parameters simultaneously, thereby preserving the CL development pattern. Hence, we first obtain the (intermediary) predictive distribution

$$f(X_{i,j}|D_I, \mu^*, \phi),$$

and subsequently apply the gamma priors used for the row parameters. In the work of Verrall (2004) a reparameterization is utilized to easily incorporate the noninformative gamma priors for the column parameters. This reparameterization relies on the symmetry of the CL method. In other words, if we let $X^*_{i,j}$ take value $X^*_{j,i}$, we are presented with a new triangle, which if developed using the CL method yields the same results as the original triangle. Furthermore, as was seen in the above subsection, using an noninformative prior replicates the CL results. Hence the predictive distribution

$$f(X^*_{i,j}|X^*_{i,0}, \ldots, X^*_{i,j-1}, \mu^*, \phi),$$

follows an overdispersed negative binomial distribution with mean provided by developing the new triangle using the CL method. The $(\mu^*_i)_{i=0}^I$ play the role of the column parameters and reparameterized gamma priors are applied in order to produce the (final) predictive distribution. For the details and the exact reparameterization of both the intermediary predictive distribution as well as the informative priors, see Verrall (2004).

The model is called the Bayesian negative binomial model, the name is attributed to the fact that prior information of the row parameters is applied to an overdispersed negative binomial model to obtain the posterior distribution. This model defines a stochastic version of the BF method. The estimates of the development pattern produced by this method mirror those used in the CL method and the initial estimates of the expected ultimate claim amounts are governed by gamma priors.

A third scenario is considered in Verrall (2004), one that is called the Bayesian overdispersed Poisson model since both priors of the row and column parameters are applied to an overdispersed Poisson model to obtain the posterior distribution. In contrast to the Bayesian negative binomial model, the Bayesian overdispersed Poisson model estimates both row and column parameters simultaneously. The row parameters can either be estimated using noninformative priors or informative priors. We do not discuss the Bayesian overdispersed Poisson model further since it does not replicate the BF claims reserves.
We note that in a Bayesian framework one can, in general, only give numerical answers using simulation techniques. As studied by many authors, the use of the Markov chain Monte Carlo simulation approach is advocated. The advantage being the ease of implementation using the software WinBUGS Spiegelhalter et al. (1996); see e.g. Scollnik (2001) for a useful reference of basic claims reserving methods implemented using WinBUGS. These simulation techniques provide the user with the entire empirical distribution, which is very desirable considering current solvency requirements. The remaining two stochastic approaches aimed at explaining the BF method provide only the second moment of the reserve distribution, however, simulation techniques and software are not required. The calculations required for the remaining approaches can be done in a spreadsheet environment.

3.2 A Distribution-Free Approach

In this section we study the work of Mack (2008), in which he argues that the CL factors should not be used for the BF cumulative development pattern. He provides a distribution-free (DF) stochastic model to represent the BF method. The model assumptions used in Mack (2008) are provided below. Note that we omit any tail factor considerations since we assume fully developed data.

Model Assumptions 3.3 (Distribution-Free Model):

- Incremental claim amounts $X_{i,j}$ are independent.
- There exist parameters $(\mu_i)_{i=0,\ldots,I}, (\gamma_j)_{j=0,\ldots,I}$ and proportionality constants $(s_j^2)_{j=0,\ldots,I}$ with

$$E[X_{i,j}] = \mu_i \gamma_j,$$

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

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

A standard assumption that governs the variance of the incremental claim amounts is the application of a constant dispersion parameter $\phi$. Taylor (2002) provided evidence against such an assumption and as such the use of $s_j^2$ avoids this by allowing variation over development years. However, the $s_j^2$ cannot be considered dispersion parameters in the same sense. Additionally, by not specifying the $\gamma_j$ parameters in the variance of the incremental claim amounts, the $\gamma_j$ are not forced to be positive. Hence the positivity constraint that prevents aggregate negative incremental claim amounts is bypassed.

These model assumptions produce the following form of the outstanding loss liabilities, which is seen to recreate the BF predictor:

$$E[R_i|D_i] = E[R_i] = (1 - \beta_{I-1})\mu_i,$$

where $\beta_j = \sum_{k=0}^{j} \gamma_j$. 

17
The parameter $\mu_i$ is estimated based on prior knowledge, hence an exogenous estimator of the expected ultimate claim amount is required, denote it $\hat{\nu}_i$ for accident year $i$. Furthermore, it is assumed that an estimate of uncertainty can be attached to these estimators. Both the estimator and the uncertainty should be independent from the data and externally obtained.

Recall that $\gamma_j$ plays the role of the development pattern, which in the case of the CL method, is estimated using the CL factors $\hat{f}_j$. Rather than estimating the development pattern solely from the claims development data, Mack (2008) also incorporates the prior information $\hat{\nu}_i$.

\[
\hat{\gamma}_j = \frac{\sum_{i=0}^{I-1} X_{i,j}}{\sum_{i=0}^{I} \hat{\nu}_i}, \quad j \in \{0, \ldots, I-1\}.
\]

The above estimation is of the same form that produces the best linear unbiased estimator of the development pattern in the case of known $\mu_i$. However, the development pattern estimated as above is not guaranteed to sum to one. Mack (2008) advocates “selecting” the pattern based on the initial estimate $\hat{\gamma}_j$. He advocates doing so by applying some smoothing techniques and paying close attention to later development years $j$ where the initial estimation is highly unstable due to lack of data. The selected development pattern is denoted $\hat{\gamma}_j^*$. For more details regarding the estimation of the development pattern, see also Mack (2006).

As with the development pattern, the estimation of the variance parameters $s^2_j$ is done following the form of an unbiased estimator in the case of known $\mu_i$. They are given below where $\hat{\nu}_i$ and $\hat{\gamma}_j^*$ supplement the fact that the $\mu_i$ are unknown.

\[
\hat{s}_j^2 = \frac{1}{I + 1 - j} \sum_{i=0}^{I-1} \left( \frac{X_{i,j} - \hat{\nu}_i \hat{\gamma}_j^*}{\hat{\nu}_i} \right)^2, \quad j \in \{0, \ldots, I-1\}.
\]

Smoothing techniques are applied in order to select $\hat{s}_j^{2*}$ for $j \in \{0, \ldots, I-1\}$ and extrapolation is performed to obtain $\hat{s}_I^{2*}$.

We focus our attention on calculating the uncertainty of the claims reserves for a single accident year $i$ produced under this model, note that the uncertainty can be quantified for aggregate accident years as well; see Mack (2008). The predictor of the outstanding loss liabilities under the DF model is given by

\[
\hat{R}_i^{DF} = (1 - \hat{\beta}_I^{1-i})\hat{\nu}_i, \quad \text{where} \quad \hat{\beta}_j^* = \sum_{k=0}^{j} \hat{\gamma}_j^*.
\]

for accident year $i \in \{1, \ldots, I\}$. Its conditional MSEP is given by

\[
\text{msep}_{R_i | D_i}(\hat{R}_i^{DF}) = E\left[ (\hat{R}_i^{DF} - R_i)^2 \mid D_i \right].
\]

Mack (2008) decouples the conditional MSEP into conditional process variance and conditional estimation error. Estimating the conditional process variance is a
trivial replacement of parameters with their estimates. Estimating the conditional estimation error is more involved and we leave it to the reader to discover the details in Mack (2008), in which guidance regarding the estimation of the uncertainty of \( \hat{\nu}_i \) and \( \hat{\beta}_{I-i}^* \) is provided by equations (5) and (7), respectively. An estimate of the conditional MSEP under the DF model for single accident year \( i \) is provided by

\[
\hat{\text{msep}}_{R_i \mid D_i}(\hat{R}_i^{DF}) = \hat{\nu}_i \left( \sum_{j > I - i} \hat{s}_{j}^2 \right) + (1 - \hat{\beta}_{I-i}^*)^2 \text{Var}(\hat{\nu}_i) + \left( \hat{\nu}_i^2 + \text{Var}(\hat{\nu}_i) \right) \text{Var}(\hat{\beta}_{I-i}^*)
\]

where the first term on the right-hand side constitutes the conditional process variance and the latter two terms, the conditional estimation error.

### 3.3 A Likelihood Approach (Papers I and II)

The likelihood approach is defined and prediction uncertainty under this approach is developed in accompanying papers Paper I and Paper II. The two papers specify the same model assumptions but derive the conditional MSEP, used to quantify the prediction uncertainty, in different ways, albeit coming to the same conclusion. The techniques described in the latter of the two papers is recommended due to the fact that the results are very easy to implement in a spreadsheet environment.

Like Verrall (2004), we assume the incremental claim amounts to be independently overdispersed Poisson distributed. Furthermore, we assume independent exogenous estimators are given for the expected ultimate claim amounts. As in Mack (2008), we do not assume distributional results for these estimators, rather an estimate of their uncertainty. For simplicity and as it is done in practice, see Swiss Solvency Test (2006), we apply a coefficient of variation to obtain such uncertainty estimates. However, the guidance provided by using loss ratios, as seen in Mack (2008), to determine the exogenous estimators and their uncertainty can easily be applied in the likelihood approach. Finally, rather than a Bayesian approach, we incorporate the uncertainty of the development pattern using asymptotic properties of MLE.

It is well known that the MLEs under the independent overdispersed Poisson assumption on the incremental claim amounts recreates the CL reserves. As in Verrall (2004), we maintain that in practice the CL development pattern is used for calculating the BF reserves, and hence incorporate this into the model assumptions. The drawback of the independent overdispersed Poisson assumption is the positivity constraint on the incremental claim amounts. This drawback is also present in the Bayesian approach, but avoided in the distribution-free approach.

**Model Assumptions 3.4 (Overdispersed Poisson (ODP) Model):**

- The increments \( X_{i,j} \) are independent overdispersed Poisson distributed and there exist positive parameters \( \gamma_0, \ldots, \gamma_I, \mu_0, \ldots, \mu_I \) and \( \phi > 0 \) such that
  \[
  E[X_{i,j}] = m_{i,j} = \mu_i \gamma_j,
  \]
  \[
  \text{Var}(X_{i,j}) = \phi m_{i,j},
  \]

19
with \( \sum_{j=0}^{I} \gamma_j = 1 \).

- \( \hat{\nu}_i \) are independent random variables that are unbiased estimators of \( \mu_i = E[C_{i,I}] \) for all \( i \).
- \( X_{i,j} \) and \( \hat{\nu}_k \) are independent for all \( i, j, k \).

From these model assumptions, we obtain

\[
E[C_{i,I} | C_{i,0}, \ldots, C_{i,I-1}] = C_{i,I-1} + \sum_{j > I-i} E[X_{i,j}] = C_{i,I-1} + (1 - \beta_{I-i}) \mu_i,
\]

where \( \beta_j = \sum_{k=0}^{j} \gamma_k \) is the cumulative development pattern. This means that the ODP model satisfies the basic BF assumptions and hence can also be used to explain the BF method.

As noted above, we estimate the \( \gamma_j \) parameters using MLE. However, since it is known that the overdispersed Poisson assumption recreates the CL reserves, we directly use the CL factors \( \hat{f}_j \); see e.g. Wüthrich and Merz (2008), Section 2.4.

\[
\hat{\gamma}_j = \prod_{k=j+1}^{I} \frac{1}{\hat{f}_k (1 - \frac{1}{\hat{f}_{k-1}})}.
\]

Equivalently, the cumulative development pattern \( \beta_j \) is estimated using the previously defined CL estimate \( \hat{\beta}_j \). Furthermore, since we require the MLE of \( \mu_i \) to estimate the dispersion parameter \( \phi \), we provide it by again making use of the CL factors,

\[
\hat{\mu}_i = C_{i,I-1} \hat{f}_{I-1} \cdots \hat{f}_{I-1}.
\]

The dispersion parameter \( \phi \) is assumed constant over all development years, this assumption is required in order to recreate the CL development pattern under the overdispersed Poisson. This is due to the fact that a constant \( \phi \) has no influence on the parameter estimation of \( \mu_i, \gamma_j \). Within the framework of GLM we use different types of residuals (Pearson, deviance, Ansccombe, etc.) to estimate \( \phi \); see e.g. McCullagh and Nelder (1989) or Fahrmeir and Tutz (2001). We could also estimate \( \phi \) using maximum likelihood, but rather, due to ease of implementation, we use Pearson residuals, given by

\[
\hat{\phi} = \frac{1}{d} \sum_{i+j \leq l} \frac{(X_{i,j} - \hat{m}_{i,j})^2}{\hat{m}_{i,j}},
\]

where \( d = \frac{(I+1)(I+2)}{2} - 2I - 1 \) is the degrees of freedom of the model and \( \hat{m}_{i,j} = \hat{\mu}_i \hat{\gamma}_j \).

Note that the predictor of the ultimate claim amount under the ODP model exactly resembles that of the BF method described in Chapter 2, hence we reuse the notation given therein. We consider the conditional MSEP of the ODP predictor
\( \hat{C}_{i,t}^{BF} \) for single accident year \( i \). We refer to Paper II for details regarding the aggregated conditional MSEP derivation. From equation (5.5) in Paper I, we have

\[
\text{msep}_{C_{i,t}|D_{i}}(\hat{C}_{i,t}^{BF}) = E \left[ \left( \hat{C}_{i,t}^{BF} - C_{i,t} \right)^2 | D_{i} \right]
= \sum_{j > I - i} \text{Var}(X_{i,j}) + \left( \sum_{j > I - i} \hat{\gamma}_j \right)^2 \text{Var}(\hat{\nu}_i) + \mu_i^2 \left( \sum_{j > I - i} \hat{\gamma}_j - \sum_{j > I - i} \gamma_j \right)^2.
\]

The first term on the right-hand side above is the conditional process variance, it represents the stochastic movement of the \( X_{i,j} \), the inherent uncertainty from our model assumptions. The latter two terms form the conditional estimation error; these terms constitute the uncertainty in the prediction of the prior estimate \( \hat{\nu}_i \) and the MLEs \( \hat{\gamma}_j \).

For the estimation of the conditional process variance, Model Assumptions 3.4 motivate the following estimator:

\[
\hat{\text{Var}}(X_{i,j}) = \hat{\phi} \hat{\nu}_i \sum_{j > I - i} \hat{\gamma}_j.
\]

To quantify the uncertainty in the prior estimate \( \hat{\nu}_i \) of the expected ultimate claim \( E[C_{i,t}] \), we use the following:

\[
\left( \sum_{j > I - i} \hat{\gamma}_j \right)^2 \hat{\text{Var}}(\hat{\nu}_i).
\]

Since \( \hat{\nu}_i \) is determined exogenously this can generally only be done using external data like market statistics and expert opinion. The regulator, for example, provides an estimate for the coefficient of variation of \( \hat{\nu}_i \), denoted by \( \hat{\text{Vco}}(\hat{\nu}_i) \), that quantifies how good the exogenous estimator \( \hat{\nu}_i \) is. Statistical estimates based on impact studies for the determination of estimates \( \hat{\text{Vco}}(\hat{\nu}_i) \) exist, for example, in the context of the Swiss Solvency Test (2006). These studies suggest that 5% to 10% is a reasonable range for \( \hat{\text{Vco}}(\hat{\nu}_i) \). Hence the term quantifying the uncertainty of the prior estimate \( \hat{\nu}_i \) can be rewritten as

\[
\left( \sum_{j > I - i} \hat{\gamma}_j \right)^2 \hat{\text{Vco}}(\hat{\nu}_i)^2 = \left( \sum_{j > I - i} \hat{\gamma}_j \right)^2 \hat{\nu}_i^2 \hat{\text{Vco}}(\hat{\nu}_i)^2.
\]

Note that an appropriate choice for \( \hat{\text{Vco}}(\hat{\nu}_i) \) is crucial for a meaningful analysis. This choice is closely related to a Bayesian setup where one chooses an appropriate prior distribution for \( \hat{\nu}_i \); see e.g. Mack (2000).

Quantifying the uncertainty of the MLEs \( \hat{\gamma}_j \) requires careful consideration. The standard approach, see England and Verrall (2002), is to estimate

\[
\left( \sum_{j > I - i} \hat{\gamma}_j - \sum_{j > I - i} \gamma_j \right)^2
\]
by the unconditional expectation
\[
E \left[ \left( \sum_{j > i} \hat{\gamma}_j - \sum_{j > i} \gamma_j \right)^2 \right] = \sum_{j > i} E \left[ \left( \hat{\gamma}_j - \gamma_j \right) \left( \hat{\gamma}_i - \gamma_i \right) \right].
\]
Neglecting that MLEs have a possible bias term we make the following approximation:
\[
\sum_{j > i, l > i} E \left[ \left( \hat{\gamma}_j - \gamma_j \right) \left( \hat{\gamma}_l - \gamma_l \right) \right] \approx \sum_{j > i, l > i} \text{Cov}(\hat{\gamma}_j, \hat{\gamma}_l).
\]
It is with regards to the estimation of \(\text{Cov}(\hat{\gamma}_j, \hat{\gamma}_l)\) that Paper I and Paper II diverge. The former making use of a generalized linear representation of the ODP model together with Taylor approximations as described in Section 5.1.2 of the paper. The latter using the asymptotic properties of MLE as described in Section 2.4 of the corresponding paper. In the case where the asymptotic properties of MLE are utilized, one needs to find the inverse of the Fisher information matrix to obtain an approximation of the dependence structure amongst the MLEs. We denote this resulting estimate \(\hat{\text{Cov}}(\hat{\gamma}_j, \hat{\gamma}_l)\).

Hence, an estimate of the conditional MSEP under the ODP model for single accident year \(i\) is provided by (see Estimator 4.1 in Paper II)
\[
\hat{\text{MSEP}}_{C_i|\mathcal{D}_i}(\hat{C}_{BF,i}) = \sum_{j > i} \hat{\phi} \hat{v}_i \hat{\gamma}_j + \left( \sum_{j > i} \hat{\gamma}_j \right)^2 \hat{\text{Var}}(\hat{v}_i) + \hat{v}_i^2 \sum_{j > i} \hat{\text{Cov}}(\hat{\gamma}_j, \hat{\gamma}_l),
\]
where the first term on the right-hand side constitutes the conditional process variance and the latter two terms, the conditional estimation error.

### 3.4 Numerical Comparison

We consider a case study to highlight the performance of the above stochastic approaches to the BF method. We compare the results not only to those obtained using the stochastic CL method introduced in Mack (1993) but also to results obtained using GLM. We analyse the dataset provided in Table 3.1, given in '000s. Furthermore, we assume the uncertainty of the prior estimate of the expected ultimate claim amounts \(\hat{v}_i\) to be given by a coefficient of variation of 5\%.

\[
\hat{\text{Var}}(\hat{v}_i) = \hat{v}_i^2 (0.05)^2, \quad i \in \{1, \ldots, I\}.
\]

The comparison of the methods is given in Table 3.2. Applying the BF method as done in practice, i.e. with the CL factor estimates, results in a reserve of 7,356,575. This reserve is matched under the likelihood approach described in accompanying papers Paper I and Paper II as well as very closely approximated using the BF Bayesian negative binomial approach. In the process of calculating the results for the distribution-free approach to the BF method, the development pattern was normalized to sum to one but no other selection was made.
3.4. Numerical Comparison

<table>
<thead>
<tr>
<th>i/j</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>( \hat{\nu}_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5,947</td>
<td>3,721</td>
<td>896</td>
<td>208</td>
<td>207</td>
<td>62</td>
<td>66</td>
<td>15</td>
<td>11</td>
<td>16</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>6,347</td>
<td>3,246</td>
<td>723</td>
<td>152</td>
<td>68</td>
<td>37</td>
<td>53</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11,365</td>
</tr>
<tr>
<td>2</td>
<td>6,269</td>
<td>2,976</td>
<td>847</td>
<td>263</td>
<td>153</td>
<td>65</td>
<td>54</td>
<td>9</td>
<td>-</td>
<td>-</td>
<td>10,963</td>
</tr>
<tr>
<td>3</td>
<td>5,863</td>
<td>2,683</td>
<td>723</td>
<td>191</td>
<td>133</td>
<td>88</td>
<td>43</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10,617</td>
</tr>
<tr>
<td>4</td>
<td>5,779</td>
<td>2,745</td>
<td>654</td>
<td>274</td>
<td>105</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11,045</td>
</tr>
<tr>
<td>5</td>
<td>6,185</td>
<td>2,828</td>
<td>573</td>
<td>245</td>
<td>105</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11,481</td>
</tr>
<tr>
<td>6</td>
<td>5,600</td>
<td>2,893</td>
<td>563</td>
<td>226</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11,414</td>
</tr>
<tr>
<td>7</td>
<td>5,288</td>
<td>2,440</td>
<td>528</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11,127</td>
</tr>
<tr>
<td>8</td>
<td>5,291</td>
<td>2,358</td>
<td>528</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10,987</td>
</tr>
<tr>
<td>9</td>
<td>5,676</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11,618</td>
</tr>
</tbody>
</table>

Table 3.1: Observed incremental payments \( X_{i,j} \) and prior estimates \( \hat{\nu}_i \).

The calculations for the Bayesian negative binomial approach were performed using WinBugs. For the results, we ran 20,000 iterations, discarding the first 10,000. In this approach we highlight two forms of uncertainty for the exogenous estimators of the expected ultimate claim amounts. The strong prior results assume the estimators have virtually no uncertainty, an assumption that allows the Bayesian negative binomial approach to closely approximate the BF reserves.

The weak prior results assumes high uncertainty of the estimators of the expected ultimate claim amounts. As stated in Verrall (2004), the Bayesian negative binomial approach with weak prior information leads the resulting reserves to closely mirror that of the CL method. A result verified in Table 3.2.

<table>
<thead>
<tr>
<th>method</th>
<th>reserves</th>
<th>process error</th>
<th>estimation error</th>
<th>prediction error (mse(1/2))</th>
<th>Vco</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF: likelihood</td>
<td>7,356,575</td>
<td>329,007</td>
<td>338,396</td>
<td>471,971</td>
<td>6.4%</td>
</tr>
<tr>
<td>BF: distribution-free</td>
<td>7,505,506</td>
<td>621,899</td>
<td>375,424</td>
<td>726,431</td>
<td>9.7%</td>
</tr>
<tr>
<td>BF: Bayesian NB (strong prior)</td>
<td>7,356,560</td>
<td>-</td>
<td>-</td>
<td>427,278</td>
<td>5.8%</td>
</tr>
<tr>
<td>CL: Bayesian NB (weak prior)</td>
<td>6,006,464</td>
<td>-</td>
<td>-</td>
<td>419,182</td>
<td>7.0%</td>
</tr>
<tr>
<td>CL: distribution-free</td>
<td>6,047,059</td>
<td>424,379</td>
<td>185,026</td>
<td>462,960</td>
<td>7.7%</td>
</tr>
<tr>
<td>GLM: overdispersed Poisson</td>
<td>6,047,059</td>
<td>298,290</td>
<td>309,563</td>
<td>429,891</td>
<td>7.1%</td>
</tr>
<tr>
<td>GLM: Gamma</td>
<td>5,947,049</td>
<td>624,811</td>
<td>926,371</td>
<td>1,117,386</td>
<td>18.8%</td>
</tr>
</tbody>
</table>

Table 3.2: Aggregate reserve and uncertainty results for the a variety of claims reserving methods.

The overdispersed Poisson method and the gamma method belong to the class of GLM, we discuss their relevance in claims reserving in the following chapter.
4. Generalized Linear Models

The use of GLM in actuarial science is well developed and broadly accepted. Not only does the framework of GLM allow for flexibility in parameter and model selection, in some cases, such as with the CL method, GLM recovers traditional methods for claims reserves estimation. For a comprehensive reference of GLM, see McCullagh and Nelder (1989).

We begin by studying the exponential dispersion family (EDF) and its role in modelling claims reserves; see Jørgensen (1987, 1997) for more on the EDF and Renshaw (1994), Haberman and Renshaw (1996), England and Verrall (2002) and Wüthrich and Merz (2008) for applications to insurance. The overdispersed Poisson model, a member of the EDF, is given special attention due to its use not only in Chapter 3, but also in the work of Renshaw (1994), Renshaw and Verrall (1998), and Verrall (2000, 2004).

Furthermore, we study a special sub-family of the EDF, the so-called Tweedie EDF. Besides containing many standard models, such as the Gaussian, overdispersed Poisson and gamma, of particular interest are the compound Poisson models; Mildenhall (1999) provides an excellent review of these models. For specific applications of the Tweedie compound Poisson model, see e.g. Jørgensen and De Souza (1994), Smyth and Jørgensen (2002), and Wüthrich (2003).

Based on Paper III we provide insight regarding the sensitivity of the claims reserves and their associated uncertainty over the Tweedie EDF. Furthermore, we develop second order Taylor approximations for the claims reserves and the conditional MSEP, used to quantify the uncertainty, for members of the Tweedie family that are difficult to obtain in practice, but are close enough to models for which claims reserves and conditional MSEP estimations are easy to determine. We provide a case study based on one of many considered datasets to highlight the performance of our approximations. This permits conclusions to be drawn with regards to the model uncertainty, specific to the type of dataset, when the underlying Tweedie EDF is assumed.

4.1 The Exponential Dispersion Family

Nelder and Wedderburn (1972) established the framework of GLM and the so-called analysis of deviance. These concepts were originally developed for exponential families of distributions, yet extended to a wider class of distributions, termed dispersion models.
Generalized Linear Models

A random variable $X_{i,j}$ follows an exponential dispersion model if it has generalized density

$$f(x; \theta_{i,j}, \phi_{i,j}, w_{i,j}) = \exp \left\{ \frac{w_{i,j}}{\phi_{i,j}} (x \theta_{i,j} - b(\theta_{i,j})) \right\} c(x; \phi_{i,j}/w_{i,j}),$$

where $w_{i,j} > 0$ denotes a known weight, $\theta_{i,j}$ is the canonical parameter, $\phi_{i,j} > 0$ the dispersion parameter. The function $b$ is a twice differentiable general function that determines the more specific family the model falls into and $c$ is a suitable normalizing constant. This generalized density can be defined with respect to Lebesgue measure or the counting measure. Moreover, the domain of $x$ depends on the choice of $b$.

As noted above, the function $b$ determines to which specific family the exponential dispersion model belongs. Likewise, one can specify the structure of the underlying unit variance function, $V$, defined as,

$$V(m) = b''((b')^{-1}(m));$$

see e.g. Jørgensen (1997), Theorem 2.11. Under these assumptions, $X_{i,j}$ has expectation and variance given by

$$E[X_{i,j}] = m_{i,j} = b'(\theta_{i,j}),$$
$$\text{Var}(X_{i,j}) = \frac{\phi_{i,j}}{w_{i,j}} b''(\theta_{i,j}) = \frac{\phi_{i,j}}{w_{i,j}} V(m_{i,j});$$

see e.g. Bühlmann and Gisler (2005), Theorem 2.2.

It is evident that the overdispersed Poisson model is obtained by defining the unit variance function as the identity function, $V(m) = m$ with $\phi_{i,j}/w_{i,j}$ playing the role of the dispersion parameter. The corresponding function $b$ is given by the exponential function, $b(\theta_{i,j}) = \exp(\theta_{i,j})$. Since we aim to describe the use of the overdispersed Poisson in claims reserving, we specify the necessary multiplicative structure of the canonical parameter $\theta_{i,j}$ by introducing accident year, or row, parameters $\mu_i$ and development year, or column, parameters $\gamma_j$. In order to attain the relationship $m_{i,j} = \mu_i \gamma_j$, we specify $\theta_{i,j}$ as

$$\theta_{i,j} = \log(\mu_i \gamma_j).$$

We estimate these model parameters using MLE. The log-likelihood function under the assumption that we independently model the incremental claim amounts $X_{i,j}$ using the overdispersed Poisson model given information $\mathcal{D}_I$ is provided by

$$l_{\mathcal{D}_I}(\mu_i, \gamma_j, \frac{\phi_{i,j}}{w_{i,j}}) = \sum_{i+j \leq I} \left( \frac{w_{i,j}}{\phi_{i,j}} (X_{i,j} \log(\mu_i \gamma_j) - \mu_i \gamma_j) + \log c(x; \phi_{i,j}/w_{i,j}) \right).$$
4.1. THE EXPONENTIAL DISPERSION FAMILY

The resulting MLEs for $\mu_i$ and $\gamma_j$ are given by the system of equations

\[
\begin{align*}
\hat{\mu}_i &= \frac{1}{\phi_i} \sum_{j=0}^{I-i} \frac{w_{i,j}}{\phi_i} \hat{X}_{i,j}, & i \in \{0, \ldots, I\}, \\
\hat{\gamma}_j &= \frac{1}{\phi_i} \sum_{i=0}^{I-j} \frac{w_{i,j}}{\phi_i} \hat{X}_{i,j}, & j \in \{0, \ldots, I\}, \\
\hat{\mu}_0 &= 1.
\end{align*}
\]

As noted before, a constant dispersion factor is required to recreate the CL reserves. The simplification that is obtained under the assumption that $\hat{\phi}_{i,j} \frac{w_{i,j}}{\phi_i} = \hat{\phi} > 0$ is nowhere more evident than in the above system of equations. The parameter $\phi$ cancels out of the equations, thereby playing no role on the estimation of the $\mu_i$ and $\gamma_j$, and the resulting system is consistent with the basic CL method. The dispersion parameter, however, does require estimation. As an estimate of $\phi$, we use $\hat{\phi}$ described in Chapter 3 using Pearson residuals. The predictor resulting from the application of maximum likelihood to estimate the model parameters $\mu_i$ and $\gamma_j$ is defined by the method described below.

The Maximum Likelihood (ML) Method:

Under the assumption that we independently model the incremental claim amounts $X_{i,j}$ as overdispersed Poisson($\mu_i \gamma_j$) and estimate the parameters using MLE, the predictor for the ultimate claim amount for accident year $i$ is given by,

\[
\hat{C}_{i,I}^{ML} = C_{i,I} - \sum_{j=I-i}^{I} \hat{\mu}_i \hat{\gamma}_j = \hat{C}_{i,I}^{CL},
\]

with associated predictor of the outstanding loss liabilities given by

\[
\hat{R}_i^{ML} = \sum_{j=I-i}^{I} \hat{\mu}_i \hat{\gamma}_j = \hat{R}_i^{CL}.
\]

The uncertainty of the ML predictor is quantified using the conditional MSEP. Notice that the predictor $\hat{R}_i^{ML}$ is $\mathcal{D}_I$-measurable and hence the conditional MSEP decouples into conditional process variance and conditional estimation error:

\[
msep_{R_i | D_I}(\hat{R}_i^{ML}) = \sum_{j=I-i}^{I} \text{Var}(X_{i,j}) + \left( \sum_{j=I-i}^{I} \hat{\mu}_i \hat{\gamma}_j - \sum_{j=I-i}^{I} \mu_i \gamma_j \right)^2.
\]

The assumption of independent overdispersed Poisson random variables motivates the following estimation of the conditional process variance:

\[
\sum_{j=I-i}^{I} \hat{\text{Var}}(X_{i,j}) = \hat{\phi} \sum_{j=I-i}^{I} \hat{\mu}_i \hat{\gamma}_j.
\]

Furthermore,

\[
\left( \sum_{j=I-i}^{I} \hat{\mu}_i \hat{\gamma}_j - \sum_{j=I-i}^{I} \mu_i \gamma_j \right)^2
\]
is estimated by

$$E \left[ \left( \sum_{j > I - i} \hat{\mu}_i \hat{\gamma}_j - \sum_{j > I - i} \mu_i \gamma_j \right)^2 \right] = \sum_{j > I - i, l > I - k} E \left[ \left( \hat{\mu}_i \hat{\gamma}_j - \mu_i \gamma_j \right) \left( \hat{\mu}_k \hat{\gamma}_l - \mu_k \gamma_l \right) \right].$$

Note that the predictor $\hat{\mu}_i \hat{\gamma}_j$ is not necessarily unbiased for $E[X_{i,j}] = \mu_i \gamma_j$. This bias is for typical claims reserving data of negligible order. One uses the approximation

$$\left( \sum_{j > I - i} \hat{\mu}_i \hat{\gamma}_j - \sum_{j > I - i} \mu_i \gamma_j \right)^2 \approx \sum_{j > I - i, l > I - k} \text{Cov}(\hat{\mu}_i \hat{\gamma}_j, \hat{\mu}_k \hat{\gamma}_l);$$

see e.g. England and Verrall (2002) and Wüthrich and Merz (2008), Section 6.4.3. Hence, an estimate of the conditional MSEP of $\hat{R}_{i,I}^{ML}$ is given by

$$\hat{\text{msep}}_{R_{i|D_i}}(\hat{R}_{i,I}^{ML}) = \sum_{j > I - i} \hat{\phi}_i \hat{\gamma}_j + \sum_{j > I - i, l > I - k} \hat{\text{Cov}}(\hat{\mu}_i \hat{\gamma}_j, \hat{\mu}_k \hat{\gamma}_l);$$

see Appendix B of Paper III for details regarding the estimation of the covariance term $\text{Cov}(\hat{\mu}_i \hat{\gamma}_j, \hat{\mu}_k \hat{\gamma}_l)$. We generalize the overdispersed Poisson model in the following section.

### 4.2 The Tweedie Exponential Dispersion Family

As mentioned above, one can specify a specific model, or family of models, in the EDF by specifying the unit variance function. The overdispersed Poisson model is given by the identity function, the Tweedie EDF by the power function.

A random variable $X_{i,j}$ is a member of the Tweedie EDF if it belongs to the EDF, where the canonical parameter $\theta_{i,j} \in \Theta_p$ and function $b$ are defined as

$$b_p(\theta_{i,j}) = \frac{1}{2 - p}((1 - p)\theta_{i,j})^{\frac{2 - p}{2 - p}}, \quad p \notin (0, 1] \cup [2],$$

$$b_1(\theta_{i,j}) = \exp(\theta_{i,j}), \quad p = 1,$$

$$b_2(\theta_{i,j}) = -\log(-\theta_{i,j}), \quad p = 2,$$

and

$$\Theta_p = \begin{cases} \mathbb{R}, & \text{for } p = 0, 1, \\ [0, \infty), & \text{for } p < 0, \\ (-\infty, 0], & \text{for } 1 < p \leq 2, \\ (-\infty, 0), & \text{for } 2 < p < \infty. \end{cases}$$

The specification on $b$ is made so that the unit variance function, $V$, has a power structure with power $p \in (-\infty, 0] \cup [1, \infty)$; see Jørgensen (1997), Proposition 4.2, regarding the possible values of $p$. That is, $b_p$ implies that

$$V(m) = m^p.$$
4.3 Model Uncertainty (Paper III)

Model Assumptions 4.1 (Tweedie EDF):
The increments $X_{i,j}$ are independent Tweedie distributed and there exist positive parameters $\gamma_0, \ldots, \gamma_I$, $\mu_0, \ldots, \mu_I$, $\phi > 0$, and $p \in (-\infty, 0] \cup [1, \infty)$ such that

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

The resulting MLEs for $\mu_i$ and $\gamma_j$ are given by the system of equations

$$
\hat{\mu}_i \sum_{j=0}^{I-i} \hat{\gamma}_j^{2-p} = \sum_{j=0}^{I-i} X_{i,j} \hat{\gamma}_j^{1-p}, \quad i \in \{1, \ldots, I\},
$$

$$
\hat{\gamma}_j \sum_{i=0}^{I-j} \hat{\mu}_i^{2-p} = \sum_{i=0}^{I-j} X_{i,j} \hat{\mu}_i^{1-p}, \quad j \in \{0, \ldots, I\},
$$

$$
\hat{\mu}_0 = 1;
$$

see e.g. Wüthrich and Merz (2008), formulas (5.49) and (5.50). From the system of equations given above it is clear that $\hat{\mu}_i$ and $\hat{\gamma}_j$ are functions of $p$. Furthermore, due to this, the predictor of the outstanding loss liabilities is also a function of $p$.

$$
\hat{R}_ML^i(p) = \hat{R}_i(p) = \sum_{j>I-i} \hat{\mu}_i(p) \hat{\gamma}_j(p).
$$

The reader is left to consult Paper III for the uncertainty of the claims reserves under the Tweedie EDF assumption. The formulations closely mirror that in the overdispersed Poisson model, given that it is a member of the Tweedie EDF with $p = 1$, but is slightly more complicated. We emphasize that both the claims reserves and the estimate of its uncertainty, once the Tweedie assumption is made, depend solely on the model parameter $p$. In the following section we study the effect of varying this parameter.

4.3 Model Uncertainty (Paper III)

The Tweedie EDF is governed by model parameter $p$. For example, the model parameters $p = 1$ corresponds to the overdispersed Poisson model and $p = 2$ to the gamma model. In this section, we study the sensitivity of the claims reserves with respect to this model parameter. Studies on model uncertainty are becoming increasingly relevant to industry as solvency requirements continue to emphasize their importance. Peters et al. (2009) and Gigante and Sigalotti (2005) have also tackled this issue of model uncertainty, the former opting for a Bayesian Markov chain Monte Carlo simulation approach averaging over $p$, and the latter addressing the issue within a GLM framework, using an iterative procedure to solve for $p$ using quasi-likelihood functions as introduced by Wedderburn (1974) and Nelder and Pregibon (1987).
**Generalized Linear Models**

We write both the claims reserves and the estimates of the model parameters as functions of $p$ and analyse the sensitivity of $\hat{R}_i(p)$ with respect to $p$ by a Taylor expansion. The second order Taylor expansion around $p$ is given by

$$\hat{R}_i'(p + \varepsilon) = \hat{R}_i(p) + \hat{R}'(p) \varepsilon + \frac{\hat{R}''(p) \varepsilon^2}{2},$$

with,

$$\hat{R}_i'(p) = \sum_{j > I - i} \left( \hat{\mu}_i'(p) \hat{\gamma}_j(p) + \hat{\mu}_i(p) \hat{\gamma}'_j(p) \right),$$

$$\hat{R}_i''(p) = \sum_{j > I - i} \left( \hat{\mu}_i''(p) \hat{\gamma}_j(p) + 2\hat{\mu}_i'(p) \hat{\gamma}'_j(p) + \hat{\mu}_i(p) \hat{\gamma}''_j(p) \right).$$

To obtain the first derivatives of the MLEs with respect to $p$, we differentiate the system of equations given in the previous section. We use simplifying notation $\hat{\mu}_i' = \hat{\mu}_i'(p)$ and $\hat{\gamma}'_i = \hat{\gamma}'_i(p)$. We obtain the following:

$$\hat{\mu}_i' \sum_{j=0}^{I-i} \hat{\gamma}_{ij}^2 - p \sum_{j=0}^{I-i} \frac{\hat{\gamma}_{ij}}{\hat{\gamma}_{ij}^p} \left[ (2 - p) \hat{\mu}_i \hat{\gamma}_j - (1 - p) X_{i,j} \right]$$

$$= \sum_{j=0}^{I-i} \frac{\log(\hat{\gamma}_{ij}) \hat{\gamma}_j}{\hat{\gamma}_{ij}^p} \left[ \hat{\mu}_i \hat{\gamma}_j - X_{i,j} \right], \quad i \in \{1, \ldots, I\},$$

$$\hat{\gamma}_j' \sum_{i=0}^{I-j} \hat{\mu}_i^2 - p \sum_{i=0}^{I-j} \frac{\hat{\mu}_i}{\hat{\mu}_i^p} \left[ (2 - p) \hat{\gamma}_j \hat{\mu}_i - (1 - p) X_{i,j} \right]$$

$$= \sum_{i=0}^{I-j} \frac{\log(\hat{\mu}_i) \hat{\mu}_i}{\hat{\mu}_i^p} \left[ \hat{\gamma}_j \hat{\mu}_i - X_{i,j} \right], \quad j \in \{0, \ldots, I\},$$

$$\hat{\mu}_0' = 0.$$

To solve the above system of equations we define a $(2I + 2) \times (2I + 2)$ matrix $A$, whose components are the following:

$$a_{i,i} = \sum_{j=0}^{I-i} \hat{\gamma}_{ij}^2 - p, \quad i \in \{1, \ldots, I\},$$

$$a_{i,I+j+1} = \frac{1}{\hat{\gamma}_{ij}} \left( (2 - p) \hat{\mu}_i \hat{\gamma}_j - (1 - p) X_{i,j} \right), \quad i \in \{0, \ldots, I\}, j \in \{0, \ldots, I - i\},$$

$$a_{I+j+1,i} = \frac{1}{\hat{\mu}_i^p} \left( (2 - p) \hat{\mu}_i \hat{\gamma}_j - (1 - p) X_{i,j} \right), \quad j \in \{0, \ldots, I\}, i \in \{0, \ldots, I - j\},$$

$$a_{I+j+1,I+j+1} = \sum_{i=0}^{I-j} \hat{\mu}_i^2 - p, \quad j \in \{0, \ldots, I\},$$

$$a_{0,0} = 1;$$

where the remaining entries of the matrix are defined to be zero. In addition to the matrix $A$, we define column vectors $\xi' = (\hat{\mu}_0', \hat{\mu}_1', \hat{\gamma}_0', \ldots, \hat{\gamma}_I')^T$ and $\alpha = \ldots$
(0, α₁, . . . , αᵢ, β₀, . . . , βⱼ)^T, where
\[
\alpha_i = \sum_{j=0}^{I-i} \frac{\log(\hat{\gamma}_j)}{\hat{\gamma}_j} (\hat{\mu}_i \hat{\gamma}_j - X_{i,j}), \quad i \in \{1, \ldots, I\},
\]
\[
\beta_j = \sum_{i=0}^{I-j} \frac{\log(\hat{\mu}_i)}{\hat{\mu}_i} (\hat{\mu}_i \hat{\gamma}_j - X_{i,j}), \quad j \in \{0, \ldots, I\}.
\]

Using matrix notation, we rewrite the system of equations that produces the first derivatives of the MLEs with respect to \( p \) as
\[
A \hat{\zeta}' = \alpha.
\]

Lemma 4.2 (see Lemma 4.1 in Paper III)
The first derivative of the MLE \( \hat{\zeta} \) is given by
\[
\hat{\zeta}' = A^{-1} \alpha,
\]
where \( A \) and \( \alpha \) are defined as given above.

We obtain the second derivatives of the MLEs with respect to \( p \) in similar fashion, producing the following Lemma:

Lemma 4.3 (see Lemma 4.2 in Paper III)
The second derivative of the MLE \( \hat{\zeta} \) is given by
\[
\hat{\zeta}'' = A^{-1} \kappa,
\]
where \( A \) is defined as given above and \( \kappa \) is defined in Section 4.2 of Paper III.

Similarly, we write the estimate of the conditional MSEP of \( \hat{R}_i(p) \) as a function of \( p \) and analyse its sensitity with respect to \( p \) by a Taylor expansion. To obtain the estimated conditional MSEP as a function of \( p \) it is necessary to express the estimate of the dispersion parameter as a function of \( p \), namely \( \hat{\phi}(p) \). We make use of the Pearson residuals to estimate the dispersion parameter and obtain the following:
\[
\hat{\phi}(p) = \frac{1}{d} \sum_{i+j \leq l} \frac{(X_{i,j} - \hat{m}_{i,j}(p))^2}{V(\hat{m}_{i,j}(p))},
\]
where \( d = \frac{(I+1)(I+2)}{2} - 2I - 1 \) is the degrees of freedom of the model and \( \hat{m}_{i,j}(p) = \hat{\mu}_i(p) \hat{\gamma}_j(p) \). Hence, the estimated conditional MSEP can be written as a function of \( p \) as follows:
\[
\text{msep}_{R_i|D_1}(\hat{R}_i(p)) = \hat{\phi}(p) \sum_{j > I-i} \hat{\mu}_i(p) \hat{\gamma}_j(p) + \sum_{l > I-k} \text{Cov}(\hat{\mu}_i(p) \hat{\gamma}_j(p), \hat{\mu}_k(p) \hat{\gamma}_l(p)).
\]
The details necessary to obtain the derivatives of the estimated conditional MSEP with respect to $p$ are provided in Lemmas 5.1 and 5.2 as well as Appendix C of Paper III. As with the claims reserves, the derivatives with respect to $p$ of the estimated conditional MSEP introduce a Taylor expansion used to approximate the estimated conditional MSEP for small deviations from the chosen value $p$.

To gauge the performance of these approximations, we provide an example. The example is centered around $p = 2$, corresponding to the assumption that the incremental claim amounts $X_{i,j}$ independently follow a gamma distribution. A further example for $p = 1$ is provided in Section 6 of Paper III. The aim is to see if the claims reserves and their corresponding uncertainty fluctuate with small changes of the assumed model parameter $p$. Furthermore, given a considerable change, we aim to see if we can rely on the first or second order Taylor approximations to accurately capture them.

The dataset and numerical calculations are all provided in Section 6 of Paper III. It is evident from Figure 4.1 that the claims reserves are rather stable, notice that the scale of the y-axis is very small. This implies that Taylor approximations are barely necessary although they naturally perform very well. This is in contrast to the estimated conditional MSEP, where the scale of the y-axis is very large, signifying that the conditional MSEP varies widely; see Figure 4.2. Furthermore, the second order Taylor approximation is a large improvement upon the first when the shift in $p$ increases. This example, verified by many more case studies that we performed, hints at the possibility that the claims reserves are rather stable regardless of the member of the Tweedie EDF that is chosen to model them; a result verified by Peters et al. (2009). The uncertainty of the predictor, on the other hand, displays no such stability, it has been shown by this example to vary widely over the family of models.
4.3. Model Uncertainty (Paper III)

Figure 4.1: True and approximated claims reserves.

Figure 4.2: True and approximated rooted conditional MSEP.
5. Multivariate Chain Ladder Models

Until now, we have only considered a single claims development triangle. We have presented several methods that, based on the triangle, have the ability to determine the claims reserves as well as quantify the uncertainty in prediction. In this chapter, we study the problem of claims reserving for multiple correlated claims development triangles. As a direct consequence of its simplicity, the CL method is frequently used to tackle this issue. Hence we focus our attention on so-called multivariate chain ladder models.

Naturally, the easiest solution is to aggregate the data and perform the analysis on the aggregate (univariate) triangle. Such an approach, however, is rather crude since the aggregate triangle usually does not conform to the same homogeneity properties as the individual triangles; see e.g. Anje (1994) and Klemmt (2005). We study two situations where such homogeneity assumptions are violated. First, in the case that each claims development triangle constitutes a separate line of business and second, when within one line of business, claims have been separated based on their size. In the latter scenario one could investigate any characteristic that influences the development of the underlying claims. We study size since we see it as the canonical characteristic applicable to almost any line of business.

In the situation of multiple lines of business, it is intuitively clear that claims from different types of insurance contracts can behave very differently, Pröhl and Schmidt (2005) elaborate on this point, providing many other reasons for studying multiple correlated claims development triangles. The study focusing on such homogeneity violations has been considered in the literature by Braun (2004), Pröhl and Schmidt (2005), Schmidt (2006), and Merz and Wüthrich (2008a). These contributions were limited to univariate predictors with associated MSEP (i.e. second moment calculation) and multivariate predictors lacking MSEP analysis. Merz and Wüthrich (2008b) bridge the gap and provide dependence structure as well as prediction uncertainty. They achieve this by defining correlation through a deterministic dependence matrix. Simulation based techniques using bootstrapping have also been considered in Brehm (2002), Kirschner et al. (2002), and Taylor and McGuire (2005, 2007). However, we focus on the analytic solutions since they often provide more insight to the problem.

Although the methods above can also be used in the second situation, a more intuitive method to deal with this source of heterogeneity is presented in Paper IV. In this method, the dependence between the different types of claims is modelled naturally through the claims settlement process.
5.1 Correlated Portfolios

To demonstrate the method introduced in Merz and Wüthrich (2008b), additional notation is required. Since we study multiple claims development triangles, we endow the cumulative claim amounts with superscript \( n \in \{1, \ldots, N\} \). We denote the observations available at time \( I \) by

\[
D_N^I = \bigcup_{n=1}^{N} \left\{ C_{i,j}^{(n)} ; i + j \leq I \right\},
\]

and the \( N \)-dimensional vector representation of the cumulative claim amounts by

\[
C_{i,j} = \left( C_{i,j}^{(1)}, \ldots, C_{i,j}^{(N)} \right)'.
\]

Furthermore, we denote by \( D(v) \) and \( D(v)^p \) the \( N \times N \) diagonal matrices of the \( N \)-dimensional vector \( v \) and that of the vector \( v \) where the entries are raised to the power \( p \), respectively. Finally, we define \( \mathbf{1} = (1, \ldots, 1)' \) to be the \( N \)-dimensional vector consisting of ones.

**Model Assumptions 5.1 (Multivariate Chain Ladder Model):**

\((C_{i,j})_{j \geq 0}\) form an \( N \)-dimensional Markov process for every accident year \( i \) where,

- there exist deterministic vectors
  \[
  f_j = \left( f_j^{(1)}, \ldots, f_j^{(N)} \right)',
  \]
  with \( f_j^{(n)} > 0, n \in \{1, \ldots, N\} \), and symmetric positive definite \( N \times N \) matrices \( \Sigma_j \) for \( j \in \{1, \ldots, I - 1\} \) such that for all \( i \in \{0, \ldots, I\} \) and \( j \in \{1, \ldots, I\} \), we have
  \[
  E[C_{i,j} | C_{i,j-1}] = D(f_{j-1}) C_{i,j-1},
  \]
  \[
  \text{Cov}(C_{i,j}, C_{i,j} | C_{i,j-1}) = D(C_{i,j-1}^{1/2}) \Sigma_{j-1} D(C_{i,j-1}^{1/2}).
  \]

- cumulative claim amounts \( C_{i,j} \) of different accident years \( i \) are independent.

As before, the parameters \( f_j^{(n)} \) are called CL factors. As a result of these model assumptions, we obtain for all \( i \),

\[
E[C_i | D_i^N] = \prod_{j=i-\ldots}^{I-1} D(f_j) C_{i,i-\ldots};
\]

see Lemma 3.3 in Merz and Wüthrich (2008b). This result mirrors the conditional expectation obtained in the univariate CL method under the assumptions presented.
5.1. Correlated Portfolios

in Mack (1993). Furthermore, it motivates the following predictor of the ultimate claim amount:

$$\hat{C}_{i,I} = \prod_{j=I-i}^{I-1} D(\hat{f}_j) C_{i,I-j},$$

where $\hat{f}_j$ is given by equation (3.9) in Merz and Wüthrich (2008b). This approach to estimating the CL factors incorporates the correlation matrices $\Sigma_j$ and hence we deem $\hat{C}_{i,I}$ to be a multivariate predictor. Furthermore, Lemma 3.5 in Merz and Wüthrich (2008b) shows that such $\hat{f}_j$ are unbiased, optimal least squares estimators for the CL factors. One could also estimate the CL factors based on the univariate CL factors defined previously. This approach is taken in Braun (2004) and Merz and Wüthrich (2008a), however, the resulting ultimate claim amount predictor can only be thought of as a univariate predictor.

Without providing all the details, Merz and Wüthrich (2008b) are able to analytically quantify the prediction uncertainty of $\sum_{n=1}^{N} \hat{C}_{i,I}$ using the conditional MSEP. Hence, for single accident year $i$, it is given by

$$\text{msep} \sum_{n=1}^{N} C^{(n)}_{i,I} \left( \sum_{n=1}^{N} C^{(n)}_{i,I} \right) = 1' \text{Var}(C_{i,I} | D^N_I) 1$$

$$+ 1' \left( \hat{C}_{i,I} - E[C_{i,I} | D^N_I] \right) \left( \hat{C}_{i,I} - E[C_{i,I} | D^N_I] \right)' 1,$$

where the first term on the right-hand side represents the conditional process variance and the last, the conditional estimation error. Lemma 4.1 in Merz and Wüthrich (2008b) provides a recursive representation for the conditional process variance. Once obtained, the conditional process variance can then be estimated by the usual replacing of parameters with their estimates. Parameter estimation is discussed in Section 5 of Merz and Wüthrich (2008b), we do not elaborate here.

The conditional estimation error lacks any such simplicity. To estimate the conditional estimation error, Merz and Wüthrich (2008b) modify the conditional resampling approach of the CL factors described in Buchwalder et al. (2006b). The modification is done in accordance with the multivariate CL factor estimates provided above. It should be noted that several approaches are available to estimate the conditional estimation error for the CL method. Extensive discussion on these approaches have arisen since the choice of which approach to use is not solely mathematical in nature, but also circumstantial; see e.g. Buchwalder et al. (2006b), Mack et al. (2006), Gisler (2006), Venter (2006), Murphy (2007) and Wüthrich and Merz (2008).

Results 4.8 and 4.10 in Merz and Wüthrich (2008b) provide the estimates of the conditional MSEP for single accident year $i$ and aggregated accident years, respectively. In the following section, we study the work presented in Paper IV. A different motivation is given to utilize a multivariate CL method with appropriately different model assumptions driving the dependence structure.
5.2 Claims Separation by Size (Paper IV)

We consider a single line of business and study the associate claims development triangle. If the underlying outstanding loss liabilities are too heterogeneous, practitioners often divide the portfolio into two subportfolios, one containing small claims, the other large claims. They then apply the CL method to each subportfolio. The importance of such separation has been studied before; see e.g. Klemmt (2005). In general, one obtains more stable behaviour in the subportfolios. The application of a multivariate CL method requires careful consideration since the separation does not satisfy the additivity property; see e.g. Anje (1994).

In Paper IV, the multivariate CL model is analysed for the small-large claims separation and a formula for the prediction uncertainty is derived. This provides those practitioners who separate small and large claims with the necessary tools required to quantify the prediction uncertainty of the resulting claims reserves.

In order to classify the individual claims, we introduce a threshold parameter $d_i$ that is chosen a priori for each accident year $i$. One could pose the question of “best” threshold value. We do not treat this question here since we believe it to be also circumstantial in nature, rather than purely mathematical. For example, in practice there is often a natural threshold determined by considerations such as internal processes, reinsurance programs, and the volume of the company.

We introduce additional notation to describe the model assumptions. We let $C_{i,j}^{(s)}$ denote the cumulative payments in the small claim layer $[0, d_i]$ (based on individual claims observations). The random variable $C_{i,j}^{(l)}$ denotes the cumulative payments in the large layer $(d_i, \infty)$, and $C_{i,j}^{(s-l)}$ denotes the payments in the layer $(d_i, \infty)$ for claims that develop into large claims in period $j$. More formally, we let $Y_{i,j}^{(k)}$ be the incremental claim amount for the $k^{th}$ claim of accident year $i$ in development year $j$. Then

$$Z_{i,j}^{(k)} = \sum_{l=0}^{j} Y_{i,j}^{(k)}$$

are the cumulative claim amounts for the $k^{th}$ claim of accident year $i$ in the first $j$ development years. Furthermore, let $n_{i,j}$ be the number of reported claims for accident year $i$ after $j$ development years. Then

$$X_{i,j} = \sum_{k=1}^{n_{i,j}} Y_{i,j}^{(k)}$$

is the incremental claim amount in cell $(i, j)$ for all claims, and

$$C_{i,j} = \sum_{l=0}^{j} X_{i,l} = \sum_{k=1}^{n_{i,j}} Z_{i,j}^{(k)}$$

is the cumulative claim amount in cell $(i, j)$. We assume that incremental claim amounts are non-negative. We further introduce a threshold parameter $d_i$ for each
accident year $i$ such that we denote $Z_{i,j}^{(k)}$ a “small claim” if $Z_{i,j}^{(k)} \leq d_i$. Hence, we define the following random variables:

\[
C_{i,j}^{(s)} = \sum_{k=1}^{n_{i,j}} \min(Z_{i,j}^{(k)}, d_i), \\
C_{i,j}^{(s\rightarrow l)} = \sum_{k=1}^{n_{i,j}} \max(Z_{i,j}^{(k)} - d_i, 0)I_{\{Z_{i,j}^{(k)} \leq d_i\}}, \quad \text{for } j \geq 1, \\
C_{i,j}^{(l)} = \sum_{k=1}^{n_{i,j}} \max(Z_{i,j}^{(k)} - d_i, 0), 
\]

where $C_{i,0}^{(s\rightarrow l)}$ is defined to be zero for all $i$. The separation was performed in layers to avoid a decrease in volume in the small claim amount that would have otherwise occurred when a small claim becomes large. Given the above cumulative variables, the following holds:

\[
C_{i,j} = C_{i,j}^{(s)} + C_{i,j}^{(l)}. 
\]

We furthermore define $C_{i,j} = (C_{i,j}^{(s)}, C_{i,j}^{(s\rightarrow l)}, C_{i,j}^{(l)})'$,

\[
D_I = \{C_{i,j}; \ i + j \leq I\}, 
\]

and

\[
B_k = \{C_{i,j}; \ i + j \leq I, 0 \leq j \leq k\} \subseteq D_I.
\]

**Model Assumptions 5.2 (Separation Chain Ladder Model):**

$(C_{i,j})_{j \geq 0}$ form a multi-dimensional Markov process for every accident year $i$ where,

- there exist deterministic CL factors $f_{j}^{(s)}, f_{j}^{(l)}, f_{j}^{(s\rightarrow l)}$ and variance parameters $\sigma_{j}^{(s)}, \sigma_{j}^{(l)}, \sigma_{j}^{(s\rightarrow l)}, \rho_{j}^{(s,s\rightarrow l)}$ such that for all $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, I - 1\}$
we have,

\[
\begin{align*}
E\left[C_{i,j+1}^{(s)} \mid C_{i,j}\right] &= f_j^{(s)} C_{i,j}, \\
E\left[C_{i,j+1}^{(s \to l)} \mid C_{i,j}\right] &= f_j^{(s \to l)} C_{i,j}, \\
E\left[C_{i,j+1}^{(l)} \mid C_{i,j}\right] &= f_j^{(l)} C_{i,j} + f_j^{(s \to l)} C_{i,j}, \\
\text{Var}\left(C_{i,j+1}^{(s)} \mid C_{i,j}\right) &= (\sigma_j^{(s)})^2 C_{i,j}^{(s)}, \\
\text{Var}\left(C_{i,j+1}^{(s \to l)} \mid C_{i,j}\right) &= (\sigma_j^{(s \to l)})^2 C_{i,j}^{(s)}, \\
\text{Var}\left(C_{i,j+1}^{(l)} \mid C_{i,j}\right) &= (\sigma_j^{(l)})^2 C_{i,j}^{(l)} + (\sigma_j^{(s \to l)})^2 C_{i,j}^{(s)}, \\
\text{Cov}\left(C_{i,j+1}^{(s)}, C_{i,j+1}^{(s \to l)} \mid C_{i,j}\right) &= \sigma_j^{(s)} \sigma_j^{(s \to l)} C_{i,j}^{(s \to l)} \rho_j^{(s \to l)}, \\
\text{Cov}\left(C_{i,j+1}^{(s \to l)}, C_{i,j+1}^{(l)} \mid C_{i,j}\right) &= \sigma_j^{(s \to l)} \sigma_j^{(l)} C_{i,j}^{(s)}, \\
\text{Cov}\left(C_{i,j+1}^{(s)}, C_{i,j+1}^{(l)} \mid C_{i,j}\right) &= \sigma_j^{(s)} \sigma_j^{(l)} C_{i,j}^{(s \to l)} \rho_j^{(s \to l)}.
\end{align*}
\]

- cumulative claim amounts \(C_{i,j}\) in different accident years \(i\) are independent.

To estimate the factors \(f_j\) we use the CL approach, it produces the following estimators:

\[
\begin{align*}
\hat{f}_j^{(s)} &= \frac{\sum_{k=0}^{I-j-1} C_{k,j+1}^{(s)}}{\sum_{k=0}^{I-j-1} C_{k,j}^{(s)}}, \\
\hat{f}_j^{(s \to l)} &= \frac{\sum_{k=0}^{I-j-1} C_{k,j+1}^{(s \to l)}}{\sum_{k=0}^{I-j-1} C_{k,j}^{(s)}}, \\
\hat{f}_j^{(l)} &= \frac{\sum_{k=0}^{I-j-1} \left( C_{k,j+1}^{(l)} - C_{k,j+1}^{(s \to l)} \right)}{\sum_{k=0}^{I-j-1} C_{k,j}^{(l)}}.
\end{align*}
\]

**Lemma 5.3 (see Lemma 4.3 in Paper IV)**

Under Model Assumptions 5.2 we have that \(\hat{f}_j^{(s)}, \hat{f}_j^{(s \to l)},\) and \(\hat{f}_j^{(l)}\), are, given \(B_j\), unbiased estimators for \(f_j^{(s)}, f_j^{(s \to l)},\) and \(f_j^{(l)}\), respectively.

In addition to the above factors being conditionally unbiased they are also uncorrelated.
5.2. Claims Separation by Size (Paper IV)

Lemma 5.4 (see Lemma 4.4 in Paper IV)
The estimators $\hat{f}(s)$, $\hat{f}(s\rightarrow l)$ and $\hat{f}(l)$ are (conditionally) uncorrelated, more precisely,

$$E[\hat{f}_j^{(x)} \hat{f}_k^{(y)} | B_j] = E[\hat{f}_j^{(x)} | B_j] E[\hat{f}_k^{(y)} | B_k],$$

for $j, k \in \{0, \ldots, I-1\}$, $j < k$, and $x, y \in \{s, s \rightarrow l, l\}$.

For the proofs of the above two lemmas, see Paper IV, Section 4.2. With the factor estimators $\hat{f}_j$, which are conditionally unbiased and uncorrelated, we recursively define predictors for the cumulative claims as follows,

$$\hat{C}_{i,j}^{(s)} = \hat{f}_j^{(s)} \hat{C}_{i,j-1}^{(s)},$$

$$\hat{C}_{i,j}^{(s\rightarrow l)} = \hat{f}_j^{(s\rightarrow l)} \hat{C}_{i,j-1}^{(s)} + \hat{C}_{i,j}^{(s\rightarrow l)},$$

$$\hat{C}_{i,j}^{(l)} = \hat{f}_j^{(l)} \hat{C}_{i,j-1}^{(l)} + \hat{C}_{i,j}^{(s\rightarrow l)} + \hat{C}_{i,j}^{(l)},$$

$$\hat{C}_{i,j} = \hat{C}_{i,j}^{(s)} + \hat{C}_{i,j}^{(l)},$$

for $j \in \{I-i+1, \ldots, I\}$, where $\hat{C}_{i,I-i}^{(x)} = C_{i,I-i}^{(x)}$ for $x \in \{s, s \rightarrow l, l\}$. This formulation leads to the following corollary:

Corollary 5.5 (see Lemma 4.5 in Paper IV)
$\hat{C}_{i,l}^{(s)}$, $\hat{C}_{i,l}^{(l)}$, and $\hat{C}_{i,l}$ are conditionally, given $C_{i,l-I}$, unbiased estimators for $E[C_{i,l}^{(s)} | D_I]$, $E[C_{i,l}^{(l)} | D_I]$, and $E[C_{i,l} | D_I]$, respectively.

The proof is given in Section 4.2 of Paper IV. Since the predictor of the ultimate claim amount is $D_I$-measurable, the conditional MSEP is provided, as usual, by the conditional process variance and conditional estimation error,

$$\text{msep}_{C_{i,l}|D_I}(\hat{C}_{i,l}) = \text{Var}(C_{i,l} | D_I) + (\hat{C}_{i,l} - E[C_{i,l} | D_I])^2,$$

for single accident year $i$. As with the work of Merz and Wüthrich (2008b), the condition process variance is estimated using a recursive approach followed by replacing unknown parameters with their estimates and the conditional estimation error is estimated using a modified conditional resampling approach as described in Buchwalder et al. (2006b). The details are provided in Section 5 and the appendices of Paper IV. Furthermore, we also refer the reader to Section 6 of Paper IV for a case study that highlights the performance of this method in comparison to the univariate CL method. In Table 5.1, we present the aggregated results for the small layer, the large layer, the aggregated multivariate model and the aggregated univariate model. We conclude that, as in our example, a large proportion of claims can have a big impact on the reserves but a small impact on the prediction uncertainty. As is evident from Table 5.1 a smaller collection of claims with high severity can make a huge impact on the prediction uncertainty. For this reason,
practitioners often subdivide these two classes of claims to get more reliable predictions. If the split is done in layers then, in general, the lower layer is rather stable, whereas the upper layer requires special care. Note that we have used the CL method in the upper layer to get feasible conditional MSEP formulas. Future research should also consider other reserving methods that are more appropriate to model the upper layer, for example methods that consider all available information.

Table 5.1: The effect of layering on the CL algorithm.

<table>
<thead>
<tr>
<th></th>
<th>reserves</th>
<th>process error</th>
<th>estimation error</th>
<th>msep$^{1/2}$</th>
<th>Vco</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small layer</td>
<td>28,446.79</td>
<td>1,641.22</td>
<td>527.52</td>
<td>1,723.91</td>
<td>6.06%</td>
</tr>
<tr>
<td>Large layer</td>
<td>23,386.24</td>
<td>3,680.26</td>
<td>4,437.39</td>
<td>5,764.96</td>
<td>24.65%</td>
</tr>
<tr>
<td>Total multivariate</td>
<td>51,833.03</td>
<td>4,211.09</td>
<td>4,527.58</td>
<td>6,183.22</td>
<td>11.93%</td>
</tr>
<tr>
<td>Total univariate</td>
<td>53,573.64</td>
<td>4,121.64</td>
<td>5,380.80</td>
<td>6,777.97</td>
<td>12.65%</td>
</tr>
</tbody>
</table>

Multivariate Chain Ladder Models
Conclusions and Future Research

We have studied the classical claims reserving problem faced by non-life insurance companies. Using stochastic modelling in a GLM framework, we have provided uncertainty prediction in the BF method as well as explored the use of the Tweedie EDF. We have focused considerable thought and effort to make the results feasible for use in practice, for the most part requiring only spreadsheet environment to perform the calculations. Furthermore, we have studied multivariate methods based on the classical CL method in an attempt to provide practitioners with the necessary tools to quantify prediction uncertainty when separating claims based on some behavioural criteria.

In instances where we have obtained analytic prediction uncertainty, future research should focus on obtaining predictive distributions. A first attempt could be made with the use of bootstrapping and Markov chain Monte Carlo simulation methods. Furthermore, tail factor considerations have not been sufficiently explored in the literature; future research aimed at studying the uncertainty of tail factors would highly benefit practitioners. Other areas of interest with potential for significant development include sophisticated dependence and inflation modelling as well as stochastic discounting.


Bibliography


Daniel H. Alai, Michael Merz and Mario V. Wüthrich (2009).
Mean square error of prediction in the Bornhuetter-Ferguson claims reserving method.
Mean Square Error of Prediction in the
Bornhuetter-Ferguson Claims Reserving Method

Daniel H. Alai  Michael Merz  Mario V. Wüthrich
Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland

Abstract

The prediction of adequate claims reserves is a major subject in actuarial practice and science. Due to their simplicity, the chain ladder (CL) and Bornhuetter-Ferguson (BF) methods are the most commonly used claims reserving methods in practice. However, in contrast to the CL method, no estimator for the conditional mean square error of prediction (MSEP) of the ultimate claim has been derived in the BF method until now, and as such, this paper aims to fill that gap. This will be done in the framework of generalized linear models (GLM) using the (overdispersed) Poisson model motivation for the use of CL factor estimates in the estimation of the claims development pattern.

Keywords: Claims Reserving, Bornhuetter-Ferguson, Overdispersed Poisson Distribution, Chain Ladder Method, Generalized Linear Models, Conditional Mean Square Error of Prediction
1 Introduction

Often in non-life insurance, claims reserves are the largest item on the liability side of the balance sheet. Therefore, given the available information about the past, the prediction of an adequate amount to face the responsibilities assumed by the non-life insurance company as well as the quantification of the uncertainties in these reserves are major issues in actuarial practice and science; see e.g. Casualty Actuarial Society (1990) and Teugels and Sundt (2004).

Due to their simplicity, the chain ladder (CL) and Bornhuetter-Ferguson (BF) methods are among the easiest claims reserving methods and, therefore, the most commonly used techniques in practice. Mack (1993) published a fundamental article on claims reserving regarding the estimation of the conditional mean square error of prediction (MSEP) in the CL method. Unfortunately, until now, no estimator for the conditional MSEP of the ultimate claim has been derived in the BF method.

The BF method goes back to Bornhuetter and Ferguson (1972). Apart from its simplicity, the BF method is a very popular claims reserving method since it is rather robust against outliers in the observations and allows for incorporating a priori knowledge from experts, premium calculations or strategic business plans. Furthermore, in contrast to the CL method, the BF method has proven to be a very robust method, in particular, against instability in the proportion of ultimate claims paid in early development years. The BF method, as it was stated in the original work of Bornhuetter and Ferguson (1972), was not formulated in a probabilistic way. The work of Mack (2000) and Verrall (2004) puts the BF method into a probabilistic framework; we discuss this further below. Before describing the method in detail we would like to mention that there are also rather skeptic and critical opinions of the use of the BF method. The purpose of this paper is not to improve the method in view of this criticism but rather to explain from a probabilistic point of view, what is done when actuaries use the BF method in its current state and to derive analytic estimators for the prediction uncertainty.

The BF method is based on the simple idea of stabilizing the BF estimate $\hat{C}_{i,j}^{BF}$ using an initial estimate $\hat{\mu}_i$ of the ultimate claim $C_{i,j}$ based on external knowledge. Then it is standard practice to use the prior estimate $\hat{\mu}_i$ with the CL factor estimates $\hat{f}_j$ to predict the ultimate claim. In this case, the CL method and the BF method only differ in the choice of the estimate for the ultimate claim (CL estimate versus prior estimate). Hence, in this regard the BF method is a variant of the CL method that uses external information to obtain an initial estimate for the ultimate claim. Mack (2000) studied this from a probabilistic point of view. In his work he analysed the stochastic model for given (deterministic) claims development patterns, which are the analogon to the CL factors, and random initial estimates $\hat{\mu}_i$. Mack (2000) then derived optimal credibility weighted averages between the CL and the BF method. However, in most practical applications the claims develop-
ment pattern and the CL factors are unknown and need to be estimated from the data. This adds an additional source of uncertainty to the problem. Verrall (2004) has studied these uncertainties using a Bayesian approach to the BF method. If one uses an appropriate Bayesian approach with improper priors and an appropriate two stage procedure one then arrives at the BF method. We use a similar procedure within generalized linear models (GLM) using maximum likelihood estimators (MLE). This framework allows for an analytic estimate for the mean square error of prediction using asymptotic properties of MLE. Note that in a Bayesian framework one can, in general, only give numerical answers using simulation techniques such as the Markov chain Monte Carlo method.

A criticism of the BF method as it is currently used is that the use of the CL estimates $\hat{f}_j$ contradicts the basic idea of independence between the last observed cumulative claims $C_{i,I-i}$ and the estimated outstanding claims liabilities $\hat{C}_{i,J}^{BF} - C_{i,I-i}$, which was fundamental to the BF method; see e.g. Mack (2006). Therefore, Mack (2006) proposed different estimators for the claims development pattern. In this paper however, we do not follow this route. We rather use the well-known fact that the (overdispersed) Poisson model leads to the same claims reserves and payout pattern as the CL model. This means that we use the (overdispersed) Poisson model motivation for the use of the CL factor estimates $\hat{f}_j$. It is then straightforward to use GLM methods for parameter estimation and to derive an estimator for the conditional MSEP of the ultimate claim in the BF method.

Organization of the paper. In Section 2 we provide the notation and data structure. In Section 3 we give a short review of the CL and BF methods and compare these two techniques. Section 4 is dedicated to the overdispersed Poisson model and its representation as a GLM. In Section 5 we give an estimation procedure for the conditional MSEP in the BF method. Finally, in Section 6 we discuss an example.

2 Notation and Data Structure

Throughout, we assume the loss data for the run-off portfolio is given by a claims development triangle of observations. However, all claims reserving methods discussed in this paper can also be applied to other shapes of loss data (e.g. claims development trapezoids). In this claims development triangle the indices $i \in \{0, 1, \ldots, I\}$ and $j \in \{0, 1, \ldots, J\}$ with $I \geq J$ refer to accident years and development years, respectively. The incremental claims (i.e. incremental payments, change of reported claim amount or number of newly reported claims) for accident year $i$ and development year $j$ are denoted by $X_{i,j}$ and cumulative claims (i.e. cumulative payments, claims incurred or total number of reported claims) of accident year $i$ up to devel-
opment year $j$ are given by

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

(2.1)

We assume that the last development year is given by $J$, i.e. $X_{i,j} \equiv 0$ for all $j > J$, and the last accident year is given by $I$. Moreover, our assumption that we consider claims development triangles implies $I = J$.

<table>
<thead>
<tr>
<th>accident year $i$</th>
<th>development year $j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0$</td>
<td>$0$ \ldots $j$ \ldots $J$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$I - j$</td>
<td>$r.v. X_{i,j}, C_{i,j}$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$I$</td>
<td>$\text{predicted r.v. } X_{i,j}, C_{i,j}$</td>
</tr>
</tbody>
</table>

Figure 1: Claims development triangle.

Usually, at time $I$ (i.e. calendar year $I$), we have observations $D_I$ in the upper claims development triangle, defined as follows,

$$D_I = \{X_{i,j} : i + j \leq I\}. \quad (2.2)$$

We need to predict the random variables in its complement

$$D_I^c = \{X_{i,j} : i + j > I, i \leq I\}. \quad (2.3)$$

Figure 1 shows the claims data structure for the claims development triangle described above.

Furthermore, let $R_i$ and $R$ denote the outstanding claims liabilities for accident year $i$ at time $I$,

$$R_i = \sum_{j=I-i+1}^{J} X_{i,j} = C_{i,I-I-i} \quad \text{for } 1 \leq i \leq I, \quad (2.4)$$

and the total outstanding claims liabilities for aggregated accident years,

$$R = \sum_{i=1}^{I} R_i, \quad (2.5)$$

respectively. The prediction of the outstanding claims liabilities $R_i$ and $R$ by the so-called claims reserves or best estimates, as well as quantifying the uncertainty in this prediction, is the classical actuarial claims reserving problem studied at every non-life insurance company.
3 Bornhuetter-Ferguson and Chain Ladder Methods

In this section we give a short review of the CL and BF methods, which are the most commonly used claims reserving methods in practice on account of their simplicity. Our review is similar to the one given in Mack (2000).

3.1 Chain Ladder Method

The classical actuarial literature often explains the CL method as a pure computational algorithm to estimate claims reserves. The first distribution-free stochastic model was proposed by Mack (1993).

Model Assumptions 3.1 (CL model):

- There exist deterministic development factors $f_0, \ldots, f_{J-1} > 0$ such that for all $0 \leq i \leq I$ and $1 \leq j \leq J$ we have
  \[ E[C_{i,j} | C_{i,0}, \ldots, C_{i,j-1}] = E[C_{i,j} | C_{i,j-1}] = f_{j-1} C_{i,j-1}. \]  
  \( (3.1) \)

- Claims $C_{i,j}$ of different accident years $i$ are independent.

An easy exercise in calculating conditional expectation leads to

\[ E[C_{i,J} | D_I] = f_{J-1} E[C_{i,J-1} | C_{i,I-i}] = \ldots = C_{i,I-i} \prod_{j=I-i}^{J-1} f_j, \]  
  \( (3.2) \)

for $1 \leq i \leq I$, where the factors $f_j$ are called CL factors or development factors. Given the observations $D_I$ and CL factors $f_j$, (3.2) gives a recursive algorithm for predicting the ultimate claim $C_{i,J}$. However, in most practical applications the CL factors $f_j$ are not known and have to be estimated from the data $D_I$. It is well known that the $D_I$-measurable estimators for the CL factors $\hat{f}_j$, defined by

\[ \hat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}}, \]  
  \( (3.3) \)

for $0 \leq j \leq J - 1$, are unbiased and uncorrelated; see e.g. Mack (1993). However, they are not independent since the squares of two successive estimators $\hat{f}_j$ and $\hat{f}_{j+1}$ are negatively correlated; see e.g. Mack et al. (2006) and Wüthrich et al. (2008).

The properties of the CL factor estimates $\hat{f}_j$ imply that, given $C_{i,I-i}$, the CL estimator of the ultimate claim $C_{i,J}$, defined by

\[ \hat{C}_{i,J}^{CL} = C_{i,I-i} \prod_{j=I-i}^{J-1} \hat{f}_j \]  
  \( (3.4) \)

for $1 \leq i \leq I$, is an unbiased estimator for $E[C_{i,J} | D_I]$. 

57
3.2 Bornhuetter-Ferguson Method

The BF method goes back to Bornhuetter and Ferguson (1972). Analogously to the CL method, the classical actuarial literature often explains the BF method as a pure computational algorithm to estimate claims reserves although there are several stochastic models that motivate the BF method.

The following stochastic model is consistent with the BF method.

Model Assumptions 3.2 (BF model):

- There exist parameters \( \mu_0, \ldots, \mu_I > 0 \) and a pattern \( \beta_0, \ldots, \beta_J > 0 \) with \( \beta_J = 1 \) such that for all \( 0 \leq i \leq I, \ 0 \leq j \leq J - 1 \) and \( 1 \leq k \leq J - j \)

\[
E[C_{i,0}] = \beta_0 \mu_i, \tag{3.5}
\]

\[
E[C_{i,j+k}|C_{i,0}, \ldots, C_{i,j}] = C_{i,j} + (\beta_{j+k} - \beta_j) \mu_i. \tag{3.6}
\]

- Claims \( C_{i,j} \) of different accident years \( i \) are independent.

These assumptions imply

\[
E[C_{i,j}] = \beta_j \mu_i \quad \text{and} \quad E[C_{i,J}] = \mu_i, \tag{3.7}
\]

which is often used to explain the BF method; see e.g. Radtke and Schmidt (2004). The sequence \( (\beta_j)_j \) denotes the claims development pattern and, if \( C_{i,j} \) are cumulative payments, \( \beta_j \) is the expected cumulative cashflow pattern (also called payout pattern). Such a pattern is often used when one needs to build market-consistent/discounted reserves, where money values differ over time.

Assumption (3.6) motivates the BF estimator for the ultimate claim \( C_{i,J} \) given by

\[
\hat{C}_{i,J}^{BF} = C_{i,I-i} + (1 - \hat{\beta}_{I-i}) \hat{\mu}_i, \tag{3.8}
\]

for \( 1 \leq i \leq I \), where \( \hat{\beta}_{I-i} \) is an appropriate estimate for \( \beta_{I-i} \) and \( \hat{\mu}_i \) is a prior estimate for the expected ultimate claim \( E[C_{i,J}] \). In practice, \( \hat{\mu}_i \) is an exogenously determined estimate (i.e. without the observations \( D_I \)) such as a plan value from a strategic business plan or the value used for premium calculations.

3.3 Comparison of BF and CL Methods

From the CL Assumptions 3.1 we obtain

\[
E[C_{i,J}] = E[C_{i,j}] \prod_{k=j}^{J-1} f_k, \tag{3.9}
\]
which implies that
\[
E[C_{i,j}] = \prod_{k=j}^{J-1} f_k^{-1} E[C_{i,J}], \quad \text{for all } 0 \leq j \leq J. \tag{3.10}
\]

If we compare this to the BF method, see e.g. (3.7), we find that \(\prod_{k=j}^{J-1} f_k^{-1}\) plays the role of \(\beta_j\). Therefore, these parameters are often viewed equally and if one knows the CL factors \(f_k\) one can construct a development pattern \((\beta_j)_j\) and vice versa. That is, in practice the \(\beta_j\) are usually estimated by
\[
\hat{\beta}_j^{(CL)} = \hat{\beta}_j = \prod_{k=j}^{J-1} \frac{1}{f_k}.
\]

where \(\hat{f}_k\) are the CL factor estimates given in (3.3). Moreover, using the estimator \(\hat{\beta}_j^{(CL)}\) for \(\beta_j\) in the BF method, we see that
\[
\tilde{C}_{i,J}^{BF} = C_{i,I-I} + \left(1 - \hat{\beta}_{I-I}^{(CL)}\right) \hat{\mu}_i,
\]
\[
\tilde{C}_{i,J}^{CL} = C_{i,I-I} + \left(1 - \hat{\beta}_{I-I}^{(CL)}\right) \tilde{C}_{i,J}^{CL},
\]

which means that the CL method and BF method only differ in the choice of the estimator for the ultimate claim \(C_{i,J}\) (prior estimate \(\hat{\mu}_i\) versus CL estimate \(\tilde{C}_{i,J}^{CL}\)). In other words, if we identify \(\hat{\beta}_j^{(CL)}\) and \(\prod_{k=j}^{J-1} \hat{f}_k^{-1}\), the BF method is a variant of the CL method that uses external information to obtain an initial estimate for the ultimate claim. The main criticism of this approach is that the use of the CL factor estimates \(\hat{f}_k\) contradicts the basic idea of independence between last observed cumulative claims \(C_{i,I-I}\) and estimated outstanding claims liabilities \(\tilde{C}_{i,J}^{BF} - C_{i,I-I}\), which was fundamental to the origin of the BF method; see e.g. Mack (2006). Therefore, Mack (2006) constructed different estimators for the claims development pattern \((\beta_j)_j\). However, we do not follow this route here. We rather concentrate on the overdispersed Poisson model motivation for the use of the CL factor estimates \(\hat{f}_k\) and utilize the fact that the overdispersed Poisson model is a GLM. It is then straightforward to use GLM methods for parameter estimation and to derive an estimator for the conditional MSEP of the ultimate claim in the BF method.

4 Overdispersed Poisson Model and Generalized Linear Models

In this section we give a brief review of the overdispersed Poisson model and its formulation in a GLM context.
4.1 Overdispersed Poisson Model

We define the overdispersed Poisson model by first considering the exponential dispersion family. Random variable $Y$ belongs to the exponential dispersion family if its density or probability distribution function can be written as

$$f_Y(y; \theta, \phi) = \exp \left\{ \frac{y \theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}.$$

The overdispersed Poisson model is a member of this family with $a(\phi) = \phi$, $b(\theta) = e^\theta$ and $c(y, \phi) = -\ln y$. It differs from the Poisson model in that the variance is not equal to the mean. This model was introduced for claims reserving in a Bayesian context by Verrall (1990, 2000, 2004) and Renshaw and Verrall (1998) and it is also used in the GLM framework; see e.g. McCullagh and Nelder (1989) and England and Verrall (2002, 2006). It is well-known in actuarial literature that the (overdispersed) Poisson model leads to the same claims reserves as the CL model. This result goes back to Hachemeister and Stanard (1975) and can be found, for example, in Mack (1991) and Verrall and England (2000). This means that although the CL model and the overdispersed Poisson model are very different, they lead to the same reserve estimates, the difference in the two models is relevant only if we estimate higher moments. In the following, we will utilize this correspondence as we do not motivate the use of the estimate $\hat{\beta}_{f_i}^{(CL)}$ by CL factor estimates $\hat{f}_k$ but rather by the MLEs in the overdispersed Poisson model. Note that CL factor estimates are used to calculate the MLEs.

Model Assumptions 4.1 (Overdispersed Poisson Model):

- The increments $X_{i,j}$ are independent overdispersed Poisson distributed and there exist positive parameters $\gamma_0, \ldots, \gamma_J$, $\mu_0, \ldots, \mu_I$ and $\phi > 0$ such that

$$E[X_{i,j}] = m_{i,j} = \mu_i \gamma_j,$$

$$\text{Var}(X_{i,j}) = \phi m_{i,j},$$

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

- $\hat{\mu}_i$ are independent random variables that are unbiased estimators of $\mu_i = E[C_{i,j}]$ for all $i$.

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

From Model Assumptions 4.1 we obtain

$$E[C_{i,j+k}|C_{i,0}, \ldots, C_{i,j}] = C_{i,j} + \sum_{l=1}^k E[X_{i,j+l}] = C_{i,j} + (\beta_{j+k} - \beta_{j})\mu_i,$$

where $\beta_j = \sum_{k=0}^j \gamma_k$. This means that the overdispersed Poisson model satisfies Model Assumptions 3.2 and can also be used to explain the BF method.
Remarks 4.2:

- The so-called dispersion parameter $\phi$ does not depend on accident year $i$ and development year $j$. The restriction in the overdispersed Poisson model is that we require $X_{i,j}$ to be non-negative.

- The parameters $\gamma_k$ define an expected incremental reporting/cashflow pattern over the development years $j$.

- The exogenous estimator $\hat{\mu}_k$ is a prior estimate for the expected ultimate claim $E[C_{k,J}]$, it is solely based on external data and expert opinion. Therefore, we assume that it is independent of the data $X_{i,j}$ (this is in the BF spirit as explained by Mack (2006)). Moreover, in order to obtain a meaningful model, we assume that it is unbiased for the expected ultimate claim. In this sense, we follow a pure BF method.

There are different methods for estimating the parameters $\mu_i$ and $\gamma_j$. In the following we use MLEs. The MLEs $\hat{\mu}_i^{(MLE)}$ and $\hat{\gamma}_j^{(MLE)}$ in the overdispersed Poisson model are found by solving

$$
\hat{\mu}_i^{(MLE)} \sum_{j=0}^{I-i} \hat{\gamma}_j^{(MLE)} = \sum_{j=0}^{I-i} X_{i,j}, \quad (4.5)
$$

$$
\hat{\gamma}_j^{(MLE)} \sum_{i=0}^{I-j} \hat{\mu}_i^{(MLE)} = \sum_{i=0}^{I-j} X_{i,j}, \quad (4.6)
$$

for all $0 \leq i \leq I$ and $0 \leq j \leq J$ under the constraint that $\sum_{j=0}^{J} \hat{\gamma}_j^{(MLE)} = 1$.

Remarks 4.3:

- Because of the multiplicative structure found in the overdispersed Poisson model, see (4.2), the parameters $\mu_i$ and $\gamma_j$ can only be determined up to a constant factor, i.e. $\hat{\mu}_i = c\mu_i$ and $\hat{\gamma}_j = \gamma_j/c$ would give the same estimate for $m_{i,j}$. Therefore, we need to impose a side constraint. In our situation this becomes that the MLEs $\hat{\gamma}_j^{(MLE)}$ form a development pattern, i.e. that $\sum_{j=0}^{J} \hat{\gamma}_j^{(MLE)} = 1$.

- If we now use these MLEs $\hat{\gamma}_j^{(MLE)}$ for the estimation of the expected incremental cashflow pattern $\gamma_j$ we obtain the following BF estimator,

$$
\hat{C}_{i,J}^{BF} = C_{i,I-i} + \left(1 - \sum_{j=0}^{I-i} \hat{\gamma}_j^{(MLE)}\right) \hat{\mu}_i; \quad (4.7)
$$
see e.g. (3.8). Moreover, under Model Assumptions 4.1, it has been proved that

$$\sum_{k=0}^{j} \hat{\gamma}_k^{(MLE)} = \hat{\beta}_j^{(CL)} = \prod_{k=j+1}^{J} \frac{1}{\hat{f}_k};$$

(4.8)

see e.g. Mack (1991) or Taylor (2000) which implies that the BF estimator,

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

(4.9)

is perfectly motivated by the overdispersed Poisson model and the use of the MLE $\hat{\gamma}_j^{(MLE)}$ for $\gamma_j$. This is exactly the BF estimator as it is commonly used in practice; see e.g. (3.12). Henceforth, in this understanding we do not motivate the use of the estimate $\hat{\beta}_i^{(CL)}$ by CL factor estimates $\hat{f}_k$ but rather by the MLEs $\hat{\gamma}_k^{(MLE)}$ in the overdispersed Poisson model. This means that (4.8) provides the essential steps for the use of $\hat{\beta}_j^{(CL)}$ in the BF method.

• Note that we use a similar two stage procedure as described by Verrall (2004). First we estimate the claims development pattern $\gamma_j$ and the exposures $\mu_i$ using MLE methods resulting in $\hat{\gamma}_j^{(MLE)}$ and $\hat{\mu}_i^{(MLE)}$. Only the simultaneous MLE of $\gamma_j$ and $\mu_i$ give the CL pattern given in (4.8). In the second step we then replace the MLE $\hat{\mu}_i^{(MLE)}$ by an external estimate $\hat{\mu}_i$. It could be argued that this is, in some sense, inconsistent, but it describes what practitioners do in the BF method to smoothen the claims reserves estimates.

4.2 Overdispersed Poisson Model as a Generalized Linear Model

Renshaw (1994) and Renshaw and Verrall (1998) were the first to implement the standard GLM techniques for the derivation of estimates for incremental data in a claims reserving context. In this section we give an brief description of the overdispersed Poisson model in the GLM framework. For more details on the overdispersed Poisson model in the GLM context and on GLMs and their statistical background we refer to England and Verrall (2002) and McCullagh and Nelder (1989) or Fahrmeir and Tutz (2001), respectively.

A specific GLM model (in a parametrization suitable for claims reserving) is fully characterized by the following three components:

a) the type of the Exponential Dispersion Family for the random component $X_{i,j}$;
b) the link function $g$ relating the expectation of the random component $X_{i,j}$ to the linear predictor $\eta_{i,j} = \Gamma_{i,j}b$, i.e.

$$g(\mathbb{E}[X_{i,j}]) = \eta_{i,j}$$

for all $0 \leq i \leq I$ and $0 \leq j \leq J$;

c) the design matrices $\Gamma_{i,j}$ for all $0 \leq i \leq I$ and $0 \leq j \leq J$.

In the overdispersed Poisson model the distribution of the random component is given by the overdispersed Poisson distribution, and for the multiplicative structure of Model 4.1 it is straightforward to choose the log-link $g(\cdot) = \log(\cdot)$ as link function. Then we have

$$\eta_{i,j} = g(m_{i,j}) = \log(\mu_i) + \log(\gamma_j).$$

(4.10)

For GLMs, it is easy to obtain MLEs of the parameters $\mu_i$ and $\gamma_j$ using standard GLM software. However, since the multiplicative structure of Model 4.1 is overparameterized it becomes necessary to set constraints which could take a number of different forms. In the last section we derived the MLEs $\hat{\gamma}_j^{(MLE)}$ under the normalization assumption $\sum_{j=0}^{J} \hat{\gamma}_j^{(MLE)} = 1$ in order to obtain a claims development pattern, see e.g. Remarks 4.3. However, in the framework of GLMs it is more natural and indeed convenient to choose the constraint $\mu_0 = 1$, hence $\log(\mu_0) = 0$ and

$$\eta_{0,j} = \log(\gamma_j) \quad \text{for all } 0 \leq j \leq J;$$

(4.11)

see e.g. (4.10). This parametrization leads to the following vector of unknown parameters

$$b = (\log(\mu_1), \ldots, \log(\mu_I), \log(\gamma_0), \ldots, \log(\gamma_J))',$n

(4.12)

and the $1 \times (I + J + 1)$ design matrices

$$\Gamma_{0,j} = (0, \ldots, 0, 0, 0, \ldots, 0, e_{I+j+1}, 0, \ldots, 0),$$

(4.13)

$$\Gamma_{i,j} = (0, \ldots, 0, e_i, 0, \ldots, 0, e_{I+j+1}, 0, \ldots, 0),$$

(4.14)

for $1 \leq i \leq I$ and $0 \leq j \leq J$, where the entries $e_i = 1$ and $e_{I+j+1} = 1$ are on the $i$-th and the $(I + j + 1)$-th position, respectively. We obtain the linear predictor

$$\eta_{i,j} = \Gamma_{i,j}b.$$ 

(4.15)

Hence, we have now reduced the dimension from $(I + 1) \times (J + 1)$ unknown parameters $m_{i,j}$ to $p = I + J + 1$ unknown parameters $\log(\mu_i)$ and $\log(\gamma_j)$.

Using standard GLM software based on the Fisher scoring method, these parameters are then estimated with the MLE method. We obtain the MLEs

$$\hat{b} = \left(\overrightarrow{\log(\mu_1)}^{GLM}, \ldots, \overrightarrow{\log(\mu_I)}^{GLM}, \overrightarrow{\log(\gamma_0)}^{GLM}, \ldots, \overrightarrow{\log(\gamma_J)}^{GLM}\right)^t,$n

(4.16)
which implies a second “payout” pattern
\[
\hat{\gamma}^{(GLM)}_0, \ldots, \hat{\gamma}^{(GLM)}_J,
\] (4.17)

where
\[
\hat{\gamma}^{(GLM)}_j = \exp\left(\log^{GLM}_{\gamma} \right)
\] (4.18)

for all \(0 \leq j \leq J\). The following relationships hold,
\[
\hat{\gamma}^{(MLE)}_j = \frac{\hat{\gamma}^{(GLM)}_j}{\sum_{l=0}^{J} \hat{\gamma}^{(GLM)}_l}
\] for all \(0 \leq j \leq J\). (4.19)

Remarks 4.4:

- Note the superscripts MLE and GLM, which are used to differentiate between the two normalizations, one natural to maximum likelihood for claims reserving and the other more practical for GLM modelling purposes.

- In multiplicative models like the overdispersed Poisson model it is natural to use the log-link \(g(\cdot) = \log(\cdot)\) as the link function since the systematic effects are additive on the scale given by the log-link function. Moreover, the log-link is the so-called canonical link function for the (overdispersed) Poisson distribution that has convenient mathematical and statistical properties; see e.g. McCullagh and Nelder (1989) or Fahrmeir and Tutz (2001).

- In the next section, relationship (4.19) will be crucial to incorporate our results from GLM theory in the derivation of an estimate of the conditional MSEP in the BF method.

- From GLM theory it is well-known that the MLE (4.16) is asymptotically multivariate normally distributed with covariance matrix \(\text{Cov}(\hat{\gamma}, \hat{\gamma})\) estimated by the inverse of the Fisher information matrix (denoted by \(H^{-1}(\hat{\gamma})\)), which is a standard output in all GLM software packages; see e.g. Panjer (2006) or Fahrmeir and Tutz (2001).

5 MSEP in the BF Method using GLM

In this section we quantify the uncertainty in the estimation of the ultimate claims \(C_{i,J}\) and \(\sum_{i=1}^{I} C_{i,J}\) by the estimators \(\hat{C}_{i,J}^{BF}\) and \(\sum_{i=1}^{I} \hat{C}_{i,J}^{BF}\), respectively, given
the observations $D_I$. More precisely, our goal is to derive an estimate of the conditional MSEP for single accident years $1 \leq i \leq I$,

$$
mse_{C_{i,J}|D_I} \left( \hat{C}_{i,J}^{BF} \right) = E \left[ \left( C_{i,J} - \hat{C}_{i,J}^{BF} \right)^2 \bigg| D_I \right],
$$

as well as an estimate of the conditional MSEP for aggregated accident years

$$
mse_{\sum_{i=1}^I C_{i,J}|D_I} \left( \sum_{i=1}^I \hat{C}_{i,J}^{BF} \right) = E \left[ \left( \sum_{i=1}^I C_{i,J} - \sum_{i=1}^I \hat{C}_{i,J}^{BF} \right)^2 \bigg| D_I \right].
$$

This is described in the next sections.

### 5.1 MSEP in the BF Method, Single Accident Year

We choose $1 \leq i \leq I$. Since the incremental claims $X_{i,j}$ are independent, the conditional MSEP (5.1) can be decoupled in the following way:

$$
E \left[ \left( \sum_{j=I-i+1}^J X_{i,j} - (1 - \hat{\beta}_{I-i}^{(CL)}) \hat{\mu}_i \right)^2 \bigg| D_I \right] = \sum_{j=I-i+1}^J \text{Var}(X_{i,j}) + E \left[ \left( \sum_{j=I-i+1}^J E[X_{i,j}] - (1 - \hat{\beta}_{I-i}^{(CL)}) \hat{\mu}_i \right)^2 \bigg| D_I \right]
$$

$$
+ 2E \left[ \left( \sum_{j=I-i+1}^J (X_{i,j} - E[X_{i,j}]) \right) \left( \sum_{j=I-i+1}^J E[X_{i,j}] - (1 - \hat{\beta}_{I-i}^{(CL)}) \hat{\mu}_i \right) \bigg| D_I \right].
$$

Note that $\hat{\mu}_i$ is independent of $X_{k,j}$ for all $k,j$, that $\hat{\beta}_{I-i}^{(CL)}$ is $D_I$-measurable and that $E[\hat{\mu}_i] = \mu_i$; see e.g. (3.11) and Model Assumptions 4.1. Therefore, the last term in the above equality disappears and we get

$$
mse_{C_{i,J}|D_I} \left( \hat{C}_{i,J}^{BF} \right) = \sum_{j=I-i+1}^J \text{Var}(X_{i,j}) + \left(1 - \hat{\beta}_{I-i}^{(CL)}\right)^2 \text{Var}(\hat{\mu}_i)
$$

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

Hence, the three terms on the right-hand side of (5.5) need to be estimated in order to get an appropriate estimate for the conditional MSEP in the BF method. The first term as a (conditional) process variance originates from the stochastic movement
of $X_{i,j}$. The second and third term on the right-hand side of (5.5) constitute the (conditional) estimation error which reflects the uncertainty in the prior estimate $\hat{\mu}_i$ and the MLEs $\hat{\gamma}_j^{(MLE)}$, respectively.

**Process Variance**

For the estimation of the (conditional) process variance, Model Assumptions 4.1 motivates the following estimator:

$$\text{Var}(X_{i,j}) = \hat{\phi} \hat{\mu}_i \sum_{j=I-i+1}^{J} \hat{\gamma}_j^{(MLE)} = \hat{\phi} \hat{\mu}_i \left(1 - \hat{\beta}_I^{(CL)}\right),$$  

(5.6)

where $\hat{\phi}$ is an estimate of the dispersion parameter $\phi$. Within the framework of GLM we use different types of residuals (Pearson, deviance, Anscombe, etc.) to estimate $\phi$; see e.g. McCullagh and Nelder (1989) or Fahrmeir and Tutz (2001). In the following we will use the Pearson residuals defined by

$$\hat{R}_{i,j}^{(p)} = \frac{X_{i,j} - \hat{m}_{i,j}}{\sqrt{\hat{m}_{i,j}}},$$  

(5.7)

where $\hat{m}_{i,j}$ is the GLM estimate of $m_{i,j}$ given by

$$\hat{m}_{i,j} = \hat{\mu}_i^{(GLM)} \hat{\gamma}_j^{(GLM)}$$  

(5.8)

for $0 \leq i + j \leq I$. The estimate of the dispersion parameter is then given by

$$\hat{\phi} = \sum_{0\leq i+j\leq I} \left(\hat{R}_{i,j}^{(p)}\right)^2 \frac{1}{N - p},$$  

(5.9)

where

$$N = \text{number of observations } X_{i,j} \text{ in } D_I, \text{ i.e. } N = |D_I|, \hspace{1cm} (5.10)$$

$$p = \text{number of estimated parameters, i.e. } p = I + J + 1. \hspace{1cm} (5.11)$$

**Estimation Error**

The (conditional) estimation error is given by the second and third term on the right-hand side of (5.5). This means that we need to quantify the volatility of the prior estimates $\hat{\mu}_i$ and the MLEs $\hat{\gamma}_j^{(MLE)}$ around the true parameters $\mu_i$ and $\gamma_j$, respectively.
Prior estimate $\hat{\mu}_i$: The second term

$$\left(1 - \hat{\beta}_{j-i}^{(CL)}\right)^2 \text{Var}(\hat{\mu}_i)$$

(5.12)

quantifies the uncertainty in the prior estimate $\hat{\mu}_i$ of the expected ultimate claim $E[C_{i,j}]$. Since $\hat{\mu}_i$ is determined exogenously, see e.g. Remarks 4.2, this can generally only be done using external data like market statistics and expert opinion. The regulator, for example, provides an estimate for the coefficient of variation of $\hat{\mu}_i$, denoted by $\text{Vco}(\hat{\mu}_i)$, that quantifies how good the exogenous estimator $\hat{\mu}_i$ is. Statistical estimates based on impact studies for the determination of estimates $\text{Vco}(\hat{\mu}_i)$ exist, for example, in the context of the Swiss Solvency Test (2006). These studies suggest that 5% to 10% is a reasonable range for $\text{Vco}(\hat{\mu}_i)$. Hence the term (5.12) is estimated by

$$\left(1 - \hat{\beta}_{j-i}^{(CL)}\right)^2 \text{Var}(\hat{\mu}_i) = \left(1 - \hat{\beta}_{j-i}^{(CL)}\right)^2 \hat{\mu}_i^2 \text{Vco}(\hat{\mu}_i)^2.$$  

(5.13)

Note that an appropriate choice for $\text{Vco}(\hat{\mu}_i)$ is crucial for a meaningful analysis. This choice is closely related to a Bayesian setup where one chooses an appropriate prior distribution for $\hat{\mu}_i$; see e.g. Mack (2000). Of course, the choice of this prior distribution and/or its coefficient of variation depends on the internal processes of the company. Ideally, this is determined using market statistics as described above (and similarly as used, for example, in the context of modelling operational risk, see Lambrigger et al. (2007)). Unfortunately, in many cases there are no market statistics available and one tries to adjust the priors using internal data. However, this approach contradicts the BF method if we interpret it in the strict sense described by Mack (2006), since the choice of the prior $\hat{\mu}_i$ should be independent from the observations $X_{k,j}$. We would like to motivate further research into this direction, i.e. (1) finding appropriate priors and (2) describe the internal processes as they are used in practice. This could lead to a new theory and method using Kalman filters, see e.g. Chapter 9 in Bühlmann and Gisler (2005) and Chapter 10 in Taylor (2000), to describe loss ratio prediction based on observations of past accident years. One then immediately loses the independence assumptions and the conditional MSEP no longer decouples in a nice way.

**MLEs** $\hat{\gamma}_j^{(MLE)}$: The estimation of the third term on the right-hand side of (5.5) requires more work. We have to study the fluctuations of the MLEs $\hat{\gamma}_j^{(MLE)}$ around the true parameters $\gamma_j$

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

(5.14)
Neglecting that MLEs have a possible bias term we estimate (5.14) by

\[
\text{Var} \left( \sum_{j=I-i+1}^{J} \hat{\gamma}_{j}^{(MLE)} \right) = \sum_{j,k=I-i+1}^{J} \text{Cov} \left( \hat{\gamma}_{j}^{(MLE)} \hat{\gamma}_{k}^{(MLE)} \right)
\]

\[
= \sum_{j,k=I-i+1}^{J} \text{Cov} \left( \frac{\hat{\gamma}_{j}^{(GLM)}}{\sum_{l=0}^{J} \hat{\gamma}_{l}^{(GLM)}}, \frac{\hat{\gamma}_{k}^{(GLM)}}{\sum_{l=0}^{J} \hat{\gamma}_{l}^{(GLM)}} \right)
\]

\[
= \sum_{j,k=I-i+1}^{J} \text{Cov} \left( \frac{1}{1 + \sum_{l \neq j} \hat{\gamma}_{l}^{(GLM)}}, \frac{1}{1 + \sum_{l \neq k} \hat{\gamma}_{l}^{(GLM)}} \right)
\]

(5.15)

see e.g. (4.19). Here, we restricted our probability space such that a solution to the above equations exist. We define

\[
\Delta_{j} = \sum_{l=0}^{J} \hat{\gamma}_{l}^{(GLM)} \quad \text{and} \quad \delta_{j} = E[\Delta_{j}]
\]

(5.16)

for \( I-i+1 \leq j \leq J \). Hence we need to calculate

\[
\text{Cov} \left( \frac{1}{1 + \Delta_{j}} \frac{1}{1 + \Delta_{k}} \right)
\]

(5.17)

for \( I-i+1 \leq j, k \leq J \). We first do a Taylor approximation around \( \delta_{j} \). To this end we define the function

\[
f(x) = \frac{1}{1 + x} \quad \text{with} \quad f'(x) = -\frac{1}{(1+x)^{2}}
\]

(5.18)

and obtain for the first order Taylor approximation around \( \delta_{j} \)

\[
f(x) \approx f(\delta_{j}) + f'(\delta_{j})(x - \delta_{j}) = \frac{1}{1 + \delta_{j}} - \frac{1}{(1 + \delta_{j})^{2}}(x - \delta_{j}).
\]

(5.19)

This implies that

\[
\text{Cov} \left( \frac{1}{1 + \Delta_{j}} \frac{1}{1 + \Delta_{k}} \right) \approx \frac{1}{(1 + \delta_{j})^{2}} \frac{1}{(1 + \delta_{k})^{2}} \text{Cov}(\Delta_{j}, \Delta_{k})
\]

\[
= \frac{1}{(1 + \delta_{j})^{2}} \frac{1}{(1 + \delta_{k})^{2}} \sum_{l \neq j} \sum_{m \neq k} \text{Cov} \left( \frac{\hat{\gamma}_{l}^{(GLM)}}{\hat{\gamma}_{j}^{(GLM)}}, \frac{\hat{\gamma}_{m}^{(GLM)}}{\hat{\gamma}_{k}^{(GLM)}} \right)
\]

(5.20)
It only remains to calculate the covariance terms on the right-hand side of (5.20). We use the following linearization (Taylor approximation for the exponential function)

\[
\frac{\hat{\gamma}_j^{(GLM)}}{\tilde{\gamma}_j^{(GLM)}} = \exp \left( \log \left( \frac{\hat{\gamma}_j^{(GLM)}}{\tilde{\gamma}_j^{(GLM)}} \right) \right) = \frac{\gamma_j}{\tilde{\gamma}_j} \exp \left( \log \left( \frac{\hat{\gamma}_j^{(GLM)}}{\tilde{\gamma}_j^{(GLM)}} \right) - \log \left( \frac{\gamma_j}{\tilde{\gamma}_j} \right) \right) \approx \frac{\gamma_j}{\tilde{\gamma}_j} \left( 1 + \log \left( \frac{\hat{\gamma}_j^{(GLM)}}{\tilde{\gamma}_j^{(GLM)}} \right) - \log \left( \frac{\gamma_j}{\tilde{\gamma}_j} \right) \right). \tag{5.21}
\]

Using (5.21) we obtain for the covariance terms on the right-hand side of (5.20)

\[
\text{Cov} \left( \frac{\hat{\gamma}_j^{(GLM)}}{\tilde{\gamma}_j^{(GLM)}}, \frac{\hat{\gamma}_m^{(GLM)}}{\tilde{\gamma}_k^{(GLM)}} \right) \approx \frac{\gamma_j \gamma_m}{\gamma_j \gamma_k} \text{Cov} \left( \log \left( \frac{\hat{\gamma}_j^{(GLM)}}{\tilde{\gamma}_j^{(GLM)}} \right), \log \left( \frac{\hat{\gamma}_m^{(GLM)}}{\tilde{\gamma}_k^{(GLM)}} \right) \right). \tag{5.22}
\]

Now, we define the slightly modified design matrices

\[
\tilde{\Gamma}_j = (0, \ldots, 0, e_{I+j+1}, 0, \ldots, 0)' \quad \text{for all } 0 \leq j \leq J, \tag{5.23}
\]

which implies that, see (4.16) and (4.18),

\[
\log \left( \frac{\hat{\gamma}_j^{(GLM)}}{\tilde{\gamma}_j^{(GLM)}} \right) = \tilde{\Gamma}_j \hat{b} \quad \text{for all } 0 \leq j \leq J. \tag{5.24}
\]

Hence

\[
\text{Cov} \left( \log \left( \frac{\hat{\gamma}_j^{(GLM)}}{\tilde{\gamma}_j^{(GLM)}} \right), \log \left( \frac{\hat{\gamma}_m^{(GLM)}}{\tilde{\gamma}_k^{(GLM)}} \right) \right) = \left( \tilde{\Gamma}_j - \tilde{\Gamma}_j \right) \text{Cov} \left( \hat{b}, \hat{b} \right) \left( \tilde{\Gamma}_m - \tilde{\Gamma}_k \right)' . \tag{5.25}
\]

Using the inverse of the Fisher information matrix $H(\hat{b})$ for the estimation of the covariance term $\text{Cov} \left( \hat{b}, \hat{b} \right)$ we obtain for (5.15)

\[
\text{Var} \left( \sum_{j=I-i+1}^{J} \tilde{\gamma}_j^{(MLE)} \right) \approx \sum_{j,k=I-i+1}^{J} \frac{1}{(1 + \delta_j)^2} \left( \frac{1}{1 + \delta_k} \right)^2 \times \sum_{l} \sum_{m} \frac{\gamma_l \gamma_m}{\gamma_j \gamma_k} \left( \tilde{\Gamma}_l - \tilde{\Gamma}_j \right) H(\hat{b})^{-1} \left( \tilde{\Gamma}_m - \tilde{\Gamma}_k \right)' ; \tag{5.26}
\]

see e.g. Remarks 4.4. Hence we define the estimator

\[
\text{Var} \left( \sum_{j=I-i+1}^{J} \tilde{\gamma}_j^{(MLE)} \right) = \sum_{j,k=I-i+1}^{J} \frac{1}{(1 + \delta_j)^2} \left( \frac{1}{1 + \delta_k} \right)^2 \times \sum_{l} \sum_{m} \frac{\hat{\gamma}_l^{(GLM)} \hat{\gamma}_m^{(GLM)}}{\tilde{\gamma}_j^{(GLM)} \tilde{\gamma}_k^{(GLM)}} \left( \tilde{\Gamma}_l - \tilde{\Gamma}_j \right) H(\hat{b})^{-1} \left( \tilde{\Gamma}_m - \tilde{\Gamma}_k \right)' . \tag{5.27}
\]

69
where we set \( \hat{\delta}_j = \sum_{l \neq j}^{J} \hat{\gamma}_l^{(GLM)} \); see e.g. (5.16). This can be rewritten in matrix notation, we define the parameter \( c_{j,k} = \hat{\gamma}_j^{(GLM)} \hat{\gamma}_k^{(GLM)} \left( \hat{\mu}_0^{(MLE)} \right)^{-4} \) and \( \hat{\gamma} = (0, \ldots, 0, \hat{\gamma}_0^{(GLM)}, \ldots, \hat{\gamma}_J^{(GLM)})' \). Furthermore, we define

\[
\Psi_{j,k} = c_{j,k} \left[ \hat{\gamma}' \hat{\mu}^{(MLE)}_i + \left( \hat{\mu}_0^{(MLE)} \right)^2 \hat{\gamma}' \hat{\mu}^{(MLE)}_i \right].
\]

Then we obtain

\[
\text{Var} \left( \sum_{j=I-i+1}^{J} \hat{\gamma}_j^{(MLE)} \right) = \sum_{j,k=I-i+1}^{J} \Psi_{j,k}. \tag{5.29}
\]

Putting the three estimates (5.6), (5.13) and (5.29) together we obtain the following estimator for the (conditional) MSEP for a single accident year:

**Estimator 5.1 (MSEP for the BF method, single accident year)**

*Under Model Assumptions 4.1 an estimator for the (conditional) MSEP for a single accident year \( I - J + 1 \leq i \leq I \) is given by*

\[
\hat{\text{msep}}_{C_i,J|D_I} \left( \hat{C}_{i,J}^{BF} \right) = \hat{\phi} \left( 1 - \hat{\beta}_{i-i}^{(CL)} \right) \hat{\mu}_i + \left( 1 - \hat{\beta}_{i-i}^{(CL)} \right)^2 \hat{\mu}_i^2 \text{Var} \left( \hat{\mu}_i \right),
\]

\[
+ \hat{\mu}_i^2 \sum_{j,k=I-i+1}^{J} \Psi_{j,k}, \tag{5.30}
\]

*see (5.6), (5.13) and (5.29).*

### 5.2 MSEP in the BF Method, Aggregated Accident Years

In this section we derive an estimate for the conditional MSEP for aggregated accident years (5.3). We start by considering two different accident years \( i < l \),

\[
\text{msep}_{C_{i,J} + C_{l,J}|D_I} \left( \hat{C}_{i,J}^{BF} + \hat{C}_{l,J}^{BF} \right) = E \left[ \left( C_{i,J} + C_{l,J} - \hat{C}_{i,J}^{BF} - \hat{C}_{l,J}^{BF} \right)^2 \right| D_I]. \tag{5.31}
\]
By the usual decomposition we find
\[
\text{msep}_{C_{i,j} + C_{l,j}}|D_I \left( \hat{C}_{i,j}^{BF} + \hat{C}_{l,j}^{BF} \right) = \text{msep}_{C_{i,j}}|D_I \left( \hat{C}_{i,j}^{BF} \right) + \text{msep}_{C_{l,j}}|D_I \left( \hat{C}_{l,j}^{BF} \right) + 2 \mu_i \mu_l \sum_{1 \leq i < l \leq I} \hat{\gamma}_i \hat{\gamma}_l \Upsilon_{i,l}. \tag{5.32}
\]
That is, we need to give an estimate for the term on the right-hand side of (5.32). Analogously to (5.14), we have to study the fluctuations of the MLEs \( \hat{\gamma}_j^{(MLE)} \) around the true parameters \( \gamma_j \). Again, neglecting the possible bias of the MLEs, we estimate this term by
\[
\text{Cov} \left( \sum_{j=I-l-i+1}^{I} \hat{\gamma}_j^{(MLE)}, \sum_{k=I-l+1}^{J} \hat{\gamma}_k^{(MLE)} \right) = \sum_{j=I-l-i+1}^{I} \sum_{k=I-l+1}^{J} \text{Cov} \left( \hat{\gamma}_j^{(MLE)}, \hat{\gamma}_k^{(MLE)} \right) \approx \sum_{j=I-l-i+1}^{I} \sum_{k=I-l+1}^{J} \frac{1}{(1 + \delta_j)^2 (1 + \delta_k)^2} \times \sum_n \sum_m \frac{\gamma_n \gamma_m}{\gamma_j \gamma_k} \left( \hat{\Gamma}_n - \hat{\Gamma}_j \right) H(b)^{-1} \left( \hat{\Gamma}_m - \hat{\Gamma}_k \right)' \; \tag{5.33}
\]
see e.g. (5.26). Again, as with the single accident year case, we restrict our probability space such that the above covariance exists. This motivates the following estimator for the covariance term (5.33)
\[
\Upsilon_{i,l} = \text{Cov} \left( \sum_{j=I-l-i+1}^{I} \hat{\gamma}_j^{(MLE)}, \sum_{k=I-l+1}^{J} \hat{\gamma}_k^{(MLE)} \right) = \sum_{j=I-l-i+1}^{I} \sum_{k=I-l+1}^{J} \Psi_{j,k}. \tag{5.34}
\]
This leads to the following estimator for the (conditional) MSEP for aggregated accident years:

**Estimator 5.2 (MSEP for the BF method, aggregated acc. years)**

*Under Model Assumptions 4.1 an estimator for the (conditional) MSEP for aggregated accident years is given by*
\[
\text{msep}_{\sum_{i=1}^{I} C_{i,j}}|D_I \left( \sum_{i=1}^{I} \hat{C}_{i,j}^{BF} \right) = \sum_{i=1}^{I} \text{msep}_{C_{i,j}}|D_I \left( \hat{C}_{i,j}^{BF} \right) + 2 \sum_{1 \leq i < l \leq I} \hat{\mu}_i \hat{\mu}_l \Upsilon_{i,l}. \tag{5.35}
\]
The natural extension of this work would be to obtain some properties of Estimators 5.1 and 5.2, for example, asymptotic behaviour. However, we omit this presently, because it would go beyond the context of this work.

6 Example and Simulation

In this section we state an example and a simulation.

6.1 Example of MSEP in the BF Method

Using the incremental claims data provided in Table 1, which are scaled incremental payments from property business, we calculate the BF reserves and the estimators for the conditional MSEP as derived for Estimator 5.1 (single accident year) and Estimator 5.2 (aggregated accident years). Furthermore, we compare these results with the estimators for the conditional MSEP of the overdispersed Poisson method/Poisson GLM method, with reserves matching those of the CL method. However, before we can estimate the conditional MSEP for the BF method, we must first specify the assumptions regarding ultimate claim estimates \( \hat{\mu}_i \).

We assume the prior estimates \( \hat{\mu}_i \) for the expected ultimate claims \( E[C_{i,j}] \) are given by Table 2. Furthermore, we have to specify the uncertainty in these estimates. To this end we assume for the coefficient of variation a flat rate of 5% for all accident years, i.e.

\[
\hat{\text{co}}(\hat{\mu}_i) = 0.05 \quad \text{for} \quad 1 \leq i \leq I. \quad (6.1)
\]

This assumption, as previously stated, has been shown to be reasonable in the Swiss Solvency Test (2006). To view the effect of changing this input see Table 5 where we have presented the BF MSEP as a function of the coefficient of variation for the expected ultimate estimates. For the dispersion parameter \( \phi \) we obtain the estimate \( \hat{\phi} = 14,714; \) (6.2)

see e.g. (5.9).

<table>
<thead>
<tr>
<th>( i ) ( j )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>621</td>
<td>372</td>
<td>896</td>
<td>208</td>
<td>207</td>
<td>62</td>
<td>66</td>
<td>15</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>1</td>
<td>6347</td>
<td>3246</td>
<td>723</td>
<td>152</td>
<td>68</td>
<td>37</td>
<td>53</td>
<td>11</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6269</td>
<td>2976</td>
<td>847</td>
<td>263</td>
<td>153</td>
<td>65</td>
<td>54</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>5863</td>
<td>2683</td>
<td>723</td>
<td>191</td>
<td>133</td>
<td>88</td>
<td>43</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>5779</td>
<td>2745</td>
<td>654</td>
<td>274</td>
<td>230</td>
<td>105</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>6185</td>
<td>2828</td>
<td>573</td>
<td>245</td>
<td>105</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>5600</td>
<td>2893</td>
<td>563</td>
<td>226</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>5288</td>
<td>2440</td>
<td>528</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>5291</td>
<td>2558</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>5676</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 1: Observed incremental claims \( X_{i,j} \) given in '000s.
Now we can calculate the estimators for the conditional MSEPs as given by Estimator 5.1 and 5.2 and obtain the values given in Table 3. If we compare these results to the results for the overdispersed Poisson GLM method in Table 4 we observe the following: a) The claims reserve estimates are rather large for the BF method, reflecting the conservative prior estimates $\hat{\mu}_i$ for the expected ultimate claims given by Table 2. b) As a consequence of a) the process standard deviations of the BF method are larger than the ones of the overdispersed Poisson GLM method. c) The totals of prior and parameter uncertainty in the estimators $\hat{\mu}_i$ and $\hat{\beta}_{CL}^j$ are slightly higher than the estimation errors in the overdispersed Poisson GLM method. As a consequence, the conditional MSEPs of the BF method are larger than the ones of the overdispersed Poisson GLM method. However, due to the conservative estimation of the claims reserves, the coefficients of variation are smaller than the ones of the overdispersed Poisson GLM model.

Table 2: Prior estimates for the expected ultimate claims.

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\mu_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11,364,606</td>
</tr>
<tr>
<td>2</td>
<td>10,962,965</td>
</tr>
<tr>
<td>3</td>
<td>10,616,762</td>
</tr>
<tr>
<td>4</td>
<td>11,044,881</td>
</tr>
<tr>
<td>5</td>
<td>11,480,700</td>
</tr>
<tr>
<td>6</td>
<td>11,413,572</td>
</tr>
<tr>
<td>7</td>
<td>11,126,527</td>
</tr>
<tr>
<td>8</td>
<td>10,986,548</td>
</tr>
<tr>
<td>9</td>
<td>11,618,437</td>
</tr>
</tbody>
</table>

Table 3: Results for the stochastic BF method.

<table>
<thead>
<tr>
<th>$i$</th>
<th>BF reserves</th>
<th>process std. dev.</th>
<th>prior std. dev.</th>
<th>parameter $\beta$ std. dev.</th>
<th>prior and parameter std. dev.</th>
<th>msep$^{\frac{1}{2}}$</th>
<th>Vco</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18,129,018</td>
<td>15,401</td>
<td>806</td>
<td>15,589</td>
<td>15,660</td>
<td>21,883</td>
<td>135.8%</td>
</tr>
<tr>
<td>2</td>
<td>26,998,449</td>
<td>19,931</td>
<td>1,350</td>
<td>17,803</td>
<td>17,624</td>
<td>26,606</td>
<td>98.5%</td>
</tr>
<tr>
<td>3</td>
<td>37,575,117</td>
<td>23,514</td>
<td>1,879</td>
<td>18,545</td>
<td>18,639</td>
<td>30,005</td>
<td>79.9%</td>
</tr>
<tr>
<td>4</td>
<td>95,434,866</td>
<td>37,473</td>
<td>4,772</td>
<td>24,168</td>
<td>24,635</td>
<td>44,845</td>
<td>47.0%</td>
</tr>
<tr>
<td>5</td>
<td>178,023,578</td>
<td>51,181</td>
<td>8,901</td>
<td>29,600</td>
<td>30,910</td>
<td>59,790</td>
<td>33.6%</td>
</tr>
<tr>
<td>6</td>
<td>341,305,272</td>
<td>70,866</td>
<td>17,065</td>
<td>35,750</td>
<td>39,614</td>
<td>81,187</td>
<td>23.8%</td>
</tr>
<tr>
<td>7</td>
<td>574,089,684</td>
<td>91,909</td>
<td>28,704</td>
<td>41,221</td>
<td>50,231</td>
<td>104,739</td>
<td>18.2%</td>
</tr>
<tr>
<td>8</td>
<td>1,318,645,410</td>
<td>139,294</td>
<td>65,932</td>
<td>53,175</td>
<td>84,703</td>
<td>163,025</td>
<td>12.4%</td>
</tr>
<tr>
<td>9</td>
<td>4,768,385,244</td>
<td>264,882</td>
<td>238,419</td>
<td>75,853</td>
<td>250,195</td>
<td>364,362</td>
<td>7.6%</td>
</tr>
</tbody>
</table>

Table 3: Results for the stochastic BF method.

6.2 Simulation

In the derivation of Estimators 5.1 and 5.2 we used various approximations. We test the strength of our results by simulation. Assuming the incremental claims $X_{i,j}$ are overdispersed Poisson distributed with parameters $\hat{\gamma}_j^{(MLE)}$ and $\hat{\mu}_i^{(MLE)}$ and dispersion parameter (6.2), see e.g. Model Assumptions 4.1, we generate 10,000 run-off
Table 4: Results for the overdispersed Poisson method.

<table>
<thead>
<tr>
<th>$i$</th>
<th>reserves std. dev.</th>
<th>process parameter std. dev.</th>
<th>msep$^{1/2}$</th>
<th>$V_{co}$ std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15,125</td>
<td>14,918</td>
<td>20,882</td>
<td>138.1%</td>
</tr>
<tr>
<td>2</td>
<td>26,257</td>
<td>19,656</td>
<td>26,093</td>
<td>99.4%</td>
</tr>
<tr>
<td>3</td>
<td>34,538</td>
<td>22,543</td>
<td>28,341</td>
<td>82.0%</td>
</tr>
<tr>
<td>4</td>
<td>85,301</td>
<td>35,428</td>
<td>41,724</td>
<td>48.9%</td>
</tr>
<tr>
<td>5</td>
<td>156,493</td>
<td>47,986</td>
<td>55,113</td>
<td>35.2%</td>
</tr>
<tr>
<td>6</td>
<td>298,120</td>
<td>64,885</td>
<td>72,761</td>
<td>25.4%</td>
</tr>
<tr>
<td>7</td>
<td>449,166</td>
<td>81,296</td>
<td>90,139</td>
<td>20.1%</td>
</tr>
<tr>
<td>8</td>
<td>1,043,242</td>
<td>123,897</td>
<td>140,462</td>
<td>13.5%</td>
</tr>
<tr>
<td>9</td>
<td>3,950,816</td>
<td>241,107</td>
<td>331,605</td>
<td>8.4%</td>
</tr>
<tr>
<td>total</td>
<td>6,047,059</td>
<td>298,290</td>
<td>429,891</td>
<td>7.1%</td>
</tr>
</tbody>
</table>

Table 5: The BF msep$^{1/2}$ as a function of $\tilde{V}_{co}(\hat{\mu}_i)$.

<table>
<thead>
<tr>
<th>$\tilde{V}_{co}(\hat{\mu}_i)$</th>
<th>msep$^{1/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>400,428</td>
</tr>
<tr>
<td>0.01</td>
<td>403,534</td>
</tr>
<tr>
<td>0.02</td>
<td>412,710</td>
</tr>
<tr>
<td>0.03</td>
<td>427,565</td>
</tr>
<tr>
<td>0.04</td>
<td>447,535</td>
</tr>
<tr>
<td>0.05</td>
<td>471,971</td>
</tr>
<tr>
<td>0.06</td>
<td>500,219</td>
</tr>
<tr>
<td>0.07</td>
<td>531,671</td>
</tr>
<tr>
<td>0.08</td>
<td>565,794</td>
</tr>
<tr>
<td>0.09</td>
<td>602,133</td>
</tr>
<tr>
<td>0.10</td>
<td>640,311</td>
</tr>
</tbody>
</table>

Table 6: Simulation results comparing empirical and estimated standard deviations of the cumulative payout pattern.

<table>
<thead>
<tr>
<th>$j$</th>
<th>$\hat{\beta}^{(CL)}_j$</th>
<th>empirical $s.d.(\hat{\beta}^{(CL)}_j)$</th>
<th>estimated $s.d.(\hat{\beta}^{(CL)}_j)$ using (5.27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>58.96%</td>
<td>0.654%</td>
<td>0.653%</td>
</tr>
<tr>
<td>1</td>
<td>88.00%</td>
<td>0.486%</td>
<td>0.484%</td>
</tr>
<tr>
<td>2</td>
<td>94.84%</td>
<td>0.375%</td>
<td>0.370%</td>
</tr>
<tr>
<td>3</td>
<td>97.01%</td>
<td>0.317%</td>
<td>0.313%</td>
</tr>
<tr>
<td>4</td>
<td>98.45%</td>
<td>0.260%</td>
<td>0.258%</td>
</tr>
<tr>
<td>5</td>
<td>99.14%</td>
<td>0.220%</td>
<td>0.219%</td>
</tr>
<tr>
<td>6</td>
<td>99.65%</td>
<td>0.177%</td>
<td>0.175%</td>
</tr>
<tr>
<td>7</td>
<td>99.75%</td>
<td>0.162%</td>
<td>0.160%</td>
</tr>
<tr>
<td>8</td>
<td>99.86%</td>
<td>0.138%</td>
<td>0.137%</td>
</tr>
<tr>
<td>9</td>
<td>100.00%</td>
<td>0.000%</td>
<td>0.000%</td>
</tr>
</tbody>
</table>

Table 6 provides the resulting empirical payout pattern, the empirical variance and the estimated variance for each development year $j$. From these results it is clear that the approximation of the variance of the cumulative payout pattern is very
close to the empirical value. This means that the approximation given in (5.26) performs very well for a typical payout pattern in practice.

**Acknowledgements**

The authors would like to thank Prof. Paul Embrechts for helpful discussion related to the paper. Furthermore, the first author is grateful to ACE Limited for providing the financial support that made this research a reality.

**Bibliography**


Daniel H. Alai, Michael Merz and Mario V. Wüthrich (2009).

Prediction uncertainty in the Bornhuetter-Ferguson claims reserving method: revisited.
Submitted.
Prediction Uncertainty in the Bornhuetter-Ferguson Claims Reserving Method: Revisited

Daniel H. Alai    Michael Merz    Mario V. Wüthrich
Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland

Abstract

We revisit the stochastic model of Alai et al. (2009) for the Bornhuetter-Ferguson claims reserving method, Bornhuetter and Ferguson (1972). We derive an estimator of its conditional mean square error of prediction (MSEP) using an approach that is based on generalized linear models and maximum likelihood estimators for the model parameters. This approach leads to simple formulas, which can easily be implemented in a spreadsheet.

Keywords: Claims Reserving, Bornhuetter-Ferguson, Overdispersed Poisson Distribution, Chain Ladder Method, Generalized Linear Models, Fisher Information Matrix, Conditional Mean Square Error of Prediction
1 Introduction

The prediction uncertainty in the Bornhuetter-Ferguson (BF) claims reserving method, Bornhuetter and Ferguson (1972), has recently been studied by several authors; see e.g. Mack (2008), Verrall (2004) and Alai et al. (2009). We revisit the model studied in Alai et al. (2009). In the present paper we provide a different method of approximating the mean square error of prediction (MSEP), which substantially simplifies the formulas.

Alai et al. (2009) maintain that in practice the chain ladder (CL) development pattern is used for calculating the BF reserves, and hence incorporate this into their model assumptions. This is done by assuming the data to be overdispersed Poisson distributed. This allows one to recreate the CL estimate of the development pattern; a result dating back to Hachemeister and Stanard (1975) and Mack (1991). This is different from the approach taken in Mack (2008), but closer to the implementation of practitioners.

Organization of the paper. In Section 2 we provide the notation, data structure as well as the model considerations. In Section 3 we give a short review of the BF method. In Section 4 we give a simplified estimation procedure for the conditional MSEP in the BF method. Finally, in Section 5 we revisit the case study presented in Alai et al. (2009) and compare our results with Mack (2008).

2 Data and Model

2.1 Setup

Let $X_{i,j}$ denote the incremental claims of accident year $i \in \{0, 1, \ldots, I\}$ and development year $j \in \{0, 1, \ldots, J\}$. We assume the data is given by a claims development triangle, i.e., $I = J$, and that after $J$ development periods all claims are settled. At time $I$, we have observations $D_I = \{X_{i,j}, i + j \leq I\}$. We are interested in predicting the corresponding lower triangle $\{X_{i,j}, i + j > I, i \leq I\}$. Furthermore, define $C_{i,j}$ to be the cumulative claims of accident year $i$ up to development year $j$. Hence,

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

2.2 Model Considerations

We adopt the overdispersed Poisson model presented in Alai et al. (2009).
Model Assumptions 2.1 (Overdispersed Poisson Model)

- The incremental claims \( X_{i,j} \) are independent overdispersed Poisson distributed and there exist positive parameters \( \gamma_0, \ldots, \gamma_I, \mu_0, \ldots, \mu_I \) and \( \phi > 0 \) with

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

and \( \sum_{j=0}^{I} \gamma_j = 1 \).

- \( \hat{\nu}_k \) are independent random variables that are unbiased estimators of the expected ultimate claim \( \mu_k = E[C_{k,I}] \) for all \( k \in \{0, \ldots, I\} \).

- \( X_{i,j} \) and \( \hat{\nu}_k \) are independent for all \( i, j, k \).

Remarks 2.2:

- The exogenous estimator \( \hat{\nu}_k \) is a prior estimate of the expected ultimate claims \( E[C_{k,I}] \), which is used for the BF method; see also Section 2 in Mack (2008).

- For MSEP considerations, an estimate of the uncertainty of the \( \hat{\nu}_k \) is required. Below, we assume that a prior variance estimate \( \hat{\text{Var}}(\hat{\nu}_i) \) is given exogenously.

- For additional model interpretations we refer to Alai et al. (2009).

2.3 Maximum Likelihood Estimators

Under Model Assumptions 2.1 the log-likelihood function for \( D_I \) is given by

\[
l_{D_I}(\mu_i, \gamma_j, \phi) = \sum_{i+j \leq I} \left( \frac{1}{\phi}(X_{i,j} \log(\mu_i \gamma_j) - \mu_i \gamma_j) + \log c(X_{i,j}; \phi) \right) \\
+ \left( \frac{1}{\phi}(X_{0,I} \log \left[ \mu_0 \left( 1 - \sum_{n=0}^{I-1} \gamma_n \right) \right] - \mu_0 \left( 1 - \sum_{n=0}^{I-1} \gamma_n \right)) + \log c(X_{0,I}; \phi) \right),
\]

where \( c(\cdot, \phi) \) is the suitable normalizing function. Notice that the substitution, \( \gamma_I = (1 - \sum_{n=0}^{I-1} \gamma_n) \) has been made in accordance with the constraint provided in Model Assumptions 2.1. The maximum likelihood estimators (MLEs) \( \hat{\mu}_i, \hat{\gamma}_j \) are found by taking the derivates with respect to \( \mu_i, \gamma_j \) and setting the resulting
equations equal to zero. They are given by,

\[ \hat{\mu}_0 = \sum_{j=0}^{I} X_{0,j}, \]
\[ \hat{\mu}_i = \sum_{j=0}^{I-i} X_{i,j}, \quad i \in \{1, \ldots, I\}, \]
\[ \hat{\gamma}_j \left( \sum_{i=1}^{I-j} \hat{\mu}_i + X_{0,I} \frac{1}{1 - \sum_{n=0}^{I-1} \hat{\gamma}_n} \right) = \sum_{i=0}^{I-j} X_{i,j}, \quad j \in \{0, \ldots, I-1\}. \] (1)

Furthermore, we define \( \hat{\gamma}_I = 1 - \sum_{n=0}^{I-1} \hat{\gamma}_n \). The \( \hat{\mu}_i, \hat{\gamma}_j \) can also be calculated with help from the well-known CL factors,

\[ \hat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}}, \]

see e.g. Corollary 2.18 and Remarks 2.19 in Wüthrich and Merz (2008), i.e.

\[ \hat{\gamma}_j = \prod_{k=j}^{I-1} \frac{1}{f_{k}^{j}} \left(1 - \frac{1}{f_{j-1}}\right), \quad \hat{\mu}_i = C_{i,I} \hat{f}_{I-1} \cdots \hat{f}_{j-1}. \] (2)

Although, as is clear in (1), \( \phi \) has no influence on the parameter estimation of \( \mu_i, \gamma_j \), an estimate of \( \phi \) is required to estimate the prediction uncertainty. As done in Alai et al. (2009), we use Pearson residuals to estimate \( \phi \):

\[ \hat{\phi} = \frac{1}{d} \sum_{i+j \leq I} \frac{(X_{i,j} - \hat{m}_{i,j})^2}{\hat{m}_{i,j}}, \] (3)

where \( d = \frac{(I+1)(I+2)}{2} - 2I - 1 \) is the degrees of freedom of the model and \( \hat{m}_{i,j} = \hat{\mu}_i \hat{\gamma}_j \).

### 2.4 Asymptotic Properties of the MLE

In order to quantify the parameter estimation uncertainty \( \hat{\gamma}_j - \gamma_j \) we use the asymptotic MLE property

\[ \sqrt{n}(\hat{\zeta} - \zeta) \xrightarrow{(d)} \mathcal{N}\left(0, H(\zeta, \phi)^{-1}\right), \quad \text{as } n \to \infty, \] (4)

with Fisher information matrix \( H(\zeta, \phi) = (h_{r,s}(\zeta, \phi))_{r,s=1,\ldots,m} \), given by

\[ h_{r,s} = -E_{\zeta} \left[ \frac{\partial^2}{\partial \zeta_r \partial \zeta_s} l_{D}(\zeta, \phi) \right], \]
for $\zeta = (\zeta_1, \ldots, \zeta_{2I+1}) = (\mu_0, \ldots, \mu_I, \gamma_0, \ldots, \gamma_{I-1})$ and $\hat{\zeta}$ the corresponding MLE. Under Model Assumptions 2.1, we obtain for the components of the Fisher information matrix:

$$h_{i+1,i+1} = \frac{\mu_i^{-1}}{\phi} \sum_{j=0}^{I-1} \gamma_j, \quad i \in \{0, \ldots, I\},$$

$$h_{I+2+j,I+2+j} = \frac{\gamma_j^{-1} I^{-j}}{\phi} \sum_{i=0}^{I-1} \mu_i + \frac{\mu_0}{\phi(1 - \sum_{n=0}^{I-1} \gamma_n)}, \quad j \in \{0, \ldots, I - 1\},$$

$$h_{I+2+j,I+2+l} = \frac{\mu_0}{\phi(1 - \sum_{n=0}^{I-1} \gamma_n)}, \quad j, l \in \{0, \ldots, I - 1\}, j \neq l,$$

$$h_{i+1,i+2+j} = \frac{1}{\phi}, \quad i \in \{1, \ldots, I\}, j \in \{0, \ldots, I - i\},$$

$$h_{I+2+j,i+1} = \frac{1}{\phi}, \quad j \in \{0, \ldots, I - 1\}, i \in \{1, \ldots, I - j\}.$$ 

The remaining entries of the $(2I+1) \times (2I+1)$ matrix $H(\zeta, \phi)$ are zero. By replacing the parameters $\zeta$ and $\phi$ by their estimates given in (1) and (3), respectively, we obtain the estimated Fisher information matrix $H(\hat{\zeta}, \hat{\phi})$. The inverse of the estimated Fisher information matrix, $H(\hat{\zeta}, \hat{\phi})^{-1}$, contains, for our purposes, unnecessary information regarding the parameters $\mu_i$. Therefore, we define the $(I+1) \times (I+1)$ matrix

$$\mathcal{G} = (g_{j,l})_{j,l=0 \ldots I},$$

with

$$g_{j,l} = \text{Cov}(\hat{\gamma}_j, \hat{\gamma}_l) = H(\hat{\zeta}, \hat{\phi})_{I+2+j,I+2+l}, \quad j, l \in \{0, \ldots, I - 1\},$$

$$g_{j,I} = g_{I,j} = \text{Cov}(\hat{\gamma}_j, \hat{\gamma}_I) = -\sum_{m=0}^{I-1} H(\hat{\zeta}, \hat{\phi})_{I+2+j,I+2+m}, \quad j \in \{0, \ldots, I - 1\},$$

$$g_{I,I} = \text{Var}(\hat{\gamma}_I) = \sum_{0 \leq m \leq I-1} H(\hat{\zeta}, \hat{\phi})_{I+2+m,I+2+m}^{-1} \quad (5)$$

The first equation of (5) gives an estimator for the covariances between the MLEs $\hat{\gamma}_j$ and $\hat{\gamma}_l$, whereas the last two equations of (5) incorporate the MLE $\hat{\gamma}_I = 1 - \sum_{n=0}^{I-1} \hat{\gamma}_n$.

### 3 The Bornhuetter-Ferguson Method

In practice, the BF predictor, which dates back to Bornhuetter and Ferguson (1972), relies on the data for the development pattern $\gamma_j$ and on external data or expert opinion for the expected ultimate claims $E[C_{i,I}]$. The ultimate claim $C_{i,I}$ of accident year $i$ under Model Assumptions 2.1 using the BF method, given $\mathcal{D}_I$, is predicted by

$$\hat{C}_{i,I}^{BF} = C_{i,I-1} + \hat{\nu}_i \sum_{j > I - i} \hat{\gamma}_j, \quad (6)$$

where $\hat{\gamma}_j$ are the MLEs produced in Section 2.3 and $\hat{\nu}_i$ is an exogenous prior estimator for the expected ultimate claim $E[C_{i,I}]$ introduced in Model Assumptions 2.1.
Note that we define the BF predictor with the CL development pattern $\hat{\gamma}_j$, which is the approach used in practice; see equation (2). A different approach for the estimation of the development pattern $\gamma_j$ is given in Mack (2008), we further discuss this in the case study in Section 5.

4 The MSEP of the Bornhuetter-Ferguson Method

We begin by considering the (conditional) MSEP of the BF predictor $\hat{C}_{i,I}^{BF}$ for single accident years $i \in \{1, \ldots, I\}$. From (5.5) in Alai et al. (2009) we have

$$
\text{mse} p_{C_{i,I}^D}(\hat{C}_{i,I}^{BF}) = E\left[ \left( \hat{C}_{i,I}^{BF} - C_{i,I} \right)^2 | D_I \right] \tag{7}
$$

The first term on the right-hand side of equation (7) is the (conditional) process variance, it represents the stochastic movement of the $X_{i,j}$, the inherent uncertainty from our model assumptions. The latter two terms form the (conditional) estimation error; these terms constitute the uncertainty in the prediction of the prior estimate $\hat{\nu}_i$ and the MLEs $\hat{\gamma}_j$. The first two terms on the right-hand side of equation (7) can be estimated by replacing unknowns with their estimates; see e.g. Sections 5.1.1 and 5.1.2 in Alai et al. (2009). The last term, however, if tackled this way would equal zero. The standard approach, see England and Verrall (2002), is to estimate

$$
\left( \sum_{j > I - i} (\hat{\gamma}_j - \gamma_j) \right)^2
$$

by the unconditional expectation

$$
E \left[ \left( \sum_{j > I - i} (\hat{\gamma}_j - \gamma_j) \right)^2 \right] = \sum_{l > I - i} E \left[ (\hat{\gamma}_j - \gamma_j)(\hat{\gamma}_l - \gamma_l) \right].
$$

Neglecting that MLEs have a possible bias term we make the following approximation:

$$
\sum_{j > I - i} E \left[ (\hat{\gamma}_j - \gamma_j)(\hat{\gamma}_l - \gamma_l) \right] \approx \sum_{l > I - i} \text{Cov}(\hat{\gamma}_j, \hat{\gamma}_l).
$$

We now deviate from Alai et al. (2009) and directly use $G$, given by equations (5) to estimate the covariance terms. Hence, an estimate of the MSEP in the BF method for single accident year $i$ is given by:
Estimator 4.1 (MSEP for the BF method, single accident year)

Under Model Assumptions 2.1 an estimator for the (conditional) MSEP for a single accident year \( i \in \{1, \ldots, I\} \) is given by

\[
\hat{\text{mse}}_{C_{i,I} | D_I} (\hat{C}_{i,I}^{BF}) = \sum_{j > I - i} \hat{\phi} \hat{\nu}_j \hat{\gamma}_j + \left( \sum_{j > I - i} \hat{\gamma}_j \right)^2 \hat{\text{Var}}(\hat{\nu}_i) + \hat{\nu}_i^2 \sum_{j > I - i} g_{j,i}.
\]

Remark 4.2 If we compare the above estimator to equation (5.30) in Alai et al. (2009) we observe that the first two terms on the right-hand side are identical. However, the last term, i.e. the uncertainty in \( \hat{\gamma}_j \), has substantially simplified and can be easily calculated.

For multiple accident years the (conditional) MSEP is defined as follows:

\[
\text{mse}_{\sum_{i=1}^I C_{i,I} | D_I} \left( \sum_{i=1}^I \hat{C}_{i,I}^{BF} \right) = E \left[ \left( \sum_{i=1}^I \hat{C}_{i,I}^{BF} - \sum_{i=1}^I C_{i,I} \right)^2 \bigg| D_I \right]
= \sum_{i=1}^I \text{mse}_{C_{i,I} | D_I} (\hat{C}_{i,I}^{BF}) + 2 \sum_{i<k} \hat{\nu}_i \hat{\nu}_k \sum_{j > I - i} \sum_{l > I - k} (\hat{\gamma}_j - \gamma_j) (\hat{\gamma}_l - \gamma_l).
\]

Similar as above, it is estimated by:

Estimator 4.3 (MSEP for the BF method, aggregated acc. years)

Under Model Assumptions 2.1 an estimator for the (conditional) MSEP for aggregated accident years is given by

\[
\text{mse}_{\sum_{i=1}^I C_{i,I} | D_I} \left( \sum_{i=1}^I \hat{C}_{i,I}^{BF} \right) = \sum_{i=1}^I \text{mse}_{C_{i,I} | D_I} (\hat{C}_{i,I}^{BF}) + 2 \sum_{i<k} \hat{\nu}_i \hat{\nu}_k \sum_{j > I - i} \sum_{l > I - k} g_{j,l}.
\]

Remark 4.4 The above estimator should be compared with equation (5.35) in Alai et al. (2009).

5 Case Study

We utilize the dataset \( \{X_{i,j} : i + j \leq I\} \) provided in Alai et al. (2009), which is shown in Table 1. We assume given external estimates \( \hat{\nu}_i \) of the ultimate claims, presented in Table 2. Furthermore, we assume the uncertainty of these estimates to be given by a coefficient of variation of 5\%. Hence,

\[
\hat{\text{Var}}(\hat{\nu}_i) = \hat{\nu}_i^2 (0.05)^2.
\]

Using equation (3), we obtain for the dispersion parameter \( \phi \), the estimate \( \hat{\phi} = 14,714 \).
Table 1: Observed incremental claims $X_{i,j}$ given in '000s.

Table 2: Prior estimates for the expected ultimate claims.

Table 3: Reserve and uncertainty results for single and aggregated accident years using the method in Alai et al. (2009).

We compare the results in Table 3 to those from Mack (2008). We start by calculating the development pattern using equation (3) in Mack (2008). We normalize these results such that the pattern sums to one. Note that the normalization is necessary due to the fact that the prior estimates $\hat{\nu}_i$ are rather conservative (as mentioned in Wüthrich and Merz (2008), Example 2.11). In Table 4 we compare the cumulative development pattern (referred to as $\hat{z}^*_j$ in Mack (2008)) with the
cumulative development pattern obtained using the method of Alai et al. (2009) (referred to as $\hat{\beta}_j$). Also shown in Table 4 are the standard errors calculated for the cumulative development patterns using the respective methods.

**Remark 5.1** The distinction is made between estimates of the development pattern $\hat{\gamma}_j$ and of the cumulative development pattern $\hat{\beta}_j$; the latter being defined as follows:

$$\hat{\beta}_j = \sum_{k=0}^{j} \hat{\gamma}_k, \quad \text{for} \ j \in \{0, \ldots, I\}.$$ 

Table 4 indicates a slower decrease of the uncertainty in our approach.

In Table 5 we provide the $s_j^2$ calculated using equation (4) in Mack (2008). The role of the $s_j^2$ are comparable to that of $\hat{\phi}$. The difference originates from the fact that the $s_j^2$ depend on the development year $j$, whereas $\hat{\phi}$ does not.

<table>
<thead>
<tr>
<th>$j$</th>
<th>$\hat{\beta}_j$</th>
<th>$\text{s.e.}(\hat{\beta}_j)$</th>
<th>$\hat{\gamma}_j$</th>
<th>$\text{s.e.}(\hat{\gamma}_j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>58.96%</td>
<td>0.653%</td>
<td>58.60%</td>
<td>1.717%</td>
</tr>
<tr>
<td>1</td>
<td>88.00%</td>
<td>0.484%</td>
<td>87.66%</td>
<td>0.616%</td>
</tr>
<tr>
<td>2</td>
<td>94.84%</td>
<td>0.370%</td>
<td>94.60%</td>
<td>0.326%</td>
</tr>
<tr>
<td>3</td>
<td>97.01%</td>
<td>0.313%</td>
<td>96.84%</td>
<td>0.271%</td>
</tr>
<tr>
<td>4</td>
<td>98.45%</td>
<td>0.258%</td>
<td>98.35%</td>
<td>0.131%</td>
</tr>
<tr>
<td>5</td>
<td>99.14%</td>
<td>0.219%</td>
<td>99.07%</td>
<td>0.054%</td>
</tr>
<tr>
<td>6</td>
<td>99.65%</td>
<td>0.175%</td>
<td>99.62%</td>
<td>0.025%</td>
</tr>
<tr>
<td>7</td>
<td>99.75%</td>
<td>0.160%</td>
<td>99.73%</td>
<td>0.018%</td>
</tr>
<tr>
<td>8</td>
<td>99.86%</td>
<td>0.137%</td>
<td>99.85%</td>
<td>0.012%</td>
</tr>
<tr>
<td>9</td>
<td>100.00%</td>
<td>0.118%</td>
<td>100.00%</td>
<td>0.010%</td>
</tr>
</tbody>
</table>

Table 4: Cumulative development pattern, a comparison.

$$\hat{s}_j^2$$ calculated from equation (3) in Mack (2008).

<table>
<thead>
<tr>
<th>$j$</th>
<th>$\hat{s}_j^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>69,990</td>
</tr>
<tr>
<td>1</td>
<td>25,848</td>
</tr>
<tr>
<td>2</td>
<td>2,450</td>
</tr>
<tr>
<td>3</td>
<td>260</td>
</tr>
<tr>
<td>4</td>
<td>376</td>
</tr>
<tr>
<td>5</td>
<td>79</td>
</tr>
<tr>
<td>6</td>
<td>918</td>
</tr>
<tr>
<td>7</td>
<td>407</td>
</tr>
<tr>
<td>8</td>
<td>180</td>
</tr>
</tbody>
</table>

Table 5: $\hat{s}_j^2$ calculated from equation (3) in Mack (2008).

Finally, we apply the same coefficient of variation to determine the standard error of the ultimates using the method in Mack (2008), namely 5%. Table 6 provides the MSEP results under the method described in Mack (2008). It should be compared to Table 3, which provides the results under the method described in Alai et al. (2009) and in this paper.

As becomes clear from comparing Tables 3 and 6, one main difference between the two methods lies in the estimated process variance. It is evident that this difference originates in the model assumptions with respect to the structure of the variance of the incremental claims. Alai et al. (2009) assume

$$\text{Var}(X_{i,j}) = \phi m_{i,j},$$

87
Table 6: Reserve and uncertainty results for single and aggregated accident years using the method in Mack (2008).

<table>
<thead>
<tr>
<th>accident year $i$</th>
<th>reserves</th>
<th>std. dev.</th>
<th>prior parameter</th>
<th>std. dev.</th>
<th>prior and parameter</th>
<th>msep$^{1/2}$</th>
<th>Vco</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17,420</td>
<td>1,431</td>
<td>871</td>
<td>1,415</td>
<td>1,661</td>
<td>2,193</td>
<td>12.6%</td>
</tr>
<tr>
<td>2</td>
<td>29,059</td>
<td>2,536</td>
<td>1,453</td>
<td>1,998</td>
<td>2,470</td>
<td>3,540</td>
<td>12.2%</td>
</tr>
<tr>
<td>3</td>
<td>40,480</td>
<td>3,996</td>
<td>2,024</td>
<td>2,607</td>
<td>3,300</td>
<td>5,183</td>
<td>12.8%</td>
</tr>
<tr>
<td>4</td>
<td>102,383</td>
<td>11,541</td>
<td>5,119</td>
<td>6,025</td>
<td>7,906</td>
<td>13,989</td>
<td>13.7%</td>
</tr>
<tr>
<td>5</td>
<td>189,802</td>
<td>32,332</td>
<td>9,490</td>
<td>15,060</td>
<td>17,801</td>
<td>36,908</td>
<td>19.4%</td>
</tr>
<tr>
<td>6</td>
<td>360,691</td>
<td>73,010</td>
<td>18,035</td>
<td>30,914</td>
<td>35,790</td>
<td>81,310</td>
<td>22.5%</td>
</tr>
<tr>
<td>7</td>
<td>600,764</td>
<td>58,940</td>
<td>20,038</td>
<td>36,319</td>
<td>47,131</td>
<td>101,541</td>
<td>16.9%</td>
</tr>
<tr>
<td>8</td>
<td>1,355,361</td>
<td>73,010</td>
<td>67,768</td>
<td>67,773</td>
<td>95,842</td>
<td>209,988</td>
<td>15.5%</td>
</tr>
<tr>
<td>9</td>
<td>3,606,91</td>
<td>89,940</td>
<td>124,477</td>
<td>199,723</td>
<td>312,600</td>
<td>659,504</td>
<td>13.7%</td>
</tr>
<tr>
<td>covariance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>173,602</td>
<td>17.3%</td>
</tr>
<tr>
<td>total</td>
<td>7,505,506</td>
<td>621,899</td>
<td>252,532</td>
<td>375,424</td>
<td>726,431</td>
<td>9.7%</td>
<td></td>
</tr>
</tbody>
</table>

whereas Mack (2008) assumes

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

Table 5 shows the volatility of the $s_j^2$, which heavily impacts the process variance. A similar picture is obtained for the parameter standard deviation, in contrast to the prior standard deviation, which almost perfectly coincide.

Finally, in Table 7, we present the MSEP results for the distribution-free CL method; see Mack (1993). To obtain the (conditional) MSEP we use the approach described in Buchwalder et al. (2006). Although in no way conclusive, the overall approach of Alai et al. (2009) is more in line with the CL MSEP figures.

Table 7: Aggregate reserve and uncertainty results for the CL method, the BF approach of Alai et al. (2009), and the BF approach of Mack (2008).

<table>
<thead>
<tr>
<th></th>
<th>reserves</th>
<th>process error</th>
<th>estimation error</th>
<th>msep$^{1/2}$</th>
<th>Vco</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL method</td>
<td>6,047,061</td>
<td>424,379</td>
<td>185,026</td>
<td>462,960</td>
<td>7.7%</td>
</tr>
<tr>
<td>BF Alai et al. (2009)</td>
<td>7,356,575</td>
<td>329,007</td>
<td>338,396</td>
<td>471,971</td>
<td>6.4%</td>
</tr>
<tr>
<td>BF Mack (2008)</td>
<td>7,505,506</td>
<td>621,899</td>
<td>375,424</td>
<td>726,431</td>
<td>9.7%</td>
</tr>
</tbody>
</table>

Bibliography


Daniel H. Alai and Mario V. Wüthrich (2009).

Taylor approximations for model uncertainty within the Tweedie exponential dispersion family.

Taylor Approximations for Model Uncertainty within the Tweedie Exponential Dispersion Family

Daniel H. Alai Mario V. Wüthrich
Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland

Abstract

The use of generalized linear models (GLM) to estimate claims reserves has become a standard method in insurance. Most frequently, the exponential dispersion family (EDF) is used; see e.g. England and Verrall (2002). We study the so-called Tweedie EDF and test the sensitivity of the claims reserves and their mean square error of predictions (MSEP) over this family. Furthermore, we develop second order Taylor approximations for the claims reserves and the MSEPs for members of the Tweedie family that are difficult to obtain in practice, but are close enough to models for which claims reserves and MSEP estimations are easy to determine. As a result of multiple case studies, we find that claims reserves estimation is relatively insensitive to which distribution is chosen amongst the Tweedie family, in contrast to the MSEP, which varies widely.

Keywords: Claims Reserving, Exponential Dispersion Family, Model Uncertainty, Power Variance Function, Tweedie’s Exponential Dispersion Models, Prediction Error.
1 Introduction

The use of generalized linear models (GLM) in actuarial science is well developed and broadly accepted. Not only does the framework of GLM allow for flexibility in parameter and model selection, in some cases, such as with the chain ladder method, GLM recovers traditional methods for claims reserves estimation. For a comprehensive reference of GLM, see McCullagh and Nelder (1989). In this paper, we study the exponential dispersion family (EDF) and its role in modelling claims reserves; see e.g. Jørgensen (1987, 1997) for more on the EDF and Renshaw (1994), Haberman and Renshaw (1996), England and Verrall (2002) and Wüthrich and Merz (2008) for applications to insurance. We focus on a special member of the EDF, the so-called Tweedie exponential dispersion model. Besides containing many standard models, such as the Gaussian, Poisson and gamma, we are particularly interested in the compound Poisson models; Mildenhall (1999) provides an excellent review of these models. For specific applications of the Tweedie compound Poisson model, see e.g. Jørgensen and De Souza (1994), Smyth and Jørgensen (2002), and Wüthrich (2003).

The family of Tweedie exponential dispersion models is controlled by a model parameter $p$. For example, $p = 1$ corresponds to the overdispersed Poisson model. In this paper, we calculate the sensitivity of the claims reserves with respect to this model parameter. Peters et al. (2009) and Gigante and Sigalotti (2005) have also tackled this issue of model uncertainty, the former opting for a Bayesian Markov Chain Monte Carlo simulation approach averaging over $p$, and the latter addressing the issue within a GLM framework, using an iterative procedure to solve for $p$ using quasi-likelihood functions as introduced by Wedderburn (1974) and Nelder and Pregibon (1987). In our approach, we directly work with the likelihood function and rather than solve for $p$, we find the claims reserves in terms of $p$, i.e. for a fixed distributional model. Besides claims reserves sensitivity, we also investigate the sensitivity of the mean square error of prediction (MSEP) with respect to $p$. Furthermore, we develop second order Taylor approximations for the claims reserves and MSEPs with respect to $p$.

We conclude that, based on multiple case studies and as shown in Peters et al. (2009), the claims reserves are rather insensitive to the choice of $p$ and hence find that there is only moderate model uncertainty when modelling within the Tweedie exponential dispersion family. In contrast however, we find that the MSEP is highly sensitive to the model parameter $p$. This has important consequences for solvency considerations and the required risk bearing capital.

Organization of the paper. In Section 2 we describe the data and the model assumptions. Maximum likelihood estimation (MLE) of the underlying model parameters is discussed in Section 3. In Sections 4 and 5 we study the sensitivity as well as derive Taylor approximations of the claims reserves and the MSEP, re-
spectively, with respect to the model parameter $p$. In Section 6, we provide one of the earlier mentioned case studies to highlight the performance of the Taylor approximations.

### 2 Data and Model

#### 2.1 Setup

Let $X_{i,j}$ denote the incremental payments of accident year $i \in \{0, 1, \ldots, I\}$ and development year $j \in \{0, 1, \ldots, J\}$. We assume that the data is given by a claims development triangle, i.e. $I = J$. This means that our data are given by an upper triangle, $D_I = \{X_{i,j}, i + j \leq I\}$, and that we are interested in predicting the incremental payments for the corresponding lower triangle $\{X_{i,j}, i + j > I\}$ at time $I$. See Figure 1 for a graphic representation of the data. The outstanding loss liabilities are given by

$$R = \sum_{i+j > I} X_{i,j}.$$ 

These are the future cashflows at time $I$. We are going to predict these outstanding loss liabilities, $R$, with a predictor $\hat{R}$, the so-called claims reserves, that is based on the information $D_I$ available at time $I$.

<table>
<thead>
<tr>
<th>accident year $i$</th>
<th>development year $j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\ldots$ $j$ $\ldots$ $J$</td>
</tr>
<tr>
<td>\vdots</td>
<td>r.v. $X_{i,j}$, $i + j \leq I$</td>
</tr>
<tr>
<td>$i$</td>
<td>predicted r.v. $X_{i,j}$, $i + j &gt; I$</td>
</tr>
<tr>
<td>$I$</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Claims development triangle.

#### 2.2 The Exponential Dispersion Family

Nelder and Wedderburn (1972) established the framework of GLM and the so-called analysis of deviance. These concepts were originally developed for exponential families of distributions, yet extended to a wider class of distributions, termed dispersion models. Here, we work within the framework of this broader family of distributions but focus on an important sub-class called the Tweedie exponential dispersion models, introduced in Tweedie (1984).

**Model Assumptions 2.1** (Exponential Dispersion Model)  
A random variable $X_{i,j}$ follows an exponential dispersion model if it has generalized
density
\[ f(x; \theta_{i,j}, \phi_{i,j}) = \exp \left\{ \frac{w_{i,j}}{\phi_{i,j}} (x \theta_{i,j} - b(\theta_{i,j})) \right\} c(x; \phi_{i,j}), \]

where \( w_{i,j} > 0 \) denotes a known weight, \( \theta_{i,j} \) is the canonical parameter, \( \phi_{i,j} > 0 \) the dispersion parameter. The function \( b \) is a twice differentiable general function that determines the more specific family the distribution falls into and \( c \) is a suitable normalizing constant. This generalized density is, e.g., either defined with respect to Lebesgue measure or the counting measure. Moreover, the domain of \( x \) depends on the choice of \( b \).

As noted above, the function \( b \) determines to which specific family the exponential dispersion distribution belongs. Likewise, one can specify the structure of the underlying unit variance function, \( V(\cdot) \), defined as,

\[ V(m) = b''((b')^{-1}(m)); \]

see e.g. Jørgensen (1997), Theorem 2.11. We focus on unit variance functions of the power variety, namely \( V(m) = m^p \) for \( p \in (-\infty, 0] \cup [1, \infty) \); see e.g. Jørgensen (1997), Proposition 4.2, regarding the possible values of \( p \). The family so defined is known as the Tweedie EDF. Specific values of \( p \) correspond to specific distributions, for example when \( p = 0, 1, 2, 3 \), we recover the Gaussian, overdispersed Poisson, gamma, and inverse Gaussian distributions, respectively. As such, the parameter \( p \) plays a central role of model uncertainty for distributions within the Tweedie class.

**Model Assumptions 2.2 (Tweedie Exponential Dispersion Model)**

A random variable \( X_{i,j} \) follows a Tweedie exponential dispersion model if it is an exponential dispersion model with parameter \( \theta_{i,j} \in \Theta_p \) and function \( b \) defined as

\[ b_p(\theta_{i,j}) = \begin{cases} 
\frac{1}{2-p} ((1-p)\theta_{i,j})^{\frac{2-p}{p-1}}, & p \notin (0, 1] \cup [2], \\
\exp(\theta_{i,j}), & p = 1, \\
-\log(-\theta_{i,j}), & p = 2,
\end{cases} \]

where

\[ \Theta_p = \begin{cases} 
\mathbb{R}, & \text{for } p = 0, 1, \\
[0, \infty), & \text{for } p < 0, \\
(-\infty, 0), & \text{for } 1 < p \leq 2, \\
(-\infty, 0], & \text{for } 2 < p < \infty.
\end{cases} \]

The specification on \( b \) is made so that the unit variance function, \( V \), has a power structure with power \( p \), that is, \( b_p(\cdot) \) implies that

\[ V(m) = m^p. \]
Under these assumptions, $X_{i,j}$ has expectation and variance given by

$$E[X_{i,j}] = m_{i,j} = b_p'(\theta_{i,j}),$$
$$\text{Var}(X_{i,j}) = \frac{\phi_{i,j}}{w_{i,j}} b_p'(\theta_{i,j}) V(m_{i,j});$$

see e.g. Bühlmann and Gisler (2005), Theorem 2.2.

In the remainder of this paper we assume that $\phi_{i,j} w_{i,j}$ is constant and define $\phi = \frac{\phi_{i,j}}{w_{i,j}}$. This assumption implies that $\phi$ cancels in the MLE of $\theta_{i,j}$, which substantially simplifies the analysis.

### 3 Maximum Likelihood Estimators

#### 3.1 Claims Reserves

We assume the $X_{i,j}$ are independent Tweedie distributed (see Model Assumptions 2.2) with $\phi = \frac{\phi_{i,j}}{w_{i,j}}$, we estimate the model parameters using MLE. The log-likelihood function for $D_I$ is given by

$$l(\theta_{i,j}, \phi) = l_{D_I}(\theta_{i,j}, \phi) = \sum_{i+j \leq I} \left( \frac{1}{\phi}(X_{i,j}\theta_{i,j} - b_p(\theta_{i,j})) + \log c(X_{i,j}; \phi) \right).$$

We have $(I + 1)^2$ unknown parameters $\theta_{i,j}$ and only $(I + 1)(I + 2)/2$ observations. Therefore we introduce additional model structure to obtain a more parsimonious model. As is standard in claims reserving modelling, we choose a multiplicative model $m_{i,j} = \mu_i \gamma_j$, where $\mu_i > 0$ is the exposure of accident year $i$ and $\gamma_j > 0$ describes the claims development pattern. This implies that the canonical parameter, $\theta_{i,j} = (b_p')^{-1}(m_{i,j})$, is given by

$$\theta_{i,j} = \log (\mu_i \gamma_j), \quad p = 1,$$
$$\theta_{i,j} = \left(\frac{\mu_i \gamma_j}{1 - p}\right)^{1-p}, \quad p \neq 1.$$

The MLEs of the parameters $\mu_i$ and $\gamma_j$ can now be obtained by setting the following system of equations equal to zero:

$$\frac{\partial}{\partial \mu_i} l(\theta_{i,j}, \phi) = \frac{1}{\phi} \sum_{j=0}^{I-i} \left( X_{i,j} \frac{\partial \theta_{i,j}}{\partial \mu_i} - \frac{\partial b_p(\theta_{i,j})}{\partial \mu_i} \right), \quad i \in \{0, \ldots, I\},$$
$$\frac{\partial}{\partial \gamma_j} l(\theta_{i,j}, \phi) = \frac{1}{\phi} \sum_{i=0}^{I-j} \left( X_{i,j} \frac{\partial \theta_{i,j}}{\partial \gamma_j} - \frac{\partial b_p(\theta_{i,j})}{\partial \gamma_j} \right), \quad j \in \{0, \ldots, I\}.$$
We need to introduce a constraint to obtain a unique solution to these equations. Using $\mu_0 = 1$, we obtain the MLEs

$$
\hat{\mu}_i \sum_{j=0}^{I-i} \hat{\gamma}_j^{2-p} = \sum_{j=0}^{I-i} x_{i,j} \hat{\gamma}_j^{1-p}, \quad i \in \{1, \ldots, I\},
$$

$$
\hat{\gamma}_j \sum_{i=0}^{I-j} \hat{\mu}_i^{2-p} = \sum_{i=0}^{I-j} x_{i,j} \hat{\mu}_i^{1-p}, \quad j \in \{0, \ldots, I\}, \tag{1}
$$

see e.g. Wüthrich and Merz (2008), formulas (5.49) and (5.50). Note that for $p = 1$, the above system corresponds to the overdispersed Poisson model, which yields the chain ladder claims reserves; see e.g. Mack (1991) or Lemma 2.16 in Wüthrich and Merz (2008). From the equations given in (1), it is clear that $\hat{\mu}_i$ and $\hat{\gamma}_j$ are functions of $p$ and consequently, model uncertainty within the Tweedie family may be expressed in terms of their derivatives with respect to $p$.

Note that in (1), $\phi$ cancels and consequently has no influence on the parameter estimation of $\mu_i$ and $\gamma_j$. However, an estimate of $\phi$ is required to estimate the prediction uncertainty. We could estimate $\phi$ with MLE but this involves an infinite summation that is often difficult to evaluate; see e.g. Peters et al. (2009). Therefore, we prefer using Pearson residuals for the estimation of $\phi$. Furthermore, Pearson residuals are standard outputs in all GLM software tools and are widely accepted in practice. We obtain the following estimate:

$$
\hat{\phi} = \frac{1}{d} \sum_{i+j \leq I} \frac{(X_{i,j} - \hat{m}_{i,j})^2}{V(\hat{m}_{i,j})},
$$

where $d = \frac{(I+1)(I+2)}{2} - 2I - 1$ is the degrees of freedom of the model and $\hat{m}_{i,j} = \hat{\mu}_i \hat{\gamma}_j$.

With these parameter estimates, we can predict the outstanding loss liabilities, $R$, with the claims reserves, $\hat{R}$, given by

$$
\hat{R} = \sum_{i+j > I} \hat{\mu}_i \hat{\gamma}_j.
$$

The claims reserves $\hat{R} = \hat{R}(p)$ depend on $p$, our aim is a sensitivity analysis in $p$.

3.2 Asymptotic Properties of the MLE

The proposition directly below yields the asymptotic behaviour of MLEs; see e.g. Lehmann (1983), Theorem 6.2.3.

Proposition 3.1

Assume $X_1, \ldots, X_n$ are i.i.d. with density $f_\xi(\cdot)$ from the exponential dispersion
family with parameters $\zeta = (\zeta_1, \ldots, \zeta_m)^T$. Furthermore, $\hat{\zeta} = (\hat{\zeta}_1, \ldots, \hat{\zeta}_m)^T$ is the MLE of $\zeta$, then,

$$\sqrt{n}(\hat{\zeta} - \zeta) \xrightarrow{d} N(0, H(\zeta)^{-1}), \text{ as } n \to \infty,$$

where we define the Fisher information matrix by $H(\zeta) = (H(\zeta))_{r,s=1,\ldots,m}$ with

$$H(\zeta)_{r,s} = -E_{\zeta} \left[ \frac{\partial^2}{\partial \zeta_r \partial \zeta_s} \log f_{\zeta}(X) \right].$$

We use the notation $\zeta = (\zeta_0, \ldots, \zeta_{2I+1})^T = (\mu_0, \ldots, \mu_I, \gamma_0, \ldots, \gamma_I)^T$. Before deriving the Fisher information matrix, $H$, we provide the following necessary partial derivatives:

$$\frac{\partial \theta_{i,j}}{\partial \mu_i} = \mu_i - p_i \gamma_j, \quad \frac{\partial b_p(\theta_{i,j})}{\partial \mu_i} = \mu_i - p_i \gamma_j,$$

$$\frac{\partial \theta_{i,j}}{\partial \gamma_j} = \mu_i - p_i \gamma_j, \quad \frac{\partial b_p(\theta_{i,j})}{\partial \gamma_j} = \mu_i - p_i \gamma_j.$$

From these, and the derivatives of the log-likelihood function with respect to the underlying parameters, we obtain

$$H(\zeta, \phi)_{r,r} = \frac{\zeta_r - p_r \sum_{s=0}^{I-1} \zeta_{I+1+s}}{\phi}, \quad r \in \{1, \ldots, I\},$$

$$H(\zeta, \phi)_{s,s} = \frac{\zeta_s - p_r \sum_{r=0}^{2I+1-s} \zeta_r}{\phi}, \quad s \in \{I+1, \ldots, 2I+1\},$$

$$H(\zeta, \phi)_{r,s} = \frac{1}{\phi} \zeta_r^{1-p} \zeta_s^{1-p}, \quad r \in \{1, \ldots, 2I+1\}, s \in \{1, \ldots, 2I+1-r\},$$

$$(r,s) \notin \{1, \ldots, I\} \times \{1, \ldots, I\}.$$

The remaining entries of the $(2I+1) \times (2I+1)$ matrix $H$ are zero. Note that we omit $\hat{\mu}_0$ in our construction of $H$ because its inclusion would imply $H$ to be singular.

We estimate the MSEP using the above result. Note that $H$ depends on $\phi$ and $\zeta$. By replacing these parameters by their estimates we obtain the estimated Fisher information matrix $\hat{H} = H(\hat{\zeta}, \hat{\phi})$, which is a function of $p$. Of specific importance are the estimates of the covariances of the MLEs, $\hat{\zeta}$. Proposition 3.1 provides the following estimator,

$$\hat{\text{Cov}}(\hat{\zeta}_r, \hat{\zeta}_s) = H(\hat{\zeta}, \hat{\phi})_{r,s}^{-1}, \quad r, s \in \{1, \ldots, 2I+1\}. \quad (3)$$

Before studying the effect of the model parameter $p$ on the MSEP, we first study the sensitivity of the claims reserves with respect to the choice of $p$. 
4 Sensitivity of the Claims Reserves with respect to $p$

As stated in previous sections, assuming a distribution from the Tweedie EDF, we can estimate the parameters $\mu_i$ and $\gamma_j$, from which we can predict the outstanding loss liabilities $R$ with the claims reserves $\hat{R} = \hat{R}(p)$. We analyze the sensitivity of $\hat{R}$ with respect to $p$ by a Taylor expansion. The second order Taylor expansion around $p$ is given by

$$\hat{R}^\ast(p + \varepsilon) = \hat{R}(p) + \hat{R}'(p) \varepsilon + \frac{\hat{R}''(p)}{2} \varepsilon^2,$$

with,

$$\hat{R}'(p) = \sum_{i+j>I} (\hat{\mu}_i'\hat{\gamma}_j + \hat{\mu}_i\hat{\gamma}_j'),$$

$$\hat{R}''(p) = \sum_{i+j>I} (\hat{\mu}_i''\hat{\gamma}_j + 2\hat{\mu}_i'\hat{\gamma}_j' + \hat{\mu}_i\hat{\gamma}_j'').$$

The first and second derivatives are provided in Lemmas 4.1 and 4.2 below.

4.1 Reserves Approximation using First Order Taylor Expansion

We begin by studying the first order Taylor expansion, given by omitting the last term in (4). To approximate the claims reserves using a first order Taylor expansion, we need to calculate the first derivative with respect to $p$ of the MLE $\hat{\zeta}$. Differentiating the equations given in (1) with respect to $p$ provides:

$$\hat{\mu}_i' I - \sum_{j=0}^{I-1} \hat{\gamma}_j^2 - p + \sum_{j=0}^{I-1} \hat{\gamma}_j' \left[ (2 - p)\hat{\mu}_i\hat{\gamma}_j - (1 - p)x_{i,j} \right]$$

$$= \sum_{j=0}^{I-1} \log(\hat{\gamma}_j) \frac{\hat{\gamma}_j}{\hat{\mu}_i} \left[ \hat{\mu}_i\hat{\gamma}_j - x_{i,j} \right], \quad i \in \{1, \ldots, I\}, \quad (5)$$

$$\hat{\gamma}_j I - \sum_{i=0}^{I-1} \hat{\mu}_i^2 - p + \sum_{i=0}^{I-1} \hat{\mu}_i' \left[ (2 - p)\hat{\gamma}_j\hat{\mu}_i - (1 - p)x_{i,j} \right]$$

$$= \sum_{i=0}^{I-1} \log(\hat{\mu}_i) \frac{\hat{\mu}_i}{\hat{\mu}_i} \left[ \hat{\gamma}_j\hat{\mu}_i - x_{i,j} \right], \quad j \in \{0, \ldots, I\},$$

$$\hat{\mu}_0' = 0.$$
To solve the above system of equations we define a \((2I + 2) \times (2I + 2)\) matrix \(A\), whose components are the following:

\[
a_{i,i} = \sum_{j=0}^{I-i} \hat{\gamma}_j \hat{\gamma}_j, \quad i \in \{1, \ldots, I\},
\]

\[
a_{i,I+j+1} = \frac{1}{\hat{\mu}_i} \left( (2 - p) \hat{\mu}_i \hat{\gamma}_j - (1 - p)x_{i,j} \right), \quad i \in \{0, \ldots, I\}, j \in \{0, \ldots, I - i\},
\]

\[
a_{I+j+1,i} = \frac{1}{\hat{\mu}_i} \left( (2 - p) \hat{\mu}_i \hat{\gamma}_j - (1 - p)x_{i,j} \right), \quad j \in \{0, \ldots, I\}, i \in \{0, \ldots, I - j\},
\]

\[
a_{I+j+1,I+j+1} = \sum_{i=0}^{I-j} \hat{\mu}_i \hat{\gamma}_j \hat{\gamma}_j \hat{\gamma}_j, \quad j \in \{0, \ldots, I\},
\]

\[
a_{0,0} = 1.
\]

The remaining entries of the matrix are zero. In addition to the matrix \(A\), we define \(\hat{\zeta}' = (\hat{\mu}_0', \ldots, \hat{\mu}_I', \hat{\gamma}_0', \ldots, \hat{\gamma}_I')^T\) and \(\alpha = (0, \alpha_1, \ldots, \alpha_I, \beta_0, \ldots, \beta_I)^T\), where

\[
\alpha_i = \sum_{j=0}^{I-i} \log(\hat{\gamma}_j) \hat{\gamma}_j \hat{\mu}_i (\hat{\mu}_i \hat{\gamma}_j - x_{i,j}), \quad i \in \{1, \ldots, I\},
\]

\[
\beta_j = \sum_{i=0}^{I-j} \log(\hat{\mu}_i) \hat{\mu}_i (\hat{\mu}_i \hat{\gamma}_j - x_{i,j}), \quad j \in \{0, \ldots, I\}.
\]

Using matrix notation, we rewrite equations given in (5) as

\[A \hat{\zeta}' = \alpha.\]

**Lemma 4.1**

The first derivate of the MLE \(\hat{\zeta}\) is given by

\[\hat{\zeta}' = A^{-1} \alpha,\]

where \(A\) and \(\alpha\) are defined as given above.

In the following section, we study the second order approximation.

### 4.2 Reserves Approximation using Second Order Taylor Expansion

What remains to be provided in order to use the second order Taylor approximation is the second derivative of the MLE \(\hat{\zeta}\) with respect to \(p\). Rather than differentiating equations given in (5), we first rewrite them using simplifying notation already
introduced above for matrix $A$. We have
\[
\hat{\mu}'_{i,i} + \sum_{j=0}^{I-i} \hat{\gamma}'_{j} a_{i,I+j+1} = \alpha_{i}, \quad i \in \{1, \ldots, I\},
\]
\[
\hat{\gamma}'_{j} a_{I+j+1,I+j+1} + \sum_{i=0}^{I-j} \hat{\mu}'_{i} a_{I+j+1,i} = \beta_{j}, \quad j \in \{0, \ldots, I\},
\]
\[
\hat{\rho}'_{0} = 0.
\]
Taking derivates we obtain:
\[
\hat{\mu}''_{i,i} + \sum_{j=0}^{I-i} \hat{\gamma}''_{j} a_{i,I+j+1} = \kappa_{i}, \quad i \in \{1, \ldots, I\},
\]
\[
\hat{\gamma}''_{j} a_{I+j+1,I+j+1} + \sum_{i=0}^{I-j} \hat{\mu}''_{i} a_{I+j+1,i} = \lambda_{j}, \quad j \in \{0, \ldots, I\},
\]
\[
\hat{\rho}''_{0} = 0,
\]
where
\[
\kappa_{i} = \alpha_{i}' - \alpha_{i}' a_{i,i} - \sum_{j=0}^{I-i} \hat{\gamma}'_{j} a_{i,I+j+1}, \quad i \in \{1, \ldots, I\},
\]
\[
\lambda_{j} = \beta_{j}' - \alpha_{I+j+1,I+j+1} \gamma'_{j} - \sum_{i=0}^{I-j} \hat{\mu}'_{i} a_{I+j+1,i}, \quad j \in \{0, \ldots, I\}.
\]
Hence, we need to find the derivates of $a_{i,i}$, $a_{i,I+j+1}$, $a_{I+j+1,I+j+1}$, $a_{I+j+1,I+j+1}$, $\alpha_{i}$, and $\beta_{j}$ with respect to $p$. They are given in Appendix A. We define column vectors,
\[
\hat{\zeta}'' = (\hat{\mu}''_{0}, \ldots, \hat{\mu}''_{I}, \hat{\gamma}''_{0}, \ldots, \hat{\gamma}''_{I})^{T}
\]
and $\kappa = (0, \kappa_{1}, \ldots, \kappa_{I}, \lambda_{0}, \ldots, \lambda_{I})^{T}$, so that we can formulate the equations given in (6) as
\[
A \hat{\zeta}'' = \kappa,
\]
where $A$ is as previously defined.

**Lemma 4.2**

The second derivate of the MLE $\hat{\zeta}$ is given by
\[
\hat{\zeta}'' = A^{-1} \kappa,
\]
where $A$ and $\kappa$ are defined as given above.

**Remark 4.3** Of course this can inductively be expanded to any other derivatives $\hat{\mu}^{(k)}_{i}$ and $\hat{\gamma}^{(k)}_{j}$, $k \geq 3$, where the right-hand sides in equations given in (6) become appropriate functions depending on $\hat{\mu}^{(l)}_{i}$ and $\hat{\gamma}^{(l)}_{j}$, $l < k$. 

102
5 Sensitivity of the MSEP with respect to $p$

Before studying the derivative of the MSEP with respect to $p$, we need to estimate the MSEP. The (conditional) MSEP for predictor $\hat{R}$ of the outstanding loss liabilities $R$ is defined as follows:

$$\text{msep}_{R|D_I}(\hat{R}) = E\left[ (\hat{R} - R)^2 \mid D_I \right] = E\left[ \left( \sum_{i+j > I} \hat{\mu}_i \hat{\gamma}_j - \sum_{i+j > I} X_{i,j} \right)^2 \mid D_I \right].$$

Since the predictor $\sum_{i+j > I} \hat{\mu}_i \hat{\gamma}_j$ is $D_I$-measurable, we decompose it into terms referred to as (conditional) process variance and (conditional) estimation error; see e.g. Wüthrich and Merz (2008), Section 3.1. We get the following decomposition:

$$\text{msep}_{R|D_I}(\hat{R}) = \text{Var}\left( \sum_{i+j > I} X_{i,j} \mid D_I \right) + \left( \sum_{i+j > I} (\hat{\mu}_i \hat{\gamma}_j - E[X_{i,j}\mid D_I]) \right)^2.$$

Furthermore, due to the independence assumptions on $X_{i,j}$, we obtain

$$\text{msep}(p) = \text{msep}_{R|D_I}(\hat{R}) = \sum_{i+j > I} \text{Var}(X_{i,j}) + \left( \sum_{i+j > I} (\hat{\mu}_i \hat{\gamma}_j - E[X_{i,j}]) \right)^2.$$

Staying within the framework of Tweedie’s EDF, we see that as with the claims reserves, the estimate of the MSEP depends on the model parameter $p$. Most often in presenting results, the square root of the MSEP is given, we also follow this convention. Denoting the estimators for the process variance by $\hat{\text{pv}}(p)$ and for the estimation error as $\hat{\text{ee}}(p)$, we decompose the estimated (conditional) MSEP as follows:

$$\left( \hat{\text{msep}}^\frac{1}{2}(p) \right)' = \frac{1}{2 \hat{\text{msep}}^\frac{1}{2}(p)} \left( \hat{\text{pv}}(p)' + \hat{\text{ee}}(p)' \right),$$

where

$$\hat{\text{msep}}^\frac{1}{2}(p) = \left( \hat{\text{pv}}(p) + \hat{\text{ee}}(p) \right)^\frac{1}{2}.$$

The first order Taylor expansion of $\hat{\text{msep}}^\frac{1}{2}(p)$ around $p$ is given by

$$\hat{\text{msep}}^\frac{1}{2}(p + \varepsilon) = \hat{\text{msep}}^\frac{1}{2}(p) + (\hat{\text{msep}}^\frac{1}{2}(p))' \varepsilon. \quad (7)$$

The derivates of the estimators of the process variance and the estimation error are provided below in Lemma 5.1 and 5.2. Note that we can also find the second order Taylor expansion,

$$\hat{\text{msep}}^\frac{1}{2}''(p + \varepsilon) = \hat{\text{msep}}^\frac{1}{2}(p) + (\hat{\text{msep}}^\frac{1}{2}(p))' \varepsilon + (\hat{\text{msep}}^\frac{1}{2}(p))'' \varepsilon^2 / 2; \quad (8)$$

this requires the calculation of $\hat{\text{pv}}(p)''$ and $\hat{\text{ee}}(p)''$, which is rather involved. Some formulas required for the second order approximation are given in Appendix C. We highlight the performance of the second order approximation in the case study of Section 6.
5.1 Process Variance

Determining an estimate of the process variance is relatively easy. Indeed,

$$\text{Var}\left( \sum_{i+j>I} X_{i,j} \right) = \sum_{i+j>I} \text{Var}(X_{i,j}) = \sum_{i+j>I} \phi(\mu_i \gamma_j)^p.$$ 

To estimate this quantity we replace the parameters by their estimates, which gives

$$\hat{p}_v(p) = \hat{\text{Var}}\left( \sum_{i+j>I} X_{i,j} \right) = \sum_{i+j>I} \hat{\phi}(\hat{\mu}_i \hat{\gamma}_j)^p.$$ 

Note that $\hat{\mu}_i$, $\hat{\gamma}_j$, as well as $\hat{\phi}$ are functions of $p$. To obtain the derivative of $\hat{p}_v(p)$, we start by obtaining the derivative of $\hat{\phi}$ with respect to $p$:

$$\hat{\phi}(p)' = \frac{1}{d} \sum_{i+j \leq I} \left( y'_{i,j} (x_{i,j} - \hat{\mu}_i \hat{\gamma}_j)^2 - 2y_{i,j} (x_{i,j} - \hat{\mu}_i \hat{\gamma}_j)(\hat{\mu}_i' \hat{\gamma}_j + \hat{\mu}_i \hat{\gamma}_j') \right),$$

where $y_{i,j} = (\hat{\mu}_i \hat{\gamma}_j)^{-p}$, and $y'_{i,j} = -y_{i,j} \log(\hat{\mu}_i \hat{\gamma}_j) - p y_{i,j}^{p+1} (\hat{\mu}_i' \hat{\gamma}_j + \hat{\mu}_i \hat{\gamma}_j')$. Using the above we obtain the following lemma.

**Lemma 5.1**

The derivate of $\hat{p}_v(p)$ with respect to $p$ is given by

$$\hat{p}_v(p)' = \frac{\partial}{\partial p} \hat{\text{Var}}\left( \sum_{i+j>I} X_{i,j} \right) = \sum_{i+j>I} \left( \frac{\hat{\phi}'}{y_{i,j}} - \frac{y'_{i,j}}{y_{i,j}^2} \phi \right),$$

where $y_{i,j}$ is defined as given above.

5.2 Estimation Error

It is standard to estimate the estimation error by its expected value; see e.g England and Verrall (2002). Hence, we estimate

$$\left( \sum_{i+j>I} (\hat{\mu}_i \hat{\gamma}_j - E[X_{i,j}]) \right)^2$$

by

$$E \left[ \left( \sum_{i+j>I} (\hat{\mu}_i \hat{\gamma}_j - E[X_{i,j}]) \right)^2 \right] = \sum_{i+j>I} \sum_{k+l>I} E \left[ (\hat{\mu}_k \hat{\gamma}_l - E[X_{k,l}]) (\hat{\mu}_i \hat{\gamma}_j - E[X_{i,j}]) \right].$$
Note that the predictor \( \hat{\mu}_i \hat{\gamma}_j \) is not necessarily unbiased for \( E[X_{i,j}] = \mu_i \gamma_j \). This bias is for typical claims and is generally negligible. One uses the approximation
\[
\left( \sum_{i+j>l} (\hat{\mu}_i \hat{\gamma}_j - E[X_{i,j}]) \right)^2 \approx \sum_{i+j>|I|} \text{Cov}(\hat{\mu}_i \hat{\gamma}_j, \hat{\mu}_k \hat{\gamma}_l);
\]
see e.g. Wüthrich and Merz (2008), Section 6.4.3. Note that this method of approximation corresponds to using the unconditional MSEP. As an estimator of the above covariances we use
\[
\text{Cov}(\hat{\mu}_i \hat{\gamma}_j, \hat{\mu}_k \hat{\gamma}_l) = \gamma_j \gamma_l \text{Cov}(\hat{\mu}_i, \hat{\mu}_k) + \hat{\mu}_k \hat{\gamma}_j \text{Cov}(\hat{\mu}_i, \hat{\gamma}_l) + \hat{\mu}_l \hat{\gamma}_j \text{Cov}(\hat{\mu}_k, \hat{\gamma}_l);
\]
see Appendix B for details. Note that the above approximation requires small relative errors of the parameter estimates. The estimated covariance terms on the right-hand side of the above equality are provided in (3). We hence obtain
\[
\hat{e}(p) = E\left[ \left( \sum_{i+j>|I|} (\hat{\mu}_i \hat{\gamma}_j - E[X_{i,j}]) \right)^2 \right] = \sum_{i+j>|I|} \text{Cov}(\hat{\mu}_i \hat{\gamma}_j, \hat{\mu}_k \hat{\gamma}_l),
\]
which has as its derivative
\[
\hat{e}(p) = \sum_{i+j>|I|} \sum_{k>|I|} (\hat{\gamma}_j \hat{\gamma}_l) \text{Cov}(\hat{\mu}_i, \hat{\mu}_k) + \hat{\gamma}_j \hat{\gamma}_l \frac{\partial}{\partial p} \text{Cov}(\hat{\mu}_i, \hat{\mu}_k)
\]
\[
+ (\hat{\mu}_i \hat{\gamma}_j + \hat{\mu}_k \hat{\gamma}_l) \text{Cov}(\hat{\mu}_i, \hat{\gamma}_l) + \hat{\mu}_l \hat{\gamma}_j \frac{\partial}{\partial p} \text{Cov}(\hat{\mu}_k, \hat{\gamma}_l)
\]
\[
+ (\hat{\mu}_i \hat{\gamma}_l + \hat{\mu}_k \hat{\gamma}_j) \text{Cov}(\hat{\mu}_k, \hat{\gamma}_j) + \hat{\mu}_l \hat{\gamma}_l \frac{\partial}{\partial p} \text{Cov}(\hat{\mu}_i, \hat{\gamma}_l).
\]
To obtain the above we need only provide the derivatives with respect to \( p \) of the covariances of the MLEs. This involves the estimated Fisher information matrix \( \hat{H} = H(\hat{\theta}, \hat{\phi}) \). Denote with \( \hat{H} \) the matrix containing the derivatives of the estimated covariances:
\[
\hat{H} = \frac{\partial}{\partial p} \hat{H}^{-1} = -\hat{H}^{-1} \frac{\partial \hat{H}}{\partial p} \hat{H}^{-1}.
\]
The derivatives of the entries of \( \hat{H} \) with respect to \( p \) are given as follows, see (2),
\[
\frac{\partial \hat{H}}{\partial p}_{r,r} = -\frac{\hat{\phi}}{\hat{\phi}^2} \sum_{i=1}^{2l+1} \hat{c}_r \hat{c}_r + \frac{1}{\hat{\phi}} \sum_{i=1}^{2l+1} \left( \frac{\partial \hat{c}_r}{\partial p} \hat{c}_r + \hat{c}_r \frac{\partial \hat{c}_r}{\partial p} \right), \quad r \in \{1, \ldots, l\},
\]
\[
\frac{\partial \hat{H}}{\partial p}_{s,s} = -\frac{\hat{\theta}}{\hat{\phi} \hat{\phi}^2} \sum_{r=1}^{2l+1-1} \hat{c}_r \hat{c}_r + \frac{1}{\hat{\phi}} \sum_{r=1}^{2l+1-1} \left( \frac{\partial \hat{c}_r}{\partial p} \hat{c}_r + \hat{c}_r \frac{\partial \hat{c}_r}{\partial p} \right), \quad s \in \{l+1, \ldots, 2l+1\},
\]
\[
\frac{\partial \hat{H}}{\partial p}_{r,s} = -\frac{\hat{\phi}}{\hat{\phi}^2} \sum_{i=1}^{2l+1-1} \hat{c}_r \hat{c}_r + \frac{1}{\hat{\phi}} \left( \frac{\partial \hat{c}_r}{\partial p} \hat{c}_r + \hat{c}_r \frac{\partial \hat{c}_r}{\partial p} \right), \quad r \in \{1, \ldots, 2l+1\},
\]
\[
s \in \{1, \ldots, 2l+1-r\} \text{ and } \{r, s\} \notin \{1, \ldots, l\} \times \{1, \ldots, l\}.
\]
where

$$\frac{\partial \hat{c}_r^p}{\partial p} = -\log \hat{c}_r \hat{c}_r^{-p} + (-p)\hat{c}_r \hat{c}_r^{-1-p},$$

for \( r \in \{1, \ldots, 2I + 1\} \). Using the above, we obtain the following lemma.

**Lemma 5.2**
The derivative of \( \hat{e}c(p) \) with respect to \( p \) is given by,

$$\hat{e}c(p)' = \sum_{i,j \geq I, k,l \geq 0} \left( \hat{c}_i \hat{c}_j \hat{H}_{i,j+1}^{-1} + \hat{c}_i \hat{c}_j \hat{H}_{i,j+1}^{-1} + \hat{c}_i \hat{c}_j \hat{H}_{i,j+1}^{-1} + \hat{c}_i \hat{c}_j \hat{H}_{i,j+1}^{-1} + \hat{c}_i \hat{c}_j \hat{H}_{i,j+1}^{-1} + \hat{c}_i \hat{c}_j \hat{H}_{i,j+1}^{-1} \right),$$

where \( \hat{H} \) and \( \hat{H} \) are defined above, \( i, k \in \{1, \ldots, I\} \) and \( j, l \in \{0, \ldots, I\} \).

Note: The second order Taylor expansion (8) is provided in Appendix C.

### 6 Case Study

We analyze the standard dataset from Wüthrich and Merz (2008), see Table 1 below. We center our examples around \( p = 1 \) and \( p = 2 \), corresponding to the overdispersed Poisson and the gamma distributions. The claims reserves and the MSEP for these models are attainable with relative ease using most standard statistical software packages. For the reserves and the MSEP for values other than \( p = 1, 2 \), statistical software R was used, but this ability is not standard. Additional non-trivial programming was done to allow us this functionality.

<table>
<thead>
<tr>
<th>( i/j )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5,947</td>
<td>3,721</td>
<td>896</td>
<td>208</td>
<td>207</td>
<td>62</td>
<td>66</td>
<td>15</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>1</td>
<td>6,347</td>
<td>3,246</td>
<td>723</td>
<td>152</td>
<td>68</td>
<td>37</td>
<td>53</td>
<td>11</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6,269</td>
<td>2,976</td>
<td>847</td>
<td>263</td>
<td>153</td>
<td>65</td>
<td>54</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5,863</td>
<td>2,683</td>
<td>723</td>
<td>191</td>
<td>133</td>
<td>88</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5,779</td>
<td>2,745</td>
<td>654</td>
<td>274</td>
<td>230</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6,185</td>
<td>2,228</td>
<td>573</td>
<td>245</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5,600</td>
<td>2,893</td>
<td>563</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5,288</td>
<td>2,440</td>
<td>528</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5,291</td>
<td>2,358</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5,676</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Observed incremental payments \( X_{i,j} \) given in ’000s.

#### 6.1 Estimating with \( p = 1 \), the Overdispersed Poisson Distribution

**Claims Reserves**

Using the statistical software R we obtain the MLEs of the underlying parameters \( \mu_i \) and \( \gamma_j \). They, as well as the first and second derivatives are found in Table 2.
Note that for $p = 1$, the estimates of the underlying parameters could also have been obtained using the classical chain ladder method; see e.g. Corollary 2.18 in Wüthrich and Merz (2008). Recall the admissible values of $p$, $p \in (-\infty, 0] \cup [1, \infty)$, implying that we study positive $\varepsilon$ when using $p = 1$ as the focal point of our approximation.

\begin{table}[h]
\centering
\begin{tabular}{|c|cccc|cccc|}
\hline
$i/j$ & $\mu_i$ & $\mu_i'$ & $\mu_i''$ & $\gamma_j$ & $\gamma_j'$ & $\gamma_j''$ \\
\hline
0 & 1.000 & 0.000 & 0.000 & & & & \\
1 & 0.957 & -0.118 & -0.203 & 6,572,762 & 418,447 & 282,213 \\
2 & 0.956 & -0.039 & -0.001 & 762,835 & 41,156 & 24,929 \\
3 & 0.875 & -0.056 & -0.011 & 241,836 & 14,891 & 6,839 \\
4 & 0.886 & -0.004 & 0.138 & 56,870 & 2,931 & 3,305 \\
5 & 0.905 & -0.091 & -0.067 & 12,002 & 603 & 828 \\
6 & 0.957 & -0.060 & -0.027 & 11,641 & 711 & 1,388 \\
7 & 0.863 & -0.055 & -0.036 & 5 & 0 & 0 \\
\hline
\end{tabular}
\caption{Estimates of parameters $\mu$ and $\gamma$ and their derivatives for the case $p = 1$.}
\end{table}

The claims reserves under the assumption that the Tweedie exponential dispersion model has $p = 1$, i.e. that it is overdispersed Poisson distributed, is $6,047,059$. In Table 3 we highlight the performance of the approximation in relation to the true values under the various levels $p$, notice the accuracy of the second order approximation. Figure 2 presents these results graphically. Moreover, the claims reserves $\hat{R}(p)$ are rather insensitive to the choice of $p$.

\begin{table}[h]
\centering
\begin{tabular}{|c|cccc|cccc|}
\hline
$p$ & Exact & 1st Order & 2nd Order & Exact & 1st Order & 2nd Order \\
& Reserve & Approx. & Approx. & $\text{ms}e_{\text{p}}$ & Approx. & Approx. \\
\hline
1.00 & 6,047,059 & 6,047,059 & 6,047,059 & 429,891 & 429,891 & 429,891 \\
1.05 & 6,043,385 & 6,043,459 & 6,043,386 & 430,943 & 429,187 & 431,047 \\
1.10 & 6,039,560 & 6,039,859 & 6,039,568 & 435,395 & 428,484 & 435,922 \\
1.15 & 6,035,777 & 6,036,259 & 6,035,603 & 441,108 & 427,781 & 441,516 \\
1.20 & 6,031,429 & 6,031,942 & 6,031,492 & 447,967 & 426,374 & 447,861 \\
1.25 & 6,027,113 & 6,027,236 & 6,027,060 & 453,986 & 424,967 & 454,516 \\
1.30 & 6,022,621 & 6,022,833 & 6,022,736 & 459,023 & 423,467 & 459,612 \\
1.35 & 6,017,951 & 6,018,285 & 6,018,180 & 465,108 & 422,067 & 465,922 \\
1.40 & 6,013,100 & 6,013,591 & 6,013,459 & 471,108 & 419,667 & 472,016 \\
1.45 & 6,008,070 & 6,008,751 & 6,008,600 & 477,108 & 417,267 & 477,861 \\
1.50 & 6,002,865 & 6,003,765 & 6,003,600 & 483,108 & 414,867 & 484,516 \\
1.55 & 5,997,497 & 6,003,603 & 6,003,459 & 489,108 & 412,467 & 490,016 \\
1.60 & 5,991,983 & 6,003,860 & 6,003,735 & 495,108 & 410,067 & 495,516 \\
1.70 & 5,980,624 & 5,996,660 & 5,996,562 & 507,108 & 405,267 & 508,516 \\
1.75 & 5,974,856 & 5,993,060 & 5,993,000 & 513,108 & 402,867 & 514,016 \\
1.80 & 5,969,088 & 5,989,461 & 5,989,400 & 519,108 & 400,467 & 519,516 \\
1.90 & 5,957,772 & 5,982,261 & 5,982,200 & 531,108 & 395,667 & 531,516 \\
1.95 & 5,952,316 & 5,978,661 & 5,978,600 & 537,108 & 393,267 & 537,516 \\
\hline
\end{tabular}
\caption{True and approximated claims reserves and $\text{MSEP}^2$.}
\end{table}
Prediction Uncertainty

Table 3 highlights the results of the MSEP\textsuperscript{1/2} approximation using \( p = 1 \). Figure 3 presents these results graphically. It is evident that the MSEP\textsuperscript{1/2} is not stable in \( p \), one cannot deviate too far from \( p = 1 \) (i.e. \( \varepsilon \) cannot be too far from 0) when using the first order approximation. Furthermore, it seems that the MSEP\textsuperscript{1/2} is near a local minimum at \( p = 1 \). Under the assumption of the overdispersed Poisson distribution, the estimates of the dispersion parameter \( \phi(p) \) and its derivatives were found to be, \( \hat{\phi}(1) = 14,714 \), \( \hat{\phi}'(1) = -197,314 \) and \( \hat{\phi}''(1) = 2,678,513 \), where the estimation was done based upon the Pearson residuals. The second order approximation far outperforms the first, but as previously stated is also more strenuous to calculate.

6.2 Estimating with \( p = 2 \), the Gamma Distribution

Claims Reserves

The MLEs of the underlying parameters and their first and second order derivates are presented in Table 4. The claims reserves under the assumption that the Tweedie exponential dispersion model has \( p = 2 \) is 5,947,049. In Table 5 we highlight the performance of the approximation for the claims reserves in relation to the true values under the various levels \( p \). Figure 4 presents these results graphically.

Prediction Uncertainty

The approximations of the MSEP using the gamma distribution are presented in Table 5. Figure 5 presents these results graphically. Under the assumption of
the gamma distribution, the estimates of the dispersion parameter, $\phi(p)$, and its
derivative were found to be, $\hat{\phi}(2) = 0.04497$, $\hat{\phi}'(2) = -0.54747$ and $\hat{\phi}''(2) = 6.72616$,
where again, the estimation was done based upon the Pearson residuals.

Remark 6.1 Notice that in addition to the fact that due to our boundary con-
dition of $\hat{\mu}_0 = 1$, all derivatives of $\hat{\mu}_0$ equal zero, also all derivatives of $\hat{\gamma}_I$ equal zero.
This is the case since $\hat{\gamma}_I = X_{0,I}$, which is $D_I$-f, i.e. $\hat{\gamma}_I$ is constant. This fact shows
up clearly in Tables 2 and 4.

Conclusions

We have studied the sensitivity of the claims reserves and the estimate of the MSEP
within the Tweedie EDF. The ability to express these quantities in terms of the

Figure 3: True and approximated MSEP$^{1/2}$.

<table>
<thead>
<tr>
<th>$i/j$</th>
<th>$\hat{\mu}_0$</th>
<th>$\hat{\mu}_0'$</th>
<th>$\hat{\mu}_0''$</th>
<th>$\hat{\gamma}_I$</th>
<th>$\hat{\gamma}_I'$</th>
<th>$\hat{\gamma}_I''$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>6,999,574</td>
<td>103,394</td>
<td>-1,316,401</td>
</tr>
<tr>
<td>1</td>
<td>0.760</td>
<td>-0.111</td>
<td>0.629</td>
<td>3,426,601</td>
<td>33,285</td>
<td>-636,971</td>
</tr>
<tr>
<td>2</td>
<td>0.900</td>
<td>-0.094</td>
<td>-0.109</td>
<td>800,954</td>
<td>259</td>
<td>-138,198</td>
</tr>
<tr>
<td>3</td>
<td>0.850</td>
<td>0.062</td>
<td>0.243</td>
<td>252,086</td>
<td>-7,201</td>
<td>-16,475</td>
</tr>
<tr>
<td>4</td>
<td>1.052</td>
<td>0.414</td>
<td>0.409</td>
<td>161,788</td>
<td>-15,077</td>
<td>-29,717</td>
</tr>
<tr>
<td>5</td>
<td>0.809</td>
<td>-0.037</td>
<td>0.240</td>
<td>77,394</td>
<td>-7,532</td>
<td>-16,475</td>
</tr>
<tr>
<td>6</td>
<td>0.811</td>
<td>0.019</td>
<td>0.241</td>
<td>61,418</td>
<td>3,227</td>
<td>-13,523</td>
</tr>
<tr>
<td>7</td>
<td>0.709</td>
<td>-0.041</td>
<td>0.127</td>
<td>13,159</td>
<td>1,090</td>
<td>-3,232</td>
</tr>
<tr>
<td>8</td>
<td>0.722</td>
<td>-0.021</td>
<td>0.137</td>
<td>13,226</td>
<td>1,409</td>
<td>-5,392</td>
</tr>
<tr>
<td>9</td>
<td>0.811</td>
<td>-0.012</td>
<td>0.153</td>
<td>15,813</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4: Estimates of parameters $\mu$ and $\gamma$ and their derivatives for the case $p = 2$. 
model parameter $p$, and then taking derivates, have allowed us to develop Taylor approximations. In our case study we highlighted the performance of these approximations and furthermore found that the reserves were rather insensitive to model selection within the Tweedie EDF, in contrast to the MSEP, which varied widely. These empirical findings confirm the results in Peters et al. (2009).
Figure 5: True and approximated $\text{MSEP}^{1/2}$.

Acknowledgements

Both authors are grateful to Prof. Dr. Paul Embrechts for providing valuable insight related to this paper. Furthermore, the first author is greatly indebted to ACE Limited for providing the financial support necessary for the completion of this work.

Bibliography


A Derivatives of $A$, $\alpha$ and $\beta$

In this appendix, we provide the derivatives of the entries of the matrix $A$ and the derivates of the $\alpha$ and $\beta$, required in Section 4.2.

$$
a'_{i,i} = \sum_{j=0}^{I-1} \gamma_j^2 \frac{\hat{\gamma}_j - (2-p)\hat{\gamma}_j'}{\hat{\gamma}_j}, \quad i \in \{1, \ldots, I\},
$$

$$
a'_{I+j+1,I+j+1} = \sum_{i=0}^{I-j} \mu_i \frac{\hat{\mu}_i - (2-p)\hat{\mu}_i'}{\hat{\mu}_i}, \quad j \in \{0, \ldots, I\},
$$

$$
a'_{i,I+j+1} = (2-p)\hat{\mu}_i \gamma_j^2 \frac{\hat{\gamma}_j - (1-p)\hat{\gamma}_j'}{\hat{\gamma}_j} + (-\hat{\mu}_i + (2-p)\hat{\mu}_i') \gamma_j^2 \frac{(-p)\hat{\gamma}_j'}{\hat{\gamma}_j} - (1-p) x_{i,j} \gamma_j^2 \frac{\hat{\gamma}_j - (1-p)\hat{\gamma}_j'}{\hat{\gamma}_j} - x_{i,j} \gamma_j^2 \frac{(1-p)\hat{\gamma}_j'}{\hat{\gamma}_j}, \quad i \in \{0, \ldots, I\}, j \in \{0, \ldots, I - i\},
$$

$$
a'_{I+j+1,i} = (2-p)\hat{\gamma}_j \mu_i \gamma_j^2 \frac{\hat{\gamma}_j - (1-p)\hat{\gamma}_j'}{\hat{\gamma}_j} + (-\hat{\gamma}_j + (2-p)\hat{\gamma}_j') \hat{\mu}_i \gamma_j^2 \frac{(-p)\hat{\gamma}_j'}{\hat{\gamma}_j} - (1-p) x_{i,j} \hat{\mu}_i \gamma_j^2 \frac{\hat{\gamma}_j - (1-p)\hat{\gamma}_j'}{\hat{\gamma}_j} - x_{i,j} \hat{\mu}_i \gamma_j^2 \frac{(1-p)\hat{\gamma}_j'}{\hat{\gamma}_j}, \quad j \in \{0, \ldots, I\}, i \in \{0, \ldots, I - j\},
$$

and,

$$
\alpha'_i = \sum_{j=0}^{I-1} \left[ \log \hat{\gamma}_j \mu_i \left( \gamma_j^2 \frac{\hat{\gamma}_j - (2-p)\hat{\gamma}_j'}{\hat{\gamma}_j} \right) + \left( \hat{\gamma}_j \mu_i + \log \hat{\gamma}_j \hat{\mu}_i \right) \gamma_j^2 \frac{\hat{\gamma}_j - (2-p)\hat{\gamma}_j'}{\hat{\gamma}_j} \right], \quad i \in \{1, \ldots, I\},
$$

$$
\beta'_j = \sum_{i=0}^{I-j} \left[ \log \hat{\mu}_i \gamma_j \left( \mu_i \gamma_j^2 \frac{\hat{\mu}_i - (2-p)\hat{\mu}_i'}{\hat{\mu}_i} \right) + \left( \hat{\mu}_i \gamma_j + \log \hat{\mu}_i \hat{\gamma}_j \right) \mu_i \gamma_j^2 \frac{\hat{\mu}_i - (2-p)\hat{\mu}_i'}{\hat{\mu}_i} \right], \quad j \in \{0, \ldots, I\}.
$$

B Covariance Approximation

In this appendix, we aim to show that,

$$
\text{Cov}(\hat{\mu}_i \hat{\gamma}_j, \hat{\mu}_k \hat{\gamma}_l) \approx \hat{\gamma}_j \hat{\gamma}_l \text{Cov}(\hat{\mu}_i, \hat{\mu}_k) + \hat{\mu}_k \hat{\gamma}_j \text{Cov}(\hat{\mu}_i, \hat{\gamma}_l) + \hat{\mu}_i \hat{\gamma}_l \text{Cov}(\hat{\gamma}_j, \hat{\gamma}_l).
$$
We begin as follows,
\[
\text{Cov}(\hat{\mu}_i \hat{\gamma}_j, \hat{\mu}_k \hat{\gamma}_l) = \text{Cov}(\exp\{\log\hat{\mu}_i \hat{\gamma}_j\}, \exp\{\log\hat{\mu}_k \hat{\gamma}_l\})
\approx \mu_i \gamma_j \mu_k \gamma_l \text{Cov}(1 + \log\hat{\mu}_i \hat{\gamma}_j - \mu_i \gamma_j, 1 + \log\hat{\mu}_k \hat{\gamma}_l - \mu_k \gamma_l)
= \mu_i \gamma_j \mu_k \gamma_l \left(\text{Cov}(\log\hat{\mu}_i, \log\hat{\mu}_k) + \text{Cov}(\log\hat{\gamma}_j, \log\hat{\gamma}_l)\right)
+ \text{Cov}(\log\hat{\gamma}_j, \log\hat{\mu}_k) + \text{Cov}(\log\hat{\gamma}_l, \log\hat{\mu}_i),
\]
where we have used the linearization \(\exp(z) \approx 1 + z\) for \(z \approx 0\). We proceed with the covariance terms remaining on the right-hand side. The calculations are analogous, hence we only provide the details for approximating \(\text{Cov}(\log\hat{\mu}_i, \log\hat{\mu}_k)\).

\[
\text{Cov}(\log\hat{\mu}_i, \log\hat{\mu}_k) = \text{Cov}(1 + \log\hat{\mu}_i - \log\mu_i, 1 + \log\hat{\mu}_k - \log\mu_k)
\approx \text{Cov}(\exp\{\log\hat{\mu}_i - \log\mu_i\}, \exp\{\log\hat{\mu}_k - \log\mu_k\})
= \frac{1}{\mu_i \mu_k} \text{Cov}(\hat{\mu}_i, \hat{\mu}_k),
\]
where we again have used the linearization \(\exp(z) \approx 1 + z\) for \(z \approx 0\). As a final step of the approximation, we replace unknown variables and unknown covariances with estimates thereof.

C MSEP Approximation using Second Order Taylor Expansion

In this section we provide the main formulas for the second order Taylor expansion (8).

\[
\hat{\text{mse}}^2 p = \text{mse}^2 p + (\text{mse}^2 p)' \varepsilon + (\text{mse}^2 p)'' \varepsilon^2.
\]

\[
(\text{mse}^2 p)' = -\frac{1}{4} \text{mse}^{-\frac{3}{2}} (p) \left(\hat{\nu}(p)' + \hat{\epsilon}(p)\right) + \frac{1}{2} \text{mse}^{-\frac{1}{2}} (p) \left(\hat{\nu}(p)'' + \hat{\epsilon}(p)''\right).
\]

\[
\hat{\nu}(p)' = \sum_{i+j>1} \left(\frac{1}{y_{i,j}} \hat{\phi}' - 2 \frac{y_{i,j} \hat{\phi}'}{y_{i,j}^2} \hat{\phi}' + \left(2 \frac{(y_{i,j})^2}{y_{i,j}^2} - \frac{y_{i,j}'}{y_{i,j}^2}\right) \hat{\phi}'\right),
\]

\[
\hat{\nu}(p)'' = \frac{1}{d} \sum_{i+j<1} \left(\frac{y_{i,j}'}{y_{i,j}} (x_{i,j} - \hat{\mu}_{i,j})^2 - 4 y_{i,j}' (x_{i,j} - \hat{\mu}_{i,j})(\hat{\mu}_{i,j}' + \hat{\mu}_{i,j}'')\right)
+ 2 y_{i,j} \left(\hat{\mu}_{i,j}' + \hat{\mu}_{i,j}'\right)^2 - (x_{i,j} - \hat{\mu}_{i,j})(\hat{\mu}_{i,j}'' + 2 \hat{\mu}_{i,j}' + \hat{\mu}_{i,j}'')\right).
\]
\[ y^p_{i,j} = -y^p_{i,j} \left( \log(\hat{\mu}_i\gamma_j) + (p+1) y^p_{i,j}(\hat{\mu}'_i\gamma_j + \hat{\mu}_i\gamma''_j) \right) \]

\[ -y^{p+1}_{p+1} \left( 2(\hat{\mu}'_i\gamma_j + \hat{\mu}_i\gamma''_j) + p(\hat{\mu}'_i\gamma_j + 2\hat{\mu}'_i\gamma_j + \hat{\mu}_i\gamma''_j) \right). \]

\[ \phi(p)^{\prime\prime} = \sum_{i+j>t} \frac{1}{1+i} \left( \gamma^p_{i,j} + 2\gamma^p_{i,j} + \gamma^p_{i,j} \right) \hat{H}_{i,k}^{-1} + 2\gamma^p_{i,j} + \gamma^p_{i,j} \hat{H}_{i,k} + \gamma^p_{i,j} \hat{H}_{i,k} \]

\[ + (\hat{\mu}'_k\gamma_j + 2\hat{\mu}'_k\gamma_j + \hat{\mu}_k\gamma''_j) \hat{H}_{i+1,j+1}^{-1} + 2(\hat{\mu}'_k\gamma_j + \hat{\mu}_k\gamma''_j) \hat{H}_{i+1,j+1} + \hat{\mu}_k\gamma''_j \hat{H}_{i+1,j+1} \]

\[ + (\hat{\mu}'_k\gamma_j + 2\hat{\mu}'_k\gamma_j + \hat{\mu}_k\gamma''_j) \hat{H}_{i+1,j+1} + 2(\hat{\mu}'_k\gamma_j + \hat{\mu}_k\gamma''_j) \hat{H}_{i+1,j+1} + \hat{\mu}_k\gamma''_j \hat{H}_{i+1,j+1}. \]

\[ \hat{g} = 2 \hat{H}^{-1} \frac{\partial \hat{H}}{\partial p} \hat{H}^{-1} \frac{\partial \hat{H}}{\partial p} \hat{H}^{-1} - \hat{H}^{-1} \frac{\partial \hat{H}}{\partial p} \hat{H}^{-1}. \]

\[ \frac{\partial \hat{H}}{\partial p}_{r,r} = \frac{-\frac{\partial^2}{\partial p^2} \sum_{s=1}^{21+1+r-r} \frac{1}{\sigma^2} \sum_{r=0}^{21+1+s} \frac{1}{\sigma^2} \sum_{r=0}^{21+1+s} \left( \frac{\partial^2 \hat{e}^2 - p + \hat{\mu} \hat{G}_2}{\partial p^2} \hat{e}^2 - p + \hat{\mu} \hat{G}_2 \right) \right) \]

\[ + \frac{1}{\sigma^2} \sum_{r=0}^{21+1+s} \left( \frac{\partial^2 \hat{e}^2 - p + \hat{\mu} \hat{G}_2}{\partial p^2} \hat{e}^2 - p + \hat{\mu} \hat{G}_2 \right), \quad r \in \{1, \ldots, I\}. \]

\[ \frac{\partial \hat{H}}{\partial p}_{s,s} = \frac{-\frac{\partial^2}{\partial p^2} \sum_{s=0}^{21+1-s} \frac{1}{\sigma^2} \sum_{r=0}^{21+1-s} \frac{1}{\sigma^2} \sum_{r=0}^{21+1-s} \left( \frac{\partial^2 \hat{e}^2 - p + \hat{\mu} \hat{G}_2}{\partial p^2} \hat{e}^2 - p + \hat{\mu} \hat{G}_2 \right) \right) \]

\[ + \frac{1}{\sigma^2} \sum_{r=0}^{21+1-s} \left( \frac{\partial^2 \hat{e}^2 - p + \hat{\mu} \hat{G}_2}{\partial p^2} \hat{e}^2 - p + \hat{\mu} \hat{G}_2 \right), \quad s \in \{I+1, \ldots, 2I+1\}. \]

\[ \frac{\partial \hat{H}}{\partial p}_{r,s} = \frac{-\frac{\partial^2}{\partial p^2} \sum_{r=0}^{21+1-r} \frac{1}{\sigma^2} \sum_{r=0}^{21+1-s} \frac{1}{\sigma^2} \sum_{r=0}^{21+1-s} \left( \frac{\partial^2 \hat{e}^2 - p + \hat{\mu} \hat{G}_2}{\partial p^2} \hat{e}^2 - p + \hat{\mu} \hat{G}_2 \right) \right) \]

\[ + \frac{1}{\sigma^2} \left( \frac{\partial^2 \hat{e}^2 - p + \hat{\mu} \hat{G}_2}{\partial p^2} \hat{e}^2 - p + \hat{\mu} \hat{G}_2 \right), \quad r \in \{1, \ldots, 2I+1\}, \quad s \in \{1, \ldots, 2I+1-r\} \text{ and } (r,s) \notin \{1, \ldots, I\} \times \{1, \ldots, I\}. \]

\[ \frac{\partial^2 \hat{\zeta}^{-p}}{\partial p^2} = -2\hat{\zeta}^{-1-p} - \log \hat{\zeta} \frac{\partial \hat{\zeta}^{-p}}{\partial p} + (-p) \left( \frac{\partial \hat{\zeta}^{-1-p}}{\partial p} + \frac{\partial^2 \hat{\zeta}^{-1-p}}{\partial p^2} \right). \]
Daniel H. Alai and
Mario V. Wüthrich (2009).
Modelling small and large claims in a chain
ladder framework.
Submitted.
Modelling Small and Large Claims in a Chain Ladder Framework

Daniel H. Alai Mario V. Wüthrich
Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland

Abstract

The chain ladder (CL) method is a commonly used algorithm that produces an estimator for insurance claims reserves. The prediction uncertainty of this estimator can be quantified by its mean square error of prediction (MSEP). The frequent use of the CL algorithm is a direct consequence of its simplicity. In order to improve claims reserving prediction, practitioners often separate small and large claims and then apply a two-dimensional CL algorithm to the two claims classes. We give a mathematical framework for this two-dimensional analysis that allows to quantify the prediction uncertainty.

Keywords: Claims Reserving, Chain Ladder Method, Conditional Mean Square Error of Prediction, Small and Large Claims.
1 Introduction

One of the main tasks of actuaries is to predict the outstanding loss liabilities, the predictor of which is referred to as the claims reserves. As well as determining the claims reserves, actuaries need to quantify the associated prediction uncertainty of these reserves; see e.g. Taylor (2000) and Teugels and Sundt (2004).

There are many methods available to actuaries that produce claims reserves as well as their associated prediction uncertainty (quantified by their MSEP); see e.g. England and Verrall (2002). Arguably, the most popular claims reserving method today is the chain ladder (CL) algorithm. The frequent use of the CL algorithm is a direct consequence of its simplicity. If the underlying outstanding loss liabilities are too heterogeneous, practitioners often divide the portfolio into two subportfolios, one containing small claims, the other large claims. They then apply the CL algorithm to each subportfolio. The importance of such separations has been studied before; see e.g. Klemmt (2005). In general, one obtains more stable behaviour in the subportfolios. The application of a two-dimensional CL algorithm needs some care because the separation does not satisfy the additivity property; see e.g. Anje (1994).

The above can be described as a multivariate CL algorithm. Prediction uncertainty for a multivariate CL method has been considered in the literature by Braun (2004), Pröhl and Schmidt (2005), Schmidt (2006), and Merz and Wüthrich (2008a). These contributions were limited to univariate estimators with associated MSEP (i.e. second moment calculation) and multivariate estimators lacking MSEP analysis. Merz and Wüthrich (2008b) bridge the gap and provide dependence structure as well as MSEP, however, they define the correlation through a deterministic dependence matrix whereas we define the dependence naturally through the claims process.

We analyze the two-dimensional CL model for the small-large claims separation and provide a formula for the prediction uncertainty for this model. We hence provide those practitioners who separate small and large claims with the necessary tools required to quantify the prediction uncertainty of the resulting claims reserves.

Organization of the paper. In Section 2 we provide the notation and data structure. In Section 3 we give a short review of the univariate CL method followed by, in Section 4, our interpretation of a multivariate CL method. Section 5 is dedicated to defining estimators for the (conditional) MSEP under the multivariate CL model assumptions. In Section 6 we provide a case study.

2 Notation and Data Structure

We assume claims development data have the form of run-off triangles. However, note that the data need not be triangular, the methods outlined in this paper
can also be used for claims development trapezoids. Claims with values less than some predetermined threshold will be referred to as small claims, those with values above the threshold, large claims. We separate the data into three triangles: one that holds information about all claims but capped at some threshold (the small layer), superscripted with \((s)\); another about the excess layer (the large layer), superscripted with \((l)\); and a third triangle that represents the development of a small claim into a large claim, superscripted \((s \rightarrow l)\).

The claims development data has indices \(i \in \{0, \ldots, I\} \) and \(j \in \{0, \ldots, J\} \), we assume \(I\) to be the last (most recent) accident year and \(J\) the last period of development. Since we have made the assumption of triangular data we have \(I = J\).

Let \(Y_{i,j}^{(k)}\) be the incremental payments for the \(k\)-th claim of accident year \(i\) in development year \(j\). Then

\[
Z_{i,j}^{(k)} = \sum_{l=0}^{j} Y_{i,l}^{(k)}
\]

are the cumulative payments for the \(k\)-th claim of accident year \(i\) in the first \(j\) development years. Furthermore, let \(n_{i,j}\) be the number of reported claims for accident year \(i\) after \(j\) development years. Then

\[
X_{i,j} = \sum_{k=1}^{n_{i,j}} Y_{i,j}^{(k)}
\]

is the incremental payment in cell \((i,j)\) for all claims, and

\[
C_{i,j} = \sum_{l=0}^{j} X_{i,l} = \sum_{k=1}^{n_{i,j}} Z_{i,j}^{(k)}
\]

is the cumulative payments in cell \((i,j)\). We assume that incremental payments are non-negative. We further introduce a threshold parameter \(d_i\) for each accident year \(i\) such that we denote \(Z_{i,j}^{(k)}\) a “small claim” if \(Z_{i,j}^{(k)} \leq d_i\). Hence, we define the following random variables:

\[
C_{i,j}^{(s)} = \sum_{k=1}^{n_{i,j}} \min(Z_{i,j}^{(k)}, d_i),
\]

\[
C_{i,j}^{(s \rightarrow l)} = \sum_{k=1}^{n_{i,j}} \max(Z_{i,j}^{(k)} - d_i, 0)\mathbb{I}_{\{Z_{i,j-1}^{(k)} \leq d_i\}}, \quad \text{for } j \geq 1,
\]

\[
C_{i,j}^{(l)} = \sum_{k=1}^{n_{i,j}} \max(Z_{i,j}^{(k)} - d_i, 0),
\]

where \(C_{i,0}^{(s \rightarrow l)}\) is defined to be zero for all \(i\). Hence \(C_{i,j}^{(s)}\) denotes the cumulative payments in the small claim layer \([0, d_i]\) (based on individual claims observations).
The random variable $C_{i,j}^{(l)}$ denotes the cumulative payments in the excess layer $(d_i, \infty)$, and $C_{i,j}^{(s->l)}$ denotes the payments in the layer $(d_i, \infty)$ for claims that develop into large claims in period $j$. Given the above cumulative variables, the following holds:

$$C_{i,j} = C_{i,j}^{(s)} + C_{i,j}^{(l)}.$$  

We furthermore define $C_{i,j} = (C_{i,j}^{(s)}, C_{i,j}^{(s->l)}, C_{i,j}^{(l)})$,

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

and

$$\mathcal{B}_k = \{C_{i,j}; i + j \leq I, 0 \leq j \leq k\} \subseteq \mathcal{D}_I.$$ 

We see that $\mathcal{B}_k = \mathcal{D}_I$, the set of all observations at time $I$. Given the information $\mathcal{D}_I$, the task at hand is to predict the complement of this set, $\mathcal{D}_I^c$, defined as

$$\mathcal{D}_I^c = \{C_{i,j}; i + j > I\}.$$ 

Figure 1 illustrates the data structure defined above.

![Claims development triangle](image)

Figure 1: Claims development triangle.

Note that the outstanding loss liabilities for accident year $i$ at time $I$ are given by $C_{i,J} - C_{i,I-1}$. The aim is to predict these liabilities using the information $\mathcal{D}_I$.

### 3 Univariate Chain Ladder Model

We recall the basic univariate CL model. We address the model assumptions, the estimation of model parameters and their properties, and finally, the predictors of the ultimate claims in the so-called distribution free CL model. For details we refer to Mack (1993).

**Model Assumptions 3.1** (Chain Ladder Model)

- There exist deterministic development factors $f_0, \ldots, f_{J-1} > 0$ and variance parameters $\sigma_0^2, \ldots, \sigma_{J-1}^2 > 0$ such that for all $i \in \{0, \ldots, I\}$ and all $j \in \{0, \ldots, J-1\}$ we have

  $$E[C_{i,j+1}|C_{i,0}, \ldots, C_{i,j}] = E[C_{i,j+1}|C_{i,j}] = f_j C_{i,j},$$

  $$\text{Var}(C_{i,j+1}|C_{i,0}, \ldots, C_{i,j}) = \text{Var}(C_{i,j+1}|C_{i,j}) = \sigma_j^2 C_{i,j}.$$
• Claims $C_{i,j}$ of different accident years $i$ are independent.

The conditional expectation of the ultimate cumulative claim $C_{i,j}$ at time $I$ is given by

$$E[C_{i,j}|D_I] = C_{i,I-i} \prod_{j=I-i}^{J-1} f_j,$$

for $i \in \{1, \ldots, I\}$. The $f_j$ are called CL factors. These are usually unknown and hence need to be estimated from the data $D_I$. The following estimators of the $f_j$ have been proven to be conditionally, on $B_j$, (as well as unconditionally) unbiased as well as uncorrelated:

$$\hat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}},$$

for $j \in \{0, \ldots, J-1\}$. The fact that these estimators are unbiased and uncorrelated leads to the following conditionally unbiased estimators, given $B_{I-i}$, of the cumulative claim $C_{i,j}$, for $i \in \{1, \ldots, I\}$ and $j \in \{I-i+1, \ldots, J\}$,

$$\widehat{C}_{i,j} = C_{i,I-i} \prod_{k=I-i}^{j-1} \hat{f}_k;$$

see Mack (1993) for details. In the next section we describe the multivariate CL algorithm.

### 4 Multivariate Chain Ladder Model

In this section we outline a multivariate CL model necessary to tackle the issue of splitting claims into multiple layers. As done in the previous section, we outline the model assumptions, parameter estimation, and ultimate cumulative claim predictors for this model.

#### 4.1 Model Assumptions

**Model Assumptions 4.1** (Multivariate Chain Ladder Model)

$(C_{i,j})_{j=0,\ldots,J} \left( C_{i,j}^{(s)}, C_{i,j}^{(s-l)}, C_{i,j}^{(l)} \right)_{j=0,\ldots,J}$ is a multi dimensional Markov process for every accident year $i$ where:

i) There exist deterministic development factors $f_j^{(s)}, f_j^{(l)}, f_j^{(s-l)}$ and variance parameters $\sigma_j^{(s)}, \sigma_j^{(l)}, \sigma_j^{(s-l)}, \rho_j^{(s-s-l)}$ such that for all $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, J\}$...
\{0, \ldots, J - 1\} we have,
\[
E \left[ C_{i,j+1}^{(s)} \mid C_{i,j} \right] = f_j^{(s)} C_{i,j}^{(s)},
\]
\[
E \left[ C_{i,j+1}^{(s\rightarrow l)} \mid C_{i,j} \right] = f_j^{(s\rightarrow l)} C_{i,j}^{(s)},
\]
\[
E \left[ C_{i,j+1}^{(l)} \mid C_{i,j} \right] = f_j^{(l)} C_{i,j}^{(l)} + f_j^{(s\rightarrow l)} C_{i,j}^{(s)}.
\]

\[
\text{Var} \left( C_{i,j+1}^{(s)} \mid C_{i,j} \right) = (\sigma_j^{(s)})^2 C_{i,j}^{(s)},
\]
\[
\text{Var} \left( C_{i,j+1}^{(s\rightarrow l)} \mid C_{i,j} \right) = (\sigma_j^{(s\rightarrow l)})^2 C_{i,j}^{(s)},
\]
\[
\text{Var} \left( C_{i,j+1}^{(l)} \mid C_{i,j} \right) = (\sigma_j^{(l)})^2 C_{i,j}^{(l)} + (\sigma_j^{(s\rightarrow l)})^2 C_{i,j}^{(s)}.
\]

\[
\text{Cov} \left( C_{i,j+1}^{(s)} , C_{i,j+1}^{(s\rightarrow l)} \mid C_{i,j} \right) = \sigma_j^{(s)} \sigma_j^{(s\rightarrow l)} C_{i,j}^{(s,s\rightarrow l)},
\]
\[
\text{Cov} \left( C_{i,j+1}^{(s\rightarrow l)} , C_{i,j+1}^{(l)} \mid C_{i,j} \right) = (\sigma_j^{(s\rightarrow l)})^2 C_{i,j}^{(s)},
\]
\[
\text{Cov} \left( C_{i,j+1}^{(s)} , C_{i,j+1}^{(l)} \mid C_{i,j} \right) = \sigma_j^{(s)} \sigma_j^{(s\rightarrow l)} C_{i,j}^{(s,s\rightarrow l)}.
\]

ii) Claims $C_{i,j}$ in different accident years $i$ are independent.

**Remarks 4.2:**

- Note that the covariance assumptions suggest that the dependence between the small and large claims layers is exactly modelled by the claims that grow into the next layer. Once a claims is large, it no longer influences the behaviour of small claims. If this is not appropriate for a specific practical application, correlations $\rho_j^{(s,l)}$ and $\rho_j^{(s\rightarrow l,l)}$ need to be introduced and incorporated. Although incorporating $\rho_j^{(s,l)}$ and $\rho_j^{(s\rightarrow l,l)}$ into the model does not present difficulties, their addition results in cumbersome formulations that cloud the overall picture.

- Note that, in general, one obtains only numerical solutions if one introduces more general forms of dependencies like accounting year effects.

- For the excess layer we also choose a CL model, since this leads to tractable MSEP formulas. Note that in practice one often uses other models for the large claims layer, such as the Bornhuetter-Ferguson method Bornhuetter and Ferguson (1972), or case estimates by claims adjusters.

- Note that the threshold $d_i$ is chosen a priori. Of course, one can then ask the question of the “best” threshold value. This depends highly on the data. We do not treat this question here because our data basis did not allow for such an analysis. In practice there is often a natural threshold determined by internal processes, reinsurance programs, the volume of the company, etc.
For example in motor third party liability insurance, one often separates the claims according to a threshold $d_i$ such that the resulting small and large claims are pure property and bodily injury claims, respectively. Such a separation is highly justified since these two classes of claims have rather different risk factors.

### 4.2 Estimation of Chain Ladder Factors and Claim Predictors

To estimate the factors $f_j$ we use the CL approach, it produces the following estimators:

\[
\hat{f}_j^{(s)} = \frac{\sum_{k=0}^{I-j-1} C_{k,j+1}^{(s)}}{\sum_{k=0}^{I-j-1} C_{k,j}^{(s)}}, \quad \hat{f}_j^{(s\rightarrow l)} = \frac{\sum_{k=0}^{I-j-1} C_{k,j+1}^{(s\rightarrow l)}}{\sum_{k=0}^{I-j-1} C_{k,j}^{(s)}} \quad \text{and} \quad \hat{f}_j^{(l)} = \frac{\sum_{k=0}^{I-j-1} \left( C_{k,j+1}^{(l)} - C_{k,j+1}^{(s\rightarrow l)} \right)}{\sum_{k=0}^{I-j-1} C_{k,j}^{(l)}}.
\]

**Lemma 4.3** Under Model Assumptions 4.1 we have that $\hat{f}_j^{(s)}$, $\hat{f}_j^{(s\rightarrow l)}$, and $\hat{f}_j^{(l)}$, are, given $B_j$, unbiased estimators for $f_j^{(s)}$, $f_j^{(s\rightarrow l)}$, and $f_j^{(l)}$, respectively.

**Proof:** We have

\[
E[\hat{f}_j^{(l)} | B_j] = \frac{\sum_{k=0}^{I-j-1} E[C_{k,j+1}^{(l)} - C_{k,j+1}^{(s\rightarrow l)} | B_j]}{\sum_{k=0}^{I-j-1} C_{k,j}^{(l)}} = \frac{\sum_{k=0}^{I-j-1} \left( f_j^{(l)} C_{k,j}^{(l)} + f_j^{(s\rightarrow l)} C_{k,j}^{(s)} - f_j^{(s\rightarrow l)} C_{k,j}^{(s)} \right)}{\sum_{k=0}^{I-j-1} C_{k,j}^{(l)}} = f_j^{(l)}.
\]

The proofs for $\hat{f}_j^{(s)}$ and $\hat{f}_j^{(s\rightarrow l)}$ are analogous.

In addition to the above factors being conditionally unbiased they are also uncorrelated.

**Lemma 4.4** The estimators $\hat{f}_j^{(s)}$, $\hat{f}_j^{(s\rightarrow l)}$ and $\hat{f}_j^{(l)}$ are (conditionally) uncorrelated, more precisely,

\[
E[\hat{f}_j^{(x)} \hat{f}_k^{(y)} | B_j] = E[\hat{f}_j^{(x)} | B_j] E[\hat{f}_k^{(y)} | B_j],
\]

for $j, k \in \{0, \ldots, J-1\}$, $j < k$, and $x, y \in \{s, s \rightarrow l, l\}$. 

125
Proof: Since $B_j \subseteq B_k$,
\[
E \left[ f_j(x) f_k(y) \middle| B_j \right] = E \left[ E \left[ f_j(x) \middle| B_k \right] \middle| B_j \right] = E \left[ f_j(x) E \left[ f_k(y) \middle| B_k \right] \middle| B_j \right] = E \left[ f_j(x) \right] E \left[ f_k(y) \middle| B_j \right]. \tag*{\square}
\]

Now that we have calculated the factor estimators $\hat{f}_j$ and found them to be conditionally unbiased and uncorrelated, we recursively define predictors for the cumulative claims as follows,
\[
\hat{C}_{i,j}^{(s)} = \hat{f}_j^{(s)} \hat{C}_{i,j-1}^{(s)},
\hat{C}_{i,j}^{(s \rightarrow l)} = \hat{f}_{j-1}^{(s \rightarrow l)} \hat{C}_{i,j-1}^{(s)},
\hat{C}_{i,j}^{(l)} = \hat{f}_{j-1}^{(l)} \hat{C}_{i,j-1}^{(l)} + \hat{C}_{i,j}^{(s \rightarrow l)},
\hat{C}_{i,j} = \hat{C}_{i,j}^{(s)} + \hat{C}_{i,j}^{(l)},
\]
for $j \in \{ I - i + 1, \ldots, J \}$, where $\hat{C}_{i,j-i}^{(x)} = C_{i,j-i}^{(x)}$ for $x \in \{s, s \rightarrow l, l\}$. This formulation leads to the following lemma.

Lemma 4.5 $\hat{C}_{i,j}^{(s)}$, $\hat{C}_{i,j}^{(l)}$, and $\hat{C}_{i,j}$ are conditionally, given $C_{i,1-i}$, unbiased estimators for $E[C_{i,j}^{(s)}|D_I]$, $E[C_{i,j}^{(l)}|D_I]$, and $E[C_{i,j}|D_I]$, respectively.

Proof: We demonstrate the result for $\hat{C}_{i,j}^{(s)}$:
\[
E \left[ \hat{C}_{i,j}^{(s)} \middle| C_{i,1-i} \right] = E \left[ \hat{C}_{i,j}^{(s)} | B_{1-i} \right] = C_{i,j-i}^{(s)} E \left[ \prod_{k=1-i}^{J-1} \hat{f}_k^{(s)} | B_{1-i} \right]
\]
\[
= C_{i,j-i}^{(s)} \prod_{k=1-i}^{J-1} E \left[ \hat{f}_k^{(s)} | B_{1-i} \right] = C_{i,j-i}^{(s)} \prod_{k=1-i}^{J-1} f_k^{(s)}
\]
\[
= E \left[ C_{i,j}^{(s)} | D_I \right]. \tag*{\square}
\]

The proofs of the remaining statements are analogous.

4.3 Estimation of Covariance Parameters

To estimate the prediction uncertainty we need to specify the covariance parameters in Model Assumptions 4.1,
\[
\sigma_j^{(s)}, \sigma_j^{(s \rightarrow l)}, \sigma_j^{(l)}, \text{ and } \rho_j^{(s,s \rightarrow l)},
\]
for $j \in \{0, \ldots, J-1\}$. We note that the following variance and covariance estimators are only used in MSEP estimation, they are not used for reserve estimation. The
claims reserves are completely determined by Lemmas 4.3-4.5 and we do not obtain implicit estimators as in Merz and Wüthrich (2008b).

For estimators of $\sigma_j^{(s)}$, $\sigma_j^{(s\rightarrow l)}$, and $\sigma_j^{(l)}$, we follow the route of the univariate CL estimates. They are given as follows,

$$
\left(\hat{\sigma}_j^{(s)}\right)^2 = \frac{1}{I - j - 1} \sum_{k=0}^{I-j-1} C_{k,j}^{(s)} \left( \frac{C_{k,j+1}^{(s)}}{C_{k,j}^{(s)}} - \hat{f}_j^{(s)} \right)^2,
$$

$$
\left(\hat{\sigma}_j^{(s\rightarrow l)}\right)^2 = \frac{1}{I - j - 1} \sum_{k=0}^{I-j-1} C_{k,j}^{(s\rightarrow l)} \left( \frac{C_{k,j+1}^{(s\rightarrow l)}}{C_{k,j}^{(s\rightarrow l)}} - \hat{f}_j^{(s\rightarrow l)} \right)^2,
$$

$$
\left(\hat{\sigma}_j^{(l)}\right)^2 = \frac{1}{I - j - 1} \sum_{k=0}^{I-j-1} C_{k,j}^{(l)} \left( \frac{C_{k,j+1}^{(l)}}{C_{k,j}^{(l)}} - \hat{f}_j^{(l)} \right)^2,
$$

for $j \in \{0, \ldots, J - 2\}$.

**Lemma 4.6** Under Model Assumptions 4.1 we have that $\left(\hat{\sigma}_j^{(s)}\right)^2$, $\left(\hat{\sigma}_j^{(s\rightarrow l)}\right)^2$ and $\left(\hat{\sigma}_j^{(l)}\right)^2$, are, given $B_j$, unbiased estimators for $\left(\sigma_j^{(s)}\right)^2$, $\left(\sigma_j^{(s\rightarrow l)}\right)^2$, and $\left(\sigma_j^{(l)}\right)^2$, respectively.

**Proof:** The proof is a straightforward calculation using the model assumptions; see e.g. Wüthrich and Merz (2008), Lemma 3.5. □

What remains is to provide an estimator for correlation $\rho_j^{(s,s\rightarrow l)}$. If $\sigma_j^{(s)}$ and $\sigma_j^{(s\rightarrow l)}$ are known, we choose as an estimator of $\rho_j^{(s,s\rightarrow l)}$,

$$
\hat{\rho}_j^{(s,s\rightarrow l)} = \frac{1}{I - j - 1} \sum_{k=0}^{I-j-1} \frac{C_{k,j}^{(s)}}{\hat{\sigma}_j^{(s)} \hat{\sigma}_j^{(s\rightarrow l)}} \left( \frac{C_{k,j+1}^{(s)}}{C_{k,j}^{(s)}} - \hat{f}_j^{(s)} \right) \left( \frac{C_{k,j+1}^{(s\rightarrow l)}}{C_{k,j}^{(s\rightarrow l)}} - \hat{f}_j^{(s\rightarrow l)} \right).
$$

**Lemma 4.7** Under Model Assumptions 4.1 we have that $\hat{\rho}_j^{(s,s\rightarrow l)}$ is, given $B_j$, an unbiased estimator for $\rho_j^{(s,s\rightarrow l)}$.

**Proof:** The proof is a straightforward calculation using the model assumptions; see e.g. Merz and Wüthrich (2008b), Appendix A. □

However, $\hat{\rho}_j^{(s,s\rightarrow l)}$ is unattainable since the $\sigma$’s are unknown. Since $\hat{\rho}_j^{(s,s\rightarrow l)}$ is unbiased, it is natural to keep the structure of this estimator and replace the $\sigma$ with estimators $\hat{\sigma}$. Hence, we use $\hat{\rho}_j^{(s,s\rightarrow l)}$, defined as follows,

$$
\hat{\rho}_j^{(s,s\rightarrow l)} = \frac{1}{I - j - 1} \sum_{k=0}^{I-j-1} \frac{C_{k,j}^{(s)}}{\hat{\sigma}_j^{(s)} \hat{\sigma}_j^{(s\rightarrow l)}} \left( \frac{C_{k,j+1}^{(s)}}{C_{k,j}^{(s)}} - \hat{f}_j^{(s)} \right) \left( \frac{C_{k,j+1}^{(s\rightarrow l)}}{C_{k,j}^{(s\rightarrow l)}} - \hat{f}_j^{(s\rightarrow l)} \right),
$$

127
as the estimator for $\rho_j^{(s,s \rightarrow l)}$ for $j \in \{0, \ldots, J-2\}$. Note that $\hat{\rho}_j^{(s,s \rightarrow l)}$ is not unbiased, and therefore can be viewed as a “plug-in” estimate, however, we only use it for the MSEP estimation.

We note that if $I = J$ we are unable to estimate $\sigma_{j-1}^{(\cdot)}$ and $\rho_{j-1}^{(\cdot)}$ from our data. We handle this, as is done in Mack (1993), by extrapolating the last parameters with a (usually) decreasing series. Hence, we choose the estimators

$$\left(\hat{\sigma}_{j-1}^{(s)}\right)^2 = \min \left(\frac{\left(\hat{\sigma}_{j-2}^{(s)}\right)^4}{\left(\hat{\sigma}_{j-3}^{(s)}\right)^2}, \left(\hat{\sigma}_{j-2}^{(s)}\right)^2, \left(\hat{\sigma}_{j-3}^{(s)}\right)^2\right),$$

where $z \in \{s, s \rightarrow l, l\}$ and

$$\hat{\rho}_{j-1}^{(x,y)} = \min \left(\frac{\left(\hat{\rho}_{j-2}^{(x,y)} \hat{\sigma}_{j-2}^{(x)} \hat{\sigma}_{j-2}^{(y)}\right)^2}{\left|\hat{\rho}_{j-3}^{(x,y)} \hat{\sigma}_{j-3}^{(x)} \hat{\sigma}_{j-3}^{(y)}\right|}, \frac{\left|\hat{\rho}_{j-2}^{(x,y)} \hat{\sigma}_{j-2}^{(x)} \hat{\sigma}_{j-2}^{(y)}\right|}{\left|\hat{\rho}_{j-3}^{(x,y)} \hat{\sigma}_{j-3}^{(x)} \hat{\sigma}_{j-3}^{(y)}\right|}, \frac{\left|\hat{\rho}_{j-2}^{(x,y)} \hat{\sigma}_{j-2}^{(x)} \hat{\sigma}_{j-2}^{(y)}\right|}{\left|\hat{\rho}_{j-3}^{(x,y)} \hat{\sigma}_{j-3}^{(x)} \hat{\sigma}_{j-3}^{(y)}\right|}\right),$$

where $x = s$ and $y = s \rightarrow l$.

### 5 Mean Square Error of Prediction

In this section we derive an estimate of the conditional MSEP of the multivariate CL model for aggregated accident years. Hence, the predictor of interest is $\sum_{i=1}^{I} \hat{C}_{i,J}$, where $\hat{C}_{i,J} = \hat{C}_{i,J}^{(s)} + \hat{C}_{i,J}^{(l)}$. The conditional MSEP defined below quantifies the uncertainty of this predictor relative to $\sum_{i=1}^{I} C_{i,J}$, given $D_I$,

$$\text{msep}_{\sum_{i=1}^{I} C_{i,J} | D_I} \left( \sum_{i=1}^{I} \hat{C}_{i,J} \right) = E \left( \left( \sum_{i=1}^{I} C_{i,J} - \sum_{i=1}^{I} \hat{C}_{i,J} \right)^2 \bigg| D_I \right).$$

Due to our predictors $\hat{C}_{i,J}$ being $D_I$-measurable, the above decouples into the following, the terms coined conditional process variance and parameter estimation error,

$$\text{msep}_{\sum_{i=1}^{I} C_{i,J} | D_I} \left( \sum_{i=1}^{I} \hat{C}_{i,J} \right) = \text{Var} \left( \sum_{i=1}^{I} C_{i,J} | D_I \right) + \left( \sum_{i=1}^{I} \hat{C}_{i,J} - E \left[ \sum_{i=1}^{I} C_{i,J} | D_I \right] \right)^2.$$

Furthermore, the conditional MSEP of $\hat{C}_{i,J}$ for single accident years $i \in \{1, \ldots, I\}$ is defined similarly and satisfies

$$\text{msep}_{C_{i,J} | D_I} (\hat{C}_{i,J}) = E \left[ (C_{i,J} - \hat{C}_{i,J})^2 | D_I \right] = \text{Var} (C_{i,J} | D_I) + \left( \hat{C}_{i,J} - E [C_{i,J} | D_I] \right)^2.$$
Again, we call the first term the conditional process variance, it is the variance inherent in our model; the second is called the parameter estimation error due to the fact that we are estimating the $f_j$'s. Hence $\hat{C}_{i,j}$ is a predictor for $C_{i,j}$ and an estimator for $E[C_{i,j}|D_I]$ at time $I$. For known CL factors the second term on the right-hand side of the above equation disappears because $E[C_{i,j}|D_I]$ is then used to predict $C_{i,j}$.

Lastly, we define the MSEP for the small and large layers as follows,

$$\text{mse}_{C_{i,j}^{(s)}|D_I}(\hat{C}_{i,j}^{(s)}) = \text{Var}(C_{i,j}^{(s)}|D_I) + (\hat{C}_{i,j}^{(s)} - E[C_{i,j}^{(s)}|D_I])^2,$$

$$\text{mse}_{C_{i,j}^{(l)}|D_I}(\hat{C}_{i,j}^{(l)}) = \text{Var}(C_{i,j}^{(l)}|D_I) + (\hat{C}_{i,j}^{(l)} - E[C_{i,j}^{(l)}|D_I])^2.$$

In the following sections we derive estimates of the conditional process variance and parameter estimation error for a single accident year $i$ and for multiple accident years for the multivariate CL model. Before presenting the results, the reader should know that $\Gamma$ terms are concerned with the estimation of the process variance and $\Delta$ terms with estimation error, these terms will be defined properly in sections 5.1-5.4.

**Univariate MSEP: Single Accident Year**

Under Model Assumptions 4.1, the conditional mean square error of predictions $\text{mse}_{C_{i,j}^{(s)}|D_I}(\hat{C}_{i,j}^{(s)})$ and $\text{mse}_{C_{i,j}^{(l)}|D_I}(\hat{C}_{i,j}^{(l)})$, for $i \in \{1, \ldots, I\}$, are estimated by

$$\hat{\text{mse}}_{C_{i,j}^{(s)}|D_I}(\hat{C}_{i,j}^{(s)}) = \Gamma_{i,j}^{(s)} + (C_{i,j}^{(s)} - E[C_{i,j}^{(s)}|D_I])^2,$$

$$\hat{\text{mse}}_{C_{i,j}^{(l)}|D_I}(\hat{C}_{i,j}^{(l)}) = \Gamma_{i,j}^{(l)} + (C_{i,j}^{(l)} - E[C_{i,j}^{(l)}|D_I])^2.$$

respectively. The $\Gamma_{i,j}^{(c)}$ are given in Appendix A by equations given in (9) and $\Delta_{i,j}(m,n)$ in Appendix B by equation (10).

The MSEP of the small claim layer is identical to the MSEP of the standard univariate CL model; see e.g. Buchwalder et al. (2006). The additional terms in the large claim MSEP estimation, all having to do with parameter estimation error, are a result of the impact of the $f_{i\rightarrow l}$ factors, i.e. the dependence between small and large claims by the growth into the upper layer.

**Multivariate MSEP: Single Accident Year**

Under Model Assumptions 4.1, the conditional mean square error of prediction
\[ \text{mse}_{P}^{}(\hat{C}_{i,j}) \text{, for } i \in \{1, \ldots, I\}, \text{ are estimated by} \]
\[ \hat{\text{mse}}_{P}^{}(\hat{C}_{i,j}) = \hat{\text{mse}}_{P}^{}(\hat{C}_{i,j}) + \hat{\text{mse}}_{P}^{}(\hat{C}_{i,j}) + 2 \left( \Gamma_{i,j}^{(s,l)} + C_{i,J-I}^{(s)} \Delta_{i,J-I}^{I,J} \right) \]
\[ + \sum_{n=I-I}^{J-I} \Delta_{i,j}(J,n). \]

The \( \Gamma_{i,j} \) are given in Appendix A by equations given in (9) and \( \Delta_{i,j}(m,n) \) in Appendix B by equation (10).

When aggregating the small and large claim layers additional process variance and estimation error terms are produced. The covariance between the small and large layers is captured by \( \Gamma_{i,j}^{(s,l)} \); the remaining terms deal with the dependence between the small claim layer factors \( f^{(s)} \) with those of the large claim layer, that is, both \( f^{(s-l)} \) and \( f^{(l)} \).

**Multivariate MSEP: Multiple Accident Years**

Under Model Assumptions 4.1, the conditional mean square error of prediction \( \text{mse}_{P}^{I} \sum_{i=1}^{I} C_{i,j} \) are estimated by
\[ \hat{\text{mse}}_{P}^{I} \sum_{i=1}^{I} C_{i,j} = \sum_{i=1}^{I} \hat{\text{mse}}_{P}^{I} C_{i,j} + 2 \sum_{1 \leq i \leq k \leq J-I} \Lambda_{i,k,j}, \]
\[ (4) \]

For a definition of \( \Lambda_{i,k,j} \) refer to equation (7) in Section 5.4.

Since we have assumed the accident years to be independent, the aggregation of multiple accident years produces no extra process variance terms. The \( \Lambda \)'s are a convenient way of representing the dependencies amongst the CL factors for the different accident years.

In the remainder of this section we describe the estimates presented in equations (1), (2), (3), and (4). The technical elements of these derivations, namely the derivation of the estimators \( \Gamma \) and \( \Delta \) are left to the appendices.

### 5.1 Derivation of an Estimator of the MSEP for the Small Claims Layer, Formula (1)

In this section we derive an estimator of the conditional MSEP of the small claims layers for accident years \( i \in \{1, \ldots, I\} \). Due to the \( \mathcal{D}_{I} \)-measurability of \( \hat{C}_{i,j}^{(s)} \), we have,
\[ \text{mse}_{P}^{C_{i,j}^{(s)}}(\hat{C}_{i,j}^{(s)}) = \text{Var}(C_{i,j}^{(s)}|\mathcal{D}_{I}) + (\hat{C}_{i,j}^{(s)} - E[C_{i,j}^{(s)}|\mathcal{D}_{I}])^2. \]
Studying the right-hand side of the above equation we see the conditional process variance and the parameter estimation error. We handle these terms separately. As an estimator of the conditional process variance we use the notation $\Gamma_{i,j}^{(x,y)}$.

$$\Gamma_{i,j}^{(x,y)} = \overline{\text{Cov}}(C_{i,j}^{(x)}, C_{i,j}^{(y)} | D_I),$$  

for $i \in \{1, \ldots, I\}$, $j \in \{I-i+1, \ldots, J\}$, and $x, y \in \{s, s \rightarrow l, l\}$. Hence to estimate $\text{Var}(C_{i,j}^{(s)} | D_I)$, we use $\Gamma_{i,j}^{(s,s)}$; see Appendix A.

When studying the estimation error, we notice the following:

$$\frac{(\hat{C}_{i,j}^{(s)} - E[C_{i,j}^{(s)} | D_I])^2}{(C_{i,I-I}^{(s)})^2} = \left(\prod_{k=I-i}^{J-1} \hat{f}_k^{(s)} - \prod_{k=I-i}^{J-1} f_k^{(s)}\right)^2.$$  

We concentrate on estimating the right-hand side of the above equation. To estimate this, it is clear that we cannot simply substitute the parameters $f_j^{(s)}$ with their corresponding estimators, this would make the parameter estimation error equal to zero; see e.g. Mack (1993), page 218. Rather, we interpret the right-hand side as the volatility of the estimators, this would make the parameter estimation error equal to zero. Hence when $m = n = J$, we have no $f^{(s \rightarrow l)}$ or $f^{(l)}$ and the above expression is equal to the right-hand side of equation (6); analogously when $m = n = I - i - 1$ we have only $f^{(l)}$. Finally, we define $\Delta_{i,j}^{(m,n)}$ to be an estimator of $\delta_{i,j}^{(m,n)}$; see Appendix B for more details on this estimation procedure.

With the introduction of the $\Gamma$'s and $\Delta$'s, an estimator for the MSEP of the small claims for a single accident is, as described in equations (1) and (2),

$$\text{msep}_{C_{i,j}^{(s)} | D_I} (\hat{C}_{i,j}^{(s)}) = \Gamma_{i,j}^{(s,s)} + (C_{i,I-I}^{(s)})^2 \Delta_{i,j}^{(J,J)},$$  

for $i \in \{1, \ldots, I\}$. Next we find an estimator of the conditional MSEP for the large claims layer.
5.2 Derivation of an Estimator of the MSEP for the Large Claims Layer, Formula (2)

As in the previous section we first examine the conditional MSEP formula for a single accident year \( i \in \{1, \ldots, I\} \). It is defined as

\[
\text{mse}_{C_{i,j}|D_I}(\hat{C}_{i,j}) = \text{Var}(C_{i,j}|D_I) + \left(\hat{C}_{i,j} - E[C_{i,j}|D_I]\right)^2.
\]

We use \( I_{i,j} \), see equation (5), to estimate the conditional process variance, again, see Appendix A for details. Before we make use of \( \Delta \)'s we need to rewrite the parameter estimation error term:

\[
\left(\hat{C}_{i,j} - E[C_{i,j}|D_I]\right)^2 = \left[\left(C_{i,j} - \sum_{k=I-i}^{J-1} f_k \prod_{l=i}^{k-1} f_l\right)^2 + \sum_{k=I-i}^{J-1} \sum_{l=I-i}^{J-1} \Delta_{i,j}(m, I-I-i)\right].
\]

Hence, large claim payments result from additional payments from already classified large claims as well as payments from newly classified large claims. Making use of our \( \delta \) notation, we expand the above as follows,

\[
\left(\hat{C}_{i,j} - E[C_{i,j}|D_I]\right)^2 = \left[\left(C_{i,j} - \sum_{k=I}^{J-1} f_k \prod_{l=i}^{k-1} f_l\right)^2 + \sum_{k=I-i}^{J-1} \sum_{l=I-i}^{J-1} \delta_{i,j}(m, I-I-i)\right].
\]

To estimate this we need only replace the \( \delta \) with their estimators \( \Delta \). Hence we get that for \( i \in \{1, \ldots, I\} \),

\[
\text{msep}_{C_{i,j}|D_I}(\hat{C}_{i,j}) = I_{i,j} + \left[\left(C_{i,j} - \sum_{k=I-i}^{J-1} f_k \prod_{l=i}^{k-1} f_l\right)^2 + \sum_{k=I-i}^{J-1} \sum_{l=I-i}^{J-1} \Delta_{i,j}(m, I-I-i)\right].
\]

For a definition of \( \Delta_{i,j}(m, n) \), we again refer to Appendix B.

5.3 Derivation of an Estimator of the MSEP for the Aggregate Portfolio for a Single Accident Year, Formula (3)

We aim to aggregate the estimator of the conditional MSEP for the two layers for a single accident year \( i \in \{1, \ldots, I\} \). It will be seen that in this derivation, elements will appear that fit into either the small or large layer estimator of the conditional MSEP, this will simplify our calculations:

\[
\text{msep}_{C_{i,j}|D_I}(\hat{C}_{i,j}) = \text{Var}(C_{i,j} + C_{i,j}|D_I) + \left(\hat{C}_{i,j} + E[C_{i,j}|D_I]\right)^2.
\]
Concentrating on the conditional process variance we see that
\[
\text{Var}(C_{i,J}^{(s)} + C_{i,J}^{(l)} | \mathcal{D}_I) = \text{Var}(C_{i,J}^{(s)} | \mathcal{D}_I) + \text{Var}(C_{i,J}^{(l)} | \mathcal{D}_I) + 2 \text{Cov}(C_{i,J}^{(s)}, C_{i,J}^{(l)} | \mathcal{D}_I).
\]
Each term on the right hand side can be estimated using a \( \Gamma \), the first are present in the respective layered estimators of the conditional MSEP. The final term is estimated by \( 2 \Gamma_{i,J}^{(s,l)} \). Similarly the parameter estimation error splits into three terms, two of which are present in the individual layer MSEP estimators, the third is estimated with use of the \( \Delta \)'s:
\[
\left( \hat{C}_{i,J}^{(s)} + \hat{C}_{i,J}^{(l)} - E[C_{i,J}^{(s)} + C_{i,J}^{(l)} | \mathcal{D}_I] \right)^2 = \left( \hat{C}_{i,J}^{(s)} - E[C_{i,J}^{(s)} | \mathcal{D}_I] \right)^2 + 2 \left( \hat{C}_{i,J}^{(s)} - E[C_{i,J}^{(s)} | \mathcal{D}_I] \right) \left( \hat{C}_{i,J}^{(l)} - E[C_{i,J}^{(l)} | \mathcal{D}_I] \right) + \left( \hat{C}_{i,J}^{(l)} - E[C_{i,J}^{(l)} | \mathcal{D}_I] \right)^2.
\]
The middle term on the right-hand side above is the only one that requires attention, the others have already been treated. We rewrite this term as follows:
\[
\left( \hat{C}_{i,J}^{(s)} - E[C_{i,J}^{(s)} | \mathcal{D}_I] \right) \left( \hat{C}_{i,J}^{(l)} - E[C_{i,J}^{(l)} | \mathcal{D}_I] \right) = C_{i,J-1}^{(s)} C_{i,J-i-1}^{(l)} \delta_{i,J}^{(J-I-1)} + \left( C_{i,J-i}^{(s)} \right)^2 \sum_{n=I-i}^{J-1} \delta_{i,J}^{(J,n)}.
\]
To estimate the above, we replace the \( \delta \)'s with their corresponding \( \Delta \)'s. As a result we obtain as an estimator of the conditional MSEP of the aggregate portfolio for a single accident year \( i \in \{1, \ldots, I\} \),
\[
\text{mse}_{C_{i,J}|\mathcal{D}_I}(\hat{C}_{i,J}) = \text{mse}_{C_{i,J}|\mathcal{D}_I}(\hat{C}_{i,J}) + \text{mse}_{C_{i,J}|\mathcal{D}_I}(\hat{C}_{i,J}) + 2 \left( \sum_{i,J}^{(s,l)} + \sum_{i,J-i}^{(s,l)} \Delta_{i,J}^{(J-I-1)} + \left( C_{i,J-i}^{(s)} \right)^2 \sum_{k=I-i}^{J-1} \Delta_{i,J}^{(J,k)} \right),
\]
as presented in equation (3). We next focus our attention on calculating an estimator of the conditional MSEP for multiple accident years.

### 5.4 Derivation of an estimator of the MSEP of the Aggregate Portfolio for Multiple Accident Years, Formula (4)

Recall the conditional MSEP of the aggregate portfolio for multiple accident years.
\[
\text{mse}_{\sum_{i=1}^{I} C_{i,J}|\mathcal{D}_I}(\sum_{i=1}^{I} \hat{C}_{i,J}) = \text{Var} \left( \sum_{i=1}^{I} \hat{C}_{i,J} | \mathcal{D}_I \right) \left( \sum_{i=1}^{I} \hat{C}_{i,J} - E \left[ \sum_{i=1}^{I} C_{i,J} | \mathcal{D}_I \right] \right)^2.
\]
Focusing first on the conditional process variance we see that, since we have assumed the accident years to be independent, the conditional process variance of the sum
is equal to the sum of the conditional process variances.

\[
\text{Var}
\left(
\sum_{i=1}^{I} C_{i,j}\mid D_{I}
\right)
= \sum_{i=1}^{I} \text{Var}
\left(
C_{i,j}\mid D_{I}
\right).
\]

Estimators of the individual process variances are present in the yearly estimators of the conditional MSEP, hence we need not further discuss them here. However, the parameter estimation error does not break down as nicely:

\[
\left(\sum_{i=1}^{I} (\hat{\lambda}_{i,j}^{(s)} + \hat{\lambda}_{i,j}^{(l)}) - E\left[\sum_{i=1}^{I} (C_{i,j}^{(s)} + C_{i,j}^{(l)}) \mid D_{I}\right]\right)^{2}
= \sum_{i=1}^{I} \left(\hat{\lambda}_{i,j}^{(s)} + \hat{\lambda}_{i,j}^{(l)} - E\left[\sum_{i=1}^{I} (C_{i,j}^{(s)} + C_{i,j}^{(l)}) \mid D_{I}\right]\right)^{2} + 2 \sum_{1 \leq i < k \leq J-1} \lambda_{i,k,\delta}.
\]

We see that on the right-hand side of the above equation, we have the sum of the yearly parameter estimation errors, estimates of which are absorbed into the yearly estimators of the conditional MSEP, and a summation of \(\lambda\)'s. The \(\lambda\) terms are estimated with the help of the \(\Delta\) notation. We provide now the explicit form of \(\lambda\):

\[
\lambda_{i,k,\delta} = \left((\hat{\lambda}_{i,j}^{(s)} - E[C_{i,j}^{(s)} \mid \delta_{i}])((\hat{\lambda}_{i,j}^{(l)} - E[C_{i,j}^{(l)} \mid \delta_{i}]) + (\hat{\lambda}_{i,j}^{(s)} - E[C_{i,j}^{(s)} \mid \delta_{i}])((\hat{\lambda}_{i,j}^{(l)} - E[C_{i,j}^{(l)} \mid \delta_{i}]) + (\hat{\lambda}_{i,j}^{(s)} - E[C_{i,j}^{(s)} \mid \delta_{i}])((\hat{\lambda}_{i,j}^{(l)} - E[C_{i,j}^{(l)} \mid \delta_{i}]) + (\hat{\lambda}_{i,j}^{(s)} - E[C_{i,j}^{(s)} \mid \delta_{i}])((\hat{\lambda}_{i,j}^{(l)} - E[C_{i,j}^{(l)} \mid \delta_{i}])\right).
\]

With the use of \(\delta\) notation, the above becomes

\[
\lambda_{i,k,\delta} = C_{i,\delta,i}^{(s)} \sum_{m=I-i}^{J} \sum_{n=I-i}^{J} \delta_{i,j}(m,n)
+ (C_{i,\delta,i}^{(s)} \hat{C}_{i,k,I-i}^{(s)} + C_{i,\delta,i}^{(l)} \hat{C}_{i,k,I-i}^{(l)}) \sum_{m=I-i}^{J} \delta_{i,j}(m, I-i - 1)
+ C_{i,\delta,i}^{(l)} \hat{C}_{i,k,I-i}^{(l)} \delta_{i,j}(I-i-1, I-i-1).
\]

To estimate \(\lambda\), denote the estimator \(\Lambda\), we need only replace the \(\delta\)'s with the appropriate estimators \(\Delta\), defined in Appendix B, equation (10) to obtain

\[
\Lambda_{i,k,\delta} = C_{i,\delta,i}^{(s)} \sum_{m=I-i}^{J} \sum_{n=I-i}^{J} \Delta_{i,j}(m,n)
+ (C_{i,\delta,i}^{(s)} \hat{C}_{i,k,I-i}^{(s)} + C_{i,\delta,i}^{(l)} \hat{C}_{i,k,I-i}^{(l)}) \sum_{m=I-i}^{J} \Delta_{i,j}(m, I-i - 1)
+ C_{i,\delta,i}^{(l)} \hat{C}_{i,k,I-i}^{(l)} \Delta_{i,j}(I-i-1, I-i-1).
\]

(7)

We now put all the pieces together and we find that, as presented in equation (4),

\[
\widehat{\text{msep}}_{\sum_{i=1}^{I} C_{i,j} \mid D_{I}} \left(\sum_{i=1}^{I} \hat{C}_{i,j}\right)
= \sum_{i=1}^{I} \widehat{\text{msep}}_{C_{i,j} \mid D_{I}}(\hat{C}_{i,j}) + 2 \sum_{1 \leq i < k \leq J-1} \Lambda_{i,k,J}.
\]

134
6 Case Study

Using a series of real data, provided to us by AXA-Winterthur, Swiss Insurance Company, we present MSEP estimates for the multivariate CL model described by Model Assumptions 4.1. The dataset is displayed in four triangles, an aggregate triangle (used for the univariate calculations), and small, small to large, and large layer triangles (used for the multivariate calculations). Due to confidentiality, the data are scaled by some factor.

When calculating the MSEP for the univariate case, we use the conditional resampling approach as set forth in Buchwalder et al. (2006). We use this method since this approach was implemented in the multivariate calculations presented in the previous section. Recall that this is a decision made when estimating the parameter estimation error. Other methods could have been implemented mirroring the standard univariate situation, for example one could use the idea for the MSEP calculations presented in Mack (1993) to estimate the multivariate MSEP.

The dataset is provided in Table 4, the aggregate table, Table 5, the small claims layer, Table 6, the claims transitioning from small to large, and Table 7, the large claims layer (these tables are found in Appendix C). The results for the univariate and multivariate analysis are presented in Table 2. Note that in this case study the threshold $d_i$ was determined naturally by internal processes. In Table 1, we provide the values of the factor estimates as well as the variance parameter estimates of our multivariate CL method.

<table>
<thead>
<tr>
<th>$j$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{f}_j^{(s)}$</td>
<td>1.386</td>
<td>1.019</td>
<td>1.006</td>
<td>1.002</td>
<td>1.002</td>
<td>1.001</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\hat{g}_j^{(s)}$</td>
<td>0.013</td>
<td>0.003</td>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>$\hat{f}_j^{(l)}$</td>
<td>1.240</td>
<td>1.152</td>
<td>1.063</td>
<td>1.083</td>
<td>1.066</td>
<td>1.105</td>
<td>1.015</td>
<td>1.047</td>
<td>1.039</td>
</tr>
<tr>
<td>$\hat{g}_j^{(s)}$</td>
<td>43.045</td>
<td>0.659</td>
<td>0.139</td>
<td>0.100</td>
<td>0.070</td>
<td>0.023</td>
<td>0.007</td>
<td>0.011</td>
<td>0.004</td>
</tr>
<tr>
<td>$\hat{g}_j^{(s)}$</td>
<td>2.345</td>
<td>0.383</td>
<td>0.524</td>
<td>0.131</td>
<td>0.064</td>
<td>0.262</td>
<td>0.012</td>
<td>0.258</td>
<td>0.043</td>
</tr>
<tr>
<td>$\hat{g}_j^{(s)}$</td>
<td>51.650</td>
<td>77.453</td>
<td>7.530</td>
<td>63.973</td>
<td>11.100</td>
<td>96.683</td>
<td>1.382</td>
<td>16.246</td>
<td>10.736</td>
</tr>
<tr>
<td>$\hat{g}_j^{(l)}$</td>
<td>0.387</td>
<td>0.347</td>
<td>0.524</td>
<td>0.447</td>
<td>0.773</td>
<td>0.279</td>
<td>0.565</td>
<td>0.245</td>
<td>0.708</td>
</tr>
<tr>
<td>$\hat{g}_j^{(s)}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\hat{g}_j^{(l)}$</td>
<td>1.008</td>
<td>1.017</td>
<td>1.010</td>
<td>1.070</td>
<td>1.005</td>
<td>1.027</td>
<td>1.006</td>
<td>1.013</td>
<td>1.000</td>
</tr>
<tr>
<td>$\hat{g}_j^{(s)}$</td>
<td>0.003</td>
<td>0.013</td>
<td>0.018</td>
<td>0.005</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\hat{g}_j^{(l)}$</td>
<td>0.000</td>
<td>0.821</td>
<td>0.024</td>
<td>0.082</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\hat{g}_j^{(s)}$</td>
<td>0.499</td>
<td>1.455</td>
<td>2.455</td>
<td>32.459</td>
<td>0.203</td>
<td>12.459</td>
<td>0.475</td>
<td>2.120</td>
<td>0.475</td>
</tr>
<tr>
<td>$\hat{g}_j^{(l)}$</td>
<td>0.319</td>
<td>0.597</td>
<td>0.163</td>
<td>0.646</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 1: Factor and variance parameter estimates.

These parameter estimates provide evidence that we obtain a natural split. Note that $\hat{f}_j^{(s)}$ is almost one after five development periods (i.e. it demonstrates very stable behaviour). In contrast to the instability of the $\hat{f}_j^{(l)}$, where there is fluctuation in the later development years. As can be seen from Table 2, in our example, the multivariate CL method resulted in lower claims reserves and MSEP, as well as a
lower coefficient of variation (defined as the $msep^{1/2}$ divided by the claims reserves). Note that we do not necessarily expect the reserves to be less in the multivariate algorithm than in the univariate one. In contrast, we do expect this for the MSEP. This is due to the layering of different kinds of risks and claims; refer to Table 3 for a deeper look at the effect of layering.

<table>
<thead>
<tr>
<th>i</th>
<th>CL Reserves</th>
<th>CL Process St.dev</th>
<th>CL Parameter St.dev</th>
<th>$msep^{1/2}$</th>
<th>Vco</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.19</td>
<td>48.47</td>
<td>49.45</td>
<td>69.23</td>
<td>36955.52%</td>
</tr>
<tr>
<td>2</td>
<td>84.74</td>
<td>114.45</td>
<td>95.95</td>
<td>149.35</td>
<td>176.24%</td>
</tr>
<tr>
<td>3</td>
<td>125.94</td>
<td>130.79</td>
<td>109.48</td>
<td>170.57</td>
<td>135.44%</td>
</tr>
<tr>
<td>4</td>
<td>273.72</td>
<td>236.39</td>
<td>155.71</td>
<td>283.07</td>
<td>103.42%</td>
</tr>
<tr>
<td>5</td>
<td>297.20</td>
<td>239.31</td>
<td>156.97</td>
<td>286.20</td>
<td>96.30%</td>
</tr>
<tr>
<td>6</td>
<td>665.80</td>
<td>653.83</td>
<td>320.43</td>
<td>728.13</td>
<td>109.36%</td>
</tr>
<tr>
<td>7</td>
<td>736.54</td>
<td>652.23</td>
<td>313.66</td>
<td>723.73</td>
<td>98.26%</td>
</tr>
<tr>
<td>8</td>
<td>962.96</td>
<td>706.23</td>
<td>330.26</td>
<td>779.64</td>
<td>80.96%</td>
</tr>
<tr>
<td>9</td>
<td>1,254.58</td>
<td>790.62</td>
<td>397.33</td>
<td>892.18</td>
<td>63.84%</td>
</tr>
<tr>
<td>10</td>
<td>1,397.45</td>
<td>798.82</td>
<td>391.29</td>
<td>917.13</td>
<td>53.57%</td>
</tr>
<tr>
<td>11</td>
<td>1,816.38</td>
<td>874.61</td>
<td>438.50</td>
<td>978.37</td>
<td>53.86%</td>
</tr>
<tr>
<td>12</td>
<td>1,711.99</td>
<td>829.47</td>
<td>391.29</td>
<td>917.13</td>
<td>53.57%</td>
</tr>
<tr>
<td>13</td>
<td>2,302.98</td>
<td>1,120.57</td>
<td>469.71</td>
<td>1,215.04</td>
<td>52.76%</td>
</tr>
<tr>
<td>14</td>
<td>2,616.03</td>
<td>1,141.34</td>
<td>464.17</td>
<td>1,232.12</td>
<td>47.10%</td>
</tr>
<tr>
<td>15</td>
<td>3,222.16</td>
<td>1,227.91</td>
<td>495.17</td>
<td>1,323.99</td>
<td>41.09%</td>
</tr>
<tr>
<td>16</td>
<td>3,870.25</td>
<td>1,231.21</td>
<td>479.28</td>
<td>1,321.21</td>
<td>34.14%</td>
</tr>
<tr>
<td>17</td>
<td>5,427.43</td>
<td>1,365.31</td>
<td>472.96</td>
<td>1,444.91</td>
<td>26.62%</td>
</tr>
<tr>
<td>18</td>
<td>26,807.29</td>
<td>2,308.32</td>
<td>651.36</td>
<td>2,398.46</td>
<td>8.95%</td>
</tr>
<tr>
<td>Total</td>
<td>53,573.64</td>
<td>4,121.64</td>
<td>5,380.80</td>
<td>6,777.97</td>
<td>12.65%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i</th>
<th>Layered Reserves</th>
<th>Lay. Process St.dev</th>
<th>Lay. Parameter St.dev</th>
<th>$msep^{1/2}$</th>
<th>Vco</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.19</td>
<td>49.18</td>
<td>45.28</td>
<td>66.83</td>
<td>34859.37%</td>
</tr>
<tr>
<td>2</td>
<td>47.22</td>
<td>95.55</td>
<td>58.20</td>
<td>111.88</td>
<td>236.95%</td>
</tr>
<tr>
<td>3</td>
<td>61.91</td>
<td>99.62</td>
<td>55.80</td>
<td>113.66</td>
<td>183.59%</td>
</tr>
<tr>
<td>4</td>
<td>168.96</td>
<td>236.36</td>
<td>114.24</td>
<td>262.52</td>
<td>155.37%</td>
</tr>
<tr>
<td>5</td>
<td>122.76</td>
<td>191.37</td>
<td>74.25</td>
<td>205.27</td>
<td>167.21%</td>
</tr>
<tr>
<td>6</td>
<td>358.19</td>
<td>371.94</td>
<td>138.83</td>
<td>397.00</td>
<td>110.84%</td>
</tr>
<tr>
<td>7</td>
<td>479.60</td>
<td>410.81</td>
<td>161.42</td>
<td>441.39</td>
<td>92.03%</td>
</tr>
<tr>
<td>8</td>
<td>843.71</td>
<td>564.84</td>
<td>244.73</td>
<td>615.57</td>
<td>72.96%</td>
</tr>
<tr>
<td>9</td>
<td>1,285.74</td>
<td>683.60</td>
<td>363.58</td>
<td>774.27</td>
<td>60.22%</td>
</tr>
<tr>
<td>10</td>
<td>2,203.29</td>
<td>581.81</td>
<td>230.77</td>
<td>625.91</td>
<td>64.29%</td>
</tr>
<tr>
<td>11</td>
<td>1,785.35</td>
<td>809.18</td>
<td>378.07</td>
<td>893.15</td>
<td>50.03%</td>
</tr>
<tr>
<td>12</td>
<td>1,555.54</td>
<td>736.98</td>
<td>307.36</td>
<td>798.50</td>
<td>51.33%</td>
</tr>
<tr>
<td>13</td>
<td>3,572.74</td>
<td>1,447.25</td>
<td>745.75</td>
<td>1,628.10</td>
<td>45.57%</td>
</tr>
<tr>
<td>14</td>
<td>2,203.68</td>
<td>1,021.00</td>
<td>348.36</td>
<td>1,078.79</td>
<td>48.95%</td>
</tr>
<tr>
<td>15</td>
<td>2,798.34</td>
<td>1,227.20</td>
<td>405.37</td>
<td>1,292.42</td>
<td>46.19%</td>
</tr>
<tr>
<td>16</td>
<td>4,029.48</td>
<td>1,445.08</td>
<td>514.06</td>
<td>1,533.79</td>
<td>38.06%</td>
</tr>
<tr>
<td>17</td>
<td>5,054.69</td>
<td>1,443.82</td>
<td>420.47</td>
<td>1,503.80</td>
<td>29.75%</td>
</tr>
<tr>
<td>18</td>
<td>26,491.37</td>
<td>2,484.69</td>
<td>641.87</td>
<td>2,566.26</td>
<td>9.69%</td>
</tr>
<tr>
<td>Total</td>
<td>51,833.03</td>
<td>4,211.09</td>
<td>4,527.38</td>
<td>6,184.22</td>
<td>11.93%</td>
</tr>
</tbody>
</table>

Table 2: Reserve and MSEP results for both the univariate and multivariate algorithms.

In Table 3, we present the aggregated results for the small layer, the large layer, the aggregated multivariate model and the aggregated univariate model. We conclude that, as in our example, a large proportion of claims can have a big impact on the reserves but a small impact on the prediction uncertainty. As is evident from Table 3 a smaller collection of claims with high severity can make a huge impact on the prediction uncertainty. For this reason, practitioners often subdivide these two classes of claims to get more reliable predictions. If the split is done in layers then, in general, the lower layer is rather stable, whereas the upper layer requires special care. Hence, we provided the mathematical tools for the MSEP calculations.
required under the separation of claims. Note that we have used the CL method in the upper layer to get feasible MSEP formulas. Future research should also consider other reserving methods more appropriate to the upper layer that consider all information available, such as payments, case reserves, etc.

<table>
<thead>
<tr>
<th>Reserves</th>
<th>Process Parameter</th>
<th>msep$^{1/2}$</th>
<th>Vco</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Layer</td>
<td>28,446.79</td>
<td>1,641.22</td>
<td>527.52</td>
</tr>
<tr>
<td>Large Layer</td>
<td>23,386.24</td>
<td>3,680.26</td>
<td>4,437.39</td>
</tr>
<tr>
<td>Total Multivariate</td>
<td>51,833.03</td>
<td>4,211.09</td>
<td>4,527.58</td>
</tr>
<tr>
<td>Total Univariate</td>
<td>53,573.64</td>
<td>4,121.64</td>
<td>5,380.80</td>
</tr>
</tbody>
</table>

Table 3: The effect of layering on the CL algorithm.

Acknowledgements

The authors would like to kindly thank AXA-Winterthur, Swiss Insurance Company for providing the dataset presented in Section 6 and Prof. Dr. Paul Embrechts for contributing valuable commentary related to this work. Furthermore, the first author would like to extend his gratitude to ACE Limited for providing the financial support necessary to complete this research.

Bibliography


### A Derivation of the Gammas (Process Variance Terms)

From equation (5), recall that,

\[
\Gamma_{i,j}^{(x,y)} = \widehat{\text{Cov}}(C_{i,j}^{(x)}, C_{i,j}^{(y)}|D_I),
\]

for \( i \in \{1, \ldots, I\} \) and \( j \in \{I - i + 1, \ldots, J\} \). We study the following covariance:

\[
\text{Cov}(C_{i,j}^{(x)}, C_{i,j}^{(y)}|D_I) = \text{Cov}(E[C_{i,j}^{(x)}|C_{i,j-1}^{(y)}], E[C_{i,j}^{(y)}|C_{i,j-1}^{(y)}]|D_I) + E[\text{Cov}(C_{i,j}^{(x)}, C_{i,j}^{(y)}|C_{i,j-1}^{(y)})|D_I].
\]

This leads us to a recursive equation, one we demonstrate for the case \( x = y = s \).

For \( j \in \{I - i + 1, \ldots, J\} \),

\[
\text{Var}(C_{i,j}^{(s)}|D_I) = \text{Var}(E[C_{i,j-1}^{(s)}|C_{i,j-1}^{(s)}]|D_I) + E[\text{Var}(C_{i,j}^{(s)}|C_{i,j-1}^{(s)})|D_I]
\]

\[
= (f_{j-1}^{(s)})^2 \text{Var}(C_{i,j-1}^{(s)}|D_I) + (s_{j-1}^{(s)})^2 E[C_{i,j-1}^{(s)}|D_I],
\]

where \( \text{Var}(C_{i,1-i}^{(s)}|D_I) = 0 \) and \( E[C_{i,1-i}^{(s)}|D_I] = C_{i,i}^{(s)} \). Hence we formulate the estimator \( \Gamma_{i,j}^{(s,s)} \), recursively, as follows:

\[
\Gamma_{i,j}^{(s,s)} = (f_{j-1}^{(s)})^2 \Gamma_{i,j-1}^{(s,s)} + (s_{j-1}^{(s)})^2 \widehat{C}_{i,j-1}^{(s)}.
\]
Solving this recursive equation, we obtain the following:

\[ \Gamma_{i,j}^{(s,s)} = \left( \hat{C}_{i,j} \right)^2 \sum_{k=I-i}^{j-i} \frac{(\hat{\sigma}_k^{(s)})^2}{\hat{C}_{i,k}^{(s)}(\hat{\sigma}_k^{(s)})^2}, \]

for \( i \in \{1, \ldots, I\} \) and \( j \in \{I - i + 1, \ldots, J\} \).

Below, we provide the recursive equations for the required \( \Gamma_{i,j}^{(x,y)} \) (which are straightforward from (8)). For \( j \in \{I - i + 1, \ldots, J\} \),

\[
\begin{align*}
\Gamma_{i,j}^{(s,s)} &= \left( \hat{f}_{i,j-1}^{(s)} \right)^2 \Gamma_{i,j-1}^{(s,s)} + (\hat{\sigma}_j^{(s)})^2 \hat{C}_{i,j-1}, \\
\Gamma_{i,l}^{(s,l)} &= \hat{f}_{i-1,j-1}^{(s)} \Gamma_{i,j-1}^{(s,s)} + \hat{f}_{i-1}^{(s)} \hat{\sigma}_{j-1}^{(s)} \Gamma_{i,j-1}^{(s,l)} + \hat{\sigma}_{j-1}^{(s)} \hat{\sigma}_{j-1}^{(s)} \hat{\sigma}_{j-1}^{(s)} \hat{C}_{i,j-1}, \\
\Gamma_{i,j}^{(l,l)} &= \left( \hat{f}_{i-1,j-1}^{(l)} \right)^2 \Gamma_{i,j-1}^{(s,l)} + \hat{f}_{i-1}^{(l)} \hat{\sigma}_{j-1}^{(l)} \Gamma_{i,j-1}^{(l,l)} + (\hat{\sigma}_{j-1}^{(l)})^2 \hat{C}_{i,j-1},
\end{align*}
\]

where \( \Gamma_{i,j}^{(x,y)} = 0 \). This means that one calculates \( \Gamma_{i,j}^{(s,s)} \) recursively, then uses this result to determine \( \Gamma_{i,j}^{(s,l)} \), and finally \( \Gamma_{i,j}^{(l,l)} \).

**B Derivation of the Deltas (Estimation Error Terms)**

Recall that the \( \Delta \)'s are used to estimate the volatility of the CL factor estimators around their true values. As already stated, we apply a conditional resampling approach, as described in Buchwalder et al. (2006) and in Section 3.2.3 of Wüthrich and Merz (2008). The conditional resampling approach successively resamples the next step of the Markov process keeping the observations \( D_I \) as fixed volume measures, see e.g. (3.18) of Wüthrich and Merz (2008). This implies that, as also seen in Murphy (1994), we obtain a product structure seen in (3.26) of Wüthrich and Merz (2008). Hence for \( i \in \{1, \ldots, I\} \) and \( j \in \{I - i + 1, \ldots, J\} \), we obtain

\[
\Delta_{i,j}(m,n) = \prod_{k=I-i}^{j-1} \left( \hat{f}_i^{(x)} \hat{f}_j^{(y)} \right) + \frac{\hat{\sigma}_i^{(x)} \hat{\sigma}_i^{(y)} \hat{\sigma}_i^{(x)}}{C_{i,k}^{(x)} C_{i,k}^{(y)}} \sum_{k=I-i}^{j-1} \sum_{k=I-i}^{j-1} C_{k,k}^{(y)} C_{k,k}^{(x)} \sqrt{C_{k,j}^{(x)} C_{k,j}^{(y)}} - \prod_{k=I-i}^{j-1} \hat{f}_i^{(x)} \hat{f}_j^{(y)},
\]

where

\[
x = \begin{cases} s, & \text{for } m > k, \\
s \rightarrow l, & \text{for } m = k, \\
l, & \text{for } m < k, \end{cases}
\]

\[
y = \begin{cases} s, & \text{for } n > k, \\
s \rightarrow l, & \text{for } n = k, \\
l, & \text{for } n < k. \end{cases}
\]

and

\[
C_{k,j}^{(u)} = \begin{cases} C_{k,j}^{(s)}, & \text{for } m \geq k, \\
C_{k,j}^{(l)}, & \text{for } m < k, \end{cases}
\]

\[
C_{k,j}^{(v)} = \begin{cases} C_{k,j}^{(s)}, & \text{for } n \geq k, \\
C_{k,j}^{(l)}, & \text{for } n < k. \end{cases}
\]
Remark B.1  Note that $\hat{\rho}_{ij}^{(x,y)}$ is nonzero only for the pair $(s, s \to l)$. Recall that this is due to Model Assumptions 4.1, furthermore note that, as stated before, one could estimate all $\hat{\rho}_{ij}^{(x,y)}$ if they were thought to have an impact. We have omitted them because we believe that in most cases they are not of a level of importance that warrants additional computational complexity.
### Table 4: Dataset: aggregated cumulative payments.

<table>
<thead>
<tr>
<th>i</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>34,308</td>
<td>47,614</td>
<td>48,929</td>
<td>49,242</td>
<td>49,439</td>
<td>49,635</td>
<td>51,141</td>
<td>51,199</td>
<td>51,235</td>
<td>51,284</td>
</tr>
<tr>
<td>1</td>
<td>35,956</td>
<td>50,858</td>
<td>52,837</td>
<td>53,168</td>
<td>53,337</td>
<td>54,217</td>
<td>54,412</td>
<td>54,444</td>
<td>54,664</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>41,025</td>
<td>58,175</td>
<td>59,649</td>
<td>60,511</td>
<td>60,882</td>
<td>61,138</td>
<td>61,481</td>
<td>61,642</td>
<td>61,977</td>
<td>62,150</td>
</tr>
<tr>
<td>4</td>
<td>50,281</td>
<td>66,627</td>
<td>68,700</td>
<td>69,097</td>
<td>69,249</td>
<td>69,814</td>
<td>69,840</td>
<td>69,860</td>
<td>69,858</td>
<td>69,912</td>
</tr>
<tr>
<td>5</td>
<td>50,420</td>
<td>67,810</td>
<td>69,130</td>
<td>69,763</td>
<td>70,249</td>
<td>70,425</td>
<td>70,470</td>
<td>70,487</td>
<td>70,540</td>
<td>70,781</td>
</tr>
<tr>
<td>6</td>
<td>54,851</td>
<td>72,975</td>
<td>74,255</td>
<td>75,256</td>
<td>75,768</td>
<td>75,840</td>
<td>76,012</td>
<td>76,180</td>
<td>76,400</td>
<td>76,416</td>
</tr>
<tr>
<td>7</td>
<td>50,343</td>
<td>70,427</td>
<td>71,862</td>
<td>72,372</td>
<td>73,248</td>
<td>73,424</td>
<td>73,281</td>
<td>73,306</td>
<td>73,872</td>
<td>73,926</td>
</tr>
<tr>
<td>8</td>
<td>50,121</td>
<td>69,529</td>
<td>71,067</td>
<td>71,902</td>
<td>73,316</td>
<td>73,753</td>
<td>73,796</td>
<td>73,870</td>
<td>73,928</td>
<td>74,033</td>
</tr>
<tr>
<td>9</td>
<td>62,933</td>
<td>87,788</td>
<td>89,612</td>
<td>90,016</td>
<td>90,427</td>
<td>90,756</td>
<td>90,834</td>
<td>90,878</td>
<td>91,643</td>
<td>92,418</td>
</tr>
<tr>
<td>10</td>
<td>60,245</td>
<td>82,981</td>
<td>84,948</td>
<td>85,741</td>
<td>86,072</td>
<td>86,541</td>
<td>86,789</td>
<td>86,948</td>
<td>87,006</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>62,166</td>
<td>88,135</td>
<td>90,203</td>
<td>91,027</td>
<td>91,204</td>
<td>91,339</td>
<td>93,319</td>
<td>93,332</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>55,272</td>
<td>78,592</td>
<td>80,562</td>
<td>81,892</td>
<td>82,488</td>
<td>82,920</td>
<td>83,010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>56,504</td>
<td>80,825</td>
<td>84,994</td>
<td>85,960</td>
<td>86,348</td>
<td>87,476</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>57,555</td>
<td>82,264</td>
<td>83,926</td>
<td>84,604</td>
<td>84,717</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>57,181</td>
<td>85,054</td>
<td>86,912</td>
<td>87,617</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>57,695</td>
<td>81,224</td>
<td>83,069</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>52,872</td>
<td>74,784</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>53,839</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>51,284</td>
<td>61,289</td>
<td>61,927</td>
<td>61,655</td>
<td>63,186</td>
<td>63,186</td>
<td>63,191</td>
<td>63,194</td>
<td>63,194</td>
<td>63,197</td>
</tr>
<tr>
<td>1</td>
<td>54,664</td>
<td>74,759</td>
<td>54,843</td>
<td>54,906</td>
<td>55,001</td>
<td>55,073</td>
<td>55,147</td>
<td>55,231</td>
<td>55,373</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>67,589</td>
<td>67,660</td>
<td>67,733</td>
<td>67,863</td>
<td>68,149</td>
<td>68,150</td>
<td>68,160</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>69,912</td>
<td>69,999</td>
<td>70,414</td>
<td>70,694</td>
<td>70,718</td>
<td>70,745</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>70,781</td>
<td>70,787</td>
<td>70,851</td>
<td>70,855</td>
<td>70,872</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>76,416</td>
<td>76,421</td>
<td>76,426</td>
<td>76,433</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>73,926</td>
<td>73,944</td>
<td>74,067</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>74,033</td>
<td>74,047</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>92,418</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>0</td>
<td>33,452</td>
<td>45,469</td>
<td>46,706</td>
<td>46,834</td>
<td>46,953</td>
<td>47,046</td>
<td>47,052</td>
<td>47,059</td>
<td>47,074</td>
<td>33,452</td>
</tr>
<tr>
<td>1</td>
<td>34,815</td>
<td>48,411</td>
<td>49,567</td>
<td>49,811</td>
<td>49,918</td>
<td>50,070</td>
<td>50,156</td>
<td>50,212</td>
<td>50,222</td>
<td>50,222</td>
</tr>
<tr>
<td>2</td>
<td>40,716</td>
<td>56,891</td>
<td>58,264</td>
<td>58,734</td>
<td>59,075</td>
<td>59,099</td>
<td>59,106</td>
<td>59,174</td>
<td>59,215</td>
<td>59,215</td>
</tr>
<tr>
<td>3</td>
<td>45,010</td>
<td>63,431</td>
<td>64,272</td>
<td>64,671</td>
<td>64,755</td>
<td>64,816</td>
<td>64,842</td>
<td>64,898</td>
<td>64,947</td>
<td>64,947</td>
</tr>
<tr>
<td>5</td>
<td>49,902</td>
<td>66,815</td>
<td>68,159</td>
<td>68,345</td>
<td>68,416</td>
<td>68,457</td>
<td>68,474</td>
<td>68,525</td>
<td>68,581</td>
<td>68,581</td>
</tr>
<tr>
<td>6</td>
<td>54,161</td>
<td>71,997</td>
<td>73,151</td>
<td>73,642</td>
<td>73,734</td>
<td>73,811</td>
<td>73,933</td>
<td>74,004</td>
<td>74,024</td>
<td>74,024</td>
</tr>
<tr>
<td>7</td>
<td>49,994</td>
<td>66,524</td>
<td>70,608</td>
<td>71,227</td>
<td>71,241</td>
<td>71,267</td>
<td>71,272</td>
<td>71,293</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>49,750</td>
<td>68,492</td>
<td>69,731</td>
<td>70,142</td>
<td>70,309</td>
<td>70,417</td>
<td>70,445</td>
<td>70,504</td>
<td>70,526</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>60,640</td>
<td>84,938</td>
<td>86,259</td>
<td>86,528</td>
<td>86,711</td>
<td>86,748</td>
<td>86,826</td>
<td>86,851</td>
<td>86,939</td>
<td>86,939</td>
</tr>
<tr>
<td>10</td>
<td>59,576</td>
<td>81,888</td>
<td>83,467</td>
<td>84,021</td>
<td>84,131</td>
<td>84,292</td>
<td>84,407</td>
<td>84,445</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>61,443</td>
<td>86,542</td>
<td>88,135</td>
<td>88,646</td>
<td>88,899</td>
<td>89,244</td>
<td>89,935</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>54,708</td>
<td>76,959</td>
<td>78,512</td>
<td>79,134</td>
<td>79,356</td>
<td>79,587</td>
<td>79,645</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>55,295</td>
<td>78,280</td>
<td>80,037</td>
<td>80,498</td>
<td>80,706</td>
<td>80,837</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>56,647</td>
<td>80,294</td>
<td>83,467</td>
<td>84,021</td>
<td>84,131</td>
<td>84,292</td>
<td>84,407</td>
<td>84,445</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>59,884</td>
<td>83,285</td>
<td>84,821</td>
<td>85,296</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>56,780</td>
<td>79,052</td>
<td>80,437</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>52,279</td>
<td>73,372</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>53,245</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>47,074</td>
<td>47,079</td>
<td>47,141</td>
<td>47,147</td>
<td>47,185</td>
<td>47,193</td>
<td>47,195</td>
<td>47,196</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>50,222</td>
<td>50,247</td>
<td>50,276</td>
<td>50,282</td>
<td>50,286</td>
<td>50,286</td>
<td>50,286</td>
<td>50,286</td>
<td>50,285</td>
<td>50,285</td>
</tr>
<tr>
<td>3</td>
<td>64,924</td>
<td>64,956</td>
<td>64,956</td>
<td>64,992</td>
<td>64,992</td>
<td>64,993</td>
<td>64,993</td>
<td>65,003</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>67,196</td>
<td>67,220</td>
<td>67,278</td>
<td>67,282</td>
<td>67,283</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>68,581</td>
<td>68,586</td>
<td>68,638</td>
<td>68,640</td>
<td>68,645</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>74,024</td>
<td>74,030</td>
<td>74,034</td>
<td>74,041</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>71,293</td>
<td>71,312</td>
<td>71,372</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>70,526</td>
<td>70,529</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>86,947</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5: Dataset: small claim layer cumulative payments.

<table>
<thead>
<tr>
<th>i</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>1,190</td>
<td>438</td>
<td>40</td>
<td>68</td>
<td>78</td>
<td>296</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>365</td>
<td>25</td>
<td>16</td>
<td>21</td>
<td>-</td>
<td>89</td>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>919</td>
<td>104</td>
<td>51</td>
<td>50</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>20</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>450</td>
<td>-</td>
<td>7</td>
<td>12</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>371</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>322</td>
<td>225</td>
<td>-</td>
<td>14</td>
<td>184</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>311</td>
<td>233</td>
<td>268</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>175</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>275</td>
<td>113</td>
<td>511</td>
<td>-</td>
<td>4</td>
<td>50</td>
<td>104</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>551</td>
<td>91</td>
<td>14</td>
<td>95</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>32</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>626</td>
<td>248</td>
<td>371</td>
<td>7</td>
<td>9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>21</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>783</td>
<td>277</td>
<td>-</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>327</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>335</td>
<td>264</td>
<td>139</td>
<td>-</td>
<td>20</td>
<td>61</td>
<td>-</td>
<td>-</td>
<td>28</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>795</td>
<td>293</td>
<td>208</td>
<td>-</td>
<td>-</td>
<td>417</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>960</td>
<td>394</td>
<td>682</td>
<td>313</td>
<td>185</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>1,043</td>
<td>572</td>
<td>36</td>
<td>61</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>919</td>
<td>14</td>
<td>292</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>635</td>
<td>248</td>
<td>72</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>-</td>
<td>1,154</td>
<td>390</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>-</td>
<td>655</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 6: Dataset: small to large claim cumulative payments.
<table>
<thead>
<tr>
<th>i, j</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>856</td>
<td>2,145</td>
<td>2,458</td>
<td>2,536</td>
<td>2,604</td>
<td>2,682</td>
<td>4,095</td>
<td>4,147</td>
<td>4,177</td>
<td>4,210</td>
</tr>
<tr>
<td>1</td>
<td>1,140</td>
<td>2,447</td>
<td>3,270</td>
<td>3,357</td>
<td>3,419</td>
<td>3,545</td>
<td>4,061</td>
<td>4,200</td>
<td>4,223</td>
<td>4,442</td>
</tr>
<tr>
<td>2</td>
<td>310</td>
<td>1,284</td>
<td>1,385</td>
<td>1,777</td>
<td>1,907</td>
<td>2,079</td>
<td>2,382</td>
<td>2,536</td>
<td>2,803</td>
<td>2,953</td>
</tr>
<tr>
<td>3</td>
<td>696</td>
<td>1,728</td>
<td>1,799</td>
<td>2,022</td>
<td>2,081</td>
<td>2,098</td>
<td>2,203</td>
<td>2,203</td>
<td>2,664</td>
<td>2,665</td>
</tr>
<tr>
<td>4</td>
<td>1,108</td>
<td>1,514</td>
<td>2,178</td>
<td>2,486</td>
<td>2,391</td>
<td>2,727</td>
<td>2,711</td>
<td>2,714</td>
<td>2,695</td>
<td>2,716</td>
</tr>
<tr>
<td>5</td>
<td>519</td>
<td>996</td>
<td>1,280</td>
<td>1,604</td>
<td>1,904</td>
<td>2,009</td>
<td>2,013</td>
<td>2,013</td>
<td>2,205</td>
<td>2,200</td>
</tr>
<tr>
<td>6</td>
<td>690</td>
<td>979</td>
<td>1,104</td>
<td>1,614</td>
<td>2,025</td>
<td>2,029</td>
<td>2,079</td>
<td>2,177</td>
<td>2,391</td>
<td>2,391</td>
</tr>
<tr>
<td>7</td>
<td>349</td>
<td>933</td>
<td>1,254</td>
<td>1,509</td>
<td>2,021</td>
<td>2,025</td>
<td>2,039</td>
<td>2,039</td>
<td>2,604</td>
<td>2,633</td>
</tr>
<tr>
<td>8</td>
<td>370</td>
<td>1,037</td>
<td>1,336</td>
<td>1,761</td>
<td>3,007</td>
<td>3,336</td>
<td>3,352</td>
<td>3,402</td>
<td>3,424</td>
<td>3,470</td>
</tr>
<tr>
<td>9</td>
<td>1,862</td>
<td>2,850</td>
<td>3,353</td>
<td>3,488</td>
<td>3,717</td>
<td>4,007</td>
<td>4,008</td>
<td>4,027</td>
<td>4,704</td>
<td>5,470</td>
</tr>
<tr>
<td>10</td>
<td>669</td>
<td>1,093</td>
<td>1,480</td>
<td>1,720</td>
<td>1,941</td>
<td>2,249</td>
<td>2,382</td>
<td>2,514</td>
<td>2,561</td>
<td>2,561</td>
</tr>
<tr>
<td>11</td>
<td>723</td>
<td>1,594</td>
<td>2,068</td>
<td>2,381</td>
<td>2,392</td>
<td>2,440</td>
<td>4,395</td>
<td>4,397</td>
<td>4,397</td>
<td>4,397</td>
</tr>
<tr>
<td>12</td>
<td>564</td>
<td>1,632</td>
<td>2,049</td>
<td>2,758</td>
<td>3,131</td>
<td>3,331</td>
<td>3,365</td>
<td>3,365</td>
<td>3,365</td>
<td>3,365</td>
</tr>
<tr>
<td>13</td>
<td>1,209</td>
<td>2,545</td>
<td>4,956</td>
<td>5,461</td>
<td>5,642</td>
<td>6,639</td>
<td>6,639</td>
<td>6,639</td>
<td>6,639</td>
<td>6,639</td>
</tr>
<tr>
<td>14</td>
<td>508</td>
<td>1,969</td>
<td>2,039</td>
<td>2,336</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
</tr>
<tr>
<td>15</td>
<td>897</td>
<td>1,769</td>
<td>2,090</td>
<td>2,321</td>
<td>2,321</td>
<td>2,321</td>
<td>2,321</td>
<td>2,321</td>
<td>2,321</td>
<td>2,321</td>
</tr>
<tr>
<td>17</td>
<td>594</td>
<td>1,412</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
</tr>
<tr>
<td>18</td>
<td>593</td>
<td>1,412</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
<td>2,388</td>
</tr>
</tbody>
</table>

Table 7: Dataset: large claim layer cumulative payments.
Curriculum Vitae

Personal Data
Name Daniel H. Alai
Date of Birth August 9, 1983
Nationality Dutch

Education
   Topic: Stochastic Claims Reserving
   Supervisor: Prof. Dr. Paul Embrechts
09/2001 – 05/2006 Bachelor of Mathematics,
   Double Honours in Actuarial Science and Statistics,
   University of Waterloo, Canada
06/2001 High School Diploma,
   Canterbury High School, Ottawa, Canada

Employment
10/2006 – 12/2009 Teaching Assistant, ETH Zurich
05/2005 – 09/2005 Research Assistant, University of Waterloo
09/2004 – 01/2005 Junior Actuary, Manulife Financial
01/2004 – 05/2004 Property & Casualty Consultant, KPMG LLP
05/2003 – 09/2003 Actuarial Life Consultant, Tillinghast-Towers Perrin
09/2002 – 01/2003 Junior Actuary, Sun Life Insurance
01/2002 – 05/2002 Junior Actuary, Sun Life Insurance