Charles University in Prague
Faculty of Mathematics and Physics

MASTER THESIS

Simona Novotová

Bayesian Approaches to Stochastic Reserving

Department of Probability and Mathematical Statistics

Supervisor of the master thesis: RNDr. Michal Pešta, Ph.D.
Study programme: Mathematics
Specialization: Financial and Insurance Mathematics

Prague 2014

I would like to dedicate this thesis to my parents for their support and encouragement throughout my writing, studies and throughout my whole life.

I would thank to my supervisor RNDr. Michal Pešta, Ph.D. for his time, patience and all valuable materials which he provided me. His expert constructive advices helped me a lot.

Finally, special thanks to Mgr. Lucia Knapčoková for her correction of my English.

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In .... date .... signature of the author

Název práce: Bayesovské přístupy ve stochastickém rezervování

Autor: Bc. Simona Novotová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Michal Pešta, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Diplomová práce řeší bayesovský přístup ke stochastickému rezervování. Rezervování je v pojišťovnictví častým a diskutovaným problémem. Text práce představuje základní aktuárské pojmy a značení a následně vysvětluje základy bayesovské statistiky a odhadování. Hlavní část práce tvoří konkrétní modely využívající bayesovský princip. Pro každý z nich je odvozen podrobný postup pro stanovení odhadu celkových škod a rezerv. Cílem práce je ukázat také praktické využití modelů a vzájemné vztahy mezi jednotlivými metodami. Jsou uvedeny i příklady aplikace metod na reálná data. Výsledky jsou shrnuty v tabulkách a navzájem porovnány. Na konci se práce věnuje vlivu volby apriorního rozdělení na výslednou výšku rezerv.

Klíčová slova: bayesovský, priorní rozdělení, posteriorní rozdělení, rezervy

Title: Bayesian Approaches to Stochastic Reserving

Author: Bc. Simona Novotová

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Michal Pešta, Ph.D., Department of Probability and Mathematical Statistics

Abstract: In the master thesis the issue of bayesian approach to stochastic reserving is solved. Reserving problem is very discussed in insurance industry. The text introduces the basic actuarial notation and terminology and explains the bayesian inference in statistics and estimation. The main part of the thesis is framed by the description of the particular bayesian models. It is focused on the derivation of estimators for the reserves and ultimate claims. The aim of the thesis is to show the practical uses of the models and the relations between them. For this purpose the methods are applied on a real data set. Obtained results are summarized in tables and the comparison of the methods is provided. Finally the impact of a prior distribution on the resulting reserves is showed.

Keywords: bayesian, prior distribution, posterior distribution, reserves

Contents

Introduction

Claims reserving problem has been an area of research for actuarial professions for many years. The correct determination of the reserves amount has a key impact on the financial position of an insurance company, especially in the non-life insurance. Improper reserves affect entire company and can lead to serious problems.

For a long time the calculation of reserves had been understood as an algorithmic method with exactly specified steps. Only about 40 years ago, actuaries discovered a connection with stochastic modelling. Since that stochastic reserving methods have been much studied and a lot of methods for claims reserves estimation have been developed.

The most difficult question in a practical application of these methods is probably a model choice and an adequacy of the models. Different methods and data sets lead to different results. Every method is based on other ideas and use other principles.

Generally, the stochastic reserving methods can be divided into several groups, for example Distributional models, Generalized Linear Models, Bootstrap methods, etc.

In this thesis, we will deal with one particular branch of the reserving methods which is known as bayesian. This branch covers a lot of models which differ in their form and assumptions but the key idea of all of them consists in bayesian statistic. The term “bayesian” was first mentioned in 1760’s but the application of bayesian methods in insurance dates from 1960’s. Bayesian models are often seen as a “bridge” between the stochastic models and the pure deterministic models.

In short, bayesian reserving methods can be characterized as the methods which allow to incorporate an existing prior information. This information can be viewed as a single value or as a whole probability distribution of a proper underlying quantity. The estimation of ultimate claims and claims reserves employs two sources of information - the prior information and the observed data. Because of this, in many situations bayesian approach gives better results in comparison with the classical statistical approach.

This thesis is structured as follows. The first chapter explains the standard notation of data used in claims reserving theory. The data—incremental or cumulative claims—are understood as random variables ordered in the development triangles. In this chapter, two elementary methods for estimation of claims reserves—the Chain-Ladder method and the Bornhuetter-Ferguson method—are also presented. Although these methods are relatively simple, the results obtained in them often form the basis of the other more difficult models. The Bornhuetter-Ferguson method is even viewed as a bayesian method.

The second chapter deals with the mathematical frame of bayesian reserving methods - the bayesian statistics. Its grain is founded on Bayes theorem. It is formulated here and the terms as the prior and the posterior distribution are defined in this chapter. The process of estimation in the bayesian inference is also described here.

The concrete bayesian reserving models are presented in the third chapter. Each model starts with the formulation of its assumptions and continues with the derivation of the estimator for the claims reserves. At the beginning, two simpler bayesian methods—Benktander-Hoviven method and Cape-Cod model—are explained. They do not assume a prior distribution but only a prior estimate for the expected ultimate claims. The second part of the chapter is devoted to strict bayesian models with the prior distribution. It shows that with the smart choice of a prior distribution the posterior distribution can be obtained in explicit form.

The fourth chapter consists in the practical examples. The described claims reserving methods are applied to the concrete development triangle. All calculation are made in the software R. The results are summarised in the tables or in the graphs and there is always a confrontation and comparison of the results from the different methods. Mainly Poisson-Gamma model is analysed and in addition the impact of the choice of the prior distribution is studied in more detail.

1. Introduction to reserving theory

This thesis deals with the claims reserving problem for non-life insurance. This branch of insurance products comprises any insurance except life insurance. Life and non-life insurance products differ for example in terms of contracts and types of claims. This implies that they are modelled rather differently. In non-life insurance, there is often a time-lag between claims occurrence and claims reporting to the insurer called reporting delay. The reported claim can also take long time to settle and several years may elapse before claims payment process is finally closed. Sometimes even closed claims need to be reopened due to new developments. This time line of non-life insurance claim is shown in Figure 1.1

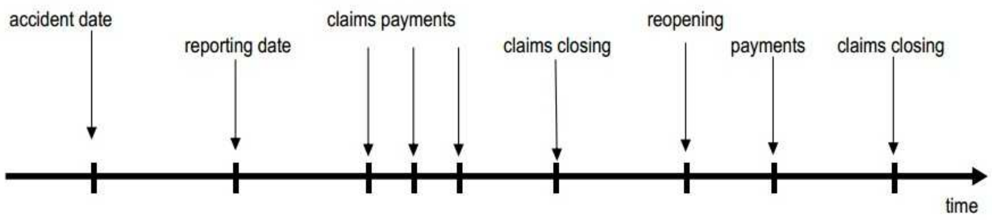

Figure 1.1: Typical time line of non-life insurance claim.

Because of these delays, the insurance company cannot settle a claim immediately after its accident day and have to build so-called claims reserves. Claims reserve represents the money which should be held by the insurer so as to be able to meet all future claims arising from policies currently in force and policies written in the past.

There are two different types of claims reserves. Reserves for claims that have been reported but have not been settled, so called RBNS (Reported But Not Settled) reserves. They are determined individually by an expert. The second type are IBNR (Incurred But Not Reported) reserves for claims that have occurred but have not been reported. They often contain also IBNeR (Incurred But Not enough Reported) reserves for not enough reported incurred claims. IBNR reserves are estimated by wide range of stochastic methods some of which are studied in the next chapters.

The notation used in this chapter came from [Mandl, Mazurová–1999] and [Wüthrich, Merz–2008].

1.1 Notation and assumptions

In this subsection we present a classical claims reserving notation and terminology and we will use it throughout the whole thesis. The main purpose of the claims reserving in non-life insurance is to set amount of IBNR reserves. When calculating them we go from information about past losses and development of payments. These data are ordered in so-called claims *development triangles* or *run-off triangles* which separate the data on two time axis. The notation of the development triangle is seen in Table 1.1.

Accident year i	Development year j					
	0	1		$\cdots$	$J-1$	J
0	$X_{0,0}$	$X_{0,1}$		$\cdots$	$X_{0,J-1}$	$X_{0,J}$
1	$X_{1,0}$	$X_{1,1}$		$\cdots$	$X_{1,J-1}$	$X_{1,J}$
$\vdots$	$\vdots$	$\vdots$				$\vdots$
		Observations $X_{i,j}$, $i+j \leq I$				
$\vdots$	$\vdots$	$\vdots$		Predicted values $X_{i,j}$, $i+j > I$		$\vdots$
$I-1$	$X_{I-1,0}$	$X_{I-1,1}$	$X_{I-1,2}$	$\cdots$		$X_{I-1,J}$
I	$X_{I,0}$	$X_{I,1}$	$X_{I,2}$	$\cdots$		$X_{I,J}$

Table 1.1: Claims development triangle

The index i corresponds to the accident year, the year of occurrence and the index j to the development year or the development period—numbers of years subsequently elapsed. I denotes the most recent accident year and J the last development year. Entries of the triangle are denoted $X_{i,j}$. $X_{i,j}$ represents all payments for claims with accident year i settled with a delay of j years and hence in the development year j and in the calendar year $i + j$. We refer to $X_{i,j}$ as *incremental payments* of accident year i and the development year j. We can also assume *cumulative payments* $C_{i,j}$ for the accident year i and the development year j. They are given by

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

and are interpreted as the payments of the accident year i settled with a delay at most j years so not later than in the development year j. The development triangle containing cumulative payments is called cumulative development triangle.

At time I, the claims development tables are split into two parts, the upper and the lower triangle. The upper triangles, denoted by D_I and $\tilde{D}_I$, contain observable incremental payments $X_{i,j}$, $i + j \leq I$ or cumulative payments $C_{i,j}$, $i + j \leq I$, respectively.

$$D_I = \{X_{i,j};\ i + j \leq I,\ 0 \leq j \leq J\},$$

$$\tilde{D}_I = \{C_{i,j};\ i + j \leq I,\ 0 \leq j \leq J\}.$$

The lower triangles D_I^C and $\tilde{D}_I^{\,C}$ include future, non-observable values of the outstanding incremental or cumulative payments and they need to be estimated.

$$D_I^C = \{X_{i,j};\ i + j > I,\ i \leq I,\ j \leq J\},$$

$$\tilde{D}_I^{\,C} = \{C_{i,j}; i + j > I,\ i \leq I,\ j \leq J\}.$$

The cumulative payment $C_{i,J}$ in the last development year is often said to be the *ultimate claim amount* or the *total claim amount* of the accident year i.

If $i + j = I$, then $C_{i,j} = C_{i,I-i}$ and $X_{i,j} = X_{i,I-i}$ are called *present cumulative* or *incremental payments*. They represent the most present observable values of payments and they are crucial in the most reserving methods.

In addition to incremental payments, $X_{i,j}$ may also represent change of reported claim amount or the number of reported claims of accident year i settled with delay of j years. Then $C_{i,j}$ denotes total incurred claims or the total number of reported claims with accident year i and reporting delay at most j years.

If $X_{i,j}$ denote incremental payments, we can define *outstanding loss liabilities* $R_{i,j}$ for the accident year i at time j which are given by

$$R_{i,j} = \sum_{k=j+1}^{J} X_{i,k} = C_{i,J} - C_{i,j}.$$

We denote

$$R_i = R_{i,I-i} = C_{i,J} - C_{i,I-i}.$$

R_i represents present the outstanding loss liabilities—reserves for the accident year i and need to be predicted. The predicted values of R_i together with $C_{i,I-i}$ give a predictor for the ultimate claim $C_{i,J}$.

In the claims reserving techniques entries of run-off triangles, $X_{i,j}$ and $C_{i,j}$ respectively, are considered to be random variables. They are observable in D_I and $\tilde{D}_I$ respectively, and all predictions come from these observations.

In this thesis we will always focus on the estimation of $C_{i,J}$. To simplify the notation and formulas we make a general assumption throughout this thesis that $I = J$ and $X_{i,j} = 0$ for all $j > J$. This assumption is adapted from [Wüthrich, Merz–2008]. Then we have to predict $C_{i,J}$ for accident years $i = 1, ..., I$.

1.2 Basic reserving methods

Before we start studies of bayesian reserving models we introduce two elementary techniques for estimating IBNR claims reserves, the Chain-Ladder (CL) method and the Bornhuetter-Ferguson (BF) method. Despite their simplicity they often give accurate results and form the basis of other more complicated models. They work with the development triangles and use conditionally expected value $E[C_{i,J} \mid \tilde{D}_I]$ as the estimator of an outcome of the random variable $C_{i,J}$.

1.2.1 Chain-Ladder method

The Chain-Ladder method is one of the most popular techniques. It was developed by T. Mack in [Mack–1993] and it is based on an assumption of proportionality of columns—development years—in a run-off triangle. This model is distribution-free and works with these two assumptions about the cumulative claims

Model Assumptions 1.2.1.

1. *Cumulative claims $C_{i,j}$ of different accident years i are independent.*

2. *There exists factors* $f_0, ..., f_{J-1} > 0$ *such that for all* $0 \le i \le I$ *and for all* $1 \le j \le J$ *we have*

$$\mathrm{E}[C_{i,j} \mid C_{i,0}, ..., C_{i,j-1}] = \mathrm{E}[C_{i,j} \mid C_{i,j-1}] = f_{j-1} C_{i,j-1}.$$

As will be seen later, the independence of the accident years is assumed in almost all reserving methods.

Factors f_j, $j = 0, ..., J-1$ are called *link ratios*, *development factors* or *CL factors*.

Under 1.2.1 we can derive that

$$\begin{aligned}\mathrm{E}[C_{i,J} \mid \tilde{D}_I] &= \mathrm{E}[C_{i,J} \mid C_{i,0}, ..., C_{i,I-i}] \\ &= \mathrm{E}[\mathrm{E}[C_{i,J} \mid C_{i,J-1}] \mid C_{i,0}, ..., C_{i,I-i}] \\ &= f_{J-1}\, \mathrm{E}[C_{i,J-1} \mid \tilde{D}_I], \quad \text{for } i = 1,\ ...,\ I.\end{aligned}$$

We can go on iteratively until we are on diagonal where $i + j = I$ and we get

$$\mathrm{E}[C_{i,J} \mid \tilde{D}_I] = f_{J-1} ... f_{I-i} C_{i,I-i}. \tag{1.1}$$

The relationship (1.1) give us technique how to predict the amount $C_{i,J}$ given cumulative observations in the upper triangle $\tilde{D}_I$. In the most practical applications, CL factors f_j, $j = 0,\ ...,\ J-1$ are not known and we need to estimate them. The Chain-Ladder method estimates these factors as follows

$$\widehat{f_j} = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}}. \tag{1.2}$$

By using (1.2) for f_j we set the Chain-Ladder estimator for the ultimate claim, denoted by $\widehat{C_{i,J}^{CL}}$, as

$$\widehat{C_{i,J}^{CL}} = \widehat{\mathrm{E}}[C_{i,J} \mid \tilde{D}_I] = \widehat{f}_{J-1} ... \widehat{f}_{I-i} C_{i,I-i}, \quad \text{for } i = 1,\ ...,\ I \tag{1.3}$$

and for the outstanding reserves $\widehat{R_{i,J}^{CL}}$ as

$$\widehat{R_{i,J}^{CL}} = \widehat{C_{i,J}^{CL}} - C_{i,I-i} = C_{i,I-i}(\widehat{f}_{J-1} ... \widehat{f}_{I-i} - 1), \quad \text{for } i = 1,\ ...,\ I.$$

It is worth to remark that under 1.2.1, estimators (1.2) and (1.3) are unbiased estimators of f_j and $\mathrm{E}[C_{i,J}]$, that is $\mathrm{E}[\widehat{f_j}] = f_j$ and $\mathrm{E}[\widehat{C_{i,J}^{CL}}] = \mathrm{E}[C_{i,J}]$. Moreover,

$$\mathrm{E}[\widehat{f_0} ... \widehat{f_j}] = \mathrm{E}[\widehat{f_0}] ... \mathrm{E}[\widehat{f_j}],$$

for all $j = 1,\ ...,\ J-1$. So, $\widehat{f_0} ... \widehat{f}_{J-1}$ are uncorrelated. Proofs of these statements and other properties can be found in [Wüthrich, Merz–2008].

In this model we have only the assumption on the first moments. We can also include variance assumption when we want to quantify uncertainties in these estimates. Such model is known as Mack Chain-Ladder model. It is presented and described in details in [Wüthrich, Merz–2008] and [England, Verrall–2002].

1.2.2 Bornhuetter-Ferguson method

The next basic method which is grounded on the run-off triangles is Bornhuetter-Ferguson method. It was published in the article "The Actuary and IBNR" [Bornhuetter, Ferguson–1972]. It is considered to be one of the bayesian techniques because it does not use only the observations from the development triangles but it also allows an expert judgement to be incorporated.

This method works with these two assumptions

Model Assumptions 1.2.2.

1. *Cumulative claims $C_{i,j}$ of different accident years i are independent.*
2. *There exists factors $\mu_0, ..., \mu_I > 0$ and pattern $\beta_0, ..., \beta_J > 0, \beta_J = 1$ such that for all $0 \leq i \leq I$, $0 \leq j \leq J-1$ and for all $1 \leq k \leq J-j$ we have*

$$\begin{aligned} \mathrm{E}[C_{i,0}] &= \beta_0 \mu_i, \\ \mathrm{E}[C_{i,j+k} \mid C_{i,0}, ..., C_{i,j}] &= \mathrm{E}[C_{i,j}] + (\beta_{j+k} - \beta_j)\mu_i. \end{aligned}$$

Moreover, the assumptions 1.2.2 implies that

$$\mathrm{E}[C_{i,j}] = \mathrm{E}[\mathrm{E}[C_{i,j} \mid C_{i,0}]] = \mathrm{E}[C_{i,0}] + (\beta_j - \beta_0)\mu_i = \beta_j \mu_i, \tag{1.4}$$

for $0 \leq i \leq I$ and $0 \leq j \leq J$ and we can formulate implied assumption for Bornhuetter-Ferguson model which are also often used for this method.

Model Assumptions 1.2.3.

1. *Cumulative claims $C_{i,J}$ of different accident years i are independent.*
2. *There exists factors $\mu_0, ..., \mu_I > 0$ and pattern $\beta_0, ..., \beta_J > 0, \beta_J = 1$ such that for all $0 \leq i \leq I$ and $0 \leq j \leq J$ we have*

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

It is important to point that assumptions 1.2.2 and 1.2.3 are not equivalent. 1.2.2 are stronger and imply 1.2.3 but not vice versa.

From relation (1.4) we have for fixed $j = J$ an equality $\mathrm{E}[C_{i,J}] = \mu_i$. So, the parameter μ_i represents the expected value of the ultimate claim $C_{i,J}$.

We want to derive estimator of the ultimate claims amount $C_{i,J}$ with the use of observations from $\tilde{D}_I$. Under 1.2.2, the conditional expected value of $C_{i,J}$ given $\tilde{D}_I$ can be expressed by the following relation

$$\begin{aligned} \mathrm{E}[C_{i,J} \mid \tilde{D}_I] &= \mathrm{E}[C_{i,J} \mid C_{i,0}, ..., C_{i,I-i}] \\ &= C_{i,I-i} + \mathrm{E}[C_{i,J} - C_{i,I-i} \mid C_{i,0}, ..., C_{i,I-i}] \\ &= \mathrm{E}[C_{i,I-i}] + (1 - \beta_{I-i})\mu_i. \end{aligned} \tag{1.5}$$

However, under implied assumptions 1.2.3 the relationship (1.5) holds true only if we assume that $C_{i,J} - C_{i,I-i}$ is independent of $C_{i,0}$, ..., $C_{i,I-i}$.

(1.5) gives us a formula for Bornhuetter-Ferguson estimator of ultimate claim

$$\widehat{C_{i,J}^{BF}} = \widehat{\mathrm{E}}[C_{i,J} \mid D_I] = C_{i,I-i} + (1 - \widehat{\beta}_{I-i})\widehat{\mu_i}, \quad 1 \leq i \leq I. \tag{1.6}$$

The parameter $\widehat{\beta}_{I-i}$ is an estimator of β_{I-i} and $\widehat{\mu_i}$ is an estimator of μ_i and $\mathrm{E}[C_{i,J}]$, respectively. Now, it only remains to obtain these estimators. We solve this problem by comparing Chain-Ladder and Bornhuetter-Ferguson method. We consider the assumptions 1.2.1 for Chain-Ladder technique. Then it holds

$$\mathrm{E}[C_{i,J}] = \mathrm{E}[\mathrm{E}[C_{i,J} \mid C_{i,J-1}]] = \mathrm{E}[C_{i,j}] \prod_{k=j}^{J-1} f_k$$

and

$$\mathrm{E}[C_{i,j}] = \mathrm{E}[C_{i,J}] \prod_{k=j}^{J-1} f_k^{-1}. \tag{1.7}$$

The term $\prod_{k=j}^{J-1} f_k^{-1}$ in (1.7) depends only on j and $\mathrm{E}[C_{i,J}]$ on i. So, we can write

$$\mathrm{E}[C_{i,j}] = \underbrace{\mathrm{E}[C_{i,J}]}_{\mu_i} \underbrace{\prod_{k=j}^{J-1} f_k^{-1}}_{\beta_j} = \mu_i \beta_j$$

and we set

$$\beta_j = \prod_{k=j}^{J-1} f_k^{-1}, \quad j = 0, \ldots, J-1. \tag{1.8}$$

where f_k ,$k = 0$, ...,$J-1$ are the developments factors from Chain-Ladder. The relation (1.8) between the CL factors and development pattern can be use only for BF model with implied assumptions 1.2.3. Under 1.2.2 this deduction cannot be done because assumptions 1.2.2 are not implied by the assumptions of Chain-Ladder.

Using the identity (1.8), the Bornhuetter-Ferguson estimator (1.6) can be expressed as follows

$$\widehat{C_{i,J}^{BF}} = C_{i,I-i} + (1 - \widehat{\beta}_{I-i})\widehat{\mu_i} = C_{i,I-i} + \left(1 - \frac{1}{\prod_{k=I-i}^{J-1} \widehat{f}_k}\right)\widehat{\mu_i}. \tag{1.9}$$

On the other hand, by rewriting (1.3) we obtain

$$\begin{aligned}
\widehat{C_{i,J}^{CL}} &= C_{i,I-i} \prod_{k=I-i}^{J-1} \widehat{f}_k = C_{i,I-i} + C_{i,I-i}\left(\prod_{k=I-i}^{J-1} \widehat{f}_k - 1\right) \\
&= C_{i,I-i} + \frac{\widehat{C_{i,J}^{CL}}}{\prod_{k=I-i}^{J-1} \widehat{f}_k}\left(\prod_{k=I-i}^{J-1} \widehat{f}_k - 1\right) \\
&= C_{i,I-i} + \left(1 - \frac{1}{\prod_{k=I-i}^{J-1} \widehat{f}_k}\right)\widehat{C_{i,J}^{CL}}. \qquad (1.10)
\end{aligned}$$

We can see that (1.10) accords to expression of the Bornhuetter-Ferguson ultimate claim (1.9) with $\widehat{\mu_i} = \widehat{C_{i,J}^{CL}}$.

So, the only difference between the CL and BF method is in the choice of the estimator for $\mathrm{E}[C_{i,J}]$. In Bornhuetter-Ferguson we use a prior estimate $\widehat{\mu_i}$. It is often a plan value from a business plan based on an expert opinion or the value used for premium calculations and it should be estimated before one has any observations $\tilde{D}_I$.

Chain-Ladder uses an opposite principle and replaces the prior estimate with the estimate $\widehat{C_{i,J}^{CL}}$ which is based only on the observed data. It means that CL and BF method constitute two extreme positions in claims reserving theory.

We have showed, that in calculation $\widehat{C_{i,J}^{BF}}$ one combines the prior information—in estimation of μ_i, with the observations—in estimation of β_j. From this reason Bornhutter-Ferguson method is considered to belong to the bayesian techniques which are the main theme of this thesis and will be presented in details in further chapters.

2. Bayesian approach to statistics

The bayesian theory constitutes a powerful branch of statistics which gives another view of statistical problems solving. It is useful in practice mostly when we have outcomes from the past. Bayesian approach offers a method of formalizing a priori beliefs and of combing them with an available observations. It gives a mathematical frame for the bayesian methods in actuarial science, specially in the claims reserving.

In this chapter, we introduce basic principles and terms of the bayesian statistics. Presented information comes from [Hušková–1985], [Bühlmann, Gisler–2005] and [Figueiredo–2004].

Suppose $\boldsymbol{X} : (\Omega, \mathcal{A}, P) \to (\mathbb{R}_n, \mathcal{B}_n)$ is a random vector, $\boldsymbol{X} = (X_1, ..., X_n)'$, with a density $r(\boldsymbol{x} \mid \boldsymbol{\theta})$ with respect to a σ- finite measure ν_n on $(\mathbb{R}_n, \mathcal{B}_n)$, where $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)'$ is a parameter with values from a set $\mathcal{H}$, a non-empty borel subset of $\mathbb{R}_k$.

In the classical approach we consider $\boldsymbol{\theta}$ to be an unknown constant or vector of constants, respectively. Only information about $\boldsymbol{X}$ is used to estimation of parameter $\boldsymbol{\theta}$. In the bayesian inference we regard $\boldsymbol{\Theta}$ as a random variable or a random vector taking values from the set $\mathcal{H}$.

So, let $\boldsymbol{\Theta} = (\Theta_1, ..., \Theta_k)'$ to be the random vector with density $q(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \mathcal{H}$ with respect to a σ- finite measure λ on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$. $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)' \in \mathcal{H}$ is a realization of $\boldsymbol{\Theta}$. The distribution of $\boldsymbol{\Theta}$ includes all information, experience and judgements which are disposable before realization of $\boldsymbol{X}$. Then the random vector $\boldsymbol{X}$ has the conditional density $r(\boldsymbol{x} \mid \boldsymbol{\theta})$ given $\boldsymbol{\Theta} = \boldsymbol{\theta}$ with respect to the ν_n. That is, for all measurable sets M and N it holds the relation

$$P[\boldsymbol{\Theta} \in M, \boldsymbol{X} \in N] = \int_M \left(\int_N r(\boldsymbol{x} \mid \boldsymbol{\theta}) \mathrm{d}\nu_n(\boldsymbol{x}) \right) q(\boldsymbol{\theta}) \mathrm{d}\lambda(\boldsymbol{\theta}).$$

The density $q(\boldsymbol{\theta})$ is called *a priori density* of $\boldsymbol{\Theta}$. We combine it with observations to give a *posterior density* $\pi(\boldsymbol{\theta} \mid \boldsymbol{x})$. This is done with Bayes theorem, which forms the basis of bayesian statistics.

Theorem 2.1 (Bayes theorem)**.** *The posterior density function for* $\boldsymbol{\Theta}$ *given* $\boldsymbol{X} = \boldsymbol{x}$*,* $\pi(\boldsymbol{\theta} \mid \boldsymbol{x})$ *is given by*

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) = \begin{cases} \frac{q(\boldsymbol{\theta}) r(\boldsymbol{x} \mid \boldsymbol{\theta})}{\int_{\mathcal{H}} q(\boldsymbol{\theta}) r(\boldsymbol{x} \mid \boldsymbol{\theta}) \mathrm{d}\lambda(\boldsymbol{\theta})} & \text{if } \int_{\mathcal{H}} q(\boldsymbol{\theta}) r(\boldsymbol{x} \mid \boldsymbol{\theta}) \mathrm{d}\lambda(\boldsymbol{\theta}) \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

For proof see [Hušková–1985].

The density $\pi(\boldsymbol{\theta} \mid \boldsymbol{x})$ is considered as an update of the previous prior distribution once the observation $\boldsymbol{x}$ have been obtained. Denominator in the expression (2.1) is a function only of $\boldsymbol{x}$, so, we can rewrite the Bayes theorem as

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) \propto q(\boldsymbol{\theta}) r(\boldsymbol{x} \mid \boldsymbol{\theta}).$$

It means that the posterior density is proportional to the product of the prior density $q(\boldsymbol{\theta})$ and $r(\boldsymbol{x} \mid \boldsymbol{\theta})$.

2.1 Choice of a priori distribution

The key issue in the bayesian theory is the choice of the prior distribution which should involve all information about $\boldsymbol{\Theta}$. In practice it never happens that we are able to identify the prior distribution exactly from the available information. There are many ways of defining it. Mostly, we assume the prior densities of a specific functional form with unknown parameters.

In the implementation of the bayes theory it might occur a problem that the posterior distribution cannot be calculated analytically. In a practical application we would like to eliminate this problem. Given a function $r(\boldsymbol{x} \mid \boldsymbol{\theta})$ we look for a prior distribution which is not only compatible with knowledge about $\boldsymbol{\Theta}$ but also leads to a posterior probability function in a closed-form expression. For this reason systems of a conjugate distributions are favoured.

Definition 2.1. *Suppose that a prior density $q(\boldsymbol{\theta})$ is from a family of densities Q. A random vector $\boldsymbol{X}$ has a density $r(\boldsymbol{x} \mid \boldsymbol{\theta})$. The family Q is said to be a conjugate with respect to the family of densities $R = \{r(\boldsymbol{x} \mid \boldsymbol{\theta}), \boldsymbol{\theta} \in \mathcal{H}\}$ if the posterior density $\pi(\boldsymbol{\theta} \mid \boldsymbol{x})$ also belongs to the family Q for all realization $\boldsymbol{x}$ of $\boldsymbol{X}$.*

It is obvious that the family of all possible densities is conjugate to R but it is not proper for practical applications. A useful conjugate family should contain as minimum densities as possible. When this is the case, computing of the posterior density from prior is not difficult.

Because of their tractability, conjugate systems are often used in bayesian reserving techniques and we will present such models in details in the next chapter.

2.2 Estimation in bayesian inference

In the classical approach when the parameter $\boldsymbol{\theta}$ is deterministic estimation and decisions about the parameter $\boldsymbol{\theta}$ are based only on the density $r(\boldsymbol{x} \mid \boldsymbol{\theta})$. In the bayesian approach, we have a random parameter $\boldsymbol{\Theta}$. Its true value $\boldsymbol{\theta}$ is a realization of the parameter and estimation of this true value is based on its posterior density (2.1).

Now, we explain basic terms and the technique of the bayesian estimation.

Definition 2.2. *Let $\mathcal{D}$ to be a set of all possible decision about the parameter realization $\boldsymbol{\theta}$. A function δ from $\mathbb{R}_n$ to $\mathcal{D}$ is called the decision function and its value in the point $\boldsymbol{x}$, $\delta(\boldsymbol{x})$, is called the decision about $\boldsymbol{\theta}$ given an observation $\boldsymbol{X} = \boldsymbol{x}$.*

The task of the bayesian estimation is to derive a decision $\delta^*(\boldsymbol{x})$ which achieves some sort of optimality criterion.

Let us consider $\mathcal{D} = \mathcal{H}$. In this situation is the optimal decision function δ^* rather our optimal bayesian estimator $\widehat{\boldsymbol{\theta}}^* = (\widehat{\theta}_1^*, ..., \widehat{\theta}_k^*)'$ of $\boldsymbol{\theta}$ which we have wanted to get.

Definition 2.3. *A loss function L is a measurable, bounded below function*

$$L : \mathcal{H} \times \mathcal{D} \to \mathbb{R},$$

where measurability is understood with respect to the σ-algebra $\mathcal{B}(\mathcal{H}) \otimes \sigma(\mathcal{D})$.

The value of the loss function $L(\boldsymbol{\theta}, \delta)$ expresses how much the decision $\delta \in \mathcal{D}$ are to be penalized when $\boldsymbol{\theta}$ is the true value of the parameter realization.

Definition 2.4. *Risk $R(\boldsymbol{\theta}, \delta)$ of the decision function δ and the loss function L when the true realization of parameter is $\boldsymbol{\theta}$ is defined as*

$$R(\boldsymbol{\theta}, \delta) = \mathrm{E}[L(\boldsymbol{\Theta}, \delta(\boldsymbol{X})) \mid \boldsymbol{\Theta} = \boldsymbol{\theta}] = \int_{\mathbb{R}_n} L(\boldsymbol{\theta}, \delta(\boldsymbol{x})) r(\boldsymbol{x} \mid \boldsymbol{\theta}) \mathrm{d}\nu_n(\boldsymbol{x}).$$

So, the risk $R(\boldsymbol{\theta}, \delta)$ is the conditional expected value of the loss function, in other words, it is the mean loss due to the choice of decision function δ when the true parameter realization is $\boldsymbol{\theta}$.

Further, we consider only the decision functions δ which satisfy $R(\boldsymbol{\theta}, \delta) < \infty$, $\forall\ \boldsymbol{\theta} \in \mathcal{H}$ and we regard Δ the set of these functions. Then, the function $R : \mathcal{H} \times \Delta \to \mathbb{R}$ is called the *risk function* and we can define the bayesian risk and the bayesian risk function.

Definition 2.5. *The random parameter $\boldsymbol{\Theta}$ has the prior density $q(\boldsymbol{\theta})$. The bayesian risk function $\rho(q, \delta)$ is defined by the following relation*

$$\rho(q, \delta) = \mathrm{E}\left[R(\boldsymbol{\Theta}, \delta)\right] = \int_{\mathcal{H}} \left(\int_{\mathbb{R}_n} L(\boldsymbol{\theta}, \delta(\boldsymbol{x})) r(\boldsymbol{x} \mid \boldsymbol{\theta}) \mathrm{d}\nu_n(\boldsymbol{x}) \right) q(\boldsymbol{\theta}) \mathrm{d}\lambda(\boldsymbol{\theta}),\ \delta \in \Delta. \tag{2.2}$$

Let $\delta_0 \in \Delta$. Then the bayesian risk of the decision function δ_0 is the value $\rho(q, \delta_0)$.

The bayesian risk constitutes our criterion for the choice of the optimal estimator of the true parameter realization $\boldsymbol{\theta}$. We want to find such an estimator $\widehat{\boldsymbol{\theta}}^*$ which minimizes it among all $\delta \in \Delta$. So

$$\widehat{\boldsymbol{\theta}}^* = \arg\min_{\delta \in \Delta} \rho(q, \delta). \tag{2.3}$$

We have to remark that the bayesian risk function (2.2) can be modified to the following form

$$\rho(q, \delta) = \int_{\mathbb{R}_n} \left\{ \left(\int_{\mathcal{H}} q(\boldsymbol{\theta}) r(\boldsymbol{x} \mid \boldsymbol{\theta}) \mathrm{d}\lambda(\boldsymbol{\theta}) \right) \left(\int_{\mathcal{H}} L(\boldsymbol{\theta}, \delta(\boldsymbol{x})) \pi(\boldsymbol{\theta} \mid \boldsymbol{x}) \mathrm{d}\lambda(\boldsymbol{\theta}) \right) \right\} \mathrm{d}\nu_n(\boldsymbol{x}).$$

This implies that the optimal bayesian estimator $\widehat{\boldsymbol{\theta}}^*$ is obtained by minimizing the posterior expected loss $\int_{\mathcal{H}} L(\boldsymbol{\theta}, \delta(\boldsymbol{x})) \pi(\boldsymbol{\theta} \mid \boldsymbol{x}) \mathrm{d}\lambda(\boldsymbol{\theta}) = \mathrm{E}[L(\boldsymbol{\Theta}, \delta(\boldsymbol{X})) \mid \boldsymbol{X} = \boldsymbol{x}]$ for a fixed $\boldsymbol{x} \in \mathbb{R}_n$ (on the assumption that $\int_{\mathcal{H}} q(\boldsymbol{\theta}) r(\boldsymbol{x} \mid \boldsymbol{\theta}) \mathrm{d}\lambda(\boldsymbol{\theta}) \neq 0$). So

$$\widehat{\boldsymbol{\theta}}^* = \arg\min_{\delta \in \mathcal{D}} \int_{\mathcal{H}} L(\boldsymbol{\theta}, \delta(\boldsymbol{x})) \pi(\boldsymbol{\theta} \mid \boldsymbol{x}) \mathrm{d}\lambda(\boldsymbol{\theta}), \quad \boldsymbol{x} \in \mathbb{R}_n. \tag{2.4}$$

Observe that (2.4) depends on the observation $\boldsymbol{x}$. However, this is not a problem because every time the decision is to be made the observation $\boldsymbol{x}$ is available.

We see, that the approach (2.3) with the bayesian risk is absolutely equivalent to the approach (2.4) with the posterior expected loss and both lead to the same optimal estimator $\widehat{\boldsymbol{\theta}}^*$.

The choice of a loss function and of a prior distribution affects the value of $\widehat{\boldsymbol{\theta}}^*$. The most common loss function is $L(\boldsymbol{\theta}, \delta(\boldsymbol{x})) = (\boldsymbol{\theta} - \delta(\boldsymbol{x}))^2$. It is called *quadratic loss function*. In such a case, the optimal bayesian estimator $\widehat{\boldsymbol{\theta}}^*$ is a posterior mean of $\boldsymbol{\theta}$ which can be expressed by the following relation

$$\widehat{\boldsymbol{\theta}}^* = \mathrm{E}[\boldsymbol{\Theta} \mid \boldsymbol{X} = \boldsymbol{x}] = \int_{\mathcal{H}} \boldsymbol{\theta} \pi(\boldsymbol{\theta} \mid \boldsymbol{x}) \mathrm{d}\lambda(\boldsymbol{\theta}).$$

For the proof see [Hušková–1985].

The use of quadratic loss function makes computations relatively straightforward. When we have the posterior density $\pi(\boldsymbol{\theta} \mid \boldsymbol{x})$, corresponding posterior expected value is usually easy to calculate. For this reason, quadratic loss function is often considered in practical applications.

There are also another types of used loss function which but we will not deal with them here. More information about them can be found in [Berger–1993].

3. Bayesian methods for claims reserving

In this chapter we investigate the main theme of this thesis—bayesian approaches to stochastic reserving—specifically to the estimation of the ultimate claims. In the recent years, these methods have been much studied and more applied in the actuarial science. They allow to include a relevant subjective judgement in a formal framework of stochastic model and that is their important benefit.

Bayesian methods for claims reserving can be characterised as methods in which an expert knowledge or other existing prior information is combined with observed data. The prior information can be given by a single value, for example by a prior estimate for the ultimate claim. In the strict bayesian approach, the prior information is described by a prior stochastic distribution of an underlying random variable such as ultimate claim or a risk parameter of accident year. The prior distribution is connected with observed data from the upper triangle D_I through the Bayes theorem (2.1) and so, a posterior distribution of the underlying quantity is obtained.

Following the posterior distribution the estimation is made. It is based on the combined sources of information therefore it may improve estimation from the classical methods and its precision may be better.

The notation of the models in this chapter is taken from [Wüthrich, Merz–2008]. We start with two models, Cape-Cod model and Benktander-Hovinen method, which do not assume a prior distribution but only a prior estimate for amount of the expected ultimate claim. So they do not belong to the strict bayesian methods but they use bayesian consideration. Both of them arise from Bornhuetter-Ferguson model and represent its extension.

3.1 Benktander-Hovinen method

In Section 1.2 we have showed that the Chain-Ladder estimator of the ultimate claims $C_{i,J}$ is based only on observations from a development triangle and it ignores a possible prior knowledge.

On the other hand, in the Bornhuetter-Ferguson approach, only the prior knowledge about the ultimate claim is used. Benktander-Hovinen (BH) method considers mixture of these two models. It was developed by Benktander and Hovinen who described it independently from each other in [Benktander–1976] and [Hovinen–1981].

The assumptions of this method correspond to BF implied assumptions 1.2.2 to which we additionally assume that the development pattern $\beta_0, \ldots, \beta_J > 0$ and the parameter μ_i are known.

We fix an accident year i, $1 \leq i \leq I$ and define a following credibility mixture

$$u_i(c) = c\,\widehat{C_{i,J}^{CL}} + (1-c)\mu_i, \quad \text{for } c \in [0,1], \tag{3.1}$$

where $\widehat{C_{i,J}^{CL}}$ is the Chain-Ladder estimator (1.3) and μ_i is the known prior estimate for $\mathrm{E}[C_{i,J}]$. Parameter c represents a credibility weight given to the Chain-Ladder estimator. When $c = 0$, (3.1) is equal to μ_i. The increasing amount of the observed data in the accident year i should imply the increase of c because estimator $\widehat{C_{i,J}^{CL}}$ is based on observations.

In the Benktander-Hovinen method we set $c = \beta_{I-i}$. Then we get Benktander-Hovinen estimator of the ultimate claim of the accident year i, $\widehat{C_{i,J}^{BH}}$, which has the form

$$\begin{aligned}\widehat{C_{i,J}^{BH}} &= C_{i,I-i} + (1-\beta_{I-i})u_i(\beta_{I-i}) \\ &= C_{i,I-i} + (1-\beta_{I-i})\left[\beta_{I-i}\,\widehat{C_{i,J}^{CL}} + (1-\beta_{I-i})\mu_i\right]. \end{aligned} \tag{3.2}$$

We assume the implied assumptions 1.2.2 and so we can use the identification (1.8) between CL factors f_j, $j = 0, ..., J-1$ and development pattern $(\beta_j)_{j=0,...,J}$, which has been derived in Subsection 1.2.2. $(\beta_j)_{j=0,...,J}$ are known and because of the relation (1.8) also the CL development factors , so $\widehat{f_j} = f_j$ and $\widehat{C_{i,J}^{CL}} = \frac{C_{i,I-i}}{\beta_{I-i}}$. This gives us the following form of the BH estimator (3.2)

$$\begin{aligned}\widehat{C_{i,J}^{BH}} &= C_{i,I-i} + (1-\beta_{I-i})\left[C_{i,I-i} + (1-\beta_{I-i})\mu_i\right] \\ &= C_{i,I-i} + (1-\beta_{I-i})\widehat{C_{i,J}^{BF}}. \end{aligned} \tag{3.3}$$

From the relationship (3.3) we can observe that the Benktander-Hovinen method is rather iterated Bornhuetter-Ferguson method which uses the BF estimator $\widehat{C}_{i,J}^{BF}$ as the new prior estimator of $\mathrm{E}[C_{i,J}]$.

We can also investigate the further iteration of BF method and look at the results it brings. We set $\widehat{C^0} = \mu_i$ and define

$$\widehat{C^{m+1}} = C_{i,I-i} + (1-\beta_{I-i})\widehat{C^m} \quad \text{for } m = 0, 1, \dots .$$

It is clear that for $m = 1$, $\widehat{C^1}$ is the BF estimator for the ultimate claim and for $m = 2$, $\widehat{C^2}$ is the BH estimator. Behaviour of $\widehat{C^m}$ for $m \to \infty$ is showed by the following theorem.

Theorem 3.1. *Under the assumptions 1.2.2 and that the development pattern $(\beta_j)_{j=0,...,J}$ is known and $\beta_{I-i} > 0$ it holds*

$$\lim_{m\to\infty} \widehat{C^m} = C_{i,J}^{CL}$$

This theorem says that if the BF method is further iterated the CL estimator is obtained. It can be proved by induction and the proof can be found in [Wüthrich, Merz–2008].

3.2 Cape-Cod model

Cape-Cod (CC) model is another reserving method which incorporates a prior judgement. In the Chain-Ladder technique presented in Section 1.2 we use diagonal value $C_{i,I-i}$ from a development triangle—the most recent observation—in the expression of estimator of the ultimate claims. If $C_{i,I-i}$ is an outlier, the CL reserving method may not give proper results. The Cape-Cod model can solve this problem by making diagonal observation more robust. It works with these assumptions

Model Assumptions 3.2.1.

1. *Cumulative claims* $C_{i,j}$ *of different accident years* i *are independent.*
2. *There exist parameters* $\pi_0,\dots,\ \pi_I,\ \kappa > 0$ *and a claims development pattern* $\beta_0,\ \dots,\ \beta_J > 0,\ \beta_J = 1$ *such that for all* $0 \le i \le I$ *is*

$$\mathrm{E}[C_{i,j}] = \kappa \pi_i \beta_j.$$

We see that these assumptions correspond with the implied assumptions 1.2.3 in the BF method with $\mu_i = \kappa\pi_i = \mathrm{E}[C_{i,J}]$. In the Cape-Cod model, parameter π_i can be interpreted as a premium received for the accident year i and we assume it is known. κ is the parameter independent of i and j and represents the average loss ratio for all accident years and we does not know its value. The development pattern $(\beta_j)_{j=0,\dots,J}$ is the same pattern as in BF model so again it holds the identity (1.8) between $(\beta_j)_{j=0,\dots,J}$ and CL factors $(f_j)_{j=0,\dots,J-1}$.

Let κ_i refers to a loss ratio for each accident year i, $i = 0,\dots,\ I$. Because $\beta_J = 1$ it can be expressed as the quotient of the expected ultimate claim $\mathrm{E}[C_{i,J}]$ and the premium received π_i

$$\kappa_i = \frac{\mathrm{E}[C_{i,J}]}{\pi_i}.$$

We have known π_i, so an estimate of κ_i, denoted $\widehat{\kappa_i}$, is given by

$$\widehat{\kappa_i} = \frac{\widehat{\mathrm{E}}[C_{i,J}]}{\pi_i}.$$

Using the CL estimator of ultimate claim as the estimator for $\mathrm{E}[C_{i,J}]$, $\widehat{\kappa_i}$ can be expressed in the form

$$\widehat{\kappa_i} = \frac{\widehat{C_{i,J}^{CL}}}{\pi_i} = \frac{C_{i,I-i}\prod_{j=I-i}^{J-1} f_j}{\pi_i} = \frac{C_{i,I-i}}{\beta_{I-i}\pi_i}. \tag{3.4}$$

Furthermore, so defined estimator $\widehat{\kappa_i}$ is an unbiased estimator for κ because

$$\mathrm{E}[\widehat{\kappa_i}] = \frac{\mathrm{E}[\widehat{C_{i,J}^{CL}}]}{\pi_i} = \frac{\mathrm{E}[C_{i,I-i}]}{\beta_{I-i}\pi_i} = \frac{\kappa\beta_{I-i}\pi_i}{\beta_{I-i}\pi_i} = \kappa.$$

However, in this model, we need an estimator of the average loss ratio κ independent of i. It can be set as a weighted average of $\widehat{\kappa_i}$ with weights $\beta_{I-i}\pi_i$. By doing this we get the estimate $\widehat{\kappa^{CC}}$ for the "robusted" overall loss ratio given by the next formula

$$\widehat{\kappa^{CC}} = \sum_{i=0}^{I} \frac{\beta_{I-i}\pi_i}{\sum_{k=0}^{I}\beta_{I-k}\pi_k}\widehat{\kappa_i} = \frac{\sum_{i=0}^{I} C_{i,I-i}}{\sum_{i=0}^{I}\beta_{I-i}\pi_i}. \tag{3.5}$$

Since $\widehat{\kappa_i}$ is unbiased for κ, $\widehat{\kappa^{CC}}$ is also the unbiased estimator of κ.

This leads to an estimator of the diagonal observation $C_{i,I-i}$ which is "robusted". It is denoted by $\widehat{C_{i,I-i}^{CC}}$ and given by the relation

$$\widehat{C_{i,I-i}^{CC}} = \widehat{\mathrm{E}}[C_{i,I-i}] = \widehat{\kappa^{CC}}\pi_i\beta_{I-i}. \tag{3.6}$$

$\widehat{C_{i,I-i}^{CC}}$ is unbiased estimator for $\mathrm{E}[C_{i,I-i}]$ and represents the smoothed diagonal value $C_{i,I-i}$. Smoothing is done over all accident years by the term $\widehat{\kappa^{CC}}$.

Now we are able to derive the Cape-Cod estimator of the ultimate claim $\widehat{C_{i,J}^{CC}}$. It is expressed by

$$\widehat{C_{i,J}^{CC}} = C_{i,I-i} - \widehat{C_{i,I-i}^{CC}} + \prod_{j=I-i}^{J-1} f_j \widehat{C_{i,I-i}^{CC}}, \quad \text{for } 1 \leq i \leq I. \tag{3.7}$$

We can observe that in the formula for the Cape-Cod estimator (3.7) the CL factors are applied to the "robusted" $\widehat{C_{i,I-i}^{CC}}$ and the difference between diagonal observation and the "robusted" diagonal value is added. So, the Cape-Cod model gives proper results also when there is an outlier on the diagonal.

There is also another way to look at the Cape-Cod estimator. We can rewrite (3.7) to the following form

$$\begin{aligned}
\widehat{C_{i,J}^{CC}} &= C_{i,I-i} + \widehat{C_{i,I-i}^{CC}} \left(\prod_{j=I-i}^{J-1} f_j - 1 \right) \\
&= C_{i,I-i} + \widehat{\kappa^{CC}} \pi_i \beta_{I-i} \left(\frac{1}{\beta_{I-i}} - 1 \right) \\
&= C_{i,I-i} + (1 - \beta_{I-i}) \widehat{\kappa^{CC}} \pi_i.
\end{aligned}$$

This form coincides with the formula of BF estimator (1.6) with the modified prior estimate $\widehat{\kappa^{CC}} \pi_i$.

3.3 Strict bayesian models

Up till now we have considered that our prior knowledge is given by a single value. Now we choose another approach and we introduce bayesian reserving models in which this prior information is represented by a prior distribution of an underlying quantity such as ultimate claim or a risk parameter

In here presented models we consider a latent random variable Θ_i. Θ_i is regarded as the risk characterization of an accident year i. Typically, we do not know the concrete values of Θ_i belonging to the accident years. However, we have knowledge about the structure of Θ_i. This information is summarized by a prior probability distribution which involves personal judgement and experience about the risk characteristic.

We suppose that risk characteristics Θ_i of different accident years i, $i = 0, \ldots, I$ are independent. Conditioning on Θ_i, we are able to say something about the structure of the incremental variables $X_{i,0}$, ..., $X_{i,J}$ of accident year i and about the cumulative variables $C_{i,0}$, ..., $C_{i,J}$, respectively. We assume that given Θ_i, $X_{i,j}$ are conditionally independent random variables with a conditional distribution which depends on Θ_i. The posterior distribution of Θ_i, given the incremental observations in the upper triangle D_I, is calculated through the Bayes theorem. It differs with dependence on a choice of the conditional distribution of $X_{i,j}$ and on a choice of the prior distribution of Θ_i. It cannot be always expressed analytically and by using some special numerical algorithms we obtain only its estimate.

In this thesis we restrict to the models when the posterior distribution can be expressed by an explicit formula.

Our goal is to find an estimator for variable $C_{i,J}$ for each accident year i. This is done by the bayesian estimation described in Section 2.2. Recall, that us till now we work with a fixed accident year i.

In all presented models, we suppose the quadratic loss function. Thus, the optimal bayesian estimator for $C_{i,J}$ is the posterior expected value $\mathrm{E}[C_{i,J} \mid D_I]$. Our first step will be the computation of the posterior distribution of Θ_i given D_I and the obtained result will help us to calculate $\mathrm{E}[C_{i,J} \mid D_I]$.

In the following subsections, we present the concrete models with the proper choice of distributions for $X_{i,j}$ and Θ_i so the posterior distribution can be obtained in an explicit form. We derive estimator for ultimate claims and show the advantages of bayesian approach.

3.3.1 Poisson model with gamma prior distribution

Poisson model with gamma prior distribution is used mainly when variables $X_{i,j}$ in a development triangle represent incremental number of reported claims in accident year i with delay of j years. The model works with these assumptions.

Model Assumptions 3.3.1.
There exist random variable Θ_i and positive constants γ_0, ..., γ_J with $\sum_{j=0}^{J} \gamma_j = 1$ such that for all $i = 0$, ..., I and $j = 0$, ..., J we have

1. *Conditionally, given $\Theta_i = \theta$, $X_{i,j}$ are independent and Poisson distributed with $\mathrm{E}[X_{i,j} \mid \Theta_i = \theta] = \theta\gamma_j$ and $\mathrm{var}[X_{i,j} \mid \Theta_i = \theta] = \theta\gamma_j$.*
2. *$(\Theta_i, (X_{i,0}, ..., X_{i,J}))'$, $i = 0$, ..., I are independent random vectors.*
3. *Θ_i is gamma distributed with shape parameter $a_i > 0$ and scale parameter $b_i > 0$.*

Observe, that in this model we do not make assumptions about the cumulative variables, as up to now, but about the incremental claims $X_{i,j}$. We assume that, the prior density of Θ_i, denoted by v_{a_i,b_i}, is the density of gamma distribution given by

$$v_{a_i,b_i}(\theta) = \frac{b_i^{a_i}}{\Gamma(a_i)} \theta^{a_i - 1} \exp(-b_i\theta), \quad \theta > 0.$$

and so the prior expected value $\mathrm{E}[\Theta_i]$ is equal to $\frac{a_i}{b_i}$. Given $\Theta_i = \theta$, $X_{i,j}$ are Poisson distributed with the parameter $\theta\gamma_j$, so they have the density $f_{i,j}^{\theta}$ with respect to counting measure given by the relation

$$f_{i,j}^{\theta}(x) = \exp(-\theta\gamma_j) \frac{(\theta\gamma_j)^x}{x!}, \quad x = 0, 1, ...$$

The assumptions 3.3.1 also imply that the expected value of $X_{i,j}$ can be expressed as

$$\mathrm{E}[X_{i,j}] = \mathrm{E}[\mathrm{E}[X_{i,j} \mid \Theta_i]] = \frac{a_i}{b_i} \gamma_j.$$

We can also derive relations for the conditional expected value and variance of $C_{i,J}$—the ultimate claim of accident year i or the total number of claims in accident year i, respectively. Under 3.3.1 it holds

$$\mathrm{E}[C_{i,J} \mid \Theta_i] = \mathrm{E}\left[\sum_{j=0}^{J} X_{i,j} \mid \Theta_i\right] = \sum_{j=0}^{J} \mathrm{E}[X_{i,j} \mid \Theta_i] = \sum_{j=0}^{J} \gamma_j \Theta_i = \Theta_i,$$

$$\mathrm{var}[C_{i,J} \mid \Theta_i] = \mathrm{var}\left[\sum_{j=0}^{J} X_{i,j} \mid \Theta_i\right] = \sum_{j=0}^{J} \mathrm{var}[X_{i,j} \mid \Theta_i] = \Theta_i.$$

In the derivation of the conditional variance of $C_{i,J}$ we use the assumptions of the conditional independence of $X_{i,j}$.

The Poisson-Gamma (Po-Ga) model assumes that $X_{i,j}$ cannot take negative values. When $X_{i,j}$ represent number of reported claims there is no problem. However, this model can be also used for development triangles of incremental payments. They may have negative values and in such a case a practical application is problematic.

Derivation of estimator for ultimate claim

Now we will use the model assumptions 3.3.1 to derive the estimator for the ultimate claim $C_{i,J}$.

As first, we calculate the posterior density of Θ_i accord to the Bayes theorem. Because the accident years—the vectors $(\Theta_i, (X_{i,0}, ..., X_{i,J}))'$, $i = 0, \ldots, I$ are independent it is sufficient to condition only with $X_{i,j}$ from accident year i. The joint conditional density of the vector $(X_{i,0}, ..., X_{i,j})'$, given $\Theta_i = \theta$, can be expressed as the product of densities $f_{i,k}^{\theta}$, $k = 0, ..., j$ because the random variables $X_{i,0}, ..., X_{i,j}$ are conditionally independent. So we get

$$f_{X_{i,0},...,X_{i,j}|\Theta_i}(x_{i,0}, ..., x_{i,j} \mid \theta) = \prod_{k=0}^{j} \exp(-\theta\gamma_k) \frac{(\theta\gamma_k)^{x_{i,k}}}{x_{i,k}!}.$$

Then according to the Bayes theorem, the posterior density of Θ_i, given $X_{i,0} = x_{i,0}, ..., X_{i,j} = x_{i,j}$, denoted by $\pi_{\Theta_i|X_{i,0},...,X_{i,j}}$, can be simplified to the following form

$$
\begin{aligned}
\pi_{\Theta_i|X_{i,0},...,X_{i,j}}(\theta \mid x_{i,0},...,x_{i,j}) &= \frac{\prod_{k=0}^{j} \exp(-\theta\gamma_k)\frac{(\theta\gamma_k)^{x_{i,k}}}{x_{i,k}!}\frac{b_i^{a_i}}{\Gamma(a_i)}\theta^{a_i-1}\exp(-b_i\theta)}{\int\limits_0^\infty \frac{b_i^{a_i}}{\Gamma(a_i)}\prod_{k=0}^{j}\exp(-\theta\gamma_k)\frac{(\theta\gamma_k)^{x_{i,k}}}{x_{i,k}!}\theta^{a_i-1}\exp(-b_i\theta)\mathrm{d}\theta} \\
&= \frac{\prod_{k=0}^{j}\exp[-(b_i+\gamma_k)\theta]\theta^{x_{i,k}+a_i-1}}{\int\limits_0^\infty \prod_{k=0}^{j}\exp[-(b_i+\gamma_k)\theta]\theta^{x_{i,k}+a_i-1}\mathrm{d}\theta} \\
&= \frac{\exp[-(b_i+\beta_j)\theta]\theta^{\sum_{k=0}^{j}x_{i,k}+a_i-1}}{\int\limits_0^\infty \theta^{\sum_{k=0}^{j}x_{i,k}+a_i-1}\exp[-(b_i+\beta_j)\theta]\mathrm{d}\theta} \\
&= \frac{\exp[-(b_i+\beta_j)\theta]\theta^{\sum_{k=0}^{j}x_{i,k}+a_i-1}}{\frac{\Gamma(\sum_{k=0}^{j}x_{i,k}+a_i)}{(b_i+\beta_j)^{\sum_{k=0}^{j}x_{i,k}+a_i}}}
\end{aligned}
\tag{3.8}
$$

We denote

$$
a_{i,j}^{post} = a_i + \sum_{k=0}^{j} x_{i,k}, \qquad b_{i,j}^{post} = b_i + \beta_j, \tag{3.9}
$$

where $\beta_j = \sum\limits_{k=0}^{j}\gamma_k$. Then the posterior density of Θ_i, given $X_{i,0} = x_{i,0}$, ...,$X_{i,j} = x_{i,j}$, can be expressed in the form

$$
\pi_{\Theta_i|X_{i,0},...,X_{i,j}}(\theta \mid x_{i,0},...,x_{i,j}) = \frac{(b_{i,j}^{post})^{a_{i,j}^{post}}}{\Gamma(a_{i,j}^{post})}\theta^{a_{i,j}^{post}-1}\exp(-b_{i,j}^{post}\theta) \quad \text{for} \quad \theta > 0 \tag{3.10}
$$

This is again the density of gamma prior distribution but with new parameters $a_{i,j}^{post}$ and $b_{i,j}^{post}$. So, the gamma distribution is the conjugate distribution for the Poisson distribution.

Observe that $a_{i,j}^{post}$ and $b_{i,j}^{post}$ are updated with information about observations $X_{i,0} = x_{i,0}$, ..., $X_{i,j} = x_{i,j}$ and depend not only on the accident year i but also on the development year j.

Specially, the conditional distribution of Θ_i given the incremental observations $D_I = d_I$, $d_I = \{x_{i,j},\ i+j \le I,\ 0 \le j \le J\}$ is gamma distribution with the shape parameter $a_{i,I-i}^{post}$ and the scale parameter $b_{i,I-i}^{post}$. This knowledge allows us to compute the posterior expected value $\mathrm{E}[\Theta_i \mid D_I]$. It holds true

$$
\mathrm{E}[\Theta_i \mid D_I = d_I] = \mathrm{E}[\Theta_i \mid X_{i,0} = x_{i,0},...,X_{i,I-i} = x_{i,I-i}] = \frac{a_{i,I-i}^{post}}{b_{i,I-i}^{post}}
$$

$$
\mathrm{E}[\Theta_i \mid D_I] = \frac{A_{i,I-i}^{post}}{b_{i,I-i}^{post}}, \tag{3.11}
$$

where $A^{post}_{i,I-i}$ is defined as $A^{post}_{i,I-i} = a_i + C_{i,I-i}$. The value $\mathrm{E}[\Theta_i \mid D_I]$ represents the optimal bayesian estimator of Θ_i. We can also rewrite (3.11) to the form

$$\begin{aligned}\mathrm{E}[\Theta_i \mid D_I] &= \frac{b_i}{b_i+\beta_{I-i}}\frac{a_i}{b_i} + \left(1-\frac{b_i}{b_i+\beta_{I-i}}\right)\frac{C_{i,I-i}}{\beta_{I-i}} \\ &= \frac{b_i}{b_i+\beta_{I-i}}\frac{a_i}{b_i} + \left(1-\frac{b_i}{b_i+\beta_{I-i}}\right)\frac{\sum_{k=0}^{I-i}\gamma_k\frac{X_{i,k}}{\gamma_k}}{\sum_{k=0}^{I-i}\gamma_k}.\end{aligned} \tag{3.12}$$

This expression implies that $\mathrm{E}[\Theta_i \mid D_I]$ can be interpreted as a credibility weighted average between the prior expected value $\mathrm{E}[\Theta_i] = \frac{a_i}{b_i}$ and the weighted average of the observation $\frac{X_{i,k}}{\gamma_k}$, $k = 0, ..., I-i$.

Now we will proceed with the derivation of relation for the posterior expected value of the ultimate claim $C_{i,J}$. Using the assumption of conditional independence of $X_{i,j}$, given Θ_i, we get the relation in the form

$$\begin{aligned}\mathrm{E}[C_{i,J} \mid D_I] &= \mathrm{E}[\mathrm{E}[C_{i,J} \mid \Theta_i, D_I] \mid D_I] \\ &= \mathrm{E}\left[\mathrm{E}\left[\sum_{j=0}^{J} X_{i,j} \mid \Theta_i, D_I\right] \mid D_I\right] \\ &= C_{i,I-i} + \mathrm{E}\left[\mathrm{E}\left[\sum_{j=I-i+1}^{J} X_{i,j} \mid \Theta_i\right] \mid D_I\right] \\ &= C_{i,I-i} + \mathrm{E}\left[\sum_{j=I-i+1}^{J} \gamma_j\Theta_i \mid D_I\right] \\ &= C_{i,I-i} + \sum_{j=I-i+1}^{J} \gamma_j\,\mathrm{E}[\Theta_i \mid D_I] \\ &= C_{i,I-i} + (1-\beta_{I-i})E[\Theta_i \mid D_I], \quad \text{for } 1 \le i \le I.\end{aligned} \tag{3.13}$$

This is the optimal bayesian estimator for ultimate claim in the Poisson-Gamma model. Substituting (3.12) into the relation (3.13) we have the estimator in the form

$$\widehat{C^{PoiGa}_{i,J}} = C_{i,I-i} + (1-\beta_{I-i})\left[\frac{b_i}{b_i+\beta_{I-i}}\frac{a_i}{b_i} + \left(1-\frac{b_i}{b_i+\beta_{I-i}}\right)\frac{C_{i,I-i}}{\beta_{I-i}}\right], \tag{3.14}$$

for $1 \le i \le I$.

We have used the conjugate distributions in the model. Because of that the computation of the estimator $\widehat{C^{PoiGa}_{i,J}}$ have been relatively easy and we have got all expression in a closed form.

In the next section we look at such cases in more details.

3.3.2 Exponential dispersion family with its associated conjugates

As have been already mentioned in Chapter 2, conjugate families of distributions are often used in the practice. The Poisson-Gamma model is only one example of such families. Now we generalize this model to the exponential dispersion family and find its associated conjugates.

Definition 3.1. *A distribution is said to belong to the exponential dispersion family (EDF), if it can be expressed as*

$$\mathrm{d}F(x) = \exp\left[\frac{x\theta - b(\theta)}{\frac{\sigma^2}{\omega}}\right] a\left(x, \frac{\sigma^2}{\omega}\right) \mathrm{d}\nu(x),\ x \in A \subseteq \mathbb{R}, \tag{3.15}$$

where $b(\cdot)$ is a real-valued twice differentiable function of θ, θ is a parameter of the family, $a(\cdot,\cdot)$ is a real-valued function of x and $\frac{\sigma^2}{\omega}$, ω and σ^2 are some real-valued constants. ν denotes either the Lebesgue measure or a counting measure.

The EDF includes a large class of families of distribution, for example, the families of the Poisson, gamma, normal and Bernoulli distributions. Each distribution within EDF is uniquely characterized by functions $b(\cdot)$ and $a(\cdot,\cdot)$. The parameter θ is called *canonical parameter*, σ^2 is referred to as the *dispersion parameter* and ω denotes a prior known weight.

The definition of exponential dispersion family can be also extended to the multidimensional case, but we will not need it in this thesis. It can be found in [Figueiredo–2004].

Now we show how EDF is used in reserving theory and introduce the exponential dispersion family model. It assumes following

Model Assumptions 3.3.2.

1. *There exists a claims development pattern $(\beta_j)_{0\le j\le J}$ with $\beta_J = 1$, $\gamma_0 = \beta_0 > 0$ and $\gamma_j = \beta_j - \beta_{j-1} > 0$ for $j = 1, ..., J$.*

2. *Conditionally, given $\Theta_i = \theta$, the $X_{i,j}$, $j = 0, ..., J$, are independent with distribution*

$$\frac{X_{i,j}}{\gamma_j \mu_i} \sim \mathrm{d}F^{\theta}_{i,j}(x) = \exp\left[\frac{x\theta - b(\theta)}{\frac{\sigma^2}{\omega_{i,j}}}\right] a\left(x, \frac{\sigma^2}{\omega_{i,j}}\right) \mathrm{d}\nu(x), \tag{3.16}$$

   where ν is the Lebesgue measure or a counting measure on $\mathbb{R}$, $b(\cdot)$ is a real-valued twice differentiable function of θ, $a(\cdot,\cdot)$ is a real-valued function of x and $\frac{\sigma^2}{\omega_{i,j}}$, $\omega_{i,j} > 0$, $\mu_i > 0$ and σ^2 are some real-valued constants and $F^{\theta}_{i,j}$ is a probability distribution on $\mathbb{R}$.

3. *The random vectors $(\Theta_i, (X_{i,0}, ..., X_{i,J}))'$, $i = 0, ..., I$ are independent and $\Theta_0, ..., \Theta_I$ are real-valued random variables with densities with respect to the Lebesgue measure*

$$u_{\mu,\tau^2}(\theta) = d(\mu, \tau^2) \exp\left[\frac{\mu\theta - b(\theta)}{\tau^2}\right],\ \theta \in H \subseteq \mathbb{R} \tag{3.17}$$

   with $\mu \equiv 1$ and $\tau^2 > 0$ and $d(\cdot,\cdot)$ is a real-valued function of μ and τ^2.

So in this model, given $\Theta_i = \theta$, $\frac{X_{i,j}}{\gamma_j \mu_i}$ have a distribution from the exponential distribution family with the canonical parameter θ, dispersion parameter σ^2 and known weights $\omega_{i,j}$. Observe, that whereas σ^2 is constant over all observations $X_{i,j}$, the weights $\omega_{i,j}$ may vary. The canonical parameter θ can take values from a set H, $H \subseteq \mathbb{R}$. Θ_i is a risk characteristic of accident year i with a prior distribution (3.17). μ and τ^2 does not depend on i, so Θ_i, $i = 0, ..., I$ are independent and identically distributed—i.i.d.

Derivation of estimator for ultimate claim

Our goal is now to set an optimal bayesian estimator for ultimate claim. We will proceed analogous to the previous section. Moreover, we would like to verify that so defined model is the example of conjugate classes of distributions and the posterior distribution of Θ_i is of type (3.17) as prior distribution.

We assume again that accident years are independent. So it suffices to concentrate on a single fixed accident year i and in the calculation of the posterior distribution of Θ_i we condition with the observation $X_{i,j}$ from this accident year.

Again $\pi_{\Theta_i|X_{i,0},...,X_{i,j}}$ regards the posterior density of Θ_i and we denote the standardized observation $\frac{X_{i,j}}{\gamma_j \mu_i}$ with $Y_{i,j}$, $Y_{i,j} = \frac{X_{i,j}}{\gamma_j \mu_i}$.

Given $\Theta_i = \theta$, the vector $(Y_{i,0}, ..., Y_{i,j})'$ has the density

$$f_{Y_{i,0},...,Y_{i,j}|\Theta_i}(y_{i,0}, ..., y_{i,j} \mid \theta) = \prod_{k=0}^{j} \exp\left[\frac{y_{i,k}\theta - b(\theta)}{\frac{\sigma^2}{\omega_{i,k}}}\right] a\left(y_{i,k}, \frac{\sigma^2}{\omega_{i,k}}\right).$$

Now with the use of the Bayes theorem we can calculate the posterior density $\pi_{\Theta_i|Y_{i,0},...,Y_{i,j}}$ and $\pi_{\Theta_i|X_{i,0},...,X_{i,j}}$ respectively.

$$
\begin{aligned}
&\pi_{\Theta_i|Y_{i,0},\dots,Y_{i,j}}(\theta \mid y_{i,0},\dots,y_{i,j}) \\
&= \frac{f_{Y_{i,0},\dots,Y_{i,j}|\Theta_i}(y_{i,0},\dots,y_{i,j} \mid \theta)u_{1,\tau^2}(\theta)}{\int_H f_{Y_{i,0},\dots,Y_{i,j}|\Theta_i}(y_{i,0},\dots,y_{i,j} \mid \theta)u_{1,\tau^2}(\theta)\mathrm{d}\theta} \\
&= \frac{\prod_{k=0}^{j} \exp\left[\frac{y_{i,k}\theta - b(\theta)}{\frac{\sigma^2}{\omega_{i,k}}}\right] a(y_{i,k}, \frac{\sigma^2}{\omega_{i,k}})u_{1,\tau^2}(\theta)}{\int_H \prod_{k=0}^{j} \exp\left[\frac{y_{i,k}\theta - b(\theta)}{\frac{\sigma^2}{\omega_{i,k}}}\right] a(y_{i,k}, \frac{\sigma^2}{\omega_{i,k}})u_{1,\tau^2}(\theta)\mathrm{d}\theta} \\
&= \frac{d(1,\tau^2)\exp\left[\frac{\mu\theta - b(\theta)}{\tau^2}\right]\prod_{k=0}^{j} \exp\left[\frac{y_{i,k}\theta - b(\theta)}{\frac{\sigma^2}{\omega_{i,k}}}\right] a(y_{i,k}, \frac{\sigma^2}{\omega_{i,k}})}{d(1,\tau^2)\int_H \exp\left[\frac{\mu\theta - b(\theta)}{\tau^2}\right]\prod_{k=0}^{j} \exp\left[\frac{y_{i,k}\theta - b(\theta)}{\frac{\sigma^2}{\omega_{i,k}}}\right] a(y_{i,k}, \frac{\sigma^2}{\omega_{i,k}})\mathrm{d}\theta} \\
&= \frac{\prod_{k=0}^{j} a(y_{i,k}, \frac{\sigma^2}{\omega_{i,k}})\exp\left[\frac{\theta - b(\theta)}{\tau^2}\right]\exp\left[\sum_{k=0}^{j} \frac{y_{i,k}\theta - b(\theta)}{\frac{\sigma^2}{\omega_{i,k}}}\right]}{\prod_{k=0}^{j} a(y_{i,k}, \frac{\sigma^2}{\omega_{i,k}})\int_H \exp\left[\frac{\theta - b(\theta)}{\tau^2}\right]\exp\left[\sum_{k=0}^{j} \frac{y_{i,k}\theta - b(\theta)}{\frac{\sigma^2}{\omega_{i,k}}}\right]\mathrm{d}\theta} \\
&= \frac{\exp\left\{\theta\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}y_{i,k}}{\sigma^2}\right] - b(\theta)\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}\right]\right\}}{\int_H \exp\left\{\theta\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}y_{i,k}}{\sigma^2}\right] - b(\theta)\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}\right]\right\}\mathrm{d}\theta} \\
&= \frac{\exp\left\{\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}\right]\left[\frac{\theta\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}y_{i,k}}{\sigma^2}\right]}{\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}} - b(\theta)\right]\right\}}{\int_H \exp\left\{\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}\right]\left[\frac{\theta\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}y_{i,k}}{\sigma^2}\right]}{\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}} - b(\theta)\right]\right\}\mathrm{d}\theta}. \qquad (3.18)
\end{aligned}
$$

From (3.18) it follows that the posterior density of Θ_i given $X_{i,0} = x_{i,0}$, ..., $X_{i,j} = x_{i,j}$ can be expressed analytically in the form

$$
\begin{aligned}
&\pi_{\Theta_i|X_{i,0},\dots,X_{i,j}}(\theta \mid x_{i,0},\dots,x_{i,j}) \\
&= \frac{\exp\left\{\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}\right]\left[\frac{\theta\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}\frac{x_{i,k}}{\gamma_k\mu_i}}{\sigma^2}\right]}{\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}} - b(\theta)\right]\right\}}{\int_H \exp\left\{\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}\right]\left[\frac{\theta\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}\frac{x_{i,k}}{\gamma_k\mu_i}}{\sigma^2}\right]}{\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}} - b(\theta)\right]\right\}\mathrm{d}\theta}. \qquad (3.19)
\end{aligned}
$$

We denote

$$
\tau^2_{post,i,j} = \left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}\right]^{-1},
$$

$$
\mu_{post,i,j} = \frac{\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}\frac{x_{i,k}}{\gamma_k\mu_i}}{\sigma^2}}{\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}}.
$$

Observe, that the integral in the denominator of (3.19) is a constant with the respect to θ and it can be expressed as a function of $\tau^2_{post,i,j}$ and $\mu_{post,i,j}$ and so we can define

$$C(\tau^2_{post,i,j}, \mu_{post,i,j}) = \frac{1}{\int_H \exp\left\{\left[\underbrace{\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}}_{(\tau^2_{post,i,j})^{-1}}\right]\left[\theta \underbrace{\frac{\left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k} x_{i,k}}{\sigma^2}\right]}{\frac{1}{\tau^2} + \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sigma^2}}}_{\mu_{post,i,j}} - b(\theta)\right]\right\} \mathrm{d}\theta}$$

This denotation allows us to simplify the complicated expression (3.19) to the following form

$$\pi_{\Theta_i|X_{i,0},\ldots,X_{i,j}}(\theta \mid x_{i,0}, \ldots, x_{i,j}) = C(\tau^2_{post,i,j}, \mu_{post,i,j}) \exp\left[\frac{\theta\mu_{post,i,j} - b(\theta)}{\tau^2_{post,i,j}}\right]. \quad (3.20)$$

Now, it is obvious that (3.20) is just the density $u_{\mu_{post,i,j},\tau^2_{post,i,j}}$ and so the conditional distribution of Θ_i given $X_{i,0} = x_{i,0}$, ...,$X_{i,j} = x_{i,j}$ is of the same type (3.17) as its prior density. This means that the class of distributions defined by (3.17) is a conjugate class of distribution to the exponential distribution family (3.16) what we wanted to show.

The new parameters $\tau^2_{post,i,j}$ and $\mu_{post,i,j}$ are updated with the information about the observations $X_{i,0}$, ...,$X_{i,j}$. We can simplify them to the following expressions

$$\tau^2_{post,i,j} = \sigma^2 \left(\frac{\sigma^2}{\tau^2} + \sum_{k=0}^{j} \omega_{i,k}\right)^{-1},$$

$$\mu_{post,i,j} = \frac{\tau^2_{post,i,j}}{\sigma^2}\left(\frac{\sigma^2}{\tau^2} + \sum_{k=0}^{j} \omega_{i,k}\frac{x_{i,k}}{\gamma_k \mu_i}\right) = \frac{\tau^2_{post,i,j}}{\sigma^2}\left(\frac{\sigma^2}{\tau^2} + \sum_{k=0}^{j} \omega_{i,k}\bar{y}_i^{(j)}\right),$$

where

$$\bar{y}_i^{(j)} = \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sum_{l=0}^{j} \omega_{i,l}} \frac{x_{i,k}}{\gamma_k \mu_i}.$$

Specially, the posterior distribution of Θ_i given $D_I = d_I$ has the density (3.17) with the parameters $\tau^2_{post,i,I-i}$ and $\mu_{post,i,I-i}$.

Now, we formulate a theorem which will help us to derive the estimator of the ultimate claims.

Theorem 3.2. *Under the model assumptions 3.3.2 we have for i, $j \geq 0$*

1. *Define $\mu(\Theta_i) = \mathrm{E}\left[\frac{X_{i,j}}{\gamma_j \mu_i} \mid \Theta_i\right]$ then it holds*

$$\mu(\Theta_i) = \mathrm{E}\left[\frac{X_{i,k}}{\gamma_k \mu_i} \mid \Theta_i\right] = b'(\Theta_i) \quad (3.21)$$

$$\mathrm{var}\left(\frac{X_{i,k}}{\gamma_k \mu_i} \mid \Theta_i\right) = \frac{\sigma^2 b''(\Theta_i)}{\omega_{i,j}}$$

2. *If* $\exp\left[\frac{\theta\mu_i - b(\theta)}{\tau^2}\right]$ *disappears on the boundary of* Θ_i *for all* μ_i, τ^2 *then the following relations hold*

$$E[X_{i,j}] = \gamma_j \mu_i \, E[\mu(\Theta_i)] = \gamma_j \mu_i,$$

$$E[\mu(\Theta_i) \mid X_{i,0} = x_{i,0}, ..., X_{i,j} = x_{i,j}] = \alpha_{i,j} \bar{y}_i^{(j)} + (1 - \alpha_{i,j})1,$$

$$E[\mu(\Theta_i) \mid X_{i,0}, ..., X_{i,j}] = \alpha_{i,j} \bar{Y}_i^{(j)} + (1 - \alpha_{i,j})1,$$

where

$$\alpha_{i,j} = \frac{\sum_{k=0}^{j} \omega_{i,k}}{\sum_{k=0}^{j} \omega_{i,k} + \frac{\sigma^2}{\tau^2}} \quad \textit{and} \quad \bar{Y}_i^{(j)} = \sum_{k=0}^{j} \frac{\omega_{i,k}}{\sum_{l=0}^{j} \omega_{i,l}} \frac{X_{i,k}}{\gamma_k \mu_i}.$$

This theorem comes from [Wüthrich, Merz–2008] and we will not present its proof here. It can be found in [Bühlmann, Gisler–2005].

The first part of it says about conditional moments of the standardized observations $\frac{X_{i,j}}{\gamma_j \mu_i}$ and such the relations for the expected value and the variance hold for every random quantity which has a distribution from the exponential dispersion family.

The second part of the theorem deals with the posterior mean of $\mu(\Theta_i)$, given the observations $X_{i,0}$, ...,$X_{i,j}$. We can notice that this posterior mean can be expressed as a credibility weighted average between the prior mean $E[\mu(\Theta_i)] = 1$ and $\bar{Y}_i^{(j)}$—the weighted average of the standardized observations $\frac{X_{i,k}}{\gamma_k \mu_i}$. The credibility weight $\alpha_{i,j}$ depends on the dispersion parameter σ^2 and on τ^2. The larger σ^2 the smaller is $\alpha_{i,j}$—the weight given to the observations. Obversely, the larger τ^2 the larger $\alpha_{i,j}$.

According to Theorem 3.2, the posterior mean of $\mu(\Theta_i)$ given D_I, denoted $\widetilde{\mu(\Theta_i)}$, is given by

$$\widetilde{\mu(\Theta_i)} = E[\mu(\Theta_i) \mid D_I] = E[\mu(\Theta_i) \mid X_{i,0}, ..., X_{i,I-i}] = \alpha_{i,I-i} \bar{Y}_i^{(I-i)} + (1 - \alpha_{i,I-i}) \tag{3.22}$$

and it is rather the optimal bayesian estimator of $\mu(\Theta_i)$ with the respect to the quadratic loss function.

Now we have all necessary information to derive optimal bayesian estimator of ultimate claims—posterior expected value $E[C_{i,J} \mid D_I] = E[C_{i,J} \mid X_{i,0}, ..., X_{i,I-i}]$. With its use, we obtain the resulting expression for $E[C_{i,J} \mid D_I]$ in the form

$$
\begin{aligned}
\mathrm{E}[C_{i,J} \mid D_I] &= \mathrm{E}[\mathrm{E}[C_{i,J} \mid \Theta_i, D_I] \mid D_I] \\
&= \mathrm{E}\left[\mathrm{E}\left[\sum_{j=0}^{J} X_{i,j} \mid \Theta_i, D_I\right] \mid D_I\right] \\
&= C_{i,I-i} + \mathrm{E}\left[\mathrm{E}\left[\sum_{j=I-i+1}^{J} X_{i,j} \mid \Theta_i\right] \mid D_I\right] \\
&= C_{i,I-i} + \mathrm{E}\left[\sum_{j=I-i+1}^{J} \gamma_j \mu_i \, \mathrm{E}\left[\frac{X_{i,j}}{\gamma_j \mu_i} \mid \Theta_i\right] \mid D_I\right] \\
&= C_{i,I-i} + \mathrm{E}\left[\sum_{j=I-i+1}^{J} \gamma_j \mu_i \mu(\Theta_i) \mid D_I\right] \\
&= C_{i,I-i} + \sum_{j=I-i+1}^{J} \gamma_j \mu_i \, \mathrm{E}\left[\mu(\Theta_i) \mid D_I\right] \\
&= C_{i,I-i} + (1 - \beta_{I-i}) \mu_i \widetilde{\mu(\Theta_i)}, \quad \text{for } 1 \leq i \leq I.
\end{aligned} \tag{3.23}
$$

(3.23) is the optimal bayesian estimator for ultimate claim in the EDF model and we denote it $\widehat{C_{i,J}^{EDF}}$. Replacing (3.22) into (3.23) we have it in the following form

$$
\widehat{C_{i,J}^{EDF}} = C_{i,I-i} + (1 - \beta_{I-i}) \mu_i \left[\alpha_{i,I-i} \bar{Y}_i^{(I-i)} + (1 - \alpha_{i,I-i})\right]. \tag{3.24}
$$

Using the analogous steps as in (3.23) we are also able to express the optimal estimator for the increments $X_{i,I-i+k}$, $k = 1, ..., i$, denoted as $\widehat{X_{i,I-i+k}^{EDF}}$

$$
\widehat{X_{i,I-i+k}^{EDF}} = \mathrm{E}[X_{i,I-i+k} \mid D_I] = \gamma_{I-i+k} \mu_i \widetilde{\mu(\Theta_i)}.
$$

3.3.3 Examples of exponential dispersion family models

As we have already mentioned, exponential dispersion family contains a lot of families of distributions. We have introduced EDF model generally. In this section, we show the classical examples of EDF models.

Instead of generally defined distribution from EDF (3.16) we assume concrete distribution from this family for the incremental claims and we derive its associated conjugates.

Poisson-Gamma model as EDF model

We dealt with the Poisson-Gamma model in Subsection 3.3.1, where we introduced it as an individual model. In fact, it is only special case of EDF model and now we show this. We reformulate Po-Ga assumptions 3.3.1 so that they coincide with the assumptions 3.3.2 of EDF model and verify if we get the same estimator of the ultimate claim.

Before we start we have to remark that the Poisson-Gamma model is the special case of EDF only when Θ_i, $i = 0, ..., I$ are a priori i.i.d., so gamma distributed with the parameters a and b which does not depend on the accident years. Given $\Theta_i = \theta$, increments $X_{i,j}$, $j = 0, ..., J$ are Poisson distributed with the parameter $\theta\gamma_j$. This means that

$$P\left[\frac{X_{i,j}}{\gamma_j} = k \mid \Theta_i = \theta\right] = P\left[X_{i,j} = k\gamma_j \mid \Theta_i = \theta\right]$$
$$= \exp(-\theta\gamma_j)\frac{(\theta\gamma_j)^{k\gamma_j}}{(k\gamma_j)!}, \quad k = \frac{l}{\gamma_j}, \quad l = 0, 1, ...$$

and the conditional density of $\frac{X_{i,j}}{\gamma_j}$, given $\Theta_i = \theta$, denoted by $f_{\frac{X_{i,j}}{\gamma_j}|\Theta_i}(x \mid \theta)$ can be expressed as

$$f_{\frac{X_{i,j}}{\gamma_j}|\Theta_i}(x \mid \theta) = \exp(-\theta\gamma_j)\frac{(\theta\gamma_j)^{x\gamma_j}}{(x\gamma_j)!}, \quad x = \frac{l}{\gamma_j}, \quad l = 0, 1, ...$$

This expression can be modified to the form which corresponds with (3.16)

$$f_{\frac{X_{i,j}}{\gamma_j}|\Theta_i}(x \mid \theta) = \exp\left[-\theta\gamma_j + x\gamma_j \log(\theta\gamma_j) - \log((x\gamma_j)!)\right]$$
$$= \exp\left[\frac{x\log(\theta) - \theta}{\gamma_j}\right] \exp\left[x\gamma_j \log(\gamma_j) - \log((x\gamma_j)!)\right].$$

We denote $\tilde{\theta} = \log(\theta)$, $b(\tilde{\theta}) = \exp(\tilde{\theta})$, $\sigma^2 = 1$, $\omega_{i,j} = \gamma_j$ and $a(x, \frac{1}{\gamma_j}) = \exp\left[x\gamma_j \log(\gamma_j) - \log((x\gamma_j)!)\right]$. Now it is obvious that, given $\Theta_i = \theta$, $\frac{X_{i,j}}{\gamma_j}$ has the distribution from exponential dispersion family with the canonical parameter $\tilde{\theta}$, the dispersion parameter 1 and with the known weights γ_j.

Now we can work with Poisson-Gamma model as with the EDF model. First, we set its conjugate prior distribution u_{μ,τ^2} with $\mu \equiv 1$

$$u_{\mu,\tau^2}(\theta) = d(\mu, \tau^2) \exp\left[\frac{\mu\log(\theta) - \theta}{\tau^2}\right] \frac{1}{\theta}$$
$$= d(\mu, \tau^2)\theta^{\frac{\mu}{\tau^2}} \exp\left(\frac{-\theta}{\tau^2}\right) \frac{1}{\theta}$$
$$= d(\mu, \tau^2)\theta^{\frac{\mu}{\tau^2}-1} \exp\left(\frac{-\theta}{\tau^2}\right), \quad \theta > 0.$$

Obviously, it is proportional to the density of gamma distribution with the parameters $a = \frac{\mu}{\tau^2} = \frac{1}{\tau^2}$ and $b = \frac{1}{\tau^2}$.

From already obtained results, the posterior density of Θ_i, given $X_{i,0} = x_{i,0}, \ldots, X_{i,j} = x_{i,j}$, is again the gamma density $u_{\mu_{post,i,j},\tau^2_{post,i,j}}$ where

$$\tau^2_{post,i,j} = \left(\frac{1}{\tau^2} + \sum_{k=0}^{j} \gamma_k\right)^{-1} = \frac{1}{\frac{1}{\tau^2} + \beta_j}$$

and

$$\mu_{post,i,j} = \tau^2_{post,i,j} \left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \gamma_k \left(\sum_{l=0}^{j} \frac{\gamma_l}{\sum_{h=0}^{j} \gamma_h} \frac{x_{i,l}}{\gamma_l} \right) \right]$$
$$= \tau^2_{post,i,j} \left[\frac{1}{\tau^2} + \sum_{k=0}^{j} \gamma_k \sum_{l=0}^{j} \frac{x_{i,l}}{\beta_j} \right]$$
$$= \tau^2_{post,i,j} \left[\frac{1}{\tau^2} + \sum_{k=0}^{j} x_{i,k} \right].$$

This corresponds with the Poisson-Gamma model again, because

$$b^{post}_{i,j} = \frac{1}{\tau^2_{post,i,j}} = \frac{1}{\tau^2} + \beta_j = b + \beta_j$$

and

$$a^{post}_{i,j} = \frac{\mu_{post,i,j}}{\tau^2_{post,i,j}} = \frac{1}{\tau^2} + \sum_{k=0}^{j} x_{i,k} = a + \sum_{k=0}^{j} x_{i,k}.$$

This implies, that the estimator for the ultimate claims is the same for the both cases and it is given by the relation

$$\widehat{C^{EDF}_{i,J}} = \widehat{C^{PoGa}_{i,J}} = C_{i,I-i} + (1 - \beta_{I-i}) \left[\frac{\beta_{I-i}}{\beta_{I-i} + \frac{1}{\tau^2}} \frac{C_{i,j}}{\beta_{I-i}} + \frac{\frac{1}{\tau^2}}{\beta_{I-i} + \frac{1}{\tau^2}} \right]$$
$$= C_{i,I-i} + (1 - \beta_{I-i}) \left[\frac{\beta_{I-i}}{\beta_{I-i} + b} \frac{C_{i,j}}{\beta_{I-i}} + \frac{b}{\beta_{I-i} + b} \right]$$

Binomial-Beta model

Another distribution which belongs to EDF is binomial distribution. In Binomial-Beta model we assume given $\Theta_i = \theta$, $X_{i,j}$, $j = 0, ..., J$, has the binomial distribution with the parameter $\gamma_j \mu_i \in \mathbb{N}$ and $\theta \in (0, 1)$. This means that

$$P[X_{i,j} = k \mid \Theta_i = \theta] = \binom{\gamma_j \mu_i}{k} \theta^k (1 - \theta)^{\gamma_j \mu_i - k}, \quad k = 0, ..., \gamma_j \mu_i$$

and

$$P\left[\frac{X_{i,j}}{\gamma_j \mu_i} = k \mid \Theta_i = \theta \right] = P[X_{i,j} = k\gamma_j \mu_i \mid \Theta_i = \theta] = \binom{\gamma_j \mu_i}{k\gamma_j \mu_i} \theta^{k\gamma_j \mu_i} (1 - \theta)^{\gamma_j \mu_i - k\gamma_j \mu_i}$$

for $k = \frac{l}{\gamma_j \mu_i}$, $l = 0, ..., \gamma_j \mu_i$.

The conditional density of $\frac{X_{i,j}}{\gamma_j \mu_i}$ given $\Theta_i = \theta$, $f_{\frac{X_{i,j}}{\gamma_j \mu_i} | \Theta_i}(x \mid \theta)$ can be written in the following form for $x = \frac{l}{\gamma_j \mu_i}$, $l = 0, ..., \gamma_j \mu_i$.

$$\begin{aligned}
& f_{\frac{X_{i,j}}{\gamma_j \mu_i} | \Theta_i}(x \mid \theta) \\
& = \binom{\gamma_j \mu_i}{x \gamma_j \mu_i} \theta^{x \gamma_j \mu_i} (1 - \theta)^{\gamma_j \mu_i - x \gamma_j \mu_i} \\
& = \exp \left\{ \log \left[\binom{\gamma_j \mu_i}{x \gamma_j \mu_i} \right] + x \gamma_j \mu_i \log(\theta) + \gamma_j \mu_i \log(1 - \theta) - x \gamma_j \mu_i \log(1 - \theta) \right\} \\
& = \exp \left[\frac{x \log(\frac{\theta}{1-\theta}) - \log(\frac{1}{1-\theta})}{\frac{1}{\gamma_j \mu_i}} \right] \binom{\gamma_j \mu_i}{x \gamma_j \mu_i} \\
& = \exp \left[\frac{x \tilde{\theta} - b(\tilde{\theta})}{\frac{1}{\omega_{i,j}}} \right] a \left(x, \frac{1}{\omega_{i,j}} \right),
\end{aligned}$$

where $\tilde{\theta} = \log(\frac{\theta}{1-\theta})$, $b(\tilde{\theta}) = \log[1 + \exp(\tilde{\theta})]$, $\omega_{i,j} = \gamma_j \mu_i$, $\sigma^2 = 1$ and $a\left(x, \frac{1}{\omega_{i,j}}\right) = \binom{\omega_{i,j}}{x \omega_{i,j}}$.

This expression of $f_{\frac{X_{i,j}}{\gamma_j \mu_i} | \Theta_i}$ corresponds with (3.16) and the expected value $\mu(\Theta_i) = \mathrm{E}[\frac{X_{i,j}}{\gamma_j \mu_i} \mid \Theta_i]$ is equal to Θ_i.

Now we define the prior distribution of Θ_i, u_{μ, τ^2}. According to the assumptions 3.3.2 we have the following prior distribution

$$\begin{aligned}
u_{\mu, \tau^2}(\theta) &= d(\mu, \tau^2) \exp\{ \frac{\mu \log(\frac{\theta}{1-\theta}) - \log(\frac{1}{1-\theta})}{\tau^2} \} \frac{1 - \theta}{\theta} \frac{1}{(1 - \theta)^2} \\
&= d(\mu, \tau^2) \theta^{\frac{\mu}{\tau^2} - 1} (1 - \theta)^{\frac{1-\mu}{\tau^2} - 1}, \quad \theta \in (0, 1).
\end{aligned}$$

It holds that u_{μ, τ^2} is proportional to $\theta^{\frac{\mu}{\tau^2} - 1} (1 - \theta)^{\frac{1-\mu}{\tau^2} - 1}$, which implies that it is a density of beta distribution with the parameters $\frac{\mu}{\tau^2} > 0$ and $\frac{1-\mu}{\tau^2} > 0$.

The prior expected value is equal to μ

$$\mathrm{E}[\Theta_i] = \frac{\frac{\mu}{\tau^2}}{\frac{\mu}{\tau^2} + \frac{1-\mu}{\tau^2}} = \frac{\frac{\mu}{\tau^2}}{\frac{1}{\tau^2}} = \mu.$$

In this case, μ cannot be identically 1 as in the assumptions of the EDF model because then the inequality $\frac{1-\mu}{\tau^2} > 0$ would not hold. Instead of $\mu \equiv 1$ we consider $\mu \in (0, 1)$.

Now we have the assumptions which correspond with EDF model assumptions and we can use the already derived results to set the posterior distribution and the estimator of the ultimate claim.

Let us take some notes about the prior weights. In this model we have $\omega_{i,j} = \gamma_j \mu_i$ which immediately implies that $\sum_{k=0}^{j} \omega_{i,j} = \beta_j \mu_i$. This simplifies the expression of $\bar{Y}_i^j$ and $\alpha_{i,j}$ to the following forms

$$\alpha_{i,j} = \frac{\beta_j}{\beta_j + \frac{\sigma^2}{\tau^2 \mu_i}}, \quad \bar{Y}_i^j = \frac{C_{i,j}}{\beta_j \mu_i}.$$

Consequently, the posterior parameters $\tau^2_{post,i,j}$ and $\mu_{post,i,j}$ as well as the formulas for $\widetilde{\mu(\Theta_i)}$ and for the estimator of ultimate claims can be written in more sophisticated forms.

Normal-Normal model

In Normal-Normal model we assume that given $\Theta_i = \theta$, $X_{i,j}$, $j = 0$, ..., J, are normal distributed with the parameters $\theta \in \mathbb{R}$ and $\sigma^2\gamma_j\mu_i > 0$, $\mathcal{N}(\theta, \sigma^2\gamma_j\mu_i)$. So, $\frac{X_{i,j}}{\gamma_j\mu_i}$ are $\mathcal{N}\left(\theta, \frac{\sigma^2}{\gamma_j\mu_i}\right)$ and the conditional density $f_{\frac{X_{i,j}}{\gamma_j\mu_i}|\Theta_i}(x \mid \theta)$ is expressed by the following relation

$$\begin{aligned} f_{\frac{X_{i,j}}{\gamma_j\mu_i}|\Theta_i}(x \mid \theta) &= \frac{1}{\sqrt{2\pi\frac{\sigma^2}{\gamma_j\mu_i}}} \exp\left[\frac{-x^2 + 2x\theta - \theta^2}{2\frac{\sigma^2}{\gamma_j\mu_i}}\right] \\ &= \exp\left(\frac{x\theta - \frac{\theta^2}{2}}{\frac{\sigma^2}{\gamma_j\mu_i}}\right) \exp\left(\frac{-x^2}{2\frac{\sigma^2}{\gamma_j\mu_i}}\right) \frac{1}{\sqrt{2\pi\frac{\sigma^2}{\gamma_j\mu_i}}}, \quad x \in \mathbb{R}. \end{aligned}$$

This expression is in the form (3.16) with $b(\theta) = \frac{\theta^2}{2}$, $\omega_{i,j} = \gamma_j\mu_i$ and $a\left(x, \frac{1}{\omega_{i,j}}\right) = \exp\left[\frac{-x^2}{2\frac{\sigma^2}{\gamma_j\mu_i}}\right] \frac{1}{\sqrt{2\pi\frac{\sigma^2}{\gamma_j\mu_i}}}$ and we verify that normal distribution belongs to EDF. Here again $\mu(\Theta_i) = \mathrm{E}[\frac{X_{i,j}}{\gamma_j\mu_i} \mid \Theta_i] = \Theta_i$.

The prior distribution of Θ_i, u_{μ,τ^2} with $\mu \equiv 1$, which is conjugate to $\mathcal{N}\left(\theta, \frac{\sigma^2}{\gamma_j}\right)$ is expressed in the following form

$$\begin{aligned} u_{1,\tau^2}(\theta) &= d(1, \tau^2) \exp\left(\frac{\theta - \frac{\theta^2}{2}}{\tau^2}\right) \\ &= d(1, \tau^2) \exp\left(\frac{1}{2\tau^2}\right) \exp\left(\frac{-1 + 2\theta - \theta^2}{2\tau^2}\right) \\ &= \tilde{d}(1, \tau^2) \exp\left[\frac{-(\theta - 1)^2}{2\tau^2}\right], \quad \theta \in \mathbb{R}. \end{aligned}$$

This is proportional to $\exp\left[\frac{-(\theta-1)^2}{2\tau^2}\right]$ so u_{1,τ^2} is the density of normal distribution $\mathcal{N}(1, \tau^2)$. The prior expected value $\mathrm{E}[\Theta_i]$ is equal to 1 and the prior variance $\mathrm{var}[\Theta_i]$ is τ^2.

From already showed in EDF model, the posterior distribution (3.20) of Θ_i is normal distribution with the posterior parameter $\mu_{post,i,j}$, $\tau^2_{post,i,j}$ and it is easy to set the estimator of the ultimate claim $C_{i,J}$. The weights are as well as in the Binomial-Beta case defined as the product of γ_j and μ_i which implies less complicated form for estimators.

Geometric-Beta model

The Geometric-Beta model is also used in the claims reserving as a special case of EDF model. Given $\Theta_i = \theta$, $\theta \in (0, 1)$ it assumes the Geometric distribution with the parameter θ for incremental claims $X_{i,j}$, $j = 0$, ..., J. This means that it holds

$$P[X_{i,j} = k \mid \Theta_i = \theta] = (1 - \theta)^k\theta, \quad \text{for } k = 0, 1, ...$$

The conditional density $f_{X_{i,j}|\Theta_i}(x \mid \theta)$ can be modified to the expression

$$f_{X_{i,j}|\Theta_i}(x \mid \theta) = (1-\theta)^x \theta = \exp\left[x \log(1-\theta) + \log(\theta)\right] = \exp\left[x\tilde{\theta} - b(\tilde{\theta})\right],$$

for $x = 0, 1, ...$

The terms $\tilde{\theta}$ and $b(\tilde{\theta})$ are defined as $\tilde{\theta} = \log(1-\theta)$ and $b(\tilde{\theta}) = -\log(1-\exp^{\tilde{\theta}})$.

Again, we can see that this density belongs to EDF. Observe, that in this model we have all γ_j, μ_i and prior weights $\omega_{i,j}$ all constant, equal to 1. The dispersion parameter σ^2 has the value 1, too.

The prior distribution of Θ_i, u_{μ,τ^2} with $\mu \equiv 1$ is according to 3.3.2 in the following form

$$\begin{aligned} u_{1,\tau^2}(\theta) &= d(1,\tau^2) \exp\left[\frac{\log(1-\theta) + \log(\theta)}{\tau^2}\right] \frac{-1}{1-\theta} \\ &= d(1,\tau^2)(1-\theta)^{\frac{1}{\tau^2}-1} \theta^{\frac{1}{\tau^2}}, \quad \theta \in (0,1). \end{aligned}$$

This is proportional to $\theta^{\frac{1}{\tau^2}}(1-\theta)^{\frac{1}{\tau^2}-1}$, so it is a density of beta distribution with the parameters $\frac{1}{\tau^2} + 1 > 0$ and $\frac{1}{\tau^2} > 0$.

Analogous as in the previous examples we have the assumptions of EDF model. This implies the beta posterior distribution for Θ_i and the estimator of the ultimate claim can be expressed easily.

4. Practical application of models

In the previous chapters we summarized the theoretical background of the bayesian claims reserving. Now we should examine the methods on a concrete example, compare, review and discuss achieved results. This is the scope of this chapter. All computations and analyses are done in the environment of R software [R Core Team–2013]. The complete source code is attached in Appendix A. In each example of concrete method we assume its corresponding assumptions for data.

4.1 Data set

For the illustration of claims reserving methods we use a data set—a cumulative development triangle displayed in Table 4.1. This data set consists of observed cumulative paid losses of accident years 1988–1997. The line of business for this data is Private Passenger Auto Liability Insurance. The data are taken from [Meyers, Shi–2011] where the more detailed description can be found.

We have to remark that the cumulative payments in the triangle contain allocated loss adjustment expenses (ALAE). ALAE are allocated to a single claim and they are typically expenses for external expertise, external lawyer, etc. In the practice it is common that data used for the claim reserves estimation contains these expenses and reserving methods are applied on such a data.

Accident	Development year j									
year i	0	1	2	3	4	5	6	7	8	9
0	4339	9617	11584	12001	12640	12966	13058	13080	13182	13183
1	4445	9090	11125	11815	12027	12078	12626	12626	12627	
2	6123	10564	12290	13579	14063	14412	14488	14497		
3	5464	9537	11481	12335	12699	12721	12725			
4	5113	8527	10344	11381	11742	12036				
5	6540	11810	14533	14998	15890					
6	5781	9519	10929	11356						
7	4356	6779	8199							
8	5250	8967								
9	5743									

Table 4.1: Observed cumulative payments

According to our notation from Chapter 1.1 we have $I = J = 9$, accident years $i = 0, ..., 9$, the development years $j = 0, ..., 9$ and the entries of the triangle—cumulative paid losses—denoted by $C_{i,j}$, $i + j \leq I$.

The corresponding incremental triangle with the incremental payments $X_{i,j}$, $i + j \leq I$ is shown in Table 4.2.

To get a graphical overview we can plot our data set—the cumulative triangle 4.1. Figure 4.1 shows the development of cumulative claims, each line in the graph corresponds to an accident year. For the better visibility we present Figure 4.2 where we can see the same development of claims but with individual

Accident year i	Development year j									
	0	1	2	3	4	5	6	7	8	9
0	4339	5278	1967	417	639	326	92	22	102	1
1	4445	4645	2035	690	212	51	548	0	1	
2	6123	4441	1726	1289	484	349	76	9		
3	5464	4073	1944	854	364	22	4			
4	5113	3414	1817	1037	361	294				
5	6540	5270	2723	465	892					
6	5781	3738	1410	427						
7	4356	2423	1420							
8	5250	3717								
9	5743									

Table 4.2: Observed incremental payments

graph for each accident year. From the both pictures it is obvious that the development of payments is less noticeable with the increase of development years.

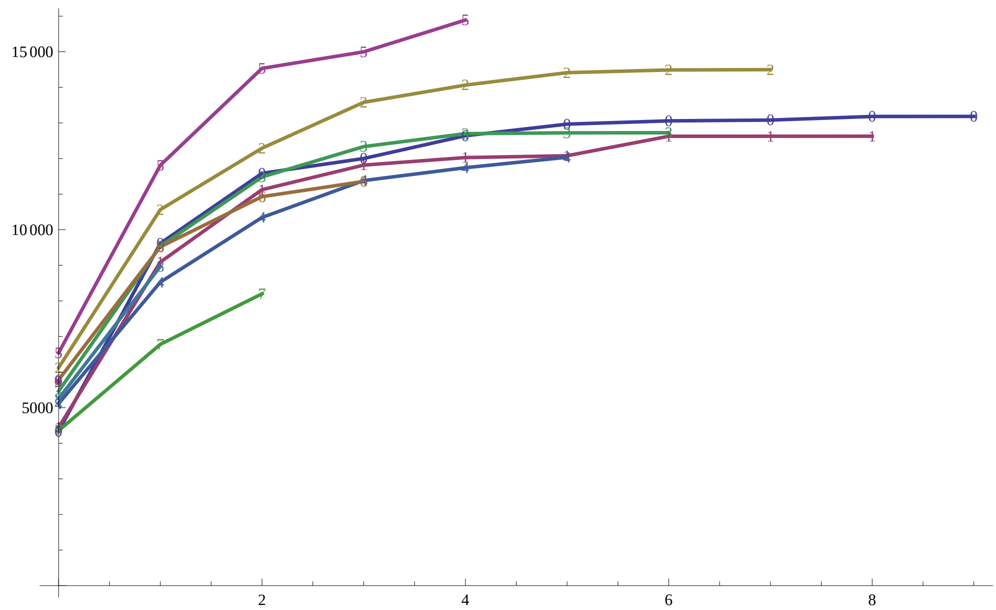

Figure 4.1: Development of cumulative payments

In addition to the development triangle we also need the information about premium received corresponding to the each accident year. The premium is denoted by π_i, $i = 0, ..., 9$ and it is listed in Table 4.3. We assume direct premium before deducting any premiums sent to reinsurers.

4.2 Chain-Ladder method

The Chain-Ladder method is based on the development factors. From the graphic representation of our data set it is not noticeable that the proportionality of the columns should not hold and so the CL model may be proper.

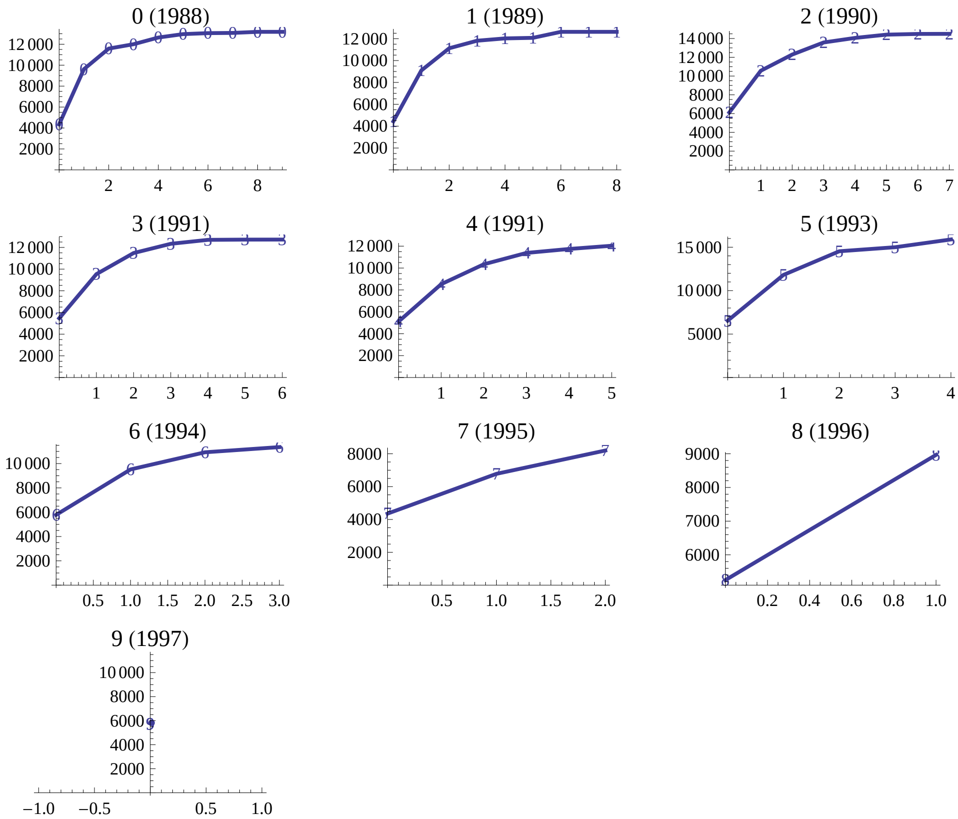

Figure 4.2: Development of cumulative payments—individual graphs

accident year	earned premium
0	18793
1	18948
2	20527
3	21278
4	20779
5	23212
6	22219
7	18314
8	17043
9	19217

Table 4.3: Table of the earned premium

Applying the formula (1.2) for development factors we obtain estimators $\widehat{f}_j$, $j = 0, ..., J-1$ listed in Table 4.4.

Now we can apply development factors to the latest known cumulative payments $C_{i,I-i}$—the diagonal values—from every accident year. We get the prediction of payments for next development periods $C_{i,j}$, $i + j \leq I$, $i \leq I$, $j \leq J$. Values $\widehat{C_{i,J}^{CL}}$, $i = 1, ..., I$, $\widehat{C_{i,9}^{CL}}$, $i = 1, ..., I$ respectively, represent the ultimate payments for the accident years. CL reserves are given as a difference of ultimate payments and the diagonal values $C_{i,I-i}$. In Table 4.5 we can see the fulfilled triangle with the observations and predicted values as well as the value of Chain-

j	0	1	2	3	4	5	6	7	8
$\widehat{f}_j$	1.780	1.199	1.063	1.039	1.016	1.014	1.001	1.004	1

Table 4.4: The values of $\widehat{f}_j$

Ladder reserves. The value of total Chain-Ladder reserve—the sum of reserves for all accident years is 14556.11.

i	0	1	2	3	4	5	6	7	8	9	CL reserves
0	4339	9617	11584	12001	12640	12966	13058	13080	13182	13183	0.00
1	4445	9090	11125	11815	12027	12078	12626	12626	12627	12628	0.96
2	6123	10564	12290	13579	14063	14412	14488	14497	14555	14556	59.19
3	5464	9537	11481	12335	12699	12721	12725	12735	12786	12787	61.82
4	5113	8527	10344	11381	11742	12036	12202	12212	12260	12261	225.36
5	6540	11810	14533	14998	15890	16152	16375	16388	16453	16455	564.54
6	5781	9519	10929	11356	11796	11991	12157	12166	12215	12216	859.56
7	4356	6779	8199	8715	9053	9202	9329	9337	9374	9375	1175.70
8	5250	8967	10755	11432	11875	12071	12238	12247	12296	12297	3330.06
9	5743	10225	12263	13035	13541	13764	13954	13965	14021	14022	8278.93
Total reserves											14556.11

Table 4.5: Fulfilled triangle and reserves for the Chain-Ladder method

4.3 Bornhuetter-Ferguson method

Bornhuetter-Ferguson method is considered to be a bayesian method and we use priori information in it.

As first, we calculate the estimators of pattern $(\beta_j)_{j=0,\dots,J}$, denoted $(\widehat{\beta}_j)_{j=0,\dots,J}$, using the identification (1.8). Values of $(\widehat{\beta}_j)_{j=0,\dots,J}$ are showed in Table 4.6.

j	0	1	2	3	4	5	6	7	8	9
$\widehat{\beta}_j$	0.410	0.729	0.875	0.930	0.966	0.982	0.995	0.996	1.000	1.000

Table 4.6: Development pattern $(\widehat{\beta}_j)_{j=0,\dots,J}$

The next step is to make a proper choice of $\widehat{\mu}_i$—a prior estimate for the expected ultimate payments $\mathrm{E}[C_{i,J}]$. This estimate may originate from widespread sources, for example from a volume measure such as premium or from external expert opinion. However, it should not be based on any observations of payments in the triangle.

We use the loss ratio method to determine $\widehat{\mu}_i$. The idea of this method is based on expected loss ratio which may differ for different accident years. More detailed description of it can be found in [Mandl, Mazurová–1999].

In this method, we denote ELR_i to be the expected loss ratio for the accident year i. Then, $\widehat{\mu}_i$ is given by the product of this ratio and premium received

$$\widehat{\mu}_i = ELR_i \pi_i.$$

The expected loss ratio should be set properly. There are a number of sources which are used for its selection, such as past data and experiences about claims or market statistics for the similar lines of business. For simplicity in our example, we assume the expected loss ratio to be constant for all accident years and we set its value to be 0.75. Such a choice is reasonable for the Auto Liability Insurance.

According to (1.9) we are now able to compute the Bornhuetter-Ferguson estimator of ultimate claims.

In Table 4.7, we can see the values of the prior estimates $\widehat{\mu}_i$, ultimate payments $\widehat{C_{i,J}^{BF}}$ and BF reserves. For comparison we also show the corresponding CL estimates.

As showed in the section 1.2, CL method is rather BF method with the prior estimate replaced by $\widehat{C_{i,J}^{CL}}$. In our example, the value of the total BF reserve is evidently higher than the CL reserves. This is caused by the fact that $\widehat{\mu}_i$ is in every accident year higher—more pessimistic and conservative—than the CL estimator $\widehat{C_{i,J}^{CL}}$.

We must point out that the loss ratio method for setting of $\widehat{\mu}_i$ is not the most used method in the practice. Usually, $\widehat{\mu}_i$ represents the strategic business plan value of the insurance company and its choice is based on an opinion of an experienced expert.

i	$\widehat{\mu}_i$	$\widehat{C_{i,J}^{BF}}$	$\widehat{C_{i,J}^{CL}}$	BF reserves	CL reserves
0	14094.75	13183.00	13183.00	0.00	0.00
1	14211.00	12628.08	12627.96	1.08	0.96
2	15395.25	14559.60	14556.19	62.60	59.19
3	15958.50	12802.15	12786.82	77.15	61.82
4	15584.25	12322.44	12261.36	286.44	225.36
5	17409.00	16487.28	16454.54	597.28	564.54
6	16664.25	12528.60	12215.56	1172.60	859.56
7	13735.50	9921.59	9374.70	1722.59	1175.70
8	12782.25	12428.45	12297.06	3461.45	3330.06
9	14412.75	14252.68	14021.93	8509.68	8278.93
Total reserves				15890.87	14556.11

Table 4.7: Reserves from the BF and the CL method

4.4 Benktander-Hovinen method

Benktander-Hovinen method is characterized as an iterated Bornhuetter-Ferguson method. We have development pattern $(\widehat{\beta}_j)_{j=0,\dots,J}$ from Table (4.6) and the prior estimate $\widehat{\mu}_i$ from Table (4.7). Using the formula (3.3) it is not difficult to calculate BH estimators $\widehat{C_{i,J}^{BH}}$, $i = 1, \dots, I$. Their values as well as corresponding BH reserves are showed in Table 4.8.

Now let us compare total reserves from the CL, BF and BH method. This is showed in Table 4.9 and we see that the BH total reserve is between the CL and the BF total reserves, because it is the mixture of the CL and BH method.

In the previous chapter we also mentioned the further iteration of the BF method which should lead to the CL estimator of ultimate claims. By applying Theorem 3.1 on our data set we get Table 4.10. We see that the differences are

i	$\widehat{C_{i,J}^{BH}}$	BH reserves
0	13183.00	0.00
1	12627.96	0.96
2	14556.21	59.21
3	12786.89	61.89
4	12262.49	226.49
5	16455.66	565.66
6	12237.59	881.59
7	9443.29	1244.29
8	12332.64	3365.64
9	14158.17	8415.17
Total reserves		14820.88

Table 4.8: Ultimate claim and reserves from the BH method

	CL method	BF method	BH method
Total reserves	14556.11	15890.87	14820.88

Table 4.9: Comparison of total reserves from the CL, BF and BH method

insignificant already after the third iteration and the differences between $\widehat{C^9}$ and $\widehat{C_{i,J}^{CL}}$ are visible only in the last two accident years.

i	$\widehat{C^2} = \widehat{C_{i,J}^{BH}}$	$\widehat{C^3}$	$\widehat{C^4}$	...	$\widehat{C^9}$	...	$\widehat{C_{i,J}^{CL}}$
0	13183.00	13183.00	13183.00	...	13183.00	...	13183.00
1	12627.96	12627.96	12627.96	...	12627.96	...	12627.96
2	14556.21	14556.19	14556.19	...	14556.19	...	14556.19
3	12786.89	12786.82	12786.82	...	12786.82	...	12786.82
4	12262.49	12261.38	12261.36	...	12261.36	...	12261.36
5	16455.66	16454.58	16454.54	...	16454.54	...	16454.54
6	12237.59	12217.11	12215.67	...	12215.56	...	12215.56
7	9443.29	9383.30	9375.78	...	9374.70	...	9374.70
8	12332.64	12306.69	12299.67	...	12297.06	...	12297.06
9	14158.17	14102.37	14069.42	...	14025.33	...	14021.93

Table 4.10: Iteration of the BF method

4.5 Cape-Cod model

Cape-Cod model consists in making the diagonal observations more robust. By using the formulas (3.4) and (3.5) we get the estimated values of the loss ratio $\widehat{\kappa}_i$, for the accident years $i = 0,..., I$ and of the "robusted" overall loss ratio $\widehat{\kappa^{CC}}$. They are listed in Table 4.11. For the better visualisation we can see these values in the graph in Figure 4.3. The value of $\widehat{\kappa^{CC}}$ is illustrated with the line.

i	π_i	$\widehat{\kappa_i}$
0	18793	0.701
1	18948	0.666
2	20527	0.709
3	21278	0.601
4	20779	0.590
5	23212	0.709
6	22219	0.550
7	18314	0.512
8	17043	0.722
9	19217	0.730
$\widehat{\kappa^{CC}}$		0.643

Table 4.11: Values of $\widehat{\kappa_i}$ and $\widehat{\kappa^{CC}}$

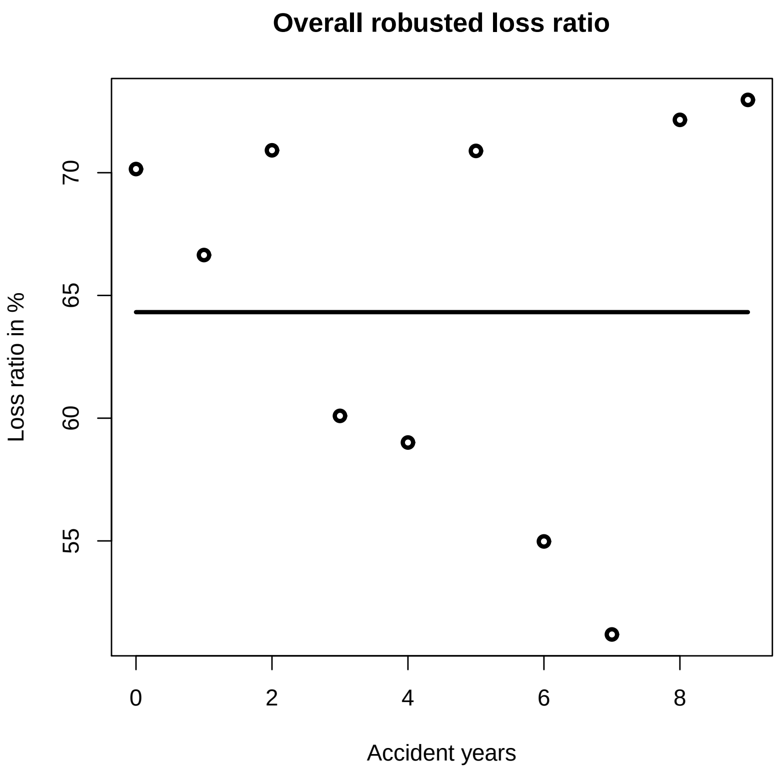

Figure 4.3: The loss ratios $\widehat{\kappa_i}$ and the robusted ratio $\widehat{\kappa^{CC}}$

Smoothed "robusted" diagonal observations $\widehat{C^{CC}_{i,I-i}}$ are obtained by (3.6). In Table 4.12 we see these values and for comparison the original diagonal observations, too.

The Cape-Cod ultimate claims and reserves are shown in Table 4.13. We can notice that the total CC reserves are smaller than the total BF and CL reserves. This can be easily explained.

In the most of the accident years smoothing have reduced the diagonal observation. We already known from the previous chapter that Cape-Cod method is rather the CL method applied to the robusted diagonal values. So it is reasonable that the total CC reserves are smaller than the total CL reserves.

When we take another look at the CC estimator $\widehat{C_{i,J}^{CC}}$, it can be understood as the BF estimator with the prior estimate $\widehat{\kappa^{CC}}\pi_i$. The prior estimate used in BF method is $\widehat{\mu_i} = ELR_i\pi_i$. $\widehat{\kappa^{CC}} = 0.643$ is below $ELR_i = 0.75$ which implies that the CC reserves are smaller that BF reserves in every accident year.

i	$C_{i,I-i}$	$\widehat{C_{i,I-i}^{CC}}$
0	13183	12087.53
1	12627	12186.30
2	14497	13149.14
3	12725	13619.70
4	12036	13119.26
5	15890	14417.57
6	11356	13285.50
7	8199	10302.16
8	8967	7993.43
9	5743	5062.42

Table 4.12: Robusted diagonal values

i	$\widehat{C_{i,J}^{CC}}$	CL reserves	BF reserves	CC reserves
0	13183.00	0	0	0
1	12627.92	0.96	1.08	0.92
2	14550.69	59.19	62.6	53.69
3	12791.16	61.82	77.15	66.16
4	12281.65	225.36	286.44	245.65
5	16402.22	564.54	597.28	512.22
6	12361.61	859.56	1172.6	1005.61
7	9676.28	1175.7	1722.59	1477.28
8	11935.51	3330.06	3461.45	2968.51
9	13040.82	8278.93	8509.68	7297.82
Total reserves		14556.1	15890.87	13627.86

Table 4.13: Ultimate claims and reserves from the Cape-Cod model

4.6 Poisson-Gamma model

Let us now illustrate the Poisson-Gamma model for our data set. This model belongs to the strict bayesian methods for claims reserving so we assume a prior gamma distribution for a risk characteristics of accident years and the conditional Poisson distribution for the incremental payments. When we look at our incremental triangle in the picture 4.2 it is obvious that all observed payments are positive and it will be no problem in the application of this model.

According to 3.3.1 we assume known positive constants γ_0, ..., γ_J with $\sum_{j=0}^{J} \gamma_j = 1$. In practice we do not usually know them and we have to choose estimators $\widehat{\gamma}_j$, $j = 0, ..., J$. Often used approach is to set

$$\widehat{\gamma}_j = \widehat{\beta}_j - \widehat{\beta}_{j-1}, \quad j = 1, ..., J,$$

where $(\widehat{\beta}_j)_{j=0,...,J}$ is the already estimated pattern from the BF method and $\widehat{\gamma}_0$ is set from the condition $\sum_{j=0}^{J} \widehat{\gamma}_j = 1$. This choice of $\widehat{\gamma}_j$ is recommended in [Wüthrich, Merz–2008]. The values of $\widehat{\gamma}_j$ can be seen in Table 4.14

j	0	1	2	3	4	5	6	7	8	9
$\widehat{\gamma}_j$	0.40957	0.31963	0.14539	0.05505	0.03606	0.01593	0.01355	0.00077	0.00399	0.00008

Table 4.14: The values of $\widehat{\gamma}_j$

The next step is to properly specify prior gamma distributions of Θ_i, for all $i = 0, \ldots, I$—to choose the shape parameters a_i and the scale parameters b_i. The issue of the prior distribution selection is very polemic aspect of bayesian theory. There is no recommended method which is considered to be the best. The assessment of the prior is often made by an expert and it contains all his judgements, beliefs and available information.

In our example, we adopt the method from [Wüthrich, Merz–2008]. In addition, we look at two different choices of the prior parameters a_i and b_i and review their influence on the resulting estimators for the ultimate claims.

We come from the coefficient of the variation of Θ_i—$\mathrm{CV}[\Theta_i]$, the expected value and the variance of Θ_i. $\mathrm{CV}[\Theta_i]$ is defined as

$$\mathrm{CV}[\Theta_i] = \frac{\sqrt{\mathrm{var}[\Theta_i]}}{\mathrm{E}[\Theta_i]},$$

it expresses the dispersion of the variable Θ_i, it is independent of the units of measurement and it is usually given in %.

From the previous chapter we have the following relations

$$\mathrm{E}[\Theta_i] = \mathrm{E}[C_{i,J}] = \frac{a_i}{b_i},$$

$$\mathrm{var}[\Theta_i] = \frac{a_i}{b_i^2},$$

$$\mathrm{CV}[\Theta_i] = \frac{1}{\sqrt{a_i}}.$$

The selection of prior parameters consists in making the fixed choice of $\mathrm{CV}[\Theta_i]$ for all accident years and in the choice of $\mathrm{E}[\Theta_i]$.

The case of $\mathrm{CV}[\Theta_i] = 5\%$

In the first assumed case, we choose $\mathrm{CV}[\Theta_i] = 5\%$ for all accident years i. Next, we set the prior expected value $\mathrm{E}[\Theta_i]$, $\mathrm{E}[C_{i,J}]$ respectively. We use the same values as in the BF method—the values of $\widehat{\mu}_i$. Having $\mathrm{E}[\Theta_i]$ and $\mathrm{CV}[\Theta_i]$ it is easy to get the prior parameters a_i and b_i. Their values are shown in Table 4.15.

i	CV$[\Theta_i]$	E$[\Theta_i]$	a_i	b_i	var$[\Theta_i]$
0	0.05	14094.75	400	0.0284	496655
1	0.05	14211.00	400	0.0281	504881
2	0.05	15395.25	400	0.0260	592534
3	0.05	15958.50	400	0.0251	636684
4	0.05	15584.25	400	0.0257	607172
5	0.05	17409.00	400	0.0230	757683
6	0.05	16664.25	400	0.0240	694243
7	0.05	13735.50	400	0.0291	471660
8	0.05	12782.25	400	0.0313	408465
9	0.05	14412.75	400	0.0278	519318

Table 4.15: The choice of the prior parameters

As was proved in the previous chapter, the posterior distribution of Θ_i, given the observations in the upper triangle, is again the gamma distribution but with the new posterior parameters $a_{i,I-i}^{post}$ and $b_{i,I-i}^{post}$. Their values are given by (3.9). In fact, we should write $\widehat{a_{i,I-i}^{post}}$ and $\widehat{b_{i,I-i}^{post}}$ because instead of the known values of γ_{I-i} their estimates have been used. But we will not use this to not overload the notation. In Table 4.16 we can see the values of the posterior parameters, as well as the posterior expected value $\mathrm{E}[\Theta_i \mid D_I]$—the optimal bayesian estimator of Θ_i.

i	$a_{i,I-i}^{post}$	$b_{i,I-i}^{post}$	$\mathrm{E}[\Theta_i \mid D_I]$
0	13583	1.0284	13208.16
1	13027	1.0281	12671.30
2	14897	1.0219	14577.52
3	13125	1.0202	12864.74
4	12436	1.0073	12346.03
5	16290	0.9887	16476.72
6	11756	0.9536	12327.54
7	8599	0.9037	9515.22
8	9367	0.7605	12317.02
9	6143	0.4373	14046.73

Table 4.16: The posterior parameters

Finally, we can set estimators of the ultimate claims $\widehat{C_{i,J}^{PoiGa}}$, which are calculated by (3.14). These estimators and the corresponding reserves are listed in Table 4.17.

The case of CV$[\Theta_i] = 0.5$ ‰

Now we make another choice of the prior parameters and derive the estimator of ultimate claims again. We change the coefficient of variation of Θ_i and we set CV$[\Theta_i]$ to be 100-times smaller than in the previous case. So, we have CV$[\Theta_i] = 0.5$ ‰ for all $i = 0, ..., I$. The prior expected values $\mathrm{E}[\Theta_i]$, $\mathrm{E}[C_{i,J}]$ respectively, remain unchanged. The parameters a_i and b_i can be again easily set

i	$\widehat{C_{i,J}^{PoiGa}}$	Po-Ga reserves
0	13183.00	0.00
1	12627.96	0.96
2	14556.28	59.28
3	12787.19	62.19
4	12262.92	226.92
5	16455.30	565.30
6	12223.44	867.44
7	9392.32	1193.32
8	12302.47	3335.47
9	14036.57	8293.57
Total reserves		14604.45

Table 4.17: The estimators of ultimate claims and reserves in the Po-Ga model

and their values are shown in Table 4.18. Observe that while the values of the prior parameters are 10000-times greater than in the previous case the variance of Θ_i is 10000-times smaller.

i	$\mathrm{CV}[\Theta_i]$	$\mathrm{E}[\Theta_i]$	a_i	b_i	$\mathrm{var}[\Theta_i]$
0	0.0005	14094.75	4000000	283.79	49.67
1	0.0005	14211.00	4000000	281.47	50.49
2	0.0005	15395.25	4000000	259.82	59.25
3	0.0005	15958.50	4000000	250.65	63.67
4	0.0005	15584.25	4000000	256.67	60.72
5	0.0005	17409.00	4000000	229.77	75.77
6	0.0005	16664.25	4000000	240.03	69.42
7	0.0005	13735.50	4000000	291.22	47.17
8	0.005	12782.25	4000000	312.93	40.85
9	0.005	14412.75	4000000	277.53	51.93

Table 4.18: The choice of the prior parameters

The posterior parameters $a_{i,I-i}^{post}$ and $b_{i,I-i}^{post}$ calculated by (3.9) are listed in Table 4.19 also with the posterior expected value $\mathrm{E}[\Theta_i \mid D_I]$.

Finally, we are able to derive the estimators of the ultimate claims $\widehat{C_{i,J}^{PoiGa}}$ and reserves by (3.14). Their values are listed in Table 4.20.

The comparison of results

As we have shown in our examples, the different prior distributions give the different values of the ultimate claims and reserves. Now we try to sum up the differences and investigate the effect of the prior choice on the resulting estimators.

Our two cases of the prior parameters choice differ only in the selection of $\mathrm{CV}[\Theta_i]$, $\mathrm{var}[\Theta_i]$ respectively. The prior expected value has always the same value. The smaller $\mathrm{CV}[\Theta_i]$ the smaller dispersion in the risk characteristic Θ_i and this

i	$a_{i,I-i}^{post}$	$b_{i,I-i}^{post}$	$E[\Theta_i \mid D_I]$
0	4013183	284.79	14091.55
1	4012627	282.47	14205.40
2	4014497	260.82	15392.05
3	4012725	251.65	15945.96
4	4012036	257.65	15571.59
5	4015890	230.73	17405.01
6	4011356	240.96	16647.09
7	4008199	292.09	13722.44
8	4008967	313.66	12781.12
9	4005743	277.94	14412.17

Table 4.19: The posterior parameters

i	$\widehat{C_{i,J}^{PoiGa}}$	Po-Ga reserves
0	13183.00	0.00
1	12628.08	1.08
2	14559.60	62.60
3	12802.09	77.09
4	12322.21	286.21
5	16487.15	597.15
6	12527.39	1171.39
7	9919.96	1720.96
8	12428.14	3461.14
9	14252.34	8509.34
Total reserves		15886.94

Table 4.20: The estimators of ultimate claims and reserves in the Po-Ga model

expresses that we are more sure about our a priori belief. This is reflected in the resulting Poisson-Gamma estimator $\widehat{C_{i,J}^{PoiGa}}$.

Let us recall the formula (3.14) for $\widehat{C_{i,J}^{PoiGa}}$

$$\widehat{C_{i,J}^{PoiGa}} = C_{i,I-i} + (1-\beta_{I-i})\left[\frac{b_i}{b_i+\beta_{I-i}}\frac{a_i}{b_i} + \left(1-\frac{b_i}{b_i+\beta_{I-i}}\right)\frac{C_{i,I-i}}{\beta_{I-i}}\right].$$

We can notice that the expression in the brackets is the credibility mixture between the prior mean $\frac{a_i}{b_i}$, which in our examples coincide with the BF prior estimate $\widehat{\mu_i}$, and the observations $\frac{C_{i,I-i}}{\beta_{I-i}}$. The weights are $\frac{b_i}{b_i+\beta_{I-i}}$ and $\frac{\beta_{I-i}}{b_i+\beta_{I-i}}$. Intuitively, the smaller value of $\mathrm{CV}[\Theta_i]$ should imply the greater weight given to $\frac{a_i}{b_i}$ and consequently the lower weight given to the observations.

We verify this in Table 4.21 where the both weights—for the prior expected value and for the observations—are shown for each case of $\mathrm{CV}[\Theta_i]$. It is obvious that already by the choice $\mathrm{CV}[\Theta_i]$ to be 5% the weights given to the prior are relatively small and the observations are dominant in the estimator of the ultimate claims. On the other hand, when the $\mathrm{CV}[\Theta_i]$ is 100-times smaller, we are very sure about a priori belief and the weights given for $\frac{a_i}{b_i}$ are almost 1 while the weights given to the observations are very low.

	CV[Θ_i] = 5%		CV[Θ_i] = 0.5 ‰	
i	$\frac{b_i}{b_i+\beta_{I-i}}$	$\frac{\beta_{I-i}}{b_i+\beta_{I-i}}$	$\frac{b_i}{b_i+\beta_{I-i}}$	$\frac{\beta_{I-i}}{b_i+\beta_{I-i}}$
0	0.0276	0.9724	0.9965	0.0035
1	0.0274	0.9726	0.9965	0.0035
2	0.0254	0.9746	0.9962	0.0038
3	0.0246	0.9754	0.9960	0.0040
4	0.0255	0.9745	0.9962	0.0038
5	0.0232	0.9768	0.9958	0.0042
6	0.0252	0.9748	0.9961	0.0039
7	0.0322	0.9678	0.9970	0.0030
8	0.0411	0.9589	0.9977	0.0023
9	0.0635	0.9365	0.9985	0.0015

Table 4.21: The credibility weights

Now we concentrate on the obtained ultimate claims and reserves and we trade off the estimators from the both assumed cases, from the CL method and from the BF method. The comparison of the ultimate claims is shown in Table 4.22. As was already mentioned, the Bornhueter-Ferguson method is based on the prior knowledge, the Chain-Ladder method uses the opposite principle and it is founded only on the observations.

The results in Table 4.22 verify that when the coefficient of the variation of the prior distribution is low—0.5 ‰ , so we consider our prior information to be proper and trustworthy, we get closer to the BF model. We have to remark that resemblance between the results from Po-Ga model and BF model is caused by a fact that we have used the same prior expected value of $C_{i,J}$, $\mathrm{E}[C_{i,J}] = \frac{a_i}{b_i} = \widehat{\mu}_i$ in the both models. Obversely, the choice of CV[Θ_i] = 5% causes that the ultimate claims are very similar to CL ultimate claims.

	$\widehat{C_{i,J}^{PoiGa}}$			
i	CV[Θ_i] = 5%	CV[Θ_i] = 0.5 ‰	$\widehat{C_{i,J}^{CL}}$	$\widehat{C_{i,J}^{BF}}$
0	13183.00	13183.00	13183.00	13183.00
1	12627.96	12628.08	12627.96	12628.08
2	14556.28	14559.59	14556.19	14559.60
3	12787.19	12802.09	12786.82	12802.15
4	12262.92	12322.21	12261.36	12322.45
5	16455.30	16487.15	16454.54	16487.29
6	12223.44	12527.39	12215.56	12528.60
7	9392.32	9919.96	9374.70	9921.59
8	12302.47	12428.14	12297.06	12428.45
9	14036.57	14252.34	14021.93	14252.68

Table 4.22: The comparison of the ultimate claims

Finally, we also compare the outstanding and the total reserves. This is shown in Table 4.23. Obviously, the case with $\mathrm{CV}[\Theta_i] = 0.5$ ‰ is very close to the BF method.

In our examples we have verified the following conclusion. The higher $\mathrm{CV}[\Theta_i]$ is chosen the more the final Po-Ga estimator is based on the observation and the closer to the CL estimator. The chosen prior expected value displays in the estimator only insignificantly.

On the other, the low $\mathrm{CV}[\Theta_i]$ represents the strong belief in the prior estimator and the credibility weight given to the prior expected value is dominant.

	Po-Ga reserves			
i	$\mathrm{CV}[\Theta_i] = 5\%$	$\mathrm{CV}[\Theta_i] = 0.5$ ‰	CL reserves	BF reserves
0	0.00	0.00	0.00	0.00
1	0.96	1.08	0.96	1.08
2	59.28	62.59	59.19	62.60
3	62.19	77.09	61.82	77.15
4	226.92	286.21	225.36	286.44
5	565.30	597.15	564.54	597.28
6	867.44	1171.39	859.56	1172.60
7	1193.32	1720.96	1175.70	1722.59
8	3335.47	3461.14	3330.06	3461.45
9	8293.57	8509.34	8278.93	8509.68
Total reserves	14604.45	15886.94	14556.11	15890.87

Table 4.23: The comparison of the total reserves

Conclusion

The goal of this thesis was to analyse the bayesian methods in the stochastic reserving. We have concentrated on the models which are often used in practice and consequently we have applied the methods on a real data set and summed up the results.

In the first two chapters of the thesis we have presented the theoretical information needed for the models definition. At the beginning, we have introduced the standard notation common in the actuarial science. We have also described the elementary Chain-ladder method and Bornhuetter-Ferguson method which are crucial in the reserving theory.

To understand the idea of the bayesian reserving methods we have become familiar with the bayesian approach to statistic in the second chapter. We have defined all the basic terms and the principle of the estimation in the bayesian inference.

In the third chapter we have presented the concrete bayesian models. In the present literature on claims reserving there are mostly listed only the formulas for the estimators of ultimate claims. From this reason we have focused on the derivation of these formulas and have explained the statistical background of the models. We have framed assumptions for each of them and have provided very detailed description of the single steps which lead to the final expression of the estimators.

We have specially analysed the exponential dispersion family and its associated conjugates. The methods based on EDF remove the biggest drawback of the bayesian approach—the intractable results and the absence of the closed formed final expression. We have shown the concrete examples of models which use EDF and have reformulated them so that they coincide with the generally defined EDF model.

Finally, we have presented the important part of this thesis - the application of the models on a real data set. All computation have been made in R software. We have set the amount of ultimate claims and reserves for each method whereby we have highlighted the characteristics of the models and have analysed the connection between them.

In the bayesian methods, the specification of a prior distribution plays a big role. This is work of the expert in the insurance company. There is not much information about the choice of prior parameters in available actuarial literature. In the thesis we have studied the impact of the prior choice on the resulting ultimate claims and reserves. We have considered different choices of the coefficient of variation for the underlying quantity and we have verified following. The higher coefficient of variation has produced the estimator based mainly on the observation and contrariwise the lower coefficient has caused the dominance of the prior expected value in the final estimator of the ultimate claims and reserves.

The bayesian approach offers much more possibilities in the estimation. It could be applied to all unknown parameters in the model and it could be interesting to investigate how this would change the estimators.

Bibliography

[Benktander–1976] Benktander, G. *An Approach to Credibility in Calculating IBNR for Casualty Excess Reinsurance*, The Actuarial Review, April, p. 7. 1976.

[Berger–1993] Berger, J. O. *Statistical Decision Theory and Bayesian Analysis* Springer, 2nd edition, 1993.

[Bornhuetter, Ferguson–1972] Bornhuetter, R.L. and Ferguson, R.E., *The Actuary and IBNR* PCAS, LIX, 1972.

[Bühlmann, Gisler–2005] Bühlmann,H. and Gisler, A. *A Course in Credibility Theory and its Applications*, Springer, Berlin, 2005.

[England, Verrall–2002] England, P. D. and Verrall. R. J. *Stochastic claims reserving in general insurance (with discussion)*,British Actuarial Journal, 8(3), 443-518, 2002.

[Figueiredo–2004] Figueiredo, M. A. T. *Lecture Notes on Bayesian Estimation and Classification*, Instituto de Telecomunicacoes and Instituto Superior Técnico, Lisboa, Portugal, 2004.

[Gesmann, Murphyy, Zhangz–2014] Gesmann, M., Murphyy, D. and Zhangz,W. *Claims reserving with R: ChainLadder-0.1.7 Package Vignette DRAFT*, `http://www.cran.r-project.org/web/packages/ChainLadder/vignettes/ChainLadder.pdf`, 2013. [Accessed 10-February-2014]

[Hovinen–1981] Hovinen, E. *Additive and Continuous IBNR*, ASTIN Colloquium Loen, Norway, 1981.

[Hušková–1985] Hušková, M. *Bayesovské metody*, Univerzita Karlova v Praze, Praha, 1985.

[Klaus–2006] Klaus, D. S. and Lehrstuhl Ffir Versicherungsmathematik and Technische Universität Dresden. *Methods and Models of Loss Reserving Based on Run-Off Triangles: A Unifying Survey*, CAS Forum, Fall, pp. 269-317, 2006.

[Mack–1993] Mack, T. *Distribution-free calculation of the standard error of chain ladder reserves estimates*, ASTIN Bulletin 23/2, 213-225, 1993.

[Mandl, Mazurová–1999] Mandl, P., Mazurová, L. *Matematické základy neživotního pojištění*, Matfyzpress, Praha, 1999.

[Meyers, Shi–2011] Meyers, G. and Shi, P. *Loss Reserving Data Pulled From NAIC Schedule P*, `http://www.casact.org/research/index.cfm?fa=loss_reserves_data`, Casualty Actuarial Society, 2011. [Accessed 15-February-2014]

[R Core Team–2013] R Core Team. *R: A Language and Environment for Statistical Computing*, `http://www.R-project.org`, R Foundation for Statistical Computing, Vienna, Austria, 2013. [Accessed 20-March-2014]

[Wüthrich, Merz–2008] Wüthrich, M. V. and Merz, M. *Stochastic claims reserving methods in insurance*, Wiley, Chichester, England, 2008.

List of Tables

List of Figures

A. Source code

This appendix contains full source code of the examples from Chapter 4.

The function to read data set and obtained triangles of cumulative and incremental payments as well as earned premium.

```
-------------------------------------------------------------------
a=read.csv(choose.files(),header=TRUE)
grp.code=unique(a$GRCODE)

# function to get triangle data
ins.line.data=function(g.code){
  b=subset(a,a$GRCODE==g.code)
  name=b$GRNAME
  grpcode=b$GRCODE
  ay=b$AccidentYear
  dev=b$DevelopmentLag

  cum_incloss=b[,6]
  cum_pdloss=b[,7]
  bulk_loss=b[,8]
  dir_premium=b[,9]
  ced_premium=b[,10]
  net_premium=b[,11]
  single=b[,12]
  posted_reserve97=b[,13]

  # get incremental paid losses
  inc_pdloss=numeric(0)
  for (i in unique(ay)){
    s=(ay==i)
    pl=c(0,cum_pdloss[s])
    ndev=length(pl)-1
    il=rep(0,ndev)
    for (j in 1:ndev){
      il[j]=pl[j+1]-pl[j]
      }
    inc_pdloss=c(inc_pdloss,il)
    }
  data.out=data.frame(name,grpcode,ay,dev,
  net_premium,dir_premium,ced_premium,
  cum_pdloss,cum_incloss,bulk_loss,inc_pdloss,single,posted_reserve97)
  return(data.out)
  }

#my data
dat=ins.line.data(grp.code[3])
dat
```

```
# upper triangle
udat=subset(dat,ay+dev<=1998)
# lower triangle
ldat=subset(dat,ay+dev>1998)

#package ChainLadder
install.packages('ChainLadder')
library(ChainLadder)

#data as triangle
cumpaid=as.triangle(udat,origin="ay", dev="dev", value="cum_pdloss")
cumpaid

#plotted data
plot(cumpaid)
plot(cumpaid,lattice=TRUE)

#incremental triangle
incpaid=cum2incr(cumpaid)
incpaid

#earned premium
premdat=subset(dat,ay+dev==1998)
premdat
prem=premdat$dir_premium
prem
-------------------------------------------------------------------
```

The computation of the development pattern in the Chain-Ladder method and consequently of ultimate claims and reserves.

```
-------------------------------------------------------------------
#####Chain-Ladder method####
#development factors
devfactors=ata(cumpaid)
devfactors
n = 10
 fac2 = sapply(1:(n-1),
function(i){
 sum(cumpaid[c(1:(n-i)),i+1])/sum(cumpaid[c(1:(n-i)),i])
 }
)
fac2
#fulfilled triangle
fullcumpaid= cbind(cumpaid)
for(k in 1:n){
 fullcumpaid[(n-k+1):n, k+1] = fullcumpaid[(n-k+1):n,k]*fac2[k]
 }
 round(fullcumpaid)
```

```
#CL ultimate claims
ultclaim=fullcumpaid[,10]
totalultclaim=sum(ultclaim)
totalultclaim

#CL reserves
diagcumpaid=getLatestCumulative(cumpaid)
diagcumpaid
res=ultclaim-diagcumpaid
res2=data.frame(res)
res2
totalres=sum(res2)
totalres
```

The Bornhuetter-Ferguson method for the data set.

```
####Bornhuetter-Ferguson method####
#computation of development pattern beta
facmin=fac2^(-1)
facmin

beta=1:10
for(k in 1:(n-1)){
beta[k] = prod(facmin[k:(n-1)])
 }
beta[n]=1
beta

#a prior estimate mi , the loss ratio method
mi=prem*0.75
mi

#BF ultimate claims
BFultclaim=1:10
for(k in 0:(n-1)){
BFultclaim[ 1+k]= cumpaid[1+k,n-k]+(1-beta[n-k])*mi[k+1]
 }
BFultclaim

#BF reserves
BFres=BFultclaim-diagcumpaid
BFres2=data.frame(BFres)
BFres2
BFtotalres=sum(BFres2)
BFtotalres
```

The Benktader-Hovinen method and iteration of the BF method.

```
--------------------------------------------------------------------
#####Benktader-Hovinen method####
#BH ultimate claims
BHultclaim=1:10
for(k in 0:(n-1)){
BHultclaim[ 1+k]= cumpaid[1+k,n-k]+(1-beta[n-k])*BFultclaim[k+1]
 }
BHultclaim
round(BHultclaim,2)

#BH reserves
BHres=BHultclaim-diagcumpaid
BHres2=data.frame(BHres)
BHres2
BHtotalres=sum(BHres2)
BHtotalres

#iteration of the BF method
Hultclaim3=1:10
for(k in 0:(n-1)){
BHultclaim3[ 1+k]= cumpaid[1+k,n-k]+(1-beta[n-k])*BHultclaim[k+1]
 }

BHultclaim4=1:10
for(k in 0:(n-1)){
BHultclaim4[ 1+k]= cumpaid[1+k,n-k]+(1-beta[n-k])*BHultclaim3[k+1]
 }

BHultclaim5=1:10
for(k in 0:(n-1)){
BHultclaim5[ 1+k]= cumpaid[1+k,n-k]+(1-beta[n-k])*BHultclaim4[k+1]
 }

BHultclaim6=1:10
for(k in 0:(n-1)){
BHultclaim6[ 1+k]= cumpaid[1+k,n-k]+(1-beta[n-k])*BHultclaim5[k+1]
 }

BHultclaim7=1:10
for(k in 0:(n-1)){
BHultclaim7[ 1+k]= cumpaid[1+k,n-k]+(1-beta[n-k])*BHultclaim6[k+1]
 }

BHultclaim8=1:10
for(k in 0:(n-1)){
BHultclaim8[ 1+k]= cumpaid[1+k,n-k]+(1-beta[n-k])*BHultclaim7[k+1]
 }
```

```
BHultclaim9=1:10
for(k in 0:(n-1)){
BHultclaim9[ 1+k]= cumpaid[1+k,n-k]+(1-beta[n-k])*BHultclaim8[k+1]
 }
```

The Cape-Cod model, computation of "robusted" loss ration and diagonal values and CC estimators for ultimate claims and reserves.

```
####Cape-Cod model####
#loss ratios ki
ki=1:10
for(k in 0:(n-1)){
ki[ 1+k]= cumpaid[1+k,n-k]/(beta[n-k]*prem[1+k])
}
ki

#robusted loss ratio kk
kk=sum(diagcumpaid)/(sum(rev(beta)*prem))
kk

#graph of loss ratios and robusted loss ratio
plot((0:9),ki*100,main = "Overall robusted loss ratio",
xlab = "Accident years",ylab = "Loss ratio in %", lwd = 3)
lines((0:9), y = rep(kk*100,10), type = "l", lwd = 3)

#robusted diagonal observations
diagrob=kk*prem*rev(beta)
diagrob

#CC ultimate claims
cc=1:10
for(k in 1:(n-1)){
cc[k+1] = prod(fac2[(10-k):(n-1)])
 }

CCultclaim1=diagcumpaid-diagrob+cc*diagrob
CCultclaim1
CCultclaim=data.frame(CCultclaim1)
CCultclaim

#CC reserves
CCres=CCultclaim-diagcumpaid
CCres2=data.frame(CCres)
CCres2
CCtotalres=sum(CCres2)
CCtotalres
```

The Poisson-Gamma model for our data set. Calculation of ultimate claims and reserving for two different choices of the prior parameters.

```
####Poisson-Gamma model####
#computation of gammas
gama=1:10
for(k in 2:n){
gama[k] = beta[k]-beta[k-1]
 }
gama[1]=1-sum(gama[2:10])
gama
sum(gama)

#the choice of CV 5%#
vco=0.05
#prior parameters
aa=rep(1/vco^2,10)
aa
bb=aa/mi
bb
aa/(bb^2)

#posterior parameters
apost2=1:10
for(k in 1:n){
apost2[k] = aa[k]+cumpaid[k,n-k+1]
 }
apost2
bpost2=1:10
for(k in 1:n){
bpost2[k] = bb[k]+sum(gama[1:(n-k+1)])
 }
bpost2
apost2/bpost2

#weights given to prior
alfa1=bb/(bb+rev(beta))
alfa1

#PoGa ultimate claims
pogaultclaim=1:10
for(k in 1:n){
pogaultclaim[ k]= cumpaid[k,n-k+1]+(1-beta[n-k+1])
*(alfa1[k]*(aa[k]/bb[k])+(1-alfa1[k])
*(cumpaid[k,n-k+1]/beta[n-k+1]))
 }
```

```
pogaultclaim

#PoGa reserves
pogares=pogaultclaim-diagcumpaid
pogares2=data.frame(pogares)
pogares2
pogatotalres=sum(pogares2)
pogatotalres

#the choice of CV = 0.05%#
vco3=0.0005
#prior parameters
aa3=rep(1/vco3^2,10)
aa3
bb3=aa3/mi
bb3
aa3/(bb3^2)

#posterior parameters
apost23=1:10
for(k in 1:n){
apost23[k] =aa3[k]+ cumpaid[k,n-k+1]
 }
apost23

bpost23=1:10
for(k in 1:n){
bpost23[k] = bb3[k]+sum(gama[1:(n-k+1)])
 }
bpost23
apost23/bpost23

#weights given to prior
alfa13=bb3/(bb3+rev(beta))
alfa13

#PoGa ultimate claims
pogaultclaim3=1:10
for(k in 1:n){
pogaultclaim3[ k]= cumpaid[k,n-k+1]+(1-beta[n-k+1])*
(alfa13[k]*(aa3[k]/bb3[k])+(1-alfa13[k])
*(cumpaid[k,n-k+1]/beta[n-k+1]))
 }
pogaultclaim3

#PoGa reserves
pogares3=pogaultclaim3-diagcumpaid
pogares23=data.frame(pogares3)
```

```
pogares23
pogatotalres3=sum(pogares23)
pogatotalres3
--------------------------------------------------------------------
```