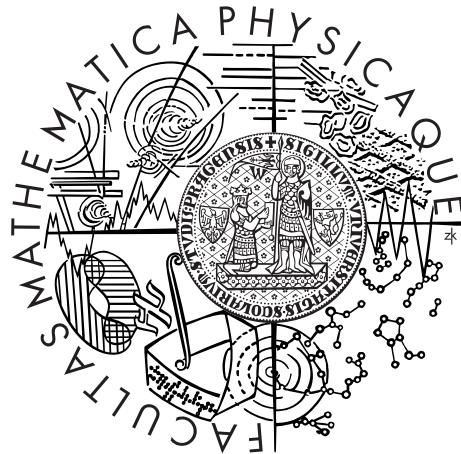


Charles University in Prague
Faculty of Mathematics and Physics

MASTER THESIS



Kateřina Malíková

Claims reserve volatility and bootstrap with application on historical data with trend in claims development

Department of Probability and Mathematical Statistics

Supervisor of the master thesis: RNDr. Michal Peřta, Ph.D.

Study programme: Mathematics

Specialization: Financial and Insurance Mathematics

Prague 2019

I thank the supervisor of my master thesis RNDr. Michal Pešta, Ph.D. for the time he spent in patient guidance of my work, for his great willingness useful advices.

I thank my loving husband for his great support and encouragement for further work. I thank him and my children for their patience and love.

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 on

Author signature

Název práce: Volatilita škodních rezerv a bootstrap s aplikací na historická data s trendem ve vývoji škod

Autor: Kateřina Malíková

Katedra: Katedra pravděpodobnosti a matematické statistiky

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

Abstrakt: Tato práce se zabývá aplikací stochastických rezervovacích metod na daná data s trendem ve vývoji škod. Popisuje metodu chain ladder a zobecněné lineární modely jako její stochastický rámec. Jsou navrženy jednoduché funkce k vyhlazení koeficientů období vzniku a vývojového období z odhadnutého modelu. Je také počítáno s extrapolací pro získání nepozorovaných koncových hodnot. Reziduální bootstrap je použit pro reparametrizovaný model s cílem získat prediktivní rozdělení společně s jeho směrodatnou chybou jako mírou volatility. Je také spočítán solventnostní kapitálový požadavek v jednoletém časovém horizontu.

Klíčová slova: stochastické rezervovací metody, bootstrap, trend, chain ladder, volatilita

Title: Claims reserve volatility and bootstrap with application on historical data with trend in claims development

Author: Kateřina Malíková

Department: Department of Probability and Mathematical Statistics

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

Abstract: This thesis deals with the application of stochastic claims reserving methods to given data with some trends in claims development. It describes the chain ladder method and the generalized linear models as its stochastic framework. Some simple functions are suggested for smoothing the origin and development period coefficients from the estimated model. The extrapolation is also considered for estimation of the unobserved tail values. The residual bootstrap is used for the reparameterized model in order to get the predictive distribution of the estimated reserve together with its standard deviation as a measure of volatility. Solvency capital requirement in one year time horizon is also calculated.

Keywords: stochastic reserving, bootstrap, trend, chain ladder, volatility

Contents

1	Introduction	2
2	Chain ladder method	4
2.1	Form of data and notation	4
2.2	Deterministic approach	5
2.3	Mack's model	6
2.4	Smoothing and tail factors	7
3	Generalized linear models	8
3.1	Generalized linear model	8
3.2	Exponential distribution family	9
3.2.1	Normal distribution	9
3.2.2	Poisson distribution	10
3.2.3	Gamma distribution	10
3.3	GLMs for claims reserving	10
3.3.1	Normal and Mack's model	11
3.3.2	Poisson and gamma model	11
3.4	Log-normal models	11
3.5	Linear predictors in GLMs	11
3.5.1	Chain ladder linear predictor	12
3.5.2	Chain ladder with calendar year parameter	12
3.5.3	Hoerl curve	12
3.5.4	Wright's model	12
3.6	Smoothing	13
3.6.1	Reparameterization	13
3.6.2	Log-linear smoothing with simple functions	14
3.6.3	Log-linear smoothing with linear functions	14
3.7	Model selection	15
3.7.1	Selection of the underlying distribution	15
3.7.2	Selection of truncation points for smoothing of the estimates	16
4	Bootstrap	18
4.1	Types and approaches	18
4.1.1	Paired vs. residual bootstrap	18
4.1.2	Parametric vs. nonparametric bootstrap	19
4.2	Approximations	19
4.3	Bootstrapping in claims reserving	20
4.4	The bootstrap process	20
4.5	Residuals	20

4.5.1	Pearson residuals	21
4.5.2	Deviance residuals	21
4.5.3	Anscombe residuals	22
5	Regulatory requirements	23
5.1	Solvency II	23
5.2	Modifications of bootstrap procedure for one year view	23
5.2.1	First modification	24
5.2.2	Second modification	24
6	Application	25
6.1	Data	25
6.2	GLM	25
6.3	Smoothing of the parameters	27
6.3.1	Simple functions	27
6.3.2	Linear functions	27
6.3.3	Reparameterization	29
6.4	Reparameterized GLM	31
6.5	Bootstrapping	31
6.6	Solvency capital requirement	32
7	Conclusion	35
	Bibliography	37
	List of Figures	39
	List of Tables	40

Chapter 1

Introduction

The aim of this work is to apply stochastic reserving methods to past data with some trend presented in development of claims in order to get an estimate of claims reserve needed to hold to cover potential losses and to investigate volatility of this estimate. The data chosen for this work relate to personal auto paid claims published in Zhang (2010) which are also available in R software which is used for application part of this work.

For most sections of the theoretical part of this work more sources of information were used which are mentioned in the text and which often use different notation and abbreviations or the same naming of different things. So the comparison of the texts requires careful reading.

This work begins with a description of the chain ladder method which serves for estimation of claims reserve needed to hold based on observed past data. The following chapter covers generalized linear models (GLMs) which are introduced at first in a general view together with some derivations of expressions of probability functions of chosen distributions with their parameters from the exponential distribution family and with an effort of linking some other stochastic methods with this GLM framework. Then the GLMs are presented as a stochastic framework for the chain ladder method providing more information about estimates than the deterministic method alone. Other settings in GLMs are also considered which do not exactly replicate the chain ladder estimates but which are also suitable for estimation in claims reserving.

The main section of the theoretical part is the description of reparameterization involving smoothing of linear coefficients of GLM, both for origin and development year, in order to eliminate the influence of outliers and to get estimates of unobserved tail values. The main structure of reparameterization matrix is taken from Björkwall et al. (2011) where also one type of specific reparameterization for development period coefficients is recommended which involves linear smoothing of the latest parameters. This work extends the possibility of smoothing by other suitable functions. From originally considered functions even the simplest ones are found to be sufficient with the advantage of the possibility of using the same reparameterization principle which does not need multiple estimation and so it does not increase the errors of estimates. These functions could be more appropriate for different datasets with different shapes of included trends than the smoothing suggested in the literature. Extrapolation of a tail of the development period coefficients is also considered.

Another chapter contains description of the bootstrap technique which replicates the past observed triangle in new pseudotriangles which are then studied in the same way as the original one to get a distribution of claims reserve. Some obligations of insurance companies for holding claims reserves including Solvency II principle are also mentioned and calculated.

In the application part most of the described theory is applied to given data with an aim to find the best estimate of claims reserve needed to hold by an insurance company with respect to possible trends presented in the data and to evaluate the volatility of the results.

Chapter 2

Chain ladder method

Insurance companies are obligated to compensate insured entities when events of loss, specified in insurance contracts which contain a list of covered risks, occur. Some of these claims are settled immediately but some of them not. For these cases insurance companies must hold some financial reserves which should cover potential losses from claims incurred in given accident years. This includes reserves for claims which are incurred but not reported (IBNR) and reserves for claims which are reported but not settled yet (RBNS) during the current accounting period. Some claims need more time to be investigated and then settled in right way. For some lines of insurance it can take even decade.

The most known and widely used approach in actuarial practice for estimating claims reserves necessary to hold is the chain ladder method. It uses observed past data. Firstly it was described as a purely deterministic method with the same approach to all datasets. Then many authors suggested some stochastic models which can reproduce the chain ladder estimates in order to have more information about the estimates. There were attempts to find links between these models, to give them one common framework, they were widely compared and discussed.

2.1 Form of data and notation

Let's start with the notation which will be used throughout this work. We assume that a dataset which consist of past observations is available. These observations could be incurred claims, paid amounts, numbers of claims. We have incremental data, each relevant to some original time period and some delay

accident year	development year					
	1	2	3	...	$I - 1$	I
1	X_{11}	X_{12}	X_{13}	...	$X_{1,I-1}$	$X_{1,I}$
2	X_{21}	X_{22}	X_{23}	...	$X_{2,I-1}$	
3	X_{31}	X_{32}	X_{33}	...		
\vdots	\vdots	\vdots	\vdots			
$I-1$	$X_{I-1,1}$	$X_{I-1,2}$				
I	$X_{I,1}$					

Table 2.1: Run-off triangle of incremental data which can be observed

period which are typically years (but also quarters etc.). The original time when the claim incurred - the origin or accident year will be denoted by $i = 1, \dots, I$, the delay time when the claim was reported or settled - the development year will be denoted by $j = 1, \dots, I$. The letter k will be used for calendar year, it means $k = i + j$. Our dataset consists of observations up to the given calendar year $k = I + 1$. So we have $n = I(I + 1)/2$ observations which can be shown in a form of triangle $\{X_{ij}; i, j \in \nabla\}$, where $\nabla = \{(i, j); i = 1, \dots, I; j = 1, \dots, I - i + 1\}$, which is shown in Table 2.1.

The aim of this work is to study volatility of an estimated reserve to be held for claims occurred in the years under review $i = 1, \dots, I$ but not reported or settled up to the given calendar year. An estimate of such a reserve can be calculated as the sum of estimates of incremental data $\{X_{ij}; i, j \in \Delta\}$ where $\Delta = \{(i, j); i = 2, \dots, I; j = I - i + 2, \dots, I\}$. This complements the original data triangle $\{X_{ij}; i, j \in \nabla\}$ to the form of square as shown in Table 2.2. But this represents only ultimate claims. Also tail values for next development years for which we have not any observations yet can be estimated.

Total reserve of ultimate claims then can be expressed by $R = \sum_{\Delta} X_{ij} = \sum_{i=2}^I R_i$ using reserves for each accident year $R_i = \sum_{j \in \Delta_i} X_{ij}$ where Δ_i is the corresponding row of the run-off triangle. The estimates of incremental claims, accident year reserves and total reserve will be denoted by \hat{X}_{ij} , $\hat{R}_i = \sum_{j \in \Delta_i} \hat{X}_{ij}$ and $\hat{R} = \sum_{\Delta} \hat{X}_{ij}$, respectively.

Data can also be given in a form of triangle of cumulative observations which means that for given accident and development year incremental observations are summed up to that development year which is denoted by $C_{ij} = \sum_{l=1}^j X_{il}$. $C_{ij}, i = 1, \dots, I, j = 1, \dots, I$ represents a random variable. We have observations of this variable for $i = 1, \dots, I; j = 1, \dots, I - i + 1$, the cumulative run-off triangle. The respective estimates will be denoted by \hat{C}_{ij} .

2.2 Deterministic approach

A basic assumption of the chain ladder method is that the development pattern is the same all the time. So observing the historical development of claims is useful for estimation of their future development. The chain ladder method uses a cumulative run-off triangle to estimate cumulative values which complement the original triangle to the form of square.

accident year	1	2	3	...	$I - 1$	I
1						
2						$X_{2,I}$
3					$X_{3,I-1}$	$X_{3,I}$
\vdots					\vdots	\vdots
$I-1$			$X_{I-1,3}$...	$X_{I-1,I-1}$	$X_{I-1,I}$
I		$X_{I,2}$	$X_{I,3}$...	$X_{I,I-1}$	$X_{I,I}$

Table 2.2: Target triangle of incremental data which is the subject of estimation

At first we need to estimate so called development factors which represent the average ratios of the growth of cumulative claims in consecutive development periods. With the chain ladder method we assume that for each accident year we have the same development factors, only the levels of amounts of claims differ.

Ratios of cumulative claims for each accident year i and development years $j + 1$ and j where the data exists are calculated by

$$f_{ij} = C_{i,j+1}/C_{ij}, \quad i, j \in \nabla - \{j; j + 1 \notin \nabla\}.$$

Weighted averages are often preferred for estimation of development factors. These averages for given development year j are calculated by

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{ij}} = \frac{\sum_{i=1}^{I-j} C_{i,j} * f_{ij}}{\sum_{i=1}^{I-j} C_{ij}}, \quad j = 1, \dots, I - 1,$$

see Wüthrich and Merz (2008). These development factors are sometimes called age-to-age factors or link ratios.

Claim amounts are then estimated for each accident year using the last known value in appropriate row of the run-off triangle by

$$\hat{C}_{ij} = C_{i,I-i+1} \hat{f}_{I-i+1} \hat{f}_{I-i+1} \dots \hat{f}_{j-1}.$$

Especially the estimates for the ultimate claims amounts we can get as

$$\hat{C}_{iI} = C_{i,I-i+1} \hat{f}_{I-i+1} \hat{f}_{I-i+1} \dots \hat{f}_{I-1}$$

from which also the estimates of outstanding claims reserves for each accident year

$$\hat{R}_i = \hat{C}_{iI} - C_{i,I-i+1} = C_{i,I-i+1} (\hat{f}_{I-i+1} \hat{f}_{I-i+1} \dots \hat{f}_{I-1} - 1)$$

can be gained. Also other quantities like incremental claims $\hat{X}_{i,j} = \hat{C}_{i,j} - \hat{C}_{i,j-1}$ can be estimated.

2.3 Mack's model

Mack's model is one of the results of the first attempts to link the chain ladder method to some statistical background. The main reason for that was to obtain the standard error of the estimates and to compare it with other claims reserving models. Mack (1993) presents the chain ladder method as a very simple and still distribution free technique.

The first assumption is that there exist some development factors $f_1, \dots, f_{I-1} > 0$ common to all accident years which multiply the cumulative values to form the expected value in the next development year under the condition that we know the previous ones. It means

$$E(C_{i,j+1} | C_{i1}, \dots, C_{ij}) = C_{ij} f_j, \quad 1 \leq i \leq I, \quad 1 \leq j \leq I - 1. \quad (2.1)$$

Other two implicit assumptions of Mack's model express independence of given data between accident years and conditional variability of cumulative claim

amounts equal to the previous observation with lower delay multiplied by some variance factor also common to all accident years:

$$\begin{aligned} \{C_{i1}, \dots, C_{iI}\} &\perp\!\!\!\perp \{C_{l1}, \dots, C_{lI}\}, & i \neq l, \\ \text{var}(C_{i,j+1}|C_{i1}, \dots, C_{ij}) &= C_{ij}\sigma_j^2, & 1 \leq i \leq I, \quad 1 \leq j \leq I-1. \end{aligned} \quad (2.2)$$

Let $\mu_{ij} = \mathbb{E}(C_{i,j+1}|C_{i1}, \dots, C_{ij})$. Development factors can be expressed as

$$f_j = \frac{\sum_{i=1}^{I-j} \mu_{i,j+1}}{\sum_{i=1}^{I-j} \mu_{ij}}, \quad j = 1, \dots, I-1.$$

The chain ladder method estimates them using the observed data by

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j} C_{i,j+1}}{\sum_{i=1}^{I-j} C_{ij}}, \quad j = 1, \dots, I-1.$$

From the first two assumptions (2.1) and (2.2) two important facts follow. Firstly that for ultimate cumulative claims

$$\mathbb{E}(C_{iI}|C_{ij}; i, j \in \nabla) = C_{i,I-i+1}f_{I-i+1} \cdot \dots \cdot f_{I-1}$$

and secondly that $\hat{C}_{iI} = C_{i,I-i}\hat{f}_{I-i}\hat{f}_{I-i+1} \dots \hat{f}_{I-1}$ is its unbiased estimator (see Mack, 1993, for proofs). Mack's model also allows estimation of standard error.

2.4 Smoothing and tail factors

Estimates of development factors from the chain ladder method could be smoothed according to actuary's expert judgement. One of the reasons for smoothing is an extrapolation to get estimates of tail factors for the development years $j = I+1, \dots, I+u$ for which we do not have any observations. Various methods can be used for this purpose. One of the common approaches used with the chain ladder development factors is exponential smoothing. It means that we get their estimates from linear regression modelling dependence of logarithm of the original chain ladder estimates decreased by one on the development year j :

$$\log(\hat{f}_j - 1) \sim j.$$

Usually we decide that the first few estimates of development factors should keep its original chain ladder value due to their different development and only the rest of them is smoothed. We can also decide not to include the last original estimates into smoothing procedure. The reason is that we have not many observations to get these estimates and that they can be somehow outlying and the development factors are not supposed to have the same character for other accident years.

Chapter 3

Generalized linear models

The simple deterministic chain ladder method for estimates of ultimate claims, even with many types of estimates of standard error, may not suffice for all purposes. Sometimes we need other characteristics, measures of uncertainty, quantiles, value at risk etc. That is the reason why actuaries tried to find some stochastic methods to get these estimates. Great effort has been devoted to specifying methods which give exactly or approximately the same results as the chain ladder method whose results seem to be useful. Generalized linear models (GLMs) provide good basis for the chain ladder method. Detailed information about GLMs is provided by McCullagh and Nelder (1989) or Olsson (2002) which are used as sources of information for a large part of this chapter.

3.1 Generalized linear model

As described in Olsson (2002) generalized linear models are generalization of general linear models. It includes also many commonly used models for counts, proportions and other types of responses. The examples are Poisson regression, probit or logit models, log-linear models etc. There is no need to assume normality or homoscedasticity. We can use other distributions from the exponential family of distributions (see the next Section 3.2). The variance is specified by a function of the mean, which allows other cases than homoscedastic only.

Instead of modelling the expected value of the dependent variable using linear predictor as in general linear model, in GLM we can model some function of this mean value which is called the link function. So we need to specify three elements:

- the distribution of the dependent variable \mathbf{Y} ,
- the link function $g(\cdot)$,
- the linear predictor $\boldsymbol{\eta} = \mathbb{X}\boldsymbol{\beta}$.

We denote $\mathbf{E}(\mathbf{Y}) = \boldsymbol{\mu}$. Our model then has the form

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbb{X}\boldsymbol{\beta}$$

where we assume a distribution from the exponential family for \mathbf{Y} . The link function $g(\cdot)$ is a monotone and differentiable function. Monotonicity is needed because of existence of the inverse function $g^{-1}(\cdot)$ which means $g^{-1}(g(\boldsymbol{\mu})) = \boldsymbol{\mu}$. There are different link functions suitable for different types of data. The most commonly used function is a canonical link function for which

$$g(\boldsymbol{\mu}) = \boldsymbol{\theta}$$

where θ is a canonical location parameter of a distribution from the exponential distribution family (see the next Section 3.2).

3.2 Exponential distribution family

Exponential distribution family is a designation for larger group of distributions. It includes many well-known distributions like Poisson, Normal, Gamma, Binomial, Inverse Gaussian etc. They have common form of the density function

$$f(y; \theta, \phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right], \quad (3.1)$$

where θ is some function of the location parameter and it is called the canonical parameter. Parameter ϕ is called the dispersion parameter. Further $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ are some functions. Function $a(\cdot)$ is often considered as unity. Otherwise when it is a function of a constant dispersion parameter, the group of these distributions can be called exponential dispersion family. Function $b(\cdot)$ is important because of its relationship with the mean value

$$E(y) = b'(\theta) \quad (3.2)$$

and the variance of the distribution

$$\text{var}(y) = a(\phi)b''(\theta). \quad (3.3)$$

So we also have a specific relationship between these two moments of the distribution. The second derivative $b''(\theta)$ is called the variance function and is denoted by $V(\mu)$ (a function of the mean). Relationships (3.2) and (3.3) can be proved using maximum likelihood theory. We will use the notation $\ell(\theta, \phi; y) = \log[f(y; \theta, \phi)]$ for the loglikelihood function. From the equation $E(\partial\ell/\partial\theta) = 0$ we get

$$E\{[y - b'(\theta)]/a(\phi)\} = 0$$

from which the relationship (3.2) yields. Similarly from the equation $E(\partial^2\ell/\partial\theta^2) + E[(\partial^2\ell/\partial\theta^2)^2] = 0$ we get

$$-\frac{b''(\theta)}{a(\phi)} + \frac{\text{var}(y)}{a^2(\phi)} = 0$$

from which the relationship (3.3) yields.

3.2.1 Normal distribution

The probability function of normal distribution for $y \in \mathbb{R}$ and $\mu \in \mathbb{R}$, $\sigma^2 > 0$ can be expressed in the form of equation (3.1) as

$$\begin{aligned} f(y; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y - \mu)^2}{2\sigma^2}\right\} \\ &= \exp\left\{-\frac{y^2 - 2y\mu - \mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right\} \\ &= \exp\left\{\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right\} \end{aligned}$$

where $E(y) = \mu$. Thus

$$\theta = \mu, \quad \phi = \sigma^2, \quad a(\phi) = \phi,$$

$$b(\theta) = \theta^2/2, \quad c(y; \phi) = -[y^2/\phi + \log(2\pi\phi)]/2$$

which means the identity canonical link, the dispersion parameter σ^2 and the variance function equal to 1.

3.2.2 Poisson distribution

The probability function of Poisson distribution for y non-negative integer and $\mu > 0$ can be expressed in the form of equation (3.1) as

$$f(y; \mu) = \frac{\mu^y}{y!} e^{-\mu} = \exp\{y \log(\mu) - \mu - \log(y!)\}$$

where $E(y) = \mu$. Thus

$$\theta = \log(\mu), \quad b(\theta) = \exp(\theta), \quad c(y; \phi) = -\log(y!), \quad a(\phi) = 1$$

which means the logarithmic canonical link and the variance function equal to the mean.

3.2.3 Gamma distribution

The probability function of gamma distribution for $y \in (0; \infty)$ and $\beta > 0$, $\alpha > 0$ can be expressed in the form of equation (3.1) as

$$\begin{aligned} f(y; \beta, \alpha) &= \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} \\ &= \exp\{\alpha \log(\beta) + (\alpha - 1) \log(y) - \beta y - \log[\Gamma(\alpha)]\} \\ &= \exp\left\{\frac{y(-\frac{\beta}{\alpha}) - [-\log(\frac{\beta}{\alpha})]}{\alpha^{-1}} + (\alpha - 1) \log(y) + \alpha \log(\alpha) - \log[\Gamma(\alpha)]\right\} \end{aligned}$$

where $E(y) = \alpha/\beta$. Thus

$$\theta = -1/\mu, \quad \phi = \alpha^{-1}, \quad a(\phi) = \phi,$$

$$b(\theta) = -\log(-\theta), \quad c(y; \phi) = (1/\phi - 1) \log(y) - \log(\phi)/\phi - \log[\Gamma(1/\phi)]$$

which means the reciprocal canonical link, the dispersion parameter α^{-1} and the variance function μ^2 .

3.3 GLMs for claims reserving

In claims reserving GLMs are used for incremental claims data where

$$E(X_{ij}) = m_{ij}, \quad \text{var}(X_{ij}) = \phi m_{ij}^p \quad (3.4)$$

where p is a parameter whose value determines specific type of distribution. Specific examples are described in Renshaw and Verrall (1998) or England and Verrall (2002).

3.3.1 Normal and Mack's model

For $p = 0$ in (3.4) we get the normal distribution. With the canonical identity link the model becomes a special case of GLMs - linear regression model.

When we reformulate original Mack's chain ladder model as a model for ratios $C_{ij}/C_{i,j-1}$ (a model with weights $C_{i,j-1}$), we get a similar model with the same estimates. Since no further assumptions about the distribution are made in Mack's model, we are limited in the way that we can estimate only the first two moments of the reserves. Some additional distribution based assumptions are needed in order to estimate some other characteristics of the reserves.

3.3.2 Poisson and gamma model

The over-dispersed Poisson (ODP) and the gamma distributions are the most often used distributions in claims reserving. They are obtained by setting $p = 1$ for the ODP or $p = 2$ for the gamma distribution in (3.4). Logarithmic link function is preferred with both these distributions, even with Gamma distribution which has different canonical link. It means

$$\log(m_{ij}) = \eta_{ij}.$$

Many authors have used these models but they differ in linear predictor structure. Different types of linear predictors are presented in Section 3.5 .

3.4 Log-normal models

Some linear predictors which are used in GLMs are taken from early works based on stochastic models underlying the chain ladder technique where log-normal model is used. It is assumed that $X_{ij}, i, j = 1, \dots, I$ are independent and

$$\log(X_{ij}) = \eta_{ij} + \epsilon_{ij},$$

where $\eta_{ij} = E[\log(X_{ij})]$ and $\epsilon_{ij} \sim N(0, \sigma^2)$ which results in

$$X_{ij} \sim \text{LN}(\eta_{ij}, \sigma^2),$$

see e.g. Björkwall et al. (2011). Random variable $Y_{ij} = \log(X_{ij})$ with normal distribution then can be considered in the framework of GLMs. In this case we use identity link function and the variance as the scale parameter $\phi = \sigma^2$ so that $m_{ij} = \eta_{ij}$ and $Y_{ij} \sim N(m_{ij}, \sigma^2)$. Some linear predictors suggested for η_{ij} in log-normal models are introduced in the next section.

3.5 Linear predictors in GLMs

There are numbers of linear predictors which can be useful in modelling claims reserves. Some of the most often used predictors are described in this chapter. They can also be modified, extended or simplified.

3.5.1 Chain ladder linear predictor

One of the first suggestions for linear predictors used with GLMs (and previously with log-normal models) was

$$\eta_{ij} = c + \alpha_i + \beta_j. \quad (3.5)$$

This is very useful when we want to get the chain ladder estimates and at the same time to obtain more stochastic information than using the deterministic approach. For the Poisson distribution ($p = 1$ in (3.4)) this log-linear model gives exactly the same estimates as the original chain ladder method, for the gamma distribution ($p = 2$ in (3.4)) the estimates are mostly really close.

3.5.2 Chain ladder with calendar year parameter

The chain ladder linear predictor can be extended by calendar year parameter

$$\eta_{ij} = c + \alpha_i + \beta_j + \gamma_k, \quad k = 2, \dots, 2k. \quad (3.6)$$

Usually some constraints are imposed on the parameters because there is a large number of them while we consider only a small number of data. In this example it could be $\alpha_1 = \beta_1 = \gamma_2 = 0$. The log-linear model with this predictor is described in Björkwall et al. (2011).

3.5.3 Hoerl curve

Another well-known linear predictor is the Hoerl curve which has the form

$$\eta_{ij} = c + \alpha_i + \beta_i \log(j) + \gamma_i j.$$

It can be rewritten on the untransformed scale by

$$\exp(\eta_{ij}) = A_i j^{\beta_i} e^{\gamma_i j}$$

where $A_i = \exp(c + \alpha_i)$. In order to estimate also the tail factors we assume that the development time j is a continuous covariate. We can simplify the model by taking only one parameter β and one parameter γ for all origin years which means we suppose that there is always the same pattern for the development of claims for each accident year. We can also leave the first few values and use this model for the development years $j > c$ for some $c > 1$.

3.5.4 Wright's model

T. S. Wright suggested the model where X_{ij} is the sum of N_{ij} independent claims whose values are Z_{ij} . This model can be reparameterized as GLM for incremental claims where

$$\begin{aligned} E(X_{ij}) &= m_{ij} = \exp\{u_{ij} + c + \alpha_i + \beta_i \log(j) + \gamma_i j + \delta k\}, \\ \text{var}(X_{ij}) &= \phi_{ij} m_{ij}. \end{aligned}$$

It is often assumed that $\phi_{ij} = \phi$. With the logarithmic link function we have

$$\log(m_{ij}) = \eta_{ij}, \quad \eta_{ij} = u_{ij} + c + \alpha_i + \beta_i \log(j) + \gamma_i j + \delta k.$$

In some sense it is an extension of the Hoerl curve by two elements: known u_{ij} terms representing technical adjustments and optional δk term for modelling possible claims inflation.

3.6 Smoothing

As in the case of deterministic methods, also with GLMs we can use smoothing and extrapolation in order to get an extra information about a tail or in order to find possible trends. Björkwall et al. (2011) presented GLMs with log-linear smoothing effects where the actuary may choose which parameters should be smoothed, how many of them are included in the smoothing procedure or how far to extrapolate. Also other smoothing functions then their suggestions are used in this work with the same reparameterization principle.

3.6.1 Reparameterization

In this section a reparameterization for GLMs expressed by (3.4) with log-link function given by (3.6) introduced in Björkwall et al. (2011) which allows for inclusion of smoothing of parameters into model selection is presented. We suppose using of constraints $\alpha_1 = \beta_1 = \gamma_2 = 0$. Then other parameters could be reparameterized by

$$\begin{aligned}\boldsymbol{\alpha}^\top &= (\alpha_2 \dots \alpha_I) &= (a_1 \dots a_q) &\begin{pmatrix} A_{12} & \dots & A_{1I} \\ \vdots & \vdots & \vdots \\ A_{q2} & \dots & A_{qI} \end{pmatrix} &= \mathbf{a}^\top \mathbf{A}, \\ \boldsymbol{\beta}^\top &= (\beta_2 \dots \beta_I) &= (b_1 \dots b_r) &\begin{pmatrix} B_{12} & \dots & B_{1I} \\ \vdots & \vdots & \vdots \\ B_{r2} & \dots & B_{rI} \end{pmatrix} &= \mathbf{b}^\top \mathbf{B}, \\ \boldsymbol{\gamma}^\top &= (\gamma_3 \dots \gamma_{I+1}) &= (g_1 \dots g_s) &\begin{pmatrix} \Gamma_{13} & \dots & \Gamma_{1,I+1} \\ \vdots & \vdots & \vdots \\ \Gamma_{s3} & \dots & \Gamma_{s,I+1} \end{pmatrix} &= \mathbf{g}^\top \mathbf{\Gamma},\end{aligned}$$

where $0 \leq q, r, s \leq I - 1$. Using matrices \mathbf{A} , \mathbf{B} and $\mathbf{\Gamma}$ we create a new matrix

$$\mathbf{D} = \begin{pmatrix} 1 & \mathbf{0}^T & \mathbf{0}^T & \mathbf{0}^T \\ \mathbf{0} & \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Gamma} \end{pmatrix}^\top \quad (3.7)$$

which dimension is $v \times w$ where $v = 3I - 2$ is the number of original parameters and $w = 1 + q + r + s$ is the number of new parameters. Using this new parameterization $\boldsymbol{\eta}$ can be expressed by

$$\boldsymbol{\eta} = (\eta_{11} \dots \eta_{1I} \ \eta_{21} \dots \eta_{2,I-1} \dots \eta_{I1})^\top = \mathbb{X}^* \boldsymbol{\theta}^* = \mathbb{X}^* (c \ \boldsymbol{\alpha} \ \boldsymbol{\beta} \ \boldsymbol{\gamma})^T = \mathbb{X}^* \mathbf{D} \boldsymbol{\theta} = \mathbb{X} \boldsymbol{\theta}$$

where $\mathbb{X}^* \boldsymbol{\theta}^*$ denotes the original linear predictor and $\mathbb{X} \boldsymbol{\theta}$ the reparameterized one.

The full GLM can be obtained with $q = r = s = I - 1$ which means that there is no smoothing and $\mathbf{A} = \mathbf{B} = \mathbf{\Gamma} = \mathbf{I}$ are identity matrices.

3.6.2 Log-linear smoothing with simple functions

Elements of matrices \mathbf{A} , \mathbf{B} and $\mathbf{\Gamma}$ could be some functions of origin, development or calendar period respectively and also of the parameters q , r or s . In this section simple functions of time are considered. The simplest case is when the parameters q , r or s are set to one and the corresponding matrix has the form $(1 \ \dots \ I)$. Other reparameterizations considered in this work involve the parameters set to 2 and the matrices of type

$$\begin{pmatrix} 1 & \dots & I \\ f(1) & \dots & f(I) \end{pmatrix}, \quad (3.8)$$

where $f(i), i = 1, \dots, I$ is a simple function, e.g. $1/i$, $\exp(i)$, $\log(i)$, \sqrt{i} , i^l , $l \in \mathbb{R}$ and others. This might be improved by keeping the first few original parameters unchanged and smoothing only the rest of them. It means to increase the value of q , r or s and to use a matrix e.g.

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 4 & \dots & I \\ 0 & 0 & 0 & f(4) & \dots & f(I) \end{pmatrix}. \quad (3.9)$$

3.6.3 Log-linear smoothing with linear functions

Björkwall et al. (2011) suggested the reparameterization with curves

$$\begin{aligned} \alpha_i &= a_{i-1}, & 2 \leq i \leq q, \\ \alpha_i &= a_{q-1} + a_q(i - q), & q + 1 \leq i \leq I, \\ \beta_j &= b_{j-1}, & 2 \leq j \leq r, \\ \beta_j &= b_{r-1} + b_r(j - r), & r + 1 \leq j \leq I, \\ \gamma_k &= g_{k-1}, & 2 \leq k \leq s, \\ \gamma_k &= g_{s-1} + g_s(k - s), & s + 1 \leq k \leq I \end{aligned} \quad (3.10)$$

where we choose the amount of smoothing in the model selection through the values of q, r, s for which we have $1 \leq q, r, s \leq I - 1$. For $q = 1$, $r = 1$ or $s = 1$ we define $a_0 = 0$, $b_0 = 0$ or $g_0 = 0$. Matrices \mathbf{A} , \mathbf{B} or $\mathbf{\Gamma}$ then can have the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & 1 & \dots & 1 \\ 0 & 0 & 0 & 0 & 1 & \dots & I - z \end{pmatrix}, \quad (3.11)$$

where z denotes one of the parameters q , r or s .

We can also try to find a reparameterization which is connected to the chain ladder development factors. We know that for the ODP distribution ($p = 1$ in (3.4)) with the log-link function and the linear predictor expressed by (3.5) we get exactly the same estimates as using the deterministic chain ladder model. So in the reparameterization of this model we do not have parameters for calendar

year meaning $s = 0$ and $\mathbf{\Gamma} = \mathbf{0}$. Theoretical development factors

$$f_j = \frac{\sum_{i=1}^{I-j} \mu_{i,j+1}}{\sum_{i=1}^{I-j} \mu_{ij}} = \frac{\sum_{i=1}^{I-j} \sum_{j=1}^{j+1} m_{il}}{\sum_{i=1}^{I-j} \sum_{j=1}^j m_{il}}, j = 1, \dots, I-1$$

can be estimated by

$$\hat{f}_j = \frac{\sum_{i=1}^{I-j} \sum_{j=1}^{j+1} \hat{m}_{il}}{\sum_{i=1}^{I-j} \sum_{j=1}^j \hat{m}_{il}}, j = 1, \dots, I-1$$

which for $q = r = I-1$ give the same results as the chain ladder estimates. When we let r to be $1 \leq r < I-1$ we allow for some smoothing as in Section 2.4.

3.7 Model selection

3.7.1 Selection of the underlying distribution

The type of the underlying distribution for GLMs is connected with the parameter p in (3.4). It can take any value, even non-integer one. Three special cases, $p = 0, 1$ or 2 representing Normal, Poisson or Gamma distribution, are described in Section 3.3. We should use residual analysis to choose between these models.

Parameter p can also be estimated in order to find the most appropriate model for given data. For that we need adjusted unstandardized residuals defined by

$$r_{ij}^{adj} = \sqrt{\frac{n}{n-q}} (C_{ij} - \hat{m}_{ij}).$$

The mean of this value squared can be expressed by

$$\mathbb{E}(r_{ij}^2) = \frac{n}{n-q} \mathbb{E}[(C_{ij} - \hat{m}_{ij})^2] = \frac{n}{n-q} \mathbb{E}[(C_{ij} - \hat{\mathbb{E}}(C_{ij}))^2].$$

Thus from that and from the expression for the variance $\text{var}(C_{ij}) = \phi m_{ij}^p$ we have the approximate relationship

$$\mathbb{E}(r_{ij}^2) \approx \text{var}(C_{ij}).$$

Then we can find the suitable value of parameter p minimizing

$$f(p, \phi) = \sum_{i,j} w_{ij} (r_{ij}^2 - \phi \hat{m}_{ij}^p)^2 \quad (3.12)$$

with respect to p and ϕ , where w_{ij} are weights often considered to be equal to 1. This method is useful just to determine suitable parameter p . Dispersion parameter ϕ is then estimated using defined parameter p together with other parameters from the model. Residual analysis is recommended to make sure we select the right model.

3.7.2 Selection of truncation points for smoothing of the estimates

When using deterministic methods to estimate development factors and reserves, truncation points, i.e. numbers of first non-smoothed parameters, are chosen by an actuary by eye from the graph or with the help of some additional information. In stochastic models we can use some model selection criterion which can help us to choose truncation points from various possible candidates. Typically we choose the values q, r and s as the argument of minimal value of some criterion which means

$$(\hat{q}, \hat{r}, \hat{s}) = \operatorname{argmin}_{(q,r,s) \in J} \operatorname{Crit}(\hat{\boldsymbol{\theta}}_{qrs})$$

for $\hat{\boldsymbol{\theta}}_{qrs}$ estimated with fixed (q, r, s) in the given model where J is the set of all the triads (q, r, s) considered with given data. The example of the set J is

$$J = \{(I-2, 1, 0), \dots, (I-2, I-1, 0), (I-1, 1, 0), \dots, (I-1, I-1, 0)\}$$

which represents the model without calendar year effect where the last accident year parameter can be linearly interpolated from the previous two with a possibility of smoothing the development year parameters with a choice of any truncation point.

There exist several criteria from which we can choose. The aim is to find a good and useful model without the necessity of having very large model with many parameters. When we use likelihood functions a good choice can be made with information criteria. Different ways of penalization of too complex model is shown by Akaike's and Bayesian information criteria which are defined by

$$\begin{aligned} \operatorname{AIC}(\hat{\boldsymbol{\theta}}_{qrs}) &= 2w - 2\ell(\hat{\boldsymbol{\theta}}_{qrs}), \\ \operatorname{BIC}(\hat{\boldsymbol{\theta}}_{qrs}) &= \log(n) w - 2\ell(\hat{\boldsymbol{\theta}}_{qrs}) \end{aligned}$$

where $w = 1 + q + r + s$ is the number of parameters in the reparameterized model and $n = I(I+1)/2$ is the number of observations for the reminder. The expression $\ell(\hat{\boldsymbol{\theta}}_{qrs})$ denotes the maximum of log-likelihood function when using (q, r, s) .

Another way of finding suitable values could be the mean squared error of prediction (MSEP):

$$\begin{aligned} \operatorname{E}[(R - \hat{R})^2] &= \operatorname{E}[(R - \operatorname{E}[R]) - (\hat{R} - \operatorname{E}[\hat{R}])^2] \\ &\approx \operatorname{E}[(R - \operatorname{E}[R]) - (\hat{R} - \operatorname{E}[\hat{R}])^2] \\ &= \operatorname{E}[(R - \operatorname{E}[R])^2] - 2\operatorname{E}[(R - \operatorname{E}[R])(\hat{R} - \operatorname{E}[\hat{R}])] \\ &\quad + \operatorname{E}[(\hat{R} - \operatorname{E}[\hat{R}])^2] \\ &\approx \operatorname{E}[(R - \operatorname{E}[R])^2] + \operatorname{E}[(\hat{R} - \operatorname{E}[\hat{R}])^2] \\ &= \operatorname{var}[R] + \operatorname{var}[\hat{R}], \end{aligned}$$

where the first approximation comes from replacing of R by \hat{R} and the second one holds when future observations are independent of past observations, see England and Verrall (2002), which represents the sum of the process variance and the estimation variance.

The root mean squared error of prediction (RMSEP) $\sqrt{\text{E} [(R - \hat{R})^2]}$ can also be calculated as the standard deviation of the full predictive distribution when it is obtained. Bootstrap method can be used to achieve this distribution. And we can replace \hat{R} by the bootstrap estimate of claims reserve calculated from the bootstrapped pseudotriangles of past claims \hat{R}^* and also R by the bootstrap value of future reserve R^{**} , see chapter 4 for explanation of bootstrap method.

Chapter 4

Bootstrap

Bootstrapping is a method which uses simulations to estimate some characteristics, relations or even to approximate the whole distribution of an estimate. A single sample of data suffice to get desired results. It uses a large number of simulations to create a pseudoreality with sets of pseudodata with the same underlying distribution as the original one. In this bootstrap world desired characteristics or relations for each pseudodata are studied and their distributions are found and it is assumed that the results achieved there can be applied to similar characteristics or relations from the real world represented by the observed data.

This approach is computer intensive but still simple. It is used especially in cases where the results cannot be achieved analytically, analytical computation is very difficult or tedious or the assumptions made are questionable. Bootstrapping is suited for situations when we do not have any idea about a model suitable for our data or when we have a model but it is too complex and we do not want to oversimplify it or even when the model is simple just to check the results. It can be helpful when we want to relax some assumptions or to find out how robust the conclusions of analysis using our model are or whether the approximations used are valid. This is a quick way to get approximate solutions.

As in Björkwall et al. (2009) we use the notation with star $*$ to denote variables in the bootstrap world distributed according to the fitted distribution or estimators derived from them when there are observations or estimates in the real world which can be substituted by these bootstrap counterparts. Notation with two stars $**$ will denote variables in the bootstrap world for which we do not have any corresponding observations in the real world.

4.1 Types and approaches

Bootstrapping is widely used and therefore we can find many modifications of this method. Basic types and approaches are presented here. More detailed description of some of their combinations for clustered data can be found in Field and Welsh (2007) together with properties of some estimates.

4.1.1 Paired vs. residual bootstrap

Basically there are two types of bootstrap: paired bootstrap and residual bootstrap, both with many adjustments. In the first case we resample the obser-

vations to make inference about them. In the second case we use the residuals of the model applied to the data for resampling process. In both cases we assume that the sample for bootstrapping contains observed values of independent and identically distributed random variables.

4.1.2 Parametric vs. nonparametric bootstrap

There are two approaches: parametric and nonparametric. In the first case we can specify a probability model for the data, in the second case it is not possible or desirable. In the parametric analysis we work with a distribution for the data which is determined by appropriate parameters. The nonparametric analysis uses the data directly through the empirical distribution. The empirical distribution function (EDF) for a random variable y is defined by

$$\hat{F}(y) = \frac{1}{n} \sum_{j=1}^n I\{y_j \leq y\},$$

which means that all sample elements have the probability n^{-1} . So unless some value occurs in the sample multiple times it has the same probability as other values.

A big attention must be held to outliers in the case of nonparametric bootstrap, because if we do not omit them, they are used in the simulated sample with the same probability as other values and especially when they are used repeatedly, they can greatly influence the approximate distribution. We should omit them from the simulation or smooth the original sample to reduce their impact.

4.2 Approximations

Relationship between the real world and the bootstrap world is described in Davison and Hinkley (1997). Generally considering a sample y_1, \dots, y_n which comes from independent and identically distributed random variables Y_1, \dots, Y_n we are interested in some population characteristic presented by parameter θ . This parameter is investigated through a statistic T . For example the inference about the mean of a distribution $\theta = \mu$ is made through the sample average $T = \frac{1}{n} \sum_{j=1}^n Y_j = \bar{Y}$. For this statistic we have an observed value t from the observed sample y_1, \dots, y_n , in the case of the mean we have $t = \frac{1}{n} \sum_{j=1}^n y_j = \bar{y}$. This value is in fact calculated from the EDF.

We describe the relationship between the characteristic θ and the population distribution using the relationship between its estimate t from the observed data and a fitted distribution, either EDF or cumulative distribution function (CDF) with estimated parameters. When a fitted distribution converges to the population distribution for increasing number of observations, $n \rightarrow \infty$, also the statistic T converges to characteristic θ from the property of consistency, unless t does not converge to t for $n \rightarrow \infty$.

When using bootstrapping we resample from the fitted distribution, either EDF or CDF with estimated parameters. We want to get Y_1^*, \dots, Y_n^* with independent variables. From simulated data we then calculate the statistic T^* . After B simulations we have B values T_1^*, \dots, T_B^* whose relationship with t we can use

to infer about properties of the relationship between T and θ . Through this procedure we will get approximate values which converge to the exact values with B increasing from the law of large numbers.

If we want to approximate distribution or quantiles of this distribution we use ordered values of the desired statistic calculated from B simulation samples. When we draw these values we can see the approximate distribution.

4.3 Bootstrapping in claims reserving

At first we need to find an appropriate type of bootstrap for claims reserving. Bootstrapping requires independent and identically distributed observations for the resampling process. Thus the paired bootstrap is not suitable for claims reserving because although the data are supposed to be independent we assume that the parameters of the distribution depend on the model covariates.

Therefore we work with the residual bootstrap in claims reserving which is based on some model for the data, usually GLM or some other model, often connected with the chain ladder estimates. Residuals are assumed to be independent and in some cases also identically distributed. When they are not, they can be adjusted to have approximately the same distribution. Various types of residuals used in actuarial literature are presented in Section 4.5.

4.4 The bootstrap process

The process of bootstrapping for claims reserving is described here. We have a data in a form of the triangle as in Table 2.1. We fit the GLM (or possibly other model) to the data so we get the fitted values. Using these values, the data and possibly other values connected with the model (see the next section) we can calculate the residuals needed for the resampling process.

Then we resample the residuals with replacement B times so we get B simulated triangle sets of residuals where B is specified with respect to the subject of investigation. When the aim is to get an approximate distribution of an estimate B must be much bigger to get results with sufficient precision, e.g. $B = 10000$, than in the case of investigating just the mean value or some other statistics, e.g. $B = 1000$ should be enough in that case.

From these simulated residual sets we calculate the simulated sets of past claims. For each bootstrap past triangle we use the same form of GLM as in the case of original data again to get to the bootstrap triangle of predicted future values from which the bootstrap reserve estimates are calculated. Desired statistics are obtained from this set of B reserve estimates whose ordered values make also the predictive distribution.

4.5 Residuals

There are several types of residuals which were used for residual bootstrap in the context of claims reserving. The most common ones are the Pearson residuals, clearly because they are well-known. But many authors tried to use

also other different types of residuals like the deviance residuals or the Anscombe residuals. Some types of standardizations or adjustments are also recommended.

4.5.1 Pearson residuals

Using the notation from previous chapters, the Pearson residuals are defined by

$$r_{ij}^P = \frac{C_{ij} - \hat{\mu}_{ij}}{\sqrt{V(\hat{\mu}_{ij})}}.$$

But these residuals have generally different variances so we need some adjustments to make them approximately equally distributed. There was suggested multiplying by a correction factor

$$r_{ij}^{Padj} = \sqrt{\frac{n}{n-p}} r_{ij}$$

or the standardization using the hat matrix

$$r_{ij}^{Pstand} = \frac{r_{ij}}{\sqrt{\hat{\phi}(1 - h_{ij})}}, \quad (4.1)$$

where h_{ij} is an element of the hat matrix \mathbf{H} which is given for GLMs by

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}.$$

Here \mathbf{X} is the model matrix of GLM and \mathbf{W} is the diagonal matrix with

$$w_{ii} = (V(\mu_{ij}) \left(\frac{\partial \eta_{ij}}{\partial \mu_{ij}} \right)^2)^{-1}$$

on the diagonal which can be rewritten by

$$w_{ii} = \mu_i^{2-p}$$

with $p = 1$ for the ODP and $p = 2$ for the gamma model (see Pinheiro et al. (2003)).

4.5.2 Deviance residuals

As mentioned in Hartl (2010) the standardized Pearson residuals have a distribution with the mean value equal to zero which means that there are lots of negative residuals which may cause also negative values in created pseudo-triangles of past claims. Especially for small expected values of these claims negative values are almost guaranteed in each pseudo-triangle which is inconsistent with our assumptions.

Therefore the deviance residuals could be used instead of the Pearsons residuals. Their use is appropriate in some cases where the Pearson residuals make negative values in expected claims, but it does not work in all cases. They are defined by

$$r_{ij}^D = \text{sign}(y - \hat{y}) * \sqrt{2 \int_{\hat{y}}^y \frac{y-t}{V(t)} dt}.$$

Nevertheless this function has not a simply expressible inverse. A numerical approach to find this inverse for a special case of a distribution with identity variance function is described in Hartl (2010).

4.5.3 Anscombe residuals

As noted in Tee et al. (2017) another good choice of residuals could be the Anscombe residuals which can avoid the skewness of the distribution of Pearson residuals in non-normal cases which then do not have required properties. The Anscombe residuals work with a transformation $A(C_{ij})$ of the variable C_{ij} instead of this variable itself. This transformation tries to create a variable with a distribution close to normal one. For GLM the Anscombe residuals are defined by

$$r_{ij}^A = \frac{A(C_{ij}) - A(\hat{\mu}_{ij})}{A'(\hat{\mu}_{ij})\sqrt{V(\hat{\mu}_{ij})}}.$$

For the case of Poisson distribution we have

$$r_{ij}^{Aodp} = \frac{\frac{3}{2}(C_{ij}^{\frac{2}{3}} - \hat{\mu}_{ij}^{\frac{2}{3}})}{\hat{\mu}_{ij}^{\frac{1}{6}}}.$$

and for the gamma distribution it is

$$r_{ij}^{Ag} = 3\left(\left(\frac{C_{ij}}{\hat{\mu}_{ij}}\right)^{\frac{1}{3}} - 1\right).$$

Chapter 5

Regulatory requirements

5.1 Solvency II

In 2009, European parliament and council have issued the Directive on the taking-up and pursuit of the business of Insurance and Reinsurance called Solvency II which all insurance and reinsurance companies operating in EU are required to follow. It defines own funds that insurance companies are obligated to hold in order to cover the Solvency Capital Requirement (SCR) for facing the risks. All possible quantifiable risks together with correlation coefficients have to be considered in the calculation of SCR, see Ohlsson and Lauzenings (2009), AISAM-ACME (2007).

For each module Value-at-Risk (VaR) measure with a 99,5% confidence level over a one-year period has to be used but the way of obtaining the probability distribution is not specifically given, see The European Parliament and the Council of the European Union (2009). This one year period is a big change from the previous ultimate point of view. That is why previously used methods had to be modified or replaced by new ones in order to meet this requirement.

SCR is calculated as VaR of Claims development result (CDR) on a confidence level of 99.5%. As presented in Merz and Wüthrich (2008) CDR at time $I + 1$ can be estimated for each accident year $i \in \{1, \dots, I\}$ by

$$CDR_i(I + 1) = \hat{R}_i^{I+1} - (X_{i,I-i+2} + \hat{R}_i^{I+2}) = \hat{C}_{i,I+u}^{I+1} - \hat{C}_{i,I+u}^{I+2}, \quad (5.1)$$

where upper indexes denote the year of calculation.

5.2 Modifications of bootstrap procedure for one year view

As already said, Solvency II specifies new time horizon for calculation of capital requirements which is one year. A possibility how to meet this requirement is modification of previously used methods in ultimate time horizon. There are two types of modifications of the bootstrap method presented in Boisseau (2006) and Boumezoued et al. (2011). The first one is more intuitive, the second one has some advantages in reducing errors.

5.2.1 First modification

One possible modification of the bootstrap procedure to meet the requirement of one year horizon is to use more steps in each bootstrap loop: firstly to create new subdiagonal of the starting triangle only, then to predict the next development. Using this we get the elements of definition of observable CDR: an estimate of claims reserves at time $I + 1$ (the beginning of the year), estimate of future payments in the next year and an estimate of claims reserve at time $I + 2$ (the end of the year $I + 1$).

Firstly we fit a GLM to the run-off triangle to get the estimates of parameters from which we obtain the estimates of the claims reserve at time $I + 1$ and we also calculate the residuals of the model. Then we resample these residuals with replacement B times to get B simulated triangles of residuals which we use for calculation of pseudo-triangles of past claims. For each such triangle we fit the GLM again in order to construct the subdiagonal of expected claims $\mu_{ij}^*, i + j = I + 2$. Incremental payments are generated using these values as the mean values of the distributions with the variances defined by $\hat{\phi}(\mu_{ij}^*)^p$.

Then the whole trapezoids (the triangles and the subdiagonals) are used in GLM fitting to get new estimates of parameters for calculation of new expected values of claims for the rest of the development. Then the sums are obtained representing the estimates of claims reserve at time $I + 2$.

5.2.2 Second modification

The first method uses GLM two times in each bootstrap loop which causes very high estimation error. Thus also other modification of the bootstrap procedure with GLMs was developed for one year horizon which tries not to increase the estimation error where it is not necessary. Another limitation of the first method is also the dependency of the subdiagonal on the upper triangle. So the random variables on the trapezoid used in GLM are not independent. The second approach tries to improve this too.

The biggest change compared to the first approach is that after the estimation of GLM parameters of the past triangle we resample the residuals B times on the whole trapezoid, meaning the upper triangle and the subdiagonal. From that we can calculate an estimate of claims reserve at time $I + 1$ and moreover the future payments in the next year. Then the GLM is used to estimate new parameters needed for calculation of an estimate of claims reserve at time $I + 2$. So we have used the resampling and the GLM in each iteration of the bootstrap only once which does not increase the estimation error and which does not create dependencies in pseudo-data used for estimation.

Chapter 6

Application

The theory described in previous chapters was applied to a data set and the results together with some description are presented in this chapter. All calculations have been made in R software and all graphical illustrations shown here are also outputs from this program.

6.1 Data

Data chosen for this study is originally from Zhang (2010). The advantage is that this data is a component of an R package called ChainLadder so it is freely available in the software used for calculations and anyone can use it again in the same form without any changes. From three available run off triangles the triangle of paid losses in personal auto insurance is chosen (see Table 6.1). The data is suitable for this work because it contains some trends in development and accident periods as will be shown and studied in subsequent sections.

6.2 GLM

Firstly GLM was applied to given data. Logarithmic link function was used which is preferred for claims reserving. The chain ladder linear predictor (3.5) was chosen for its frequent use and assumed sufficiency. It means the model

$$\log[E(C_{ij})] = c + \alpha_i + \beta_j$$

for $i, j \in \nabla$ with condition $\alpha_1 = \beta_1 = 0$ was estimated. Only two values $p = 1$ and $p = 2$, which are the most frequently used in claims reserving, were considered for the model specification in (3.4). From these two possibilities the first one seems to be better. It was concluded from residual graphs where the standardized Pearson residuals (4.1) are plotted against fitted values and origin and development period. These plots are shown at this work only for $p = 1$ (see Figure 6.1). Here the residuals are more homoscedastic and more or less evenly distributed around zero. Another reason for choosing the ODP distribution is lower value of the function (3.12) with equal weights, where lower value should be preferred for estimation of p .

From now on only the case of $p = 1$ is considered. Estimated total reserve calculated as the sum of expected future claims from the model with mentioned setting is written in Table 6.3 for comparison with other results.

$i \setminus j$	1	2	3	4	5	6	7	8	9	10
1	101125	108796	56697	38489	22743	12819	7761	2763	2160	231
2	102541	100672	57464	42505	25750	12016	6385	2480	710	
3	114932	112772	70416	47422	22218	10239	5612	1613		
4	114452	113309	73311	39597	19310	9269	4077			
5	115597	128014	71604	39275	17886	10362				
6	127760	131656	67559	38805	20945					
7	135616	126678	64792	40271						
8	127177	117072	73723							
9	128631	118172								
10	126288									

Table 6.1: Run-off triangle of personal auto paid claims

$i \setminus j$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	109154	108435	62447	37765	21383	10951	5233	2377	1039	440	182	74	29	12	4
2	110108	109383	62993	38095	21570	11047	5279	2398	1048	444	184	74	30	12	5
3	113207	112462	64766	39168	22177	11358	5428	2466	1077	457	189	77	30	12	5
4	115850	115087	66278	40082	22695	11623	5554	2523	1103	467	193	78	31	12	5
5	118333	117553	67698	40941	23181	11872	5674	2577	1126	477	197	80	32	13	5
6	120755	119960	69084	41779	23656	12115	5790	2630	1149	487	201	82	33	13	5
7	123162	122351	70461	42612	24127	12357	5905	2682	1172	497	205	83	33	13	5
8	125574	124747	71841	43446	24600	12599	6021	2735	1195	507	209	85	34	13	5
9	128005	127162	73232	44288	25076	12843	6137	2788	1218	516	214	87	34	14	5
10	130463	129604	74638	45138	25558	13089	6255	2841	1242	526	218	88	35	14	5

Table 6.2: Expected values of personal auto paid claims

6.3 Smoothing of the parameters

Estimated coefficients for original and development period can be smoothed and in the case of development period also extrapolated to get a tail values of expected claims. Two possibilities of smoothing of the model coefficients are considered here.

6.3.1 Simple functions

The first possibility for smoothing is to use some simple functions $f_\alpha(x)$ and $f_\beta(x)$ in models

$$\mathbb{E} \alpha_i = a_0 + a_1 * i + a_2 * f_\alpha(i),$$

$$\mathbb{E} \beta_j = b_0 + b_1 * j + b_2 * f_\beta(j).$$

Different types of functions were tried to smooth the particular coefficients alpha or beta, see the Figure 6.2. For the origin period coefficients we are not interested in extrapolation and prediction of a tail so we prefer just the best fit of existing values. Three of the smoothing curves of the coefficients alpha have almost identical shape. They include the logarithm, the reciprocal function and the square root. For these three fits the residual standard errors do not differ till the fourth position after the decimal point. The lowest value is found for the reciprocal function, but it can be expected that other conditions like smoothing the beta coefficients will have greater influence on the final results and volatility.

In the case of the development period coefficients we are interested in smoothing especially the latest development years, in extrapolation and tail prediction. The coefficients corresponding to the first development years may be kept unchanged since they can have different development then the rest of the parameters and their estimates are based on more observations so they are not so affected by outliers. The same functions as in the case of the coefficients alpha were tried to smooth these parameters, see Figure 6.3. It can be seen that there is no break or deviation even for the first years. But when comparing the plots with different numbers of unchanged coefficients, subjectively the best choice is to keep estimates for the first three or four development years. Smoothing with the first three coefficients unchanged is shown in Figure 6.3. Other possibilities mean too heavy or on the contrary almost no tail. Options with the reciprocal function, the logarithm and the square root are again very similar which is also proved by similar values of the standard error of prediction for future five development years. These values prefer the reciprocal function again but we will focus again more on the volatility of final results.

6.3.2 Linear functions

The second possibility for smoothing used in this work is the reparameterization described in Björkwall et al. (2011) and contained in Chapter 3.6.3 of this work. Different options of numbers of unchanged first coefficients were tried for both alpha and beta coefficients. It does not seem very meaningful to smooth a greater number of original period parameters with a linear line since this curve does not seem to smooth the data very well. A better option is to choose larger

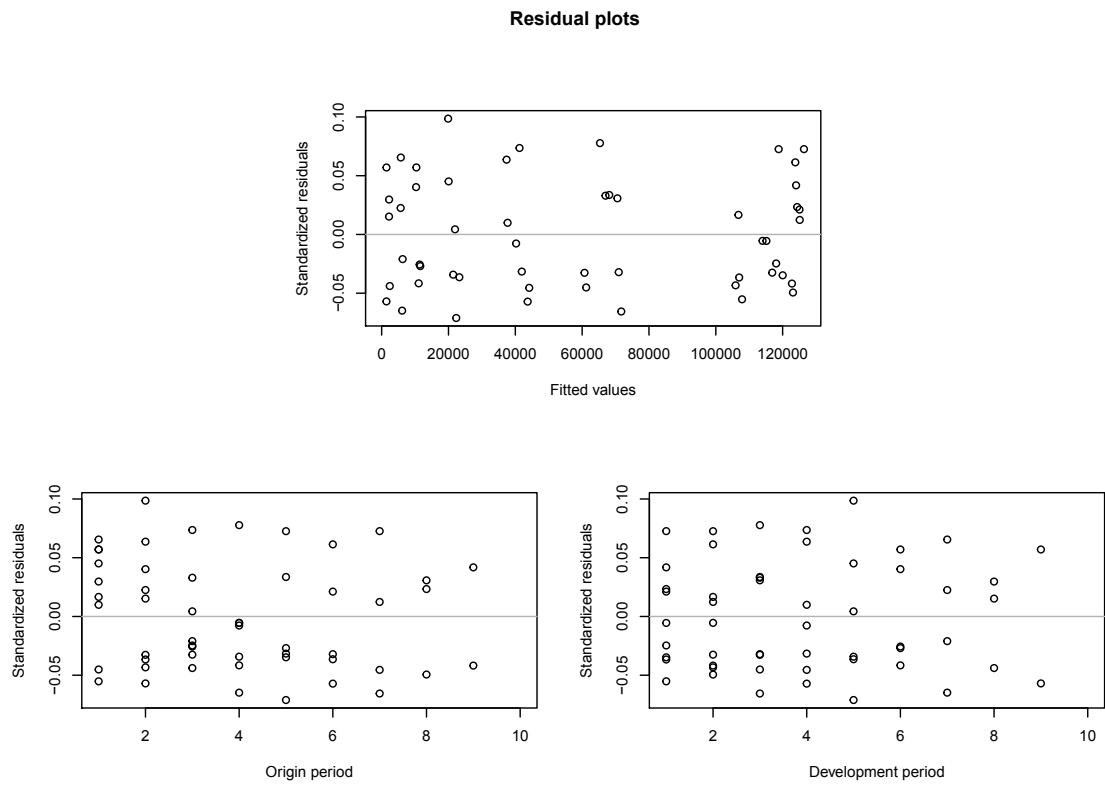


Figure 6.1: GLM Residual plots

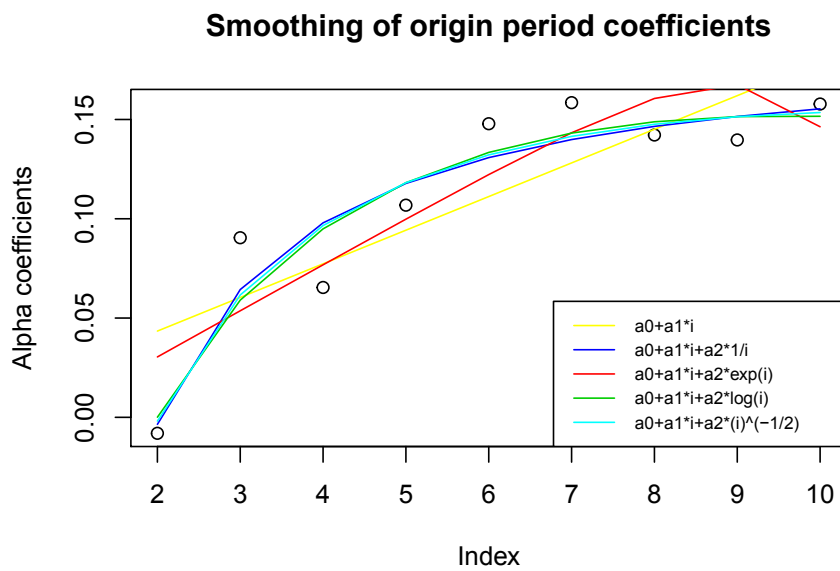


Figure 6.2: Smoothing of alpha coefficients using simple functions

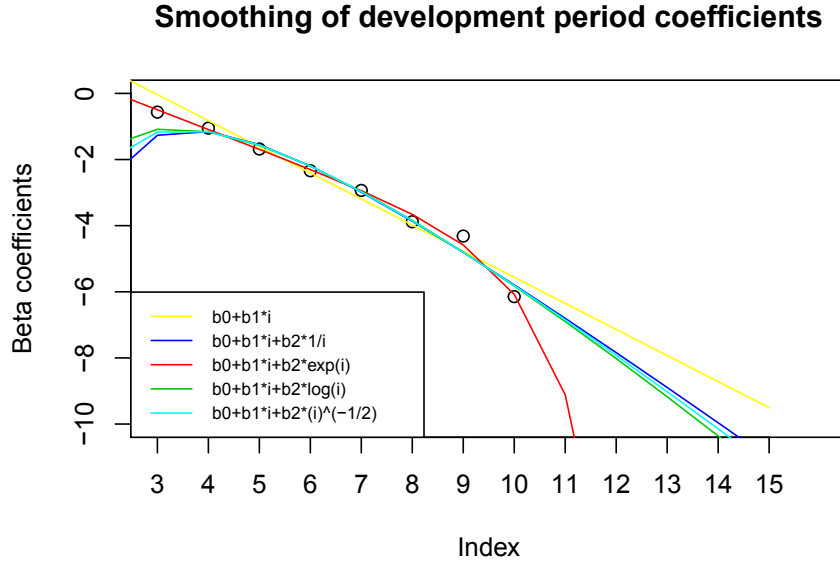


Figure 6.3: Smoothing of beta coefficients using simple functions with extrapolation

number of the parameter q in order to keep more coefficients unchanged, see Figure 6.4 where the model with the same smoothing of development coefficients was used with the number of unchanged coefficients r chosen to be the middle value of possible ones.

For the development coefficients it may makes much more sense to use linear functions to smooth the latest values and to extrapolate into the tail. For our data the parameters seem to decrease linearly, see Figure 6.5 with q for estimation of alpha coefficients chosen to be the middle value of possible ones. These lines provide a good fit for estimated coefficients except the last one, which might be seen as an outlier since we have only one observation for estimation of this parameter. Subjectively larger values of r should be preferred. The choice of over-dispersed Poisson distribution makes it impossible to use information criterions like AIC. The choice of the most appropriate reparameterization can be based on bootstrap estimation of mean squared error of prediction, see Section 3.7.2.

6.3.3 Reparameterization

For reparameterized GLM we need a matrix \mathbf{D} defined by (3.7), where matrices \mathbf{A} and \mathbf{B} can be given by (3.11) or (3.8) or matrix \mathbf{B} even by (3.9). This matrix \mathbf{D} is used in calculation of a new model matrix from the old one by

$$\mathbb{X} = \mathbb{X}^* \mathbf{D}. \quad (6.1)$$

More possible choices of reparameterizations of both original and development coefficients were tried given the conclusions from the previous section. Then the final model was chosen with respect to the overall results from bootstrapping meaning especially the volatility of estimated reserve.

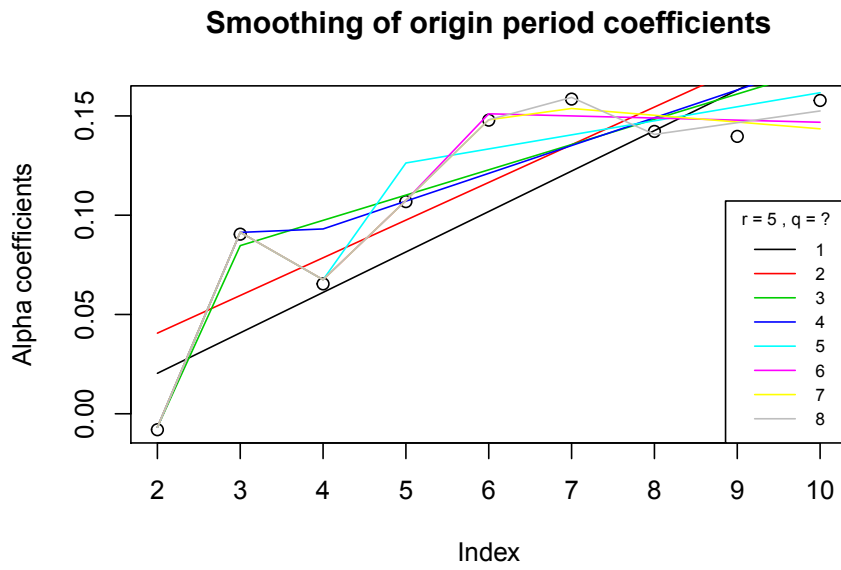


Figure 6.4: Smoothing of alpha coefficients using linear functions

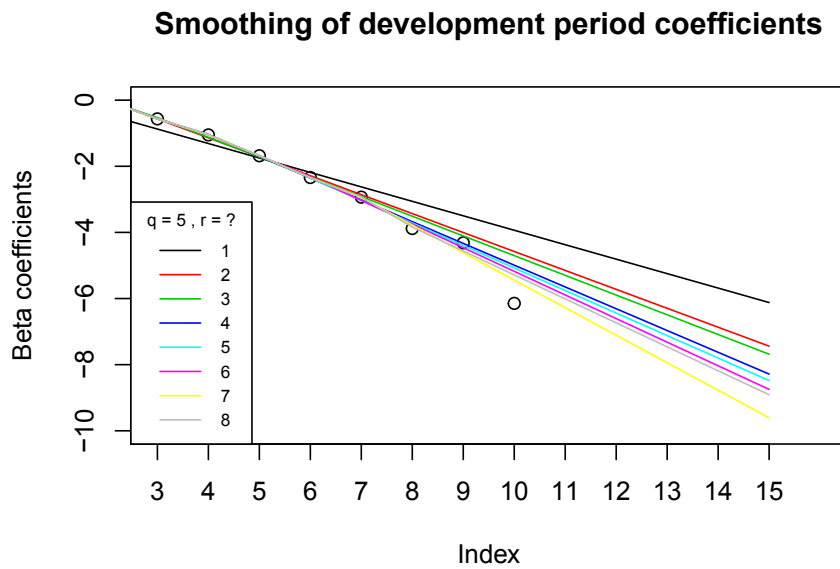


Figure 6.5: Smoothing of beta coefficients using linear functions with extrapolation

When focusing only on linear functions for smoothing of both sets of coefficient, $r = 4$ or 7 should be preferred for the development period coefficients as well as $q = 1, 3$ or 8 for the origin period coefficients with respect to volatility of final results. It includes also smaller values of q which was not expected. But these model produce too large MSEP and it is better to choose at least one set of coefficients to be smoothed by other simple functions. It turned out that the reciprocal function is really the best possibility for smoothing of the original period coefficients, as it was assumed earlier. For the development period coefficients the logarithm seems to be the best with respect to the volatility of estimated total reserve. MSEP connected with these models are much better. The best reparameterization of all which were considered in this work was obtained with the combination of the reciprocal function for smoothing alpha coefficients and the logarithm for smoothing beta coefficients with the first three original ones (with five extrapolated tail values) which means we use the matrices

$$\mathbf{A} = \begin{pmatrix} 1 & \dots & I \\ 1 & \dots & 1/I \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 4 & \dots & 15 \\ 0 & 0 & \log(4) & \dots & \log(15) \end{pmatrix}.$$

6.4 Reparameterized GLM

New GLM with new model matrix reparameterized by equation (6.1) was applied to original data with other settings being the same as in the original model, it means with the logarithmic link function, the chain ladder linear predictor and $p = 1$ which was concluded as better choice with the first model estimation. Only one other change is inclusion of the next five development years into estimation in order to get results for unobserved tail. Expected claims for all origin and development years are presented in Table 6.2 for comparison with original data. The estimates of the mean reserve with and without the tail values (for better comparison with previous result) are written in Table 6.3. New plots of the standardized Pearson residuals against fitted values and origin and development period do not show any strong violations from requirements (see Figure 6.6) so the model is accepted as suitable for given data.

6.5 Bootstrapping

Residual bootstrap was used to obtain predictive distribution of claims reserve. The most common standardized Pearson residuals were chosen for this purpose. Standardization presented in equation (4.1) may cause a problem with non-positive values in bootstrapped pseudotriangles of past claims for which we can not use the logarithmic link function. This issue is solved by setting the value 1 instead of these values. For our data these issues do not occur very often and absolute values of these cases are not so high (it happens often on places of

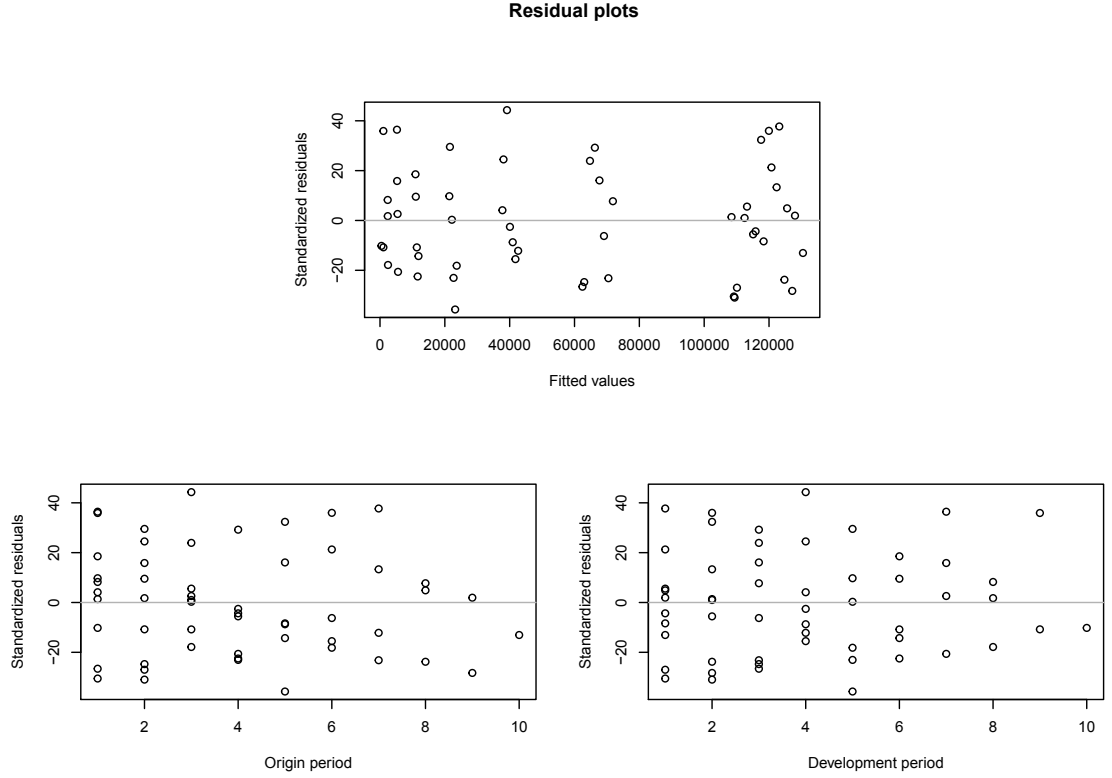


Figure 6.6: Reparameterized GLM Residual plots

lower values in the run-off triangle) so it does not increase the expected reserve so significantly like non-standardized residuals decrease it.

We need to obtain a full predictive distribution in order to get the standard deviation as an estimate of MSEP as a measure of volatility and in order to get VaR. That is the reason why higher number of bootstrap iterations was used, specifically $B = 10000$. The value of expected claims reserve is presented together with other earlier mentioned estimates in Table 6.3 and also together with some characteristics of this estimate in Table 6.4. The histogram and the EDF of claims reserve are plotted in Figure 6.7.

Original	624246.8
Reparameterized	640930.7
Rep. with a tail	644227.9
Bootstrap mean	644524.1

Table 6.3: Expected claims reserve

6.6 Solvency capital requirement

SCR is calculated as VaR on 99.5% level of CDR defined by (5.1). The second modification of bootstrap method for one year view described in Section 5.2.2 was chosen for estimation of CDR due to its better properties. The non-standardized Pearson residuals were used for residual bootstrap since we do not have any

min	579940.5
1st kv.	631769.0
mean	644524.1
3rd kv.	656747.5
VaR(0.995)	691032.4
max	716828.0
sd	18280.5

Table 6.4: Estimated claims reserve characteristics

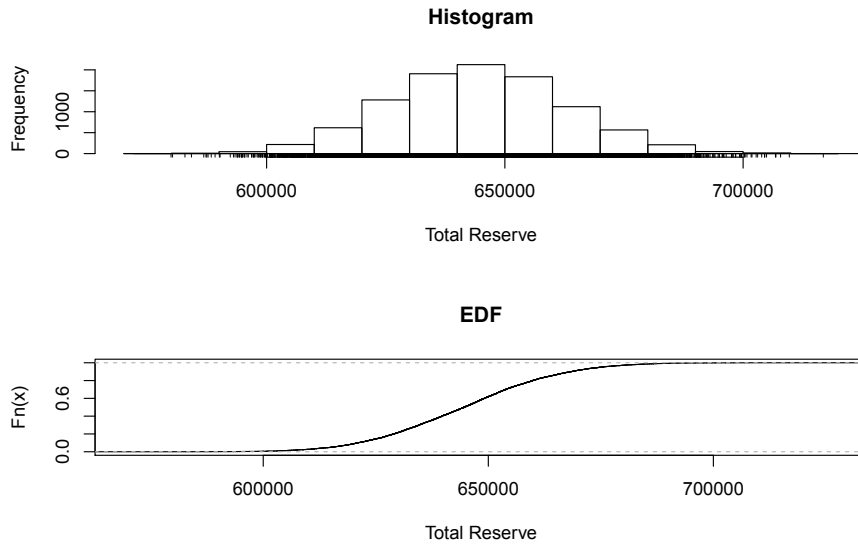


Figure 6.7: Histogram and EDF of claims reserve

values of the hat matrix for the cases on the subdiagonal. Number $B = 10000$ of bootstrap replicates was used in order to get predictive distribution which is plotted in Figure 6.8 together with the histogram of bootstrapped values which is very symmetric. Some characteristics of the predictive distribution are also shown in Table 6.5. SCR is equal to 111389.8.

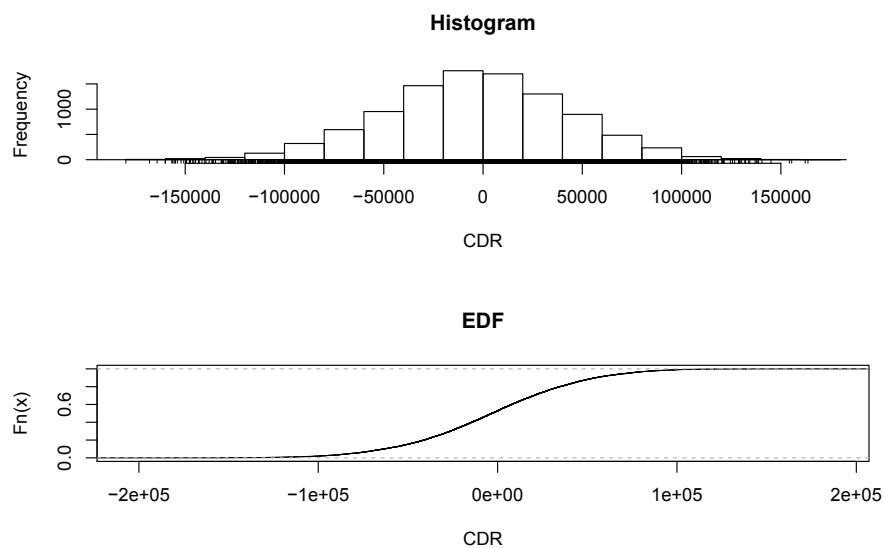


Figure 6.8: Histogram and EDF of CDR

min	-179896.4
1st kv.	-33390.4
mean	-3739.7
3rd kv.	26978.0
VaR(0.995)	111389.8
max	163649.7
sd	45842.6

Table 6.5: Estimated CDR characteristics

Chapter 7

Conclusion

The theory of the chain ladder method, GLMs and bootstrap method was summarized from different sources of information in this work together with some simple derivations for a purpose of its use in calculation of the best estimate of funds needed to hold by an insurance company to cover the risks. The data used here represent personal auto paid claims published in Zhang (2010) and available in R software. This dataset contains trends in claims development which was investigated too.

The over-dispersed Poisson distribution was chosen as the underlying distribution of given data based on comparison of residual graphs and calculated values of a function convenient for this comparison. Then the GLM with origin and development years as predictors, which in combination with chosen distribution gives the same estimates as the chain ladder method, was used to calculate expected future claims. Then the estimates of parameters, both for origin and development years, were smoothed in order to reduce the impact of outliers in the years with small numbers of observations and also to estimate the tail values of development period without any observations.

Reparameterization used in Björkwall et al. (2011) was used together with suggestions described there and also with other possible smoothing functions suitable for the data presented in this thesis. It turned out that the best model with respect to standard deviation of estimated reserve includes smoothing of the origin period coefficients using the reciprocal function and the development period coefficient using the logarithm with the first three original coefficients. Extrapolation of a tail of other five development years was included too.

Residual bootstrap with 10000 of replicates was used to obtain the predictive distribution of future claims together with standard deviation as a measure of volatility which was compared with other considered models. The 99.5-quantile was calculated as the estimate of own funds for covering potential losses. The bootstrap method was also extended to make it possible to estimate the claims development result with its predictive distribution and to calculate solvency capital requirement according to Solvency II requirements that came with one year view on reserving risk.

The results were presented in the form of tables and figures as outputs from the R software where all calculations were made. Some functions predefined in the software were used but some of them needed necessary adjustments or re-programming for this particular problem and a large part of them especially for

smoothing purposes and searching for predictive distribution of claims development result was defined as new functions. The best model according to used criteria and sometimes also according to subjective decisions was found which results should be suitable for managing the reserving risk.

Bibliography

- AISAM-ACME. Aisam-acme study on non-life long tailed liabilities. *Scandinavian Actuarial Journal*, 2007.
- S. Björkwall, O. Hössjer, and E. Ohlsson. Non-parametric and parametric bootstrap techniques for age-to-age development factor methods in stochastic claims reserving. *Scandinavian Actuarial Journal*, 2009(4):306–311, 2009.
- S. Björkwall, O. Hössjer, E. Ohlsson, and R. Verrall. A generalized linear model with smoothing effects for claims reserving. *Insurance: Mathematics and Economics*, 49(1):27–37, 2011.
- J. P. Boisseau. Solvabilité 2 et mesure de volatilité dans les provisions pour sinistres à payer. *Mémoire Centre D’études Actuarielles*, 0(0):0, 2006.
- A. Boumezoued, Y. Angoua, L Devineau, and J.-P. Boisseau. One-year reserve risk including a tail factor: closed formula and bootstrap approaches. *Working paper, HAL*, (00605329), 2011.
- A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, Cambridge, U.K., 1997. ISBN 0-521-57391-2.
- P. D. England and R. J. Verrall. Stochastic claims reserving in general insurance (with discussion). *British Actuarial Journal*, 8(3):443–518, 2002.
- C. A. Field and A. H. Welsh. Bootstrapping clustered data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3):369–390, 2007.
- T. Hartl. Bootstrapping generalized linear models for development triangles using deviance residuals. *Casualty Actuarial Society E-forum*, 2010.
- T. Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bull*, 23(2):213–225, 1993.
- P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Second Edition. Chapman and Hall, London, 1989. ISBN 0-412-31760-5.
- M. Merz and M. V. Wüthrich. Modelling the claims development result for solvency purposes. *Casualty Actuarial Society E-Forum*, pages 542–568, 2008.
- E. Ohlsson and J. Lauzeningks. The one-year non-life insurance risk. *Insurance: Mathematics and Economics*, 45(2):203–208, 2009.
- U. Olsson. *Generalized Linear Models*. Studentlitteratur, Sweden, 2002. ISBN 91-44-03141-6.

- P. J. R. Pinheiro, J. M. Andrade e Silva, and M. d. L. Centeno. Bootstrap methodology in claim reserving. *Journal of Risk and Insurance*, 70(4):701–714, 2003.
- A. E. Renshaw and R. J. Verrall. A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, 4(4):903–923, 1998.
- L. Tee, M. Käärik, and R. Viin. On comparison of stochastic reserving methods with bootstrapping. *Risks*, 5(2):0, 2017.
- The European Parliament and the Council of the European Union. *Directive 2009/138/EC of the European Parliament and of the Council*. The European Parliament and the Council of the European Union, 2009. URL <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02009L0138-20140523>.
- M. V. Wüthrich and M. Merz. *Stochastic claims reserving methods in insurance*. Willey, A, 2008. ISBN 0.
- Y. Zhang. A general multivariate chain ladder model. *Insurance: Mathematics and Economics*, 46(3):588–599, 2010.

List of Figures

6.1	GLM Residual plots	28
6.2	Smoothing of alpha coefficients using simple functions	28
6.3	Smoothing of beta coefficients using simple functions with extrapolation	29
6.4	Smoothing of alpha coefficients using linear functions	30
6.5	Smoothing of beta coefficients using linear functions with extrapolation	30
6.6	Reparameterized GLM Residual plots	32
6.7	Histogram and EDF of claims reserve	33
6.8	Histogram and EDF of CDR	34

List of Tables

2.1	Run-off triangle of incremental data which can be observed	4
2.2	Target triangle of incremental data which is the subject of estimation	5
6.1	Run-off triangle of personal auto paid claims	26
6.2	Expected values of personal auto paid claims	26
6.3	Expected claims reserve	32
6.4	Estimated claims reserve characteristics	33
6.5	Estimated CDR characteristics	34