
Technische Universität München
Department of Mathematics


# Statistical Methods for the Automatization of Basic Loss Model Calibration

Master's Thesis

by


Kimberly Konrad


in cooperation with Munich Reinsurance Company

| | |
|---|---|
| Supervisor: | Prof. Dr. Claudia Czado |
| Advisors: | Prof. Dr. Claudia Czado, Alexander Kreuzer |
| External Advisor: | Dr. Ralph Schuster (Munich Re) |
| Submission Date: | 30.09.2019 |

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

Munich, 30.09.2019

# Abstract

In this thesis we provide a method to detect trends and change points in the loss triangles of basic loss portfolios in order to ensure an appropriate assessment of the claims reserve and the premium risk and reserve risk based on these data. The adequate determination of these key financial figures is one of the most important issues of an insurance company and if such structural changes are disregarded that may distort the results.
We develop an approach build upon the theory of linear models and change point analysis to solve this problem. The aim is to identify structural changes based on statistical methods to automatize the calibration of basic loss models.
First of all, we present the necessary fundamentals and the framework of the basic loss model in the reserving and the risk context. Then, we bring these parts together and apply the considerations to the claims data to detect structural changes. We use two commonly used reserving models as starting point, integrate the possibility of trends and change points and choose the best model in terms of model fit and complexity. Finally, we verify this approach by a simulation study and show the achieved results.

# Zusammenfassung

In dieser Arbeit liefern wir eine Methode Trends und Change Points in den Schadendreiecken von Basisschadenportfolios zu erkennen, um eine angemessene Bewertung der Schadenrückstellung und des Prämien- und Reserverisikos auf der Grundlage dieser Daten zu gewährleisten. Die adäquate Bestimmung dieser Finanzkennzahlen ist eines der wichtigsten Themen eines Versicherungsunternehmens und werden solche strukturellen Veränderungen nicht berücksichtigt, kann das die Ergebnisse verzerren.
Wir entwickeln einen Ansatz basierend auf der Theorie von linearen Modellen und Change Point Analyse, um dieses Problem zu lösen. Ziel ist es, strukturelle Veränderungen auf der Grundlage statistischer Methoden zu identifizieren und damit die Kalibrierung von Basisschadenmodellen zu automatisieren. Zunächst werden die notwendigen Grundlagen und der Rahmen des Basisschadenmodells im Reservierungs- und Risikokontext dargestellt. Dann fügen wir alles zusammen und wenden die Überlegungen auf die Schadendaten an, um strukturelle Veränderungen zu erkennen. Wir verwenden zwei gängige Reservierungsmodelle als Ausgangspunkt, integrieren die Möglichkeit von Trends und Change Points und wählen das beste Modell in Hinblick auf Modellanpassung und Komplexität. Abschließend verifizieren wir diese Methode durch eine Simulationsstudie und zeigen die erzielten Ergebnisse.

# Acknowledgments

First of all, I would like to express my deepest gratitude to Prof. Claudia Czado, Ph.D. for making it possible to write this thesis. I want to thank her for her helpful advice, encouragement and excellent supervision during the last months. I would also like to thank Alexander Kreuzer for his support and the time he invested into advising me.
I am very grateful to Munich Re for the cooperation and providing the exciting topic. Especially to Dr. Ralph Schuster I want to express my infinite gratitude for his engagement, expertise and great support.
Last but not least, I thank my family and my boyfriend for their endless support during and before this thesis.

# Contents

# Notation

Within this thesis we use the following notation and abbreviations:

| | |
|---|---|
| $E[Y]$ | expected value of random variable $Y$ |
| $Cov[Y, Z]$ | covariance of random variables $Y$ and $Z$ |
| $Var[Y]$ | variance of random variable $Y$ |
| $\Theta$ | parameter set |
| $dim(\Theta)$ | number of parameters in $\Theta$ |
| $S(\ \cdot\ )$ | sum of squared residuals |
| $\mathcal{N}(\mu, \sigma^2)$ | normal distribution with mean $\mu$ and variance $\sigma^2$ |
| $\mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$ | multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$ |
| $L(\ \Theta \mid \boldsymbol{y}\ )$ | likelihood function |
| $l(\ \Theta \mid \boldsymbol{y}\ )$ | log likelihood function |
| $\psi(\cdot)$ | mean function |
| $AIC(\cdot)$ | Akaike information criterion |
| $BIC(\cdot)$ | Bayesian information criterion |
| $PL(\cdot)$ | penalized likelihood |
| | |
| $UY$ | underwriting year |
| $DY$ | development year |
| $S_{i,k}$ | incremental losses of UY $i$ in DY $k$ |
| $C_{i,k}$ | cumulative losses of UY $i$ in DY $k$ |
| $v_i$ | premium of UY $i$ |
| $R_i$ | reserve in underwriting year $i$ |
| $R$ | total reserve |
| $M_{i,k}$ | incremental loss ratio of UY $i$ in DY $k$ |
| $F_{i,k}$ | development factor of UY $i$ in DY $k$ |
| $m_k$ | expected loss ratio in DY $k$ |
| $f_k$ | expected development factor in DY $k$ |
| $s_k^2$ | additive variance parameter in DY $k$ |
| $\sigma_k^2$ | multiplicative variance parameter in DY $k$ |
| $\mathcal{D}$ | given loss triangle |
| SCR | Solvency Capital Requirement |
| CDR | claims development result |
| $mse(\hat{Y})$ | mean squared error of prediction of random variable $Y$ |

# List of Figures

# List of Tables

# Chapter 1

# Introduction

*If you want to know the future, look at the past.*

Albert Einstein

In every field of science you should understand the past to predict the future and this concept also plays an important role for insurance companies.
Insurance companies can be classified on the basis of the respective insurance product in life insurance, health insurance and non-life insurance (see e.g. Chapter 16 Cipra, 2010). But the principle is always the same, people buy an insurance policy for a fixed premium to protect against an unknown future loss and the insurance company guarantees in return to pay the covered claim amount in case a loss happens. That's why all insurance companies aim to predict claims as accurately as possible. This involves on the one hand claims of future business for pricing purposes and on the other hand of the current business for reserving purposes. Due to the differences in insured events the used methods and models to predict the claims differ for all classes of insurance and we focus on a non-life insurance company in this thesis.
The claims settlement in non-life business can take many years. Consequently, the insurance company is obliged to hold a claims reserve in order to be able to pay all outstanding claims, which are not yet fully settled. This reserve is an important position on the liability side on the balance sheet of the company and accordingly has significant impact on the company's equity (Forte et al, 2012). It is therefore subject to financial regulation and has high practical relevance for insurance companies, which results in a wide range of actuarial literature with various methods to calculate the reserves. For a summary of important stochastic claims reserving models for non-life insurance we refer for example to England and Verrall (2002) or Wüthrich and Merz (2008).
Besides the assessment of these expected future obligations, it is also required to quantify the uncertainty of this prediction (Merz and Wüthrich, 2008). Reason for that are the regulatory requirements for insurance companies by Solvency II. This supervisory regime came into effect in 2016 and specifies, among other requirements, a risk-based capital the insurers have to hold to reduce the risk of insolvency, called Solvency Capital Requirement (SCR) (BaFin, 2016). In order to calculate the SCR insurance companies can choose between using the standard formula as a one-size-fits-all solution or an own internal model

that is better adapted to the company. Following the Solvency II Directive (2009), the SCR has to comprise different risk categories and one of them is the non-life underwriting risk. This risk consists of catastrophe risk, which is arising out of losses related to extreme or exceptional events, and premium risk and reserve risk resulting from fluctuations in the claims settlement. Reserve risk quantifies the risk of insufficient reserves, whereas the premium risk comprises the risk that the premium is not sufficient for the upcoming year (Ohlsson and Lauzeningks, 2009). In this thesis, we do not consider catastrophe losses but concentrate on reserves as well as premium and reserve risk from non-life business caused by smaller and frequent losses, which we call basic losses.

The underlying assumption in the calibration of a basic loss model is that the future loss development follows a similar pattern compared to past losses. The question arises, what happens in case this assumption does not hold because of a structural change in the claims history of the portfolio. A structural change in the loss development can either appear as gradual change in form of a trend in data or an abrupt structural break, where the loss pattern before and after the break is not consistent. In practice such changes are omnipresent.

A break in a data set is called change point and the detection of change points has become increasingly important in modern statistics due to a wide range of application fields (see e.g. Eckley et al, 2011). Using data that is not representative for the future observations to estimate and predict can lead to a bad performance of the prediction model.

In the claims prediction process bad predictions can cause serious consequences for the insurance company. In case of an underestimation of the reserve or the risk capital the insurance company may not be able to settle the claims, the ruin probability of the company is higher than pretended and the solvability of the insurance company is not guaranteed anymore. On the other hand an overestimation of the reserve leads to an overestimation in liabilities and results in an underestimation of the profit and the equity (Dimitrijevic, 2015). And an overestimation of the risk capital leads to a dispensable stressed solvency situation.

The objective of this thesis is to provide an approach to detect structural changes in the claims data of basic loss portfolios to ensure an adequate assessment of the claims reserve and the premium and reserve risk. Currently, changes and breaks are qualitatively judged based on portfolio know-how, but objective analytical methods are not yet established. The results of this thesis provide the possibility to optimize the basic loss model calibration with the help of an automatized approach to detect structural changes.

Based on the theoretical foundation of linear regression models and change point analysis we propose two methods to analyze the loss history for structural changes and evaluate whether the data is disturbed by trends or change points or not. The first one serves as statistical tool to analyze the loss pattern step by step, but to automatize the model selection we also present a second method. The idea is to formulate the loss development of a basic loss portfolio in form of linear regression models and extend it by integrating structural changes. By allowing for trends and change points in the model we will get several model candidates exhibiting different model complexity. We will choose the most appropriate model and thus identify change points, trends and an appropriate model calibration considering the structural breaks in data history.

The change point detection results in identification of the part of the data that is rep-

resentative for the future loss development and consequently can be used to predict the future claims adequately.
The thesis is organized as follows. In Chapter 2 linear regression models together with two fundamental parameter estimation methods are introduced. We start with a classical linear regression model, before we relax the assumptions of constant variances and consider a weighted regression model. Furthermore, we present two special cases and information criteria, which will be used later to select the optimal model. This chapter builds the theoretical basis of the approach we will propose.
Chapter 3 begins with a short introduction to loss reserving and explains the common representation of historical claims data in the structure of a loss triangle. We proceed and present two commonly used loss reserving models, which provide the basis for later extension by change points and trends. To do this we show that both models can be expressed in form of a weighted linear regression model. In the last section of this chapter we discuss the selection of the optimal model to predict the reserve and quantify the prediction uncertainty, which contains the analysis of the loss history. At this point we explain possible reasons for structural changes in the claims data to give a motivation for the detection of changes and trends. This finalizes the framework of the basic loss model calibration in the reserving context.
Then, in Chapter 4 we introduce the requirements of Solvency II for the premium and reserve risk and show, how these can be linked to the previously discussed reserving model in an internal model context. This yields the setting of the basic loss model calibration in the SCR context.
Thereafter, we outline the theory of change point analysis in Chapter 5 starting with the problem formulation. We present a solution to this problem together with two algorithms, which can be used for an efficient computation. This provides a useful inspiration for the approach we will apply to detect change points and trends. But before we develop this approach in the next chapter we have to state the penalized likelihood method and formally define the change point problem in a linear regression model.
Finally, in Chapter 6 we present the application on the observed loss data. We develop a method to analyze the loss history development year by development year, i.e. step by step, to investigate whether the losses show a similar loss pattern or trends and change points appear in the loss development. After that, we propose a closely linked method for the selection and estimation of an overall basic loss model.
In Chapter 7 we provide a simulation study to evaluate the proposed method for model selection and estimation. We consider different scenarios and validate the detected change points. Additionally, we investigate the accuracy of the predicted reserve based on the selected model especially compared to the predicted reserve with the original reserving method. By this procedure we can also quantify the consequences of disregarding trends and change points in the past for the prediction of future losses.
The last chapter concludes the thesis summing up all results and provides an outlook on further research possibilities.
All analyses are performed with the statistical software R (R Core Team, 2019).

# Chapter 2

# Linear regression models

As the basis for the following chapters, this chapter deals with linear regression models, which are one of the most fundamental and commonly used tools in statistical modeling. We first introduce and define the classical linear regression model and present two parameter estimation methods for this model. Afterwards in Section 2.2 we consider weighted linear regression models as a generalization of the classical linear regression model and discuss the parameter estimation now in this model. This model allows for error terms with unequal variances, which is an essential condition for our later application. We follow Fahrmeir et al (2007) and Rao and Toutenbourg (1999) to present the theory.
Based on the results we consider two simple but important special cases. These models together with the information criteria, we will finally introduce in this chapter, form the theoretical basis for this thesis.

## 2.1 The classical linear regression model

The aim of a regression model is to explain a random variable $Y$ from the influence of $k$ variables $x_1, \ldots, x_k$ and a random error $\epsilon$ in the following form:

$$Y = f(x_1, \ldots, x_k) + \epsilon. \tag{2.1}$$

The random variable $Y$ is usually called response or dependent variable and the known variables $x_1, \ldots, x_k$ are denoted as covariates, independent variables or predictors. While $f(x_1, \ldots, x_k)$ describes the systematic component, the unobservable variable $\epsilon$ is a random disturbance used to explain the difference between the observations and the model. When function $f$ is linear, equation (2.1) is called a linear regression model and the dependent variable $Y$ can be expressed as

$$Y = \beta_0 + \beta_1 x_1 + \ldots \beta_k x_k + \epsilon$$

for some unknown regression parameters $\beta_0, \ldots, \beta_k$. A set of $n$ observations $y_i$ with known covariates $x_{i,1}, \ldots, x_{i,k}$ can be represented as

$$y_i = \beta_0 + \beta_1 x_{i,1} + \ldots \beta_k x_{i,k} + \epsilon_i, \qquad i = 1, \ldots, n,$$

or in matrix notation as

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ defines the vector of observations, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)^T$ the vector of regression parameters, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$ the vector of random errors and design matrix $X$ is defined as

$$X = \begin{pmatrix} 1 & x_{1,1} & \ldots & x_{1,k} \\ 1 & x_{2,1} & \ldots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \ldots & x_{n,k} \end{pmatrix}.$$

In the classical linear regression model the random errors $\epsilon_i$ are assumed to be independent and normally distributed with

$$E[\epsilon_i] = 0 \qquad \text{and} \qquad Var[\epsilon_i] = \sigma^2 \qquad \text{for all } i = 1, \ldots, n.$$

The assumption of constant variances $\sigma^2$ for all error terms is called homoscedasticity. Summing up, the assumptions of the classical linear regression model yield the following definition.

**Definition 2.1.** *(Classical linear regression model)*
*The classical linear regression model explains the random response $Y_i$ as linear function of the covariates*

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \ldots \beta_k x_{i,k} + \epsilon_i, \qquad i = 1, \ldots, n,$$

*with independent and normally distributed random variables*

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2).$$

The expectation of $Y_i$ is by linearity given by

$$E[Y_i] = \beta_0 + \beta_1 x_{i,1} + \ldots \beta_k x_{i,k}, \qquad i = 1, \ldots, n.$$

The classical linear regression model can be written in matrix notation as

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \text{with} \quad \boldsymbol{\epsilon} \sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2 I_n),$$

where $\boldsymbol{Y} = (Y_, \ldots, Y_n)^T$ is the vector of random variables, $I_n$ denotes the identity matrix of order $n$ and $\mathcal{N}_n$ denotes the multivariate normal distribution of a $n$-dimensional random vector. For more details to the univariate and multivariate normal distribution, the reader is referred to Fahrmeir et al (2007, Anhang B). Therefore, it holds

$$\boldsymbol{Y} \sim \mathcal{N}_n(X\boldsymbol{\beta}, \sigma^2 I_n).$$

We consider two methods to estimate the unknown regression parameters based on observations $\boldsymbol{y}$ of $\boldsymbol{Y}$ under the assumptions of the classical linear regression model, the ordinary

least squares estimation and the maximum likelihood estimation.

## Ordinary least squares estimation

The ordinary least squares estimate $\hat{\boldsymbol{\beta}}$ is defined as the vector that minimizes the sum of squared residuals

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = (\boldsymbol{y} - X\boldsymbol{\beta})^T (\boldsymbol{y} - X\boldsymbol{\beta}), \tag{2.2}$$

whereby $\hat{y}_i$ denote the fitted values

$$\hat{y}_i = \beta_0 + \beta_1 x_{i,1} + \ldots + \beta_k x_{i,k}, \quad i = 1, \ldots, n.$$

Hence, this method does not require a distribution assumption. Differentiating (2.2) by $\boldsymbol{\beta}$ and equating to zero yields

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 2X^T X\boldsymbol{\beta} - 2X^T\boldsymbol{y} = 0 \qquad \Leftrightarrow \qquad X^T X\boldsymbol{\beta} = X^T\boldsymbol{y}.$$

The equations on the right-hand side are known as normal equation and if $X$ is of full rank, then we obtain

$$\hat{\boldsymbol{\beta}} = \left(X^T X\right)^{-1} X^T\boldsymbol{y} \tag{2.3}$$

as unique solution of the normal equations.

## Maximum likelihood estimation

The maximum likelihood method is based on the likelihood function $L(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y})$ defined by

$$L\left(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}\right) = \prod_{i=1}^{n} f\left(y_i \mid \boldsymbol{\beta}, \sigma^2\right)$$

for a model with parameters $\boldsymbol{\beta}$ and $\sigma^2$, a density function $f(\cdot \mid \boldsymbol{\beta}, \sigma^2)$ and i.i.d. observations $\boldsymbol{y}$ from f. So obviously the maximum likelihood method needs a distribution assumption in contrast to the least squares method. The objective is to estimate the parameters by maximizing the likelihood function. By reason that the logarithm is an increasing function, instead of differentiating the likelihood function one can consider the logarithm of the likelihood function

$$l(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}) = \log L(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}),$$

called log likelihood. In a classical linear model the random variables $Y_i$ are assumed to be normally distributed and the probability density function of a normally distributed random variable $Y \sim \mathcal{N}(\mu, \sigma^2)$ is

$$f\left(y \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}.$$

Therefore, for model given in Definition 2.1 the likelihood function is specified by

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}) &= \prod_{i=1}^{n} f\left(y_i \mid \boldsymbol{\beta}, \sigma^2\right) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left(y_i - (\beta_0 + \beta_1 x_{i,1} + \ldots + \beta_k x_{i,k})\right)^2}{2\sigma^2}\right\} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{y} - X\boldsymbol{\beta}\right)^T\left(\boldsymbol{y} - X\boldsymbol{\beta}\right)\right\} \end{aligned}$$

and the corresponding log likelihood is

$$l(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}) = \sum_{i=1}^{n} \log\left(f(y_i \mid \boldsymbol{\beta}, \sigma)\right) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(\boldsymbol{y} - X\boldsymbol{\beta}\right)^T\left(\boldsymbol{y} - X\boldsymbol{\beta}\right).$$

The maximum likelihood estimates (MLE) of $\boldsymbol{\beta}$ and $\sigma^2$ are derived by equating the first derivatives of the log likelihood to zero:

$$\begin{aligned} \text{(i)}\ \frac{\partial l\left(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}\right)}{\partial \boldsymbol{\beta}} = 0 \quad &\Leftrightarrow \quad \frac{1}{2\sigma^2} 2X^T\left(\boldsymbol{y} - X\boldsymbol{\beta}\right) = 0 \\ &\Leftrightarrow \quad \boldsymbol{\beta} = \left(X^T X\right)^{-1} X^T \boldsymbol{y} \\ \text{(ii)}\ \frac{\partial l\left(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}\right)}{\partial \sigma^2} = 0 \quad &\Leftrightarrow \quad -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\left(\boldsymbol{y} - X\boldsymbol{\beta}\right)^T\left(\boldsymbol{y} - X\boldsymbol{\beta}\right) = 0 \\ &\Leftrightarrow \quad \sigma^2 = \frac{1}{n}\left(\boldsymbol{y} - X\boldsymbol{\beta}\right)^T\left(\boldsymbol{y} - X\boldsymbol{\beta}\right) \end{aligned}$$

Note that equating the first derivative of the log likelihood by $\boldsymbol{\beta}$ to zero yields again the normal equations. It follows that under the assumptions of the classical linear regression model, as given in Definition 2.1, the maximum likelihood estimation and the least squares estimation yield the same result for the regression parameters $\boldsymbol{\beta}$.

## 2.2 The weighted linear regression model

We relax the assumptions of the classical linear model and consider the linear regression model

$$\boldsymbol{Y} = X\boldsymbol{\beta}_w + \boldsymbol{\epsilon} \quad \text{with} \quad E[\boldsymbol{\epsilon}] = \boldsymbol{0} \quad \text{and} \quad Cov[\boldsymbol{\epsilon}] = \sigma_w^2 W$$

for a known positive definite matrix $W$.
The expected value remains unchanged compared to the classical linear regression model. We denote the expected value as of now by a linear function $\psi(\,\cdot\,)$ with

$$\psi(i) := E[Y_i] = \beta_{w,0} + \beta_{w,1} x_{i,1} + \ldots \beta_{w,k} x_{i,k}, \qquad i = 1, \ldots, n.$$

Unlike before, at the moment we do not require a normal error distribution and moreover, in this model the condition $Cov[\boldsymbol{\epsilon}] = \sigma_w^2 I_n$ is replaced by the assumption $Cov[\boldsymbol{\epsilon}] = \sigma_w^2 W$.

This means we allow additionally for error terms that are not independent and homoscedastic.
In the case the error variables have still not equal variances, but are at least uncorrelated, $\boldsymbol{\epsilon}$ is called heteroscedastic. For these random errors it holds

$$E[\epsilon_i \epsilon_j] = \begin{cases} \sigma_i^2 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

This can be modeled by defining the covariance matrix $W$ as a diagonal matrix

$$W = \text{diag}(w_1, \ldots, w_n).$$

Then it follows

$$Var[\epsilon_i] = \sigma_i^2 = \sigma_w^2 w_i, \qquad i = 1, \ldots, n,$$

and we get the model

$$\boldsymbol{Y} = X\boldsymbol{\beta}_w + \boldsymbol{\epsilon} \quad \text{with} \quad E[\boldsymbol{\epsilon}] = \mathbf{0} \quad \text{and} \quad Cov[\boldsymbol{\epsilon}] = \sigma_w^2 \text{diag}(w_1, \ldots, w_n) \tag{2.4}$$

for known weights $w_1, \ldots, w_n$. These weights define how much variability for each observation is allowed and we will see that weights affect the influence of each observation on the parameter estimation. As outlined in the previous section, for least squares estimation no distribution assumption is needed, so we can estimate the regression parameters $\boldsymbol{\beta}_w$ by a least squares approach.

## Weighted least squares estimation

To find the least squares estimator $\boldsymbol{\beta}_w$, a simple transformation can be used. Multiplication of $W^{-1/2} = \text{diag}(1/\sqrt{w_1}, \ldots, 1/\sqrt{w_n})$ to the left of model (2.4) yields the transformed model equation

$$W^{-1/2}\boldsymbol{Y} = W^{-1/2}X\boldsymbol{\beta_w} + W^{-1/2}\boldsymbol{\epsilon}.$$

This can be rewritten in the form of a classical linear regression model

$$\boldsymbol{Y}^* = X^*\boldsymbol{\beta}_w + \boldsymbol{\epsilon}^* \quad \text{with} \quad E[\boldsymbol{\epsilon}^*] = \mathbf{0} \quad \text{and} \quad Cov[\boldsymbol{\epsilon}^*] = \sigma_w^2 I_n,$$

where $\boldsymbol{Y}^* = W^{-1/2}\boldsymbol{Y}$, $X^* = W^{-1/2}X$ and $\boldsymbol{\epsilon}^* = W^{-1/2}\boldsymbol{\epsilon}$. The least squares method for a classical linear regression model was outlined in previous section. Hence, from (2.3), we get the least squares estimate

$$\begin{aligned} \hat{\boldsymbol{\beta}}_w &= \left(X^{*T}X^*\right)^{-1} X^{*T}\boldsymbol{y}^* \\ &= \left(X^T W^{-1/2} W^{-1/2} X\right)^{-1} X^T W^{-1/2} W^{-1/2}\boldsymbol{y} \\ &= \left(X^T W^{-1} X\right)^{-1} X^T W^{-1}\boldsymbol{y}, \end{aligned}$$

where $\boldsymbol{y}^* = W^{-1/2}\boldsymbol{y}$. This estimate $\hat{\boldsymbol{\beta}}_w$ is minimizing the weighted sum of squared residuals

$$S_w(\boldsymbol{\beta}_w) = \sum_{i=1}^{n} \frac{1}{w_i}(y_i - \hat{y}_i)^2 = \left(\boldsymbol{y} - X\boldsymbol{\beta}_w\right)^T W^{-1} \left(\boldsymbol{y} - X\boldsymbol{\beta}_w\right). \tag{2.5}$$

## Maximum likelihood estimation

For maximum likelihood estimation we use the model defined in (2.4), but now we again have to specify an error distribution and consider the following model.

**Definition 2.2.** *(Weighted linear regression model)*
*The linear regression model*

$$Y_i = \beta_{w,0} + \beta_{w,1}x_{i,1} + \ldots \beta_{w,k}x_{i,k} + \epsilon_i \quad \textit{with} \quad \epsilon_i \sim \mathcal{N}(0, \sigma_w^2 w_i), \qquad i = 1, \ldots, n,$$

*with known weights $w_1, \ldots, w_n$ is called weighted linear regression model.*

Under the assumptions in a weighted linear model it holds

$$\boldsymbol{Y} \sim \mathcal{N}_n\left(X\boldsymbol{\beta}_w, \sigma_w^2 W\right),$$

therefore, the likelihood function is given by

$$L_w(\boldsymbol{\beta}_w, \sigma_w^2 \mid \boldsymbol{y}) \;=\; \frac{1}{\sqrt{(2\pi\sigma_w^2)^n \det(W)}} \exp\left\{-\frac{1}{2\sigma_w^2}\left(\boldsymbol{y} - X\boldsymbol{\beta}_w\right)^T W^{-1}\left(\boldsymbol{y} - X\boldsymbol{\beta}_w\right)\right\}$$

and the corresponding log likelihood is

$$l_w(\boldsymbol{\beta}_w, \sigma_w^2 \mid \boldsymbol{y}) = -\frac{n}{2}\log(2\pi\sigma_w^2) - \frac{1}{2}\sum_{i=1}^{n}\log(w_i) - \frac{1}{2\sigma_w^2}\left(\boldsymbol{y} - X\boldsymbol{\beta}_w\right)^T W^{-1}\left(\boldsymbol{y} - X\boldsymbol{\beta}_w\right).$$

Differentiating the log likelihood by $\boldsymbol{\beta}_w$ and $\sigma_w^2$ and equating the derivatives to zero yields

$$\begin{aligned}
\text{(i)} \quad \frac{\partial l_w\left(\boldsymbol{\beta}_w, \sigma_w^2 \mid \boldsymbol{y}\right)}{\partial \boldsymbol{\beta}_w} = 0 \quad &\Leftrightarrow \quad \frac{1}{2\sigma_w^2} 2X^T W^{-1}\left(\boldsymbol{y} - X\boldsymbol{\beta}_w\right) = 0 \\
&\Leftrightarrow \quad \boldsymbol{\beta}_w = \left(X^T W^{-1} X\right)^{-1} X^T W^{-1}\boldsymbol{y} \\
\text{(ii)} \quad \frac{\partial l_w\left(\boldsymbol{\beta}_w, \sigma_w^2 \mid \boldsymbol{y}\right)}{\partial \sigma_w^2} = 0 \quad &\Leftrightarrow \quad -\frac{n}{2\sigma_w^2} + \frac{1}{2(\sigma_w^2)^2}\left(\boldsymbol{y} - X\boldsymbol{\beta}\right)^T W^{-1}\left(\boldsymbol{y} - X\boldsymbol{\beta}\right) = 0 \\
&\Leftrightarrow \quad \sigma_w^2 = \frac{1}{n}\left(\boldsymbol{y} - X\boldsymbol{\beta}_w\right)^T W^{-1}\left(\boldsymbol{y} - X\boldsymbol{\beta}_w\right)
\end{aligned}$$

The maximum likelihood estimate $\hat{\boldsymbol{\beta}}_w$ equals again the least squares estimate in the weighted linear regression model under the normal distribution assumption. The maximum likelihood estimates are given by

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_w &= \left(X^T W^{-1} X\right)^{-1} X^T W^{-1}\boldsymbol{y} \qquad (2.6) \\
\hat{\sigma}_w^2 &= \frac{1}{n}\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}_w\right)^T W^{-1}\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}_w\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_i}(y_i - \hat{\psi}(i))^2, \qquad (2.7)
\end{aligned}$$

where $\hat{\psi}(i)$ denote the estimated expected values

$$\hat{\psi}(i) = \hat{\beta}_{w,0} + \hat{\beta}_{w,1}x_{i,1} + \dots \hat{\beta}_{w,k}x_{i,k}, \qquad i = 1, \dots, n.$$

Since these values maximize by definition the likelihood function, it follows that the maximum value of the log likelihood is given by

$$\begin{aligned} l_w(\hat{\boldsymbol{\beta}}_w, \hat{\sigma}_w^2 \mid \boldsymbol{y}) &= -\frac{n}{2}\log(2\pi\hat{\sigma}_w^2) - \frac{1}{2}\sum_{i=1}^{n}\log(w_i) - \frac{1}{2\hat{\sigma}_w^2}\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}_w\right)^T W^{-1}\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}_w\right) \\ &= -\frac{n}{2}\log(2\pi\hat{\sigma}_w^2) - \frac{1}{2}\sum_{i=1}^{n}\log(w_i) - \frac{n}{2}. \end{aligned} \tag{2.8}$$

To conclude, maximum likelihood estimation can be applied to derive parameters for a model that fit best to the data. The corresponding maximum value of the log likelihood is therefore an important statistical value, which will be used several times in this thesis.

## 2.3 Simple weighted regression models

In this section we introduce two models, which we will need later, and use the results from the previous section to derive explicit formulas for the maximum likelihood estimates.

**Definition 2.3** (Intercept-only model)**.** *A regression model defined by*

$$Y_i = a + \epsilon_i \quad \textit{with} \quad \epsilon_i \sim \mathcal{N}(0, \sigma_w^2 w_i), \qquad i = 1, \dots, n,$$

*with a single constant regression parameter a for all $i = 1, \dots, n$ is called intercept-only model.*

The intercept-only model describes response variables, which all have equal means and differ only because of random disturbance. It has only a single parameter and no prediction variable. Hence, using (2.6) and (2.7), the maximum likelihood estimates are given by

$$\hat{a} = \left(X^T W^{-1} X\right)^{-1} X^T W^{-1} \boldsymbol{y} = \left(\sum_{i=1}^{n}\frac{1}{w_i}\right)^{-1}\sum_{i=1}^{n}\frac{y_i}{w_i} = \frac{\sum_{i=1}^{n}\frac{y_i}{w_i}}{\sum_{i=1}^{n}\frac{1}{w_i}}, \tag{2.9}$$

$$\hat{\sigma}_w^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_i}(y_i - \hat{a})^2. \tag{2.10}$$

**Definition 2.4** (Simple linear regression model)**.** *A linear regression model*

$$Y_i = a + b \cdot x_i + \epsilon_i \qquad \textit{with} \quad \epsilon_i \sim \mathcal{N}(0, \sigma_w^2 w_i), \qquad i = 1, \dots, n,$$

*with a single predictor variable $x_i$ for all $i = 1, \dots, n$ is called simple linear regression model.*

A weighted linear regression model with a single covariate explains the relationship by a line with a y-intercept $a$ and a slope $b$. The maximum likelihood estimates for this model are given by

$$\hat{a} = \frac{\overline{x_w^2} \cdot \overline{y_w} - \overline{x_w} \cdot \overline{y_w}}{\overline{1_w} \cdot \overline{x_w^2} - \left(\overline{x_w}\right)^2}, \qquad \text{and} \qquad \hat{b} = \frac{-\overline{x_w} \cdot \overline{y_w} + \overline{1_w} \cdot \overline{xy_w}}{\overline{1_w} \cdot \overline{x_w^2} - \left(\overline{x_w}\right)^2}, \tag{2.11}$$

$$\hat{\sigma}_w^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{w_i} (y_i - \hat{\mu}_i)^2 \qquad \text{with} \qquad \hat{\mu}_i = \hat{a} + \hat{b} x_i, \tag{2.12}$$

with weighted sums

$$\overline{x_w} = \sum_{i=1}^{n} \frac{x_i}{w_i}, \qquad \overline{y_w} = \sum_{i=1}^{n} \frac{y_i}{w_i}, \qquad \overline{xy_w} = \sum_{i=1}^{n} \frac{x_i y_i}{w_i},$$

$$\overline{1_w} = \sum_{i=1}^{n} \frac{1}{w_i} \qquad \text{and} \qquad \overline{x_w^2} = \sum_{i=1}^{n} \frac{x_i^2}{w_i}.$$

*Proof.* There is a single prediction variable $x_i$ for all $i = 1, \ldots, n$, hence, the simple linear model is represented by design matrix and regression parameters

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\beta}_w = \begin{pmatrix} a \\ b \end{pmatrix}.$$

By (2.6) it holds

$$\begin{aligned} \left(\hat{a}, \hat{b}\right)^T = \hat{\boldsymbol{\beta}}_w &= \left(X^T W^{-1} X\right)^{-1} X^T W^{-1} \boldsymbol{y} = \\ &= \begin{pmatrix} \sum_{i=1}^{n} \frac{1}{w_i} & \sum_{i=1}^{n} \frac{x_i}{w_i} \\ \sum_{i=1}^{n} \frac{x_i}{w_i} & \sum_{i=1}^{n} \frac{x_i^2}{w_i} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{n} \frac{y_i}{w_i} \\ \sum_{i=1}^{n} \frac{x_i y_i}{w_i} \end{pmatrix} \\ &= \frac{1}{\sum_{i=1}^{n} \frac{1}{w_i} \cdot \sum_{i=1}^{n} \frac{x_i^2}{w_i} - \left(\sum_{i=1}^{n} \frac{x_i}{w_i}\right)^2} \begin{pmatrix} \sum_{i=1}^{n} \frac{x_i^2}{w_i} & -\sum_{i=1}^{n} \frac{x_i}{w_i} \\ -\sum_{i=1}^{n} \frac{x_i}{w_i} & \sum_{i=1}^{n} \frac{1}{w_i} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n} \frac{y_i}{w_i} \\ \sum_{i=1}^{n} \frac{x_i y_i}{w_i} \end{pmatrix} \end{aligned}$$

and multiplication yields the result above. The maximum likelihood estimate for $\sigma_w^2$ follows immediately by (2.7). □

Both described models assume that a single model is valid for all data points, where the single model is in the form

$$Y_i = \psi(i) + \epsilon_i \qquad \text{with} \quad \epsilon_i \sim \mathcal{N}(0, \sigma_w^2 w_i), \qquad i = 1, \ldots, n.$$

In practice this is not adequate if the data indicates a partition in more than one interval with different behavior. In Chapter 5 we will consider models, which allow for changes in the behavior of the response.

## 2.4 Information criteria AIC and BIC

Given a set of data and a set of candidate models the next step is to select the best model. In order to do this, it is obviously important to consider at first the goodness-of-fit, but also the complexity of a model. Two commonly used criteria for model selection based on a set of observations, which balance model fit and penalize the complexity of a model, are the Akaike information criterion and the Bayesian information criterion.
In this section we will define these criteria and present the process of model selection using information criteria. To do this, we follow Fabozzi et al (2014) and Ding et al (2018). Then, we apply it to the weighted regression models.

The Akaike information criterion was proposed by Akaike (1973, 1974) as criterion for model selection and is defined for a set of data $\boldsymbol{y} = \{y_1, \dots, y_n\}$ as follows.

**Definition 2.5.** *The Akaike information criterion (AIC) for a model $\mathcal{M}$ with parameter set $\Theta$ is*

$$AIC(\mathcal{M}) := -2 \log L(\hat{\Theta} \mid \boldsymbol{y}) + 2p,$$

*where $\log L(\hat{\Theta})$ is the maximized value of the log likelihood with parameters $\hat{\Theta}$ that maximize the likelihood function and $p$ the numbers of parameters of model $\mathcal{M}$.*

The first component measures the model fit, where by the negative sign a high likelihood leads to a low AIC. In order to avoid overfitting it is not enough to use the likelihood alone for model selection. If we add more parameters to our model, we will always get a higher maximum likelihood, but not necessarily a better model because of the increase in complexity. That is what the second term of the AIC is for, it penalizes additional parameters. Hence, a low AIC refers to a high maximum likelihood and a low number of parameters, that is why the model with the lowest AIC is considered as the best model among the candidate models.
The Bayesian information criterion (E. Schwarz, 1978), also known as the Schwarz information criterion, is similar to the Akaike information criterion, but it has a harder penalty depending on the sample size.

**Definition 2.6.** *The Bayesian information criterion (BIC) for a model $\mathcal{M}$ with parameter set $\Theta$ is*

$$BIC(\mathcal{M}) := -2 \log L(\hat{\Theta} \mid \boldsymbol{y}) + p \log(n),$$

*where $\log L(\hat{\Theta})$ is the maximum log likelihood with parameters $\hat{\Theta}$, $p$ the numbers of parameters of model $\mathcal{M}$ and $n$ the number of observations.*

Again the first term of the BIC represents the model fit and the second term is a penalty for model complexity. Therefore, consistently models with lower BIC are preferred and the model selection process using information criteria consists of three steps:

**Step 1:** Fit the observed data to each candidate model $\mathcal{M}$ by estimation of the model parameters $\Theta$.

**Step 2:** Calculate the information criterion for each fitted model.

**Step 3:** Compare the candidate models via information criteria and select the model with lowest.

In case of a weighted linear regression model $\mathcal{M}_w$, as stated in Definition 2.2, the parameter set $\Theta$ is given by

$$\Theta = \{\beta_{w,0}, \beta_{w,1}, \ldots, \beta_{w,k}, \sigma_w^2\} = \{\boldsymbol{\beta}_w, \sigma_w^2\}$$

and we get the information criteria

$$\begin{aligned} AIC\,(\mathcal{M}_w) &= -2\log L_w(\hat{\Theta} \mid \boldsymbol{y}) + 2 \cdot dim(\Theta) && (2.13)\\ BIC\,(\mathcal{M}_w) &= -2\log L_w(\hat{\Theta} \mid \boldsymbol{y}) + dim(\Theta) \cdot \log(n). && (2.14) \end{aligned}$$

By using (2.8), the first term of both criteria is given by

$$-2\log L_w(\hat{\Theta} \mid \boldsymbol{y}) = n\log(2\pi\hat{\sigma}_w^2) + n + \sum_{i=1}^{n} \log(w_i) \qquad (2.15)$$

and we have the maximum likelihood estimate for $\sigma_w^2$

$$\hat{\sigma}_w^2 = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{w_i}(y_i - \hat{\psi}(i))^2$$

for the fitted values $\hat{\psi}(i)$, for all $i = 1, \ldots, n$, based on maximum likelihood estimates $\hat{\beta}_w$ given by (2.6). In Chapter 6 we will discuss how to use this model selection approach with regard to loss data. But for this purpose we continue with the required foundations and deal with loss reserving in the next chapter.

# Chapter 3

# Loss reserving

We start this chapter with a brief introduction to loss reserving and specify the problem, which includes the commonly used formulation of the loss history as a triangular array. Thereafter, we consider two loss reserving models, which provide estimates for the claims reserve based on the given loss triangle and we will show that these estimators are unbiased. To present these fundamental concepts we follow Mack (2002) and Wüthrich and Merz (2008) and for the definition of the loss development by triangles we follow Schmidt (2006).
Using the outlined theory of linear regression models in the previous chapter, we explain how the introduced reserving models are linked to linear regression and compare the presented estimated model parameters to the results from a weighted least squares and maximum likelihood estimation. Finally, we discuss how to select the best model to predict the reserve for the given loss history.

## 3.1 Introduction to loss reserving

We consider a portfolio of non-life insurance policies. If a covered event happens in a policy, the process of the claim settlement can extend over many years. At first, it can take time until the insured discovers and reports the claim, for instance asbestos claims or faults in construction can manifest themselves many years after they occurred. After reporting, it can still take up to several years until the full claim amount can be determined, examples for this are claims in case of bodily injury with long period of treatment; or the settlement can be delayed due to legal proceedings.
That is why the insurance company has to build a reserve for outstanding liabilities from not yet settled claims, called claims reserve. For all reported but not yet fully settled losses a so-called case reserve is assessed by a claim handler based on his/her experience and adjusted until the loss is closed. Until the claims are settled future payments are expected, and it is possible that the case reserve is not sufficient. So on the one hand a reserve for incurred (and reported) but not enough reserved claims, denoted IBNeR reserve, has to be build in addition to the case reserves and on the other hand a claims reserve for incurred but not yet reported (IBNyR) claims is required. We denote the IBNeR claims reserve and IBNyR claims reserve together as IBNR claims reserve and consequently the total

claims reserve of a portfolio is defined as

$$\text{total claims reserve} = \text{case reserves} + \text{IBNR reserve}.$$

Mathematically speaking, the total claims reserve of a portfolio is the estimated sum of expected future payments based on the available information at that time (Institute of Actuaries, 1997).
Whereas case reserves are determined on an individual claims basis, the IBNR reserve is set on the aggregated basis of a portfolio. For an adequate estimation of the IBNR reserve actuarial models and the history of loss payments are required. Instead of paid losses, it is also possible to use incurred losses as data basis, where incurred losses denote the sum of paid losses and case reserves.
The used loss history in the actuarial models takes account of the development of losses over time. This development can be described in incremental or in cumulative form. Incremental data shows the initial loss amount in the first year and then the successive increments of each calendar year. Cumulative data reports the total summed loss amount in all years also starting with the initial amount in the first year. Note that in general paid loss increments are positive and cumulative paid losses are monotone increasing, but for incurred losses negative increments are possible. This can be explained by the yearly adjustment of the case reserves, if the first indication was too high, it is reasonable to reduce the case reserve and that leads to a negative incremental case reserve.
The following models can be applied to both types of data and in theory both types should lead to the same expected ultimate loss in the end, but in practice there is a gap between the result of both projections (Quarg and Mack, 2004).

## 3.2 Loss triangles

We model a portfolio of outstanding claims by a set of random variables describing the incremental basic losses. Due to the long development of claims over time, older losses covered under policies that were written several years ago exhibit more data history than very recent losses covered under more recent policies. This leads to the typical triangular representation of the loss history, called loss development triangle or run-off triangle, where the rows of the triangle refer to the origin years and the columns, denoted as development years, represent the development over time. Commonly used as origin year is the policy year under which they are covered, denoted as underwriting year, or the calendar year the claim occurred, called accident year. (Institute of Actuaries, 1997)
In the following, we group the losses by underwriting years, which yields the following definition.

**Definition 3.1.** *The random variables $S_{i,k}$, $i, k = 1, \ldots, n$ denote the total incremental losses of underwriting year $i$ in development year $k$.*

This means each row in the loss triangle displays the claims arising from policies written in the same underwriting year (UY) and each column shows the development year (DY), which specifies the period in years until the respective claims are fully settled starting from the associated underwriting year. Furthermore, losses $S_{i,k}$ on a diagonal represent

the same calendar year $i+k-1$.
For the most recent underwriting year, denoted by $n$, only the losses occurred and reported during the first development year have been observed, since this corresponds to the most recent calendar year $n$. Therefore, for the previous underwriting year the losses of development year one and two are known, etc., while for the oldest available underwriting year, denoted by underwriting year 1, the whole loss development is known. Future loss development will arise in the next diagonals until all losses are finally settled. Thus, the current loss triangle will gradually turn into a rectangle. Consequently, the full loss development is split into two triangles, the upper triangle with observed losses

$$\{S_{i,k} \mid i + k - 1 \leq n\}$$

and the lower triangle with future losses

$$\{S_{i,k} \mid i + k - 1 > n\}.$$

This is illustrated in Table 1. The observations in the upper triangle are available at the

(a) Observed upper triangle

| UY | development year $k$ | | | | |
|---|---|---|---|---|---|
| i | 1 | 2 | ... | $n-1$ | $n$ |
| 1 | $S_{1,1}$ | $S_{1,2}$ | ... | $S_{1,n-1}$ | $S_{1,n}$ |
| 2 | $S_{2,1}$ | $S_{2,2}$ | ... | $S_{2,n-1}$ | |
| $\vdots$ | $\vdots$ | $\vdots$ | $\iddots$ | | |
| $n-1$ | $S_{n-1,1}$ | $S_{n-1,2}$ | | | |
| $n$ | $S_{n,1}$ | | | | |

(b) Unobserved lower triangle

| UY | development year $k$ | | | | |
|---|---|---|---|---|---|
| i | 1 | 2 | ... | $n-1$ | $n$ |
| 1 | | | | | |
| 2 | | | | | $S_{2,n}$ |
| $\vdots$ | | | | $\iddots$ | $\vdots$ |
| $n-1$ | | | ... | $S_{n-1,n-1}$ | $S_{n-1,n}$ |
| $n$ | | $S_{n,2}$ | ... | $S_{n,n-1}$ | $S_{n,n}$ |

Table 1: Upper and lower loss development triangle with incremental losses

end of calendar year $n$, while the lower triangle needs to be predicted. The future losses define the loss reserve the insurance company needs to hold for the outstanding liabilities.

**Definition 3.2.** *The loss reserve for underwriting year $i$ at the end of calendar year $n$ is given by*

$$R_i = \sum_{k=n+2-i}^{n} S_{i,k}, \qquad i = 2, \ldots, n, \tag{3.1}$$

*and $R_1 = 0$. The total reserve $R$ for the outstanding claims at the end of calendar year $n$ is defined as the sum of reserves for these underwriting years, i.e.*

$$R = \sum_{i=2}^{n} R_i.$$

In case the incremental losses $S_{i,k}$ represent incurred losses this total reserve defines actually the IBNR reserve of the portfolio and to get the total claims reserve we have to add the case reserves from the observed incurred losses. These case reserves are reported but even though they are part of the future payments. The predicted reserve based on paid losses $S_{i,k}$ already defines the total claims reserve of the portfolio.
Instead of incremental losses a loss triangle can also display cumulative losses.

**Definition 3.3.** *The random variables $C_{i,k}$, $i,k = 1,\dots,n$, denote the cumulative loss amounts of underwriting years $i$ after $k$ development years. It holds*

$$C_{i,k} = \sum_{j=1}^{k} S_{i,j}, \qquad i,k = 1,\dots,n.$$

The other way around it is obviously that

$$S_{i,k} = C_{i,k} - C_{i,k-1} \qquad \text{for all} \qquad i,k = 1,\dots,n,$$

where $C_{i,0} := 0$. Therefore, the reserve for underwriting year $i$, specified in Definition 3.2, can also be expressed in form of

$$R_i = C_{i,n} - C_{i,n+1-i}, \qquad i = 2,\dots,n, \tag{3.2}$$

where the losses $C_{i,n+1-i}$, $i = 1,\dots,n$, define the latest known cumulative loss amounts for the underwriting years, which are represented by the last loss diagonal of the observed upper triangle

$$\mathcal{D} := \{C_{i,k} \mid i + k - 1 \leq n\}.$$

The cumulative losses $C_{i,n}$ describe the ultimate loss amount for underwriting years $i = 1,\dots,n$. It follows that the loss reserve can be predicted either by estimating the lower incremental or cumulative loss triangle. Prediction of the lower triangle yields a projection of the loss development as illustrated for the cumulative loss development in Table 2. Therefore, Definition 3.2 and equation (3.2) yield the following corollary.

**Corollary 3.4.** *We denote the predicted loss reserve based on $\mathcal{D}$ for underwriting year $i = 2,\dots,n$ by $\hat{R}_i$ and it can be estimated by*

$$\hat{R}_i = \sum_{k=n+2-i}^{n} \hat{S}_{i,k},$$

*where $\hat{S}_{i,k}$, $i + k - 1 > n$, denote the predicted incremental losses, or equivalently by*

$$\hat{R}_i = \hat{C}_{i,n} - C_{i,n+1-i},$$

*where $\hat{C}_{i,n}$, $i = 2,\dots,n$, denote the predicted ultimate cumulative losses.*

So we aim to predict the total claims reserves $R$ by predicting the future losses for all underwriting years $i = 1,\dots,n$ given the loss history $\mathcal{D}$.

| UY | development year $k$ | | | | |
|---|---|---|---|---|---|
| i | 1 | 2 | ... | $n-1$ | $n$ |
| 1 | $C_{1,1}$ | $C_{1,2}$ | ... | $C_{1,n-1}$ | $C_{1,n}$ |
| 2 | $C_{2,1}$ | $C_{2,2}$ | ... | $C_{2,n-1}$ | $\hat{C}_{2,n}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\ddots$ | $\vdots$ |
| $n$–1 | $C_{n-1,1}$ | $C_{n-1,2}$ | ... | $\hat{C}_{n-1,n-1}$ | $\hat{C}_{n-1,n}$ |
| $n$ | $C_{n,1}$ | $\hat{C}_{n,2}$ | ... | $\hat{C}_{n,n-1}$ | $\hat{C}_{n,n}$ |

Table 2: Cumulative loss development with observed upper triangle and predicted lower triangle

## 3.3 Loss reserving methods

In this section we consider two stochastic loss reserving methods. The objective of a loss reserving method is in general to predict the loss reserves and there are various different statistical approaches for this estimation. One of the oldest and most popular methods is the chain ladder method. The original chain ladder method was a simple but accurate algorithmic technique to estimate the reserves, but it provides only a point estimator and gives no indication of the uncertainty in the predictions. So Mack (1993) developed a stochastic model underlying the chain ladder algorithm, which allows quantifying also the uncertainties in the prediction.
The stochastic chain ladder model is a distribution-free model and is based on the assumption that the loss development shows for all underwriting years a similar multiplicative behavior from development year to development year. That is why we call it the multiplicative model in the following.
Additionally, we consider another distribution-free stochastic reserving model, which assumes instead of a similar multiplicative loss pattern an additive loss pattern with similar ratios per development year, so we call it the additive model.
Based on both models of Mack (2002) we aim to predict the loss reserve, so after defining the models we will derive model parameters to predict the future development.

### The additive model

All loss reserving methods are based on the assumption that the development of the losses of every underwriting year follows a common development pattern. One simple way to model similar loss development is to assume that the underwriting years are independent and their behavior is only affected by different volume measures.
We use the known premiums $v_i$ of all underwriting years $i = 1, \ldots, n$ as volume measures. The volume adjusted random variable $\frac{S_{i,k}}{v_i}$, $i, k = 1, \ldots, n$, is denoted by $M_{i,k}$ and defines the incremental loss ratio of underwriting year $i$ in development year $k$. The additive model focuses on loss ratios as central parameters to describe the loss development and

the following assumptions.

**Assumption 3.1** (Assumptions of the additive model)**.**

*(A1) All incremental losses are independent random variables, i.e. the incremental losses of different underwriting years $i, j = 1, \dots, n$*

$$\{S_{i,1}, ..., S_{i,n}\}, \{S_{j,1}, ..., S_{j,n}\}, i \neq j$$

*are independent and the increments of different development years $k, l = 1, \dots, n$*

$$\{S_{1,k}, ..., S_{n,k}\}, \{S_{1,l}, ..., S_{n,l}\}, k \neq l$$

*are independent.*

*(A2) There exist finite expected loss ratios $m_1, \dots, m_n$ such that*

$$E\left[\frac{S_{i,k}}{v_i}\right] = E\left[M_{i,k}\right] = m_k$$

*holds for all $i = 1, \dots, n, k = 1, \dots, n$.*

*(A3) There exist constant variance parameters $s_1^2, \dots, s_n^2$ such that*

$$Var\left[\frac{S_{i,k}}{v_i}\right] = Var[M_{i,k}] = \frac{s_k^2}{v_i}$$

*holds for all $i, k = 1, \dots, n$.*

Under the model assumptions (A2) and (A3) we can determine the first two moments of the incremental losses $S_{i,k}$ by

$$E\left[S_{i,k}\right] = v_i \cdot m_k \qquad \text{for all } i, k = 1 \dots, n \tag{3.3}$$

and

$$Var\left[S_{i,k}\right] = s_k^2 \cdot v_i \qquad \text{for all } i, k = 1 \dots, n. \tag{3.4}$$

**Theorem 3.5.** *Under the Assumptions 3.1 unbiased estimators for the mean loss ratios $m_1, \dots, m_n$ are given by*

$$\hat{m}_k = \frac{\sum_{i=1}^{n+1-k} S_{i,k}}{\sum_{i=1}^{n+1-k} v_i}, \qquad k = 1, \dots, n.$$

*Proof.* The expected value of the estimator $\hat{m}_k$ is by definition of $\hat{m}_k$ and independence of the underwriting years given by

$$E\left[\hat{m}_k\right] \;=\; E\left[\frac{\sum_{i=1}^{n+1-k} S_{i,k}}{\sum_{i=1}^{n+1-k} v_i}\right] \overset{\text{(A1)}}{=} \frac{\sum_{i=1}^{n+1-k} E\left[S_{i,k}\right]}{\sum_{i=1}^{n+1-k} v_i}.$$

Using (3.3) for $E\left[S_{i,k}\right]$ yields

$$E\left[\hat{m}_k\right] = \frac{\sum_{i=1}^{n+1-k} v_i \cdot m_k}{\sum_{i=1}^{n+1-k} v_i} = m_k,$$

which completes the proof of the theorem. □

In the previous proof and in the following we frequently use the characteristics of the expected value and the variance. For more details see e.g. Klenke (2013).

**Theorem 3.6.** *Under the Assumptions 3.1 the variances of the estimators $\hat{m}_k$ are given by*

$$Var\left[\hat{m}_k\right] = \frac{s_k^2}{\sum_{i=1}^{n+1-k} v_i}, \qquad k = 1, \dots, n-1.$$

*Proof.* By independence of underwriting years and Theorem 3.5 it holds

$$Var[\hat{m}_k] \;=\; Var\left[\frac{\sum_{i=1}^{n+1-k} S_{i,k}}{\sum_{i=1}^{n+1-k} v_i}\right] \overset{(A1)}{=} \frac{\sum_{i=1}^{n+1-k} Var[S_{i,k}]}{\left(\sum_{i=1}^{n+1-k} v_i\right)^2}.$$

Hence, from (3.4) we obtain

$$Var[\hat{m}_k] \;=\; \frac{\sum_{i=1}^{n+1-k} v_i \cdot s_k^2}{\left(\sum_{i=1}^{n+1-k} v_i\right)^2} = \frac{s_k^2}{\sum_{i=1}^{n+1-k} v_i}$$

and the theorem follows. □

**Theorem 3.7.** *Under the model assumptions of the additive method unbiased estimators for the variance parameters $s_1^2, \dots, s_{n-1}^2$ are given by*

$$\hat{s}_k^2 = \frac{1}{n-k} \sum_{i=1}^{n+1-k} v_i \left(\frac{S_{i,k}}{v_i} - \hat{m}_k\right)^2, \qquad k = 1, \dots, n-1.$$

*Proof.* To show that $\hat{s}_k^2$ is unbiased we take the expectation and get by independence assumption (A1)

$$E[\hat{s}_k^2] \;=\; \frac{1}{n-k} \sum_{i=1}^{n+1-k} v_i \cdot E\left[\left(\frac{S_{i,k}}{v_i} - \hat{m}_k\right)^2\right]. \tag{3.5}$$

At first, we consider only

$$E\left[\left(\frac{S_{i,k}}{v_i} - \hat{m}_k\right)^2\right] \;=\; \underbrace{E\left[\left(\frac{S_{i,k}}{v_i}\right)^2\right]}_{Var\left[\frac{S_{i,k}}{v_i}\right] + E\left[\frac{S_{i,k}}{v_i}\right]^2} - 2E\left[\frac{S_{i,k}}{v_i} \cdot \hat{m}_k\right] + \underbrace{E\left[\hat{m}_k^2\right]}_{Var\left[\hat{m}_k\right] + E[\hat{m}_k]^2}.$$

Hence, using Theorem 3.6 and (A3), we obtain

$$\begin{aligned} E\left[\left(\frac{S_{i,k}}{v_i} - \hat{m}_k\right)^2\right] &= \frac{s_k^2}{v_i} + m_k^2 - 2E\left[\frac{S_{i,k}}{v_i} \cdot \hat{m}_k\right] + \frac{s_k^2}{\sum_{j=1}^{n+1-k} v_j} + m_k^2 \\ &= s_k^2 \left(\frac{1}{v_i} + \frac{1}{\sum_{j=1}^{n+1-k} v_j}\right) + 2m_k^2 - 2E\left[\frac{S_{i,k}}{v_i} \cdot \hat{m}_k\right]. \end{aligned} \tag{3.6}$$

Next we consider only the last term on the right-hand side of (3.6) and we get

$$E\left[\frac{S_{i,k}}{v_i}\cdot\hat{m}_k\right] = E\left[\frac{S_{i,k}}{v_i}\cdot\frac{\sum_{j=1}^{n+1-k}S_{j,k}}{\sum_{j=1}^{n+1-k}v_j}\right] = \frac{\sum_{j=1}^{n+1-k}E\left[S_{i,k}\cdot S_{j,k}\right]}{v_i\cdot\sum_{j=1}^{n+1-k}v_j}.$$

Since by independence of underwriting years it holds

$$E[S_{i,k}\cdot S_{j,k}] = \begin{cases} E[S_{i,k}]\cdot E[S_{j,k}], & i\neq j, \\ E[S_{i,k}^2], & \text{else} \end{cases}$$

and it is

$$E[S_{i,k}^2] = Var[S_{i,k}] + E[S_{i,k}]^2,$$

we get

$$\begin{aligned}\frac{\sum_{j=1}^{n+1-k}E\left[S_{i,k}\cdot S_{j,k}\right]}{v_i\cdot\sum_{j=1}^{n+1-k}v_j} &= \frac{Var[S_{i,k}]+\sum_{j=1}^{n+1-k}E\left[S_{i,k}\right]\cdot E[S_{j,k}]}{v_i\cdot\sum_{j=1}^{n+1-k}v_j} = \\ &= \frac{v_i\cdot s_k^2 + v_i\cdot m_k^2\cdot\sum_{j=1}^{n+1-k}v_j}{v_i\cdot\sum_{j=1}^{n+1-k}v_j} = \frac{s_k^2}{\sum_{j=1}^{n+1-k}v_j} + m_k^2,\end{aligned}$$

where the second step follows by (3.3) and (3.4). Hence, we have

$$E\left[\frac{S_{i,k}}{v_i}\cdot\hat{m}_k\right] = \frac{s_k^2}{\sum_{j=1}^{n+1-k}v_j} + m_k^2.$$

Putting this back in equation (3.6) gives

$$\begin{aligned}E\left[\left(\frac{S_{i,k}}{v_i}-\hat{m}_k\right)^2\right] &= s_k^2\left(\frac{1}{v_i}+\frac{1}{\sum_{j=1}^{n+1-k}v_j}\right) + 2m_k^2 - 2\left(\frac{s_k^2}{\sum_{j=1}^{n+1-k}v_j}+m_k^2\right) \\ &= s_k^2\left(\frac{1}{v_i}-\frac{1}{\sum_{j=1}^{n+1-k}v_j}\right).\end{aligned}$$

Thus, it holds by (3.5)

$$\begin{aligned}E[\hat{s}_k^2] &= \frac{1}{n-k}\sum_{i=1}^{n+1-k}v_i\cdot s_k^2\left(\frac{1}{v_i}-\frac{1}{\sum_{j=1}^{n+1-k}v_j}\right) = \\ &= \frac{s_k^2}{n-k}\left(\sum_{i=1}^{n+1-k}1-\frac{\sum_{i=1}^{n+1-k}v_i}{\sum_{j=1}^{n+1-k}v_j}\right) = s_k^2,\end{aligned}$$

since the remaining terms cancel out, which completes the proof. □

As a result of Theorem 3.5 an unbiased estimator for $E[S_{i,k}]$ is given by

$$\hat{S}_{i,k} = v_i\cdot\hat{m}_k \quad \text{for all } k\geq n+2-i. \tag{3.7}$$

Furthermore, conditional on the loss history $\mathcal{D} = \{C_{i,k} \mid i + k + 1 \leq n\}$ we get for underwriting years $i = 2, \ldots, n$

$$\begin{aligned} E\left[C_{i,n} \mid \mathcal{D}\right] &= C_{i,n-i+1} + E\left[S_{i,n-i+2} + \ldots + S_{i,n} \mid \mathcal{D}\right] = C_{i,n-i+1} + \sum_{k=n-i+2}^{n} E\left[S_{i,k}\right] \\ &= C_{i,n-i+1} + v_i \cdot \sum_{k=n-i+2}^{n} m_k. \end{aligned} \tag{3.8}$$

The unbiased estimator for this is given by

$$\hat{C}_{i,n} = C_{i,n-i+1} + v_i \cdot \sum_{k=n-i+2}^{n} \hat{m}_k. \tag{3.9}$$

Using Corollary 3.4 yields finally the predicted reserve in the additive model

$$\hat{R}_i = v_i \cdot \sum_{k=n-i+2}^{n} \hat{m}_k, \qquad i = 2, \ldots, n. \tag{3.10}$$

## The multiplicative model

The underlying assumption of the multiplicative method is that the underwriting years show a similar development pattern based on development factors. Let $F_{i,k}$ denote the individual development factor $\frac{C_{i,k}}{C_{i,k-1}}$ of underwriting year $i = 1, \ldots, n$ from development year $k-1$ to development year $k$, then the basic assumption is that the development factors are similar for all underwriting years.
The development factors define how the cumulative losses are linked and the principle focus of the multiplicative model lies on them. To be more precise, the model assumptions of the multiplicative method are the following:

**Assumption 3.2** (Assumptions of the multiplicative model)**.**

*(M1) All cumulative losses of different underwriting years are independent, i.e. for* $i, j = 1, \ldots, n$

$$\{C_{i,1}, ..., C_{i,n}\}, \{C_{j,1}, ..., C_{j,n}\}, i \neq j$$

*are independent sets of random variables.*

*(M2) There exist finite development factors* $f_2, \ldots, f_n$ *such that*

$$E\left[\frac{C_{i,k}}{C_{i,k-1}} \,\middle|\, C_{i,1}, \ldots C_{i,k-1}\right] = E\left[F_{i,k} \mid C_{i,1}, \ldots C_{i,k-1}\right] = f_k$$

*holds for all* $i = 1, \ldots, n, k = 2, \ldots, n$.

*(M3) There exist constant variance parameters* $\sigma_2^2, \ldots, \sigma_n^2$ *such that*

$$Var\left[\frac{C_{i,k}}{C_{i,k-1}} \;\middle|\; C_{i,1}, \ldots C_{i,k-1}\right] = Var\left[F_{i,k} \mid C_{i,1}, \ldots, C_{i,k-1}\right] = \frac{\sigma_k^2}{C_{i,k-1}}$$

*holds for all* $i = 2, \ldots, n, k = 2, \ldots, n$.

The assumptions (M2) and (M3) lead to

$$E\left[C_{i,k} \mid C_{i,1}, \ldots C_{i,k-1}\right] = C_{i,k-1} \cdot f_k \qquad \text{for all } i = 1 \ldots, n, k = 2, \ldots, n. \tag{3.11}$$

and

$$Var\left[C_{i,k} \mid C_{i,1}, \ldots C_{i,k-1}\right] = C_{i,k-1} \cdot \sigma_k^2 \qquad \text{for all } i = 1 \ldots, n, k = 2, \ldots, n. \tag{3.12}$$

**Theorem 3.8.** *Under the Assumptions 3.2 of the multiplicative model it holds that*

$$\hat{f}_k = \frac{\sum_{i=1}^{n+1-k} C_{i,k}}{\sum_{i=1}^{n+1-k} C_{i,k-1}}$$

*are unbiased and uncorrelated estimators for development factor* $f_k = E[F_{i,k} \mid C_{i,1}, \ldots, C_{i,k-1}]$ *for all* $k = 2, \ldots, n$.

*Proof.* The estimator $\hat{f}_k$ is unbiased as $E[\hat{f}_k] = E\left[E[\hat{f}_k \mid C_{i,1}, \ldots C_{i,k-1}]\right]$ and

$$\begin{aligned} E[\hat{f}_k \mid C_{i,1}, \ldots C_{i,k-1}] &= E\left[\frac{\sum_{i=1}^{n+1-k} C_{i,k}}{\sum_{i=1}^{n+1-k} C_{i,k-1}} \;\middle|\; C_{i,1}, \ldots C_{i,k-1}\right] \\ &= \frac{\sum_{i=1}^{n+1-k} E\left[C_{i,k} \mid C_{i,1}, \ldots C_{i,k-1}\right]}{\sum_{i=1}^{n+1-k} C_{i,k-1}} \end{aligned}$$

by independence assumption (M1) and measurability of the previous cumulative losses $C_{i,k-1}$. Hence, from (3.11), we get

$$E[\hat{f}_k \mid C_{i,1}, \ldots C_{i,k-1}] = \frac{\sum_{i=1}^{n+1-k} C_{i,k-1} \cdot f_k}{\sum_{i=1}^{n+1-k} C_{i,k-1}} = f_k,$$

which shows that the estimators are unbiased. It remains to show, that they are uncorrelated. For this we define

$$\mathcal{B}_k = \{C_{i,j} \mid j \leq k, i + j - 1 \leq n\}, \qquad 1 \leq k \leq n.$$

Then it holds for all $j < k$ by unbiasedness and measurability

$$E\left[\hat{f}_j \cdot \hat{f}_k\right] = E\left[E\left[\hat{f}_j \cdot \hat{f}_k \mid \mathcal{B}_k\right]\right] = E\left[\hat{f}_j \cdot E\left[\hat{f}_k \mid \mathcal{B}_k\right]\right] = E\left[\hat{f}_j\right] \cdot \hat{f}_k = \hat{f}_j \cdot \hat{f}_k,$$

which completes the proof of the theorem. □

**Theorem 3.9.** *The variance of estimator $\hat{f}_k$ in the multiplicative model is given by*

$$Var\left[\hat{f}_k \mid C_{i,1}, \dots C_{i,k-1}\right] = \frac{\sigma_k^2}{\sum_{i=1}^{n+1-k} C_{i,k-1}}, \qquad k = 2, \dots, n-1.$$

*Proof.* Again using independence of the underwriting years and (3.12) yields immediately

$$Var\left[\hat{f}_k \mid C_{i,1}, \dots C_{i,k-1}\right] \overset{(M1)}{=} \frac{\sum_{i=1}^{n+1-k} Var\left[C_{i,k} \mid C_{i,1}, \dots C_{i,k-1}\right]}{\left(\sum_{i=1}^{n+1-k} C_{i,k-1}\right)^2} = \frac{\sigma_k^2}{\sum_{i=1}^{n+1-k} C_{i,k-1}}.$$

□

**Theorem 3.10.** *Under the Assumptions 3.2 it holds*

$$\hat{\sigma}_k^2 = \frac{1}{n-k} \sum_{i=1}^{n+1-k} C_{i,k-1} \left(\frac{C_{i,k}}{C_{i,k-1}} - \hat{f}_k\right)^2$$

*is an unbiased estimator for variance parameter $\sigma_k^2 = Var[F_{i,k} \mid C_{i,1}, \dots, C_{i,k-1}] \cdot C_{i,k-1}$ for all $k = 2, \dots, n-1$.*

*Proof.* The proof for unbiasedness of the variance parameter proceeds similar as described for the additive method. It holds by the tower property

$$E\left[\hat{\sigma}_k^2\right] = E\left[E\left[\hat{\sigma}_k^2 \mid C_{i,1}, \dots C_{i,k-1}\right]\right]$$

and by independence assumption (M1)

$$E\left[\hat{\sigma}_k^2 \mid C_{i,1}, \dots C_{i,k-1}\right] = \frac{1}{n-k} \sum_{i=1}^{n-k+1} C_{i,k-1} \cdot E\left[\left(\frac{C_{i,k}}{C_{i,k-1}} - \hat{f}_k\right)^2 \,\middle|\, C_{i,1}, \dots C_{i,k-1}\right]. \tag{3.13}$$

We consider only

$$E\left[\left(\frac{C_{i,k}}{C_{i,k-1}} - \hat{f}_k\right)^2 \,\middle|\, C_{i,1}, \dots C_{i,k-1}\right] = E\left[\left(\frac{C_{i,k}}{C_{i,k-1}}\right)^2 \,\middle|\, C_{i,1}, \dots C_{i,k-1}\right] + \tag{3.14}$$
$$+ E\left[\hat{f}_k^2 \,\middle|\, C_{i,1}, \dots C_{i,k-1}\right] - 2 \cdot E\left[\frac{C_{i,k}}{C_{i,k-1}} \cdot \hat{f}_k \,\middle|\, C_{i,1}, \dots C_{i,k-1}\right].$$

The first term of (3.14) is given by

$$E\left[\left(\frac{C_{i,k}}{C_{i,k-1}}\right)^2 \,\middle|\, C_{i,1}, \dots C_{i,k-1}\right] =$$
$$= Var\left[\frac{C_{i,k}}{C_{i,k-1}} \,\middle|\, C_{i,1}, \dots C_{i,k-1}\right] + E\left[\frac{C_{i,k}}{C_{i,k-1}} \,\middle|\, C_{i,1}, \dots C_{i,k-1}\right]^2 = \frac{\sigma_k^2}{C_{i,k-1}} + f_k^2,$$

where the second step follows by model assumptions (M2) and (M3). The last term of (3.14) is by Theorems 3.10 and 3.9

$$\begin{aligned} E\left[\hat{f}_k^2 \mid C_{i,1},\dots C_{i,k-1}\right] &= Var\left[\hat{f}_k \mid C_{i,1},\dots C_{i,k-1}\right] + E\left[\hat{f}_k \mid C_{i,1},\dots C_{i,k-1}\right]^2 \\ &= \frac{\sigma_k^2}{\sum_{i=1}^{n+1-k} C_{i,k-1}} + f_k^2. \end{aligned}$$

Next, we calculate the remaining part on the right-hand side of the above equality (3.14)

$$\begin{aligned} E\left[\frac{C_{i,k}}{C_{i,k-1}} \cdot \hat{f}_k \,\middle|\, C_{i,1},\dots C_{i,k-1}\right] &= E\left[\frac{C_{i,k}}{C_{i,k-1}} \cdot \frac{\sum_{j=1}^{n+1-k} C_{j,k}}{\sum_{j=1}^{n+1-k} C_{j,k-1}} \,\middle|\, C_{i,1},\dots C_{i,k-1}\right] \\ &= \frac{\sum_{j=1}^{n+1-k} E\left[C_{i,k} \cdot C_{j,k} \mid C_{i,1},\dots C_{i,k-1}\right]}{C_{i,k-1} \cdot \sum_{j=1}^{n+1-k} C_{j,k-1}}, \end{aligned} \tag{3.15}$$

where we used linearity of the conditional expectation and measurability of $C_{i,k-1}$. Again by independence of underwriting years it holds

$$E\left[C_{i,k} \cdot C_{j,k} \mid C_{i,1},\dots C_{i,k-1}\right] = \begin{cases} E\left[C_{i,k} \mid C_{i,1},\dots C_{i,k-1}\right] \cdot E\left[C_{j,k} \mid C_{i,1},\dots C_{i,k-1}\right], & i \neq j, \\ E\left[C_{i,k}^2 \mid C_{i,1},\dots C_{i,k-1}\right], & \text{else.} \end{cases}$$

Together with

$$E\left[C_{i,k}^2 \mid C_{i,1},\dots C_{i,k-1}\right] = Var\left[C_{i,k} \mid C_{i,1},\dots C_{i,k-1}\right] + E\left[C_{i,k} \mid C_{i,1},\dots C_{i,k-1}\right]^2,$$

this yields for (3.15)

$$\begin{aligned} &\frac{Var\left[C_{i,k} \mid C_{i,1},\dots C_{i,k-1}\right] + \sum_{j=1}^{n+1-k} E\left[C_{i,k} \mid C_{i,1},\dots C_{i,k-1}\right] E\left[C_{j,k} \mid C_{i,1},\dots C_{i,k-1}\right]}{C_{i,k-1} \cdot \sum_{j=1}^{n+1-k} C_{j,k-1}} = \\ &= \frac{C_{i,k-1}\sigma_k^2 + \sum_{j=1}^{n+1-k} C_{i,k-1} f_k \cdot C_{j,k-1} f_k}{C_{i,k-1} \cdot \sum_{j=1}^{n+1-k} C_{j,k-1}} = \frac{\sigma_k^2}{\sum_{j=1}^{n+1-k} C_{j,k-1}} + f_k^2, \end{aligned}$$

where in the second step we used (3.12) and (3.11) once again. Putting the three parts of (3.14) together gives

$$\begin{aligned} &E\left[\left(\frac{C_{i,k}}{C_{i,k-1}} - \hat{f}_k\right)^2 \,\middle|\, C_{i,1},\dots C_{i,k-1}\right] = \\ &\qquad = \frac{\sigma_k^2}{C_{i,k-1}} + f_k^2 - 2\left(\frac{\sigma_k^2}{\sum_{j=1}^{n+1-k} C_{j,k-1}} + f_k^2\right) + \frac{\sigma_k^2}{\sum_{i=1}^{n+1-k} C_{i,k-1}} + f_k^2 \\ &\qquad = \sigma_k^2\left(\frac{1}{C_{i,k-1}} - \frac{1}{\sum_{j=1}^{n+1-k} C_{j,k-1}}\right). \end{aligned}$$

Using this for equation (3.13) results in

$$E\left[\hat{\sigma}_k^2 \mid C_{i,1},\dots C_{i,k-1}\right] = \sigma_k^2,$$

which proves the claim. □

Now with the result of Theorem 3.10 we get

$$\hat{C}_{i,k} = C_{i,k-1} \cdot \hat{f}_k \quad \text{for all } k \geq n - i + 1 \tag{3.16}$$

as unbiased estimator for $E\left[C_{i,k} \mid C_{i,1}, \dots C_{i,k-1}\right]$.
Furthermore, conditional on the loss history $\mathcal{D} = \{C_{i,k} \mid 1 \leq i + k - 1 \leq n\}$ it holds for underwriting years $i = 2, \dots, n$

$$\begin{aligned} E\left[C_{i,n} \mid \mathcal{D}\right] &= E\left[E\left[C_{i,n} \mid C_{i,1}, \dots C_{i,n-1}\right] \mid \mathcal{D}\right] = E\left[C_{i,n-1} \cdot f_n \mid \mathcal{D}\right] = \\ &= E\left[C_{i,n-1} \mid \mathcal{D}\right] \cdot f_n = \dots = \left[C_{i,n-i+1} \mid \mathcal{D}\right] \cdot f_{n-i+2} \cdot \dots \cdot f_n = \\ &= C_{i,n-i+1} \cdot \prod_{k=n-i+2}^{n} f_k. \end{aligned} \tag{3.17}$$

Therefore, unbiased estimators for the ultimate cumulative losses are

$$\hat{C}_{i,n} = C_{i,n-i+1} \cdot \prod_{k=n-i+2}^{n} \hat{f}_k, \qquad i = 2, \dots, n, \tag{3.18}$$

and finally by Corollary 3.4 the predicted reserve in the multiplicative model based on loss triangle $\mathcal{D}$ is given by

$$\hat{R}_i = C_{i,n-i+1} \cdot \left( \prod_{k=n-i+2}^{n} \hat{f}_k - 1 \right), \qquad i = 2, \dots, n. \tag{3.19}$$

## 3.4 Reserving models as weighted linear regression models

In this section we show that the additive and multiplicative reserving models can be interpreted as weighted linear regression models. For this purpose we use the results derived in Chapter 2. There we introduced weighted linear regression models by some linear relationship

$$Y_i = \beta_{w,0} + \beta_{w,1} x_{i,1} + \dots \beta_{w,k} x_{i,k} + \epsilon_i, \qquad i = 1, \dots, n,$$

and error terms $\epsilon_i$ with

$$E[\epsilon_i] = 0 \qquad \text{and} \qquad Var[\epsilon_i] = \sigma_w^2 w_i$$

for known weights $w_1, \dots, w_n$.
We want to describe the additive and multiplicative reserving model in such a form and apply the parameter estimation methods discussed in the previous chapter to derive parameter estimators. Then, we compare the estimators to the unbiased estimators given in the previous Section 3.3.

## Additive regression model

For a fixed development year $k \in \{1, \dots, n\}$ by Assumptions 3.1 of the additive model the loss ratios $M_{i,k}$ of the upper triangle can be formulated as

$$M_{i,k} = m_k + \epsilon_{i,k}, \qquad i = 1, \dots, n-k+1, \tag{3.20}$$

with some independent random variables $\epsilon_{i,k}$ so that

$$E[\epsilon_i] = 0 \qquad \text{and} \qquad Var[\epsilon_i] = \frac{s_k^2}{v_i}.$$

Note that assumption $E[\epsilon_i] = 0$ implies $E[M_{i,k}] = m_k$, which corresponds to (A2), $Var[\epsilon_{i,k}] = \frac{s_k^2}{v_i}$ gives (A3) and independence is required for (A1). This model fits to the linear regression model defined in (2.4) for the loss ratios as response and known weights given by $w_i = 1/v_i$ with known premiums $v_1, \dots, v_n$. From Section 2.2 we know that we can estimate $m_k$ by a weighted least squares estimation using (2.5). Minimizing the weighted sum of squared residuals

$$S_w(m_k) = \sum_{i=1}^{n-k+1} v_i \left(M_{i,k} - m_k\right)^2$$

leads

$$\begin{aligned} \frac{\partial S_w(m_k)}{\partial m_k} = 0 \qquad &\Leftrightarrow \qquad \sum_{i=1}^{n-k+1} 2v_i \left(M_{i,k} - m_k\right)(-1) = 0 \\ &\Leftrightarrow \qquad m_k = \frac{\sum_{i=1}^{n-k+1} v_i M_{i,k}}{\sum_{i=1}^{n-k+1} v_i} = \frac{\sum_{i=1}^{n-k+1} S_{i,k}}{\sum_{i=1}^{n-k+1} v_i}. \end{aligned}$$

It follows that the weighted least squares estimators correspond to the unbiased estimators $\hat{m}_k$ given in Theorem 3.5. Under the assumption of normally distributed errors we can also state the maximum likelihood estimators. By formula (2.9) for an intercept-only model, we have

$$\hat{m}_k = \frac{\sum_{i=1}^{n-k+1} v_i M_{i,k}}{\sum_{i=1}^{n-k+1} v_i},$$

which also leads to the parameter estimators in the presented reserving model of Mack (2002). This is not surprising as we already showed in Section 2.2 that maximum likelihood estimation yields the same result as least squares estimation for normally distributed variables. The maximum likelihood estimator for the variance is by formula (2.10) given by

$$\hat{s}_k^2 = \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} v_i (M_{i,k} - \hat{m}_k)^2.$$

## Multiplicative regression model

Similarly, we can define the multiplicative model for a fixed development year $k \in \{2, \dots, n\}$ in the form

$$F_{i,k} = f_k + \epsilon_{i,k}, \quad i = 1, \dots, n-k+1, \tag{3.21}$$

again with independent random errors $\epsilon_{i,k}$, which satisfy

$$E[\epsilon_i] = 0 \qquad \text{and} \qquad Var[\epsilon_i] = \frac{\sigma_k^2}{C_{i,k-1}}$$

for given previous losses $C_{i,k-1}$. This can be derived from the model assumptions (M1)-(M3) in the multiplicative model. Hence, also the multiplicative model can be connected to weighted regression by using weights $w_i = 1/C_{i,k-1}$ for all $i = 1, \dots, n-k+1$ and development factors as response. Minimizing the weighted sum of squared residuals

$$S_w(f_k) = \sum_{i=1}^{n-k+1} C_{i,k-1} \left(F_{i,k} - f_k\right)^2$$

yields again the unbiased estimators of Mack's model

$$\hat{f}_k = \frac{\sum_{i=1}^{n-k+1} C_{i,k}}{\sum_{i=1}^{n-k+1} C_{i,k-1}}$$

as defined in Theorem 3.10. We know that also the maximum likelihood estimation results in the same estimators if normally distributed errors are assumed, which is easily confirmed by applying formula (2.9) again. Formula (2.10) again gives the corresponding maximum likelihood estimator for the variance parameter

$$\hat{\sigma}_k^2 = \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} C_{i,k-1}(F_{i,k} - \hat{f}_k)^2.$$

In conclusion, we showed that both the additive and the multiplicative model can be expressed in forms of a linear regression with heteroscadastic error terms. That means we can estimate parameters based on the least squares approach and under the assumption of a normal distribution also based on maximum likelihood estimation.
We will use this result later and additionally extend the regression model in case an intercept-only model is not sufficient to characterize the loss development. A motivation for this extension is given in the next section, but the approach will be presented later. Before we can gain further insight on this matter in Chapter 6, we finalize this chapter with some specifications of the loss model, which is required as basis for the next chapter.

## 3.5 Model Selection

We introduced two different models to describe the loss reserve and both are stochastic models, which allows to quantify the uncertainties. But before uncertainties can be investigated in the next chapter, one model for the portfolio of basic losses has to be fixed based on that the parameters are estimated and the reserve is predicted. This choice contains the remaining model decisions we outline in this section. We start with the choice between the additive and multiplicative model as underlying reserving method.

### Selection of the reserving model

Table 3 shows a comparison of both models and it gives an overview on the estimated parameters in both models. Both introduced reserving models of Mack (2002) are stochastic, distribution-free models and assume a common development pattern for all underwriting years. To choose one of both models we have to check the model assumptions for the observed loss triangle $\mathcal{D}$.
A first indication which model fits better can be examined by analyzing the observed loss development by using two different plots. The first plot helps to assess if the loss

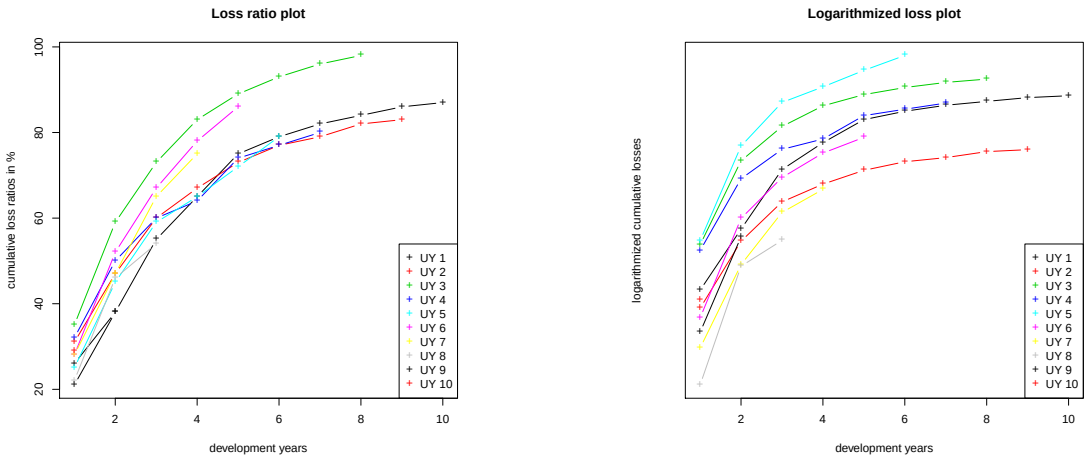



Figure 1: Plots of cumulative loss ratios $C_{i,k}/v_i$ (left) and logarithmized cumulative losses $\log(C_{i,k})$ (right) per underwriting year $i$ against the development years $k = 1, \ldots, n-i+1$.

pattern is compatible with the additive model and the second one associated with the multiplicative model. To validate the assumption that all underwriting years have similar incremental loss ratios per development year it is helpful to plot the loss ratios versus the development years but in cumulative form. This means we plot the losses in cumulative form $C_{i,k}$ (instead of incremental losses) divided by the premiums $v_i$ per underwriting year $i = 1, \ldots, n$ for all observed development years $k = 1, \ldots, n-i+1$. The cumulative loss ratios $C_{i,k}/v_i$ correspond to the sum of incremental loss ratios $M_{i,1} + \ldots + M_{i,k}$ up to development year $k$. Parallel lines in this plot indicate similar additive loss development. The multiplicative assumption can be assessed via the plot of the logarithmized cumulative losses $\log(C_{i,k})$ for all observed losses $C_{i,k}$ against the development years $k$. Again

| | | **Additive** | **Multiplicative** |
|---|---|---|---|
| Data | Used data: | incremental losses & premiums | cumulative losses |
| | Central object: | loss ratios | development factors |
| Assumptions | Independence: | underwriting years & development years | underwriting years |
| | Mean: | $E\left[S_{i,k}\right] = v_i \cdot m_k$ | $E\left[C_{i,k} \mid C_{i,1}, \dots C_{i,k-1}\right] = C_{i,k-1} \cdot f_k$ |
| | Variance: | $Var\left[S_{i,k}\right] = s_k^2 \cdot v_i$ | $Var\left[C_{i,k} \mid C_{i,1}, \dots C_{i,k-1}\right] = C_{i,k-1} \cdot \sigma_k^2$ |
| Unbiased estimated parameters | Mean: | $\hat{m}_k = \frac{\sum_{i=1}^{n-k+1} S_{i,k}}{\sum_{i=1}^{n-k+1} v_i}$ | $\hat{f}_k = \frac{\sum_{i=1}^{n-k+1} C_{i,k}}{\sum_{i=1}^{n-k+1} C_{i,k-1}}$ |
| | Variance: | $\hat{s}_k^2 = \frac{1}{n-k} \sum_{i=1}^{n+1-k} v_i \left(\frac{S_{i,k}}{v_i} - \hat{m}_k\right)^2$ | $\hat{\sigma}_k^2 = \frac{1}{n-k} \sum_{i=1}^{n+1-k} C_{i,k-1} \left(\frac{C_{i,k}}{C_{i,k-1}} - \hat{f}_k\right)^2$ |
| Prediction | Losses: | $\hat{S}_{i,k} = v_i \cdot \hat{m}_k$ | $\hat{C}_{i,k} = C_{i,k-1} \cdot \hat{f}_k$ |
| | Ultimate: | $\hat{C}_{i,n} = C_{i,n+1-i} + v_i \sum_{k=n+2-i}^{n} \hat{m}_k$ | $\hat{C}_{i,n} = C_{i,n+1-i} \cdot \prod_{k=n+2-i}^{n} \hat{f}_k$ |
| | Reserve: | $\hat{R}_i = v_i \sum_{k=n+2-i}^{n} \hat{m}_k$ | $\hat{R}_i = C_{i,n+1-i} \cdot \left(\prod_{k=n+2-i}^{n} \hat{f}_k - 1\right)$ |

Table 3: Comparison between additive and multiplicative reserving model

parallelism indicates development according to the model assumptions. These two plots are shown in Figure 1 for an exemplary triangle $\mathcal{D}$ with 10 underwriting years.
In both plots the triangular form of the data is recognizable as each line represents the observed development of one underwriting year. These plots give an impression of the total model fit in the additive and the multiplicative model. In the plots in Figure 1 there is no clear indication that the assumptions are violated neither in the additive nor the

multiplicative model, since the development is at least roughly parallel.
In order to make sure that the assumptions hold for each development year $k = 1, \ldots, n$ the incremental loss ratios and development factors of each development year should be considered in a separate plot. To check whether the observed loss ratios are similar per development year $k$ or not, i.e. to evaluate assumption

$$E\left[\frac{S_{i,k}}{v_i}\right] = E\left[M_{i,k}\right] = m_k, \quad i = 1, \ldots, n-k+1,$$

we plot the incremental loss ratios $M_{i,k}$ of this development year versus the underwriting years $i$. For the multiplicative model we plot the development factors $F_{i,k}$ versus the underwriting years $i$ to verify the assumption

$$E\left[\frac{C_{i,k}}{C_{i,k-1}} \;\middle|\; C_{i,1}, \ldots C_{i,k-1}\right] = E\left[F_{i,k} \mid C_{i,1}, \ldots C_{i,k-1}\right] = f_k, \qquad i = 1, \ldots, n-k+1.$$

For the exemplary loss triangle used for Figure 1 these plots are shown in Figure 2. In the

(a) Loss ratios per DYs (b) Development factors per DYs

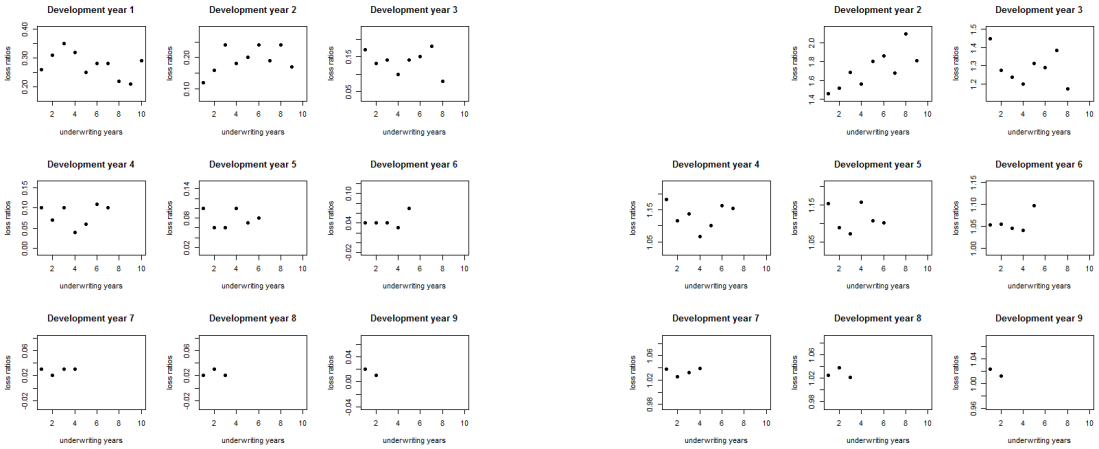



Figure 2: Plots of incremental loss ratios $M_{i,k}$ (left) and development factors $F_{i,k}$ (right) versus the underwriting years $i$ for each development year $k$.

optimal case, all loss ratios and development factors are approximately at the same level per development year, which indicates equal mean values. Otherwise, these plots can hint violations in the model assumptions. The question arises of how this can be objectively evaluated and how to deal with it in case model violations are identified. Consequently, we will consider the analysis of the loss history in the next paragraph of this section more detailed.
Furthermore, the assumptions of independence in both models has to be verified. For this we refer to Mack (2002), who elaborates in Section 3.2.7 of his book a method to check this condition.
Sometimes the additive method fits better the early loss development until the major part of the losses developed and for later development years the loss development is better described by a multiplicative model. Thus, it can be also reasonable to use a mixture of both

methods. This is possible since the modeling of the loss development proceeds column by column. Hence, we can assume the additive model for the first development years until some development year, called transition point $t$, $1 \leq t \leq n$, which specifies for all columns up to and including this point the additive model and afterwards a multiplicative model. The first development year is always modeled additive, since the multiplicative method provides parameters as of the second development year. For loss reserving purposes this is sufficient, but modeling the uncertainty requires additionally the consideration of the first loss of the upcoming underwriting year.

## Analysis of the loss history

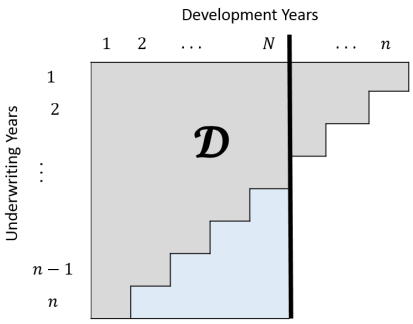
As discussed in this chapter, the aim of a loss reserving model is to use the observed loss development to predict the future losses. This is based on the assumption that older years are similar to newer years in terms of the assumptions concluded in Table 3. If this is violated, then the models can produce unreliable results. It follows the importance to make sure that the data is not disturbed by structural changes in the loss development.
Reasons for changes in the loss pattern can be changes in claims management or a change in legislation that impacts the loss amounts. Such a change in a calendar year could affect all losses in this calendar year and all subsequent losses, which divides the triangle into two parts with divergent loss patterns. It is also possible that instead of a calendar year effect, a break in behavior before and after an underwriting year is visible in the loss triangle. This can happen when the composition of the portfolio changes.
Another scenario that can cause insufficient estimates is data affected by trends. If the data of a development year, which is assumed to have a constant mean, shows a significant increasing or decreasing behavior from older to newer underwriting years instead, this should not be neglected. For example, the ongoing improvement in safety in cars can lead to fewer car accidents, or (to stay with the example of a car insurance) a portfolio can change for the worse due to a gradual increase in young drivers over time.
These are just a few of the possible examples that can produce structural changes and violate the model assumptions, what makes clear why the loss history has to be analyzed to investigate whether disturbances are apparent in the data. If not, then we ensured the model assumption is not violated and the original reserving models introduced in this chapter are usable without any adjustment. But otherwise, in the presence of trends or change points or even both, the model assumption of equal means per development year does not hold.
In practice, the loss history is analyzed and often a starting year different to the first available underwriting year is determined. This year is used for the calibration of the model and the data before that year is discarded. The idea is to avoid the influence of the very old years with varying pattern and to use only the part of data that is assumed to be representative for the future development. This choice is usually based on the expert judgment of an actuary and at least change points can be handled by this procedure. But no analytical method has been established yet, which objectively detects changes and select the optimal starting point. Furthermore, this adjustment does not solve the issue with trends. In most commonly used standard methods there is no possibility to take trends into account. In Chapter 5 we will introduce a statistical method that detects changes

in data and thereafter in Chapter 6 we present how to apply these ideas to our models to investigate whether any changes are contained and extend the approach by trends in order to propose a common solution for both problems. There we will discuss in detail how to adjust the model to get appropriate results.
But before we can predict the reserve and quantify its uncertainty in the next chapter, we have to examine if all losses are settled after $n$ DYs as previously assumed or if we have to extend or can reduce the number of material development years.

## Assessment of the ultimate development year

Sometimes the loss development shows a slightly different form than introduced in Table 1. Given the loss history $\mathcal{D}$ with the loss development of the past $n$ underwriting years, we assumed that all losses fully settled after $n$ development years with the typical lower triangle to predict.
In practice it is possible that the number of development years until all losses of an underwriting year are fully settled, called ultimate development year and denoted by $N$, is not equal to the given number of available underwriting years $n$. This is shown in Figure 3. Case (a) holds, if a long history of underwriting years is given and these years indicate

(a) Shorter loss development




(b) Longer loss development

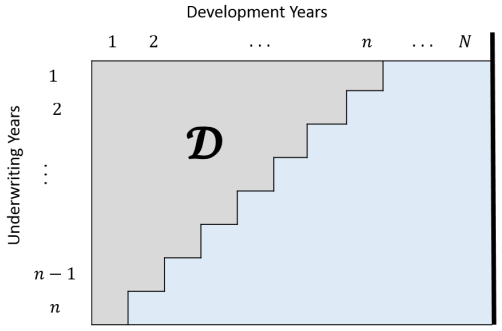



Figure 3: Ultimate development year for a short loss development (a) vs. for a long loss development (b)

all losses are already settled before development year $n$ after $N < n$ years and there is no further development after development year $N$. In this case we do not expect losses after $N$ development years for newer underwriting years, too, so we can neglect the development years $N + 1, \ldots, n$. The outstanding losses we have to predict are then given by a smaller triangle, as pictured by the light blue part in the figure.
If in contrast even for older underwriting years still outstanding losses are expected, since the loss development is still persisting, it is not enough to consider only the loss settlement until development year $n$. We have to extend the prediction of future losses until fully developed after development year $N > n$, as shown in case (b) of the figure.
To choose an appropriate ultimate development year, it is useful to plot the observed cumulative losses $C_{i,k}$ versus the development years $k = 1, \ldots, n$ as displayed in Figure 4. It shows two different examples with obviously different observed loss development. While the left plot indicates a short loss development, since the cumulative losses do not

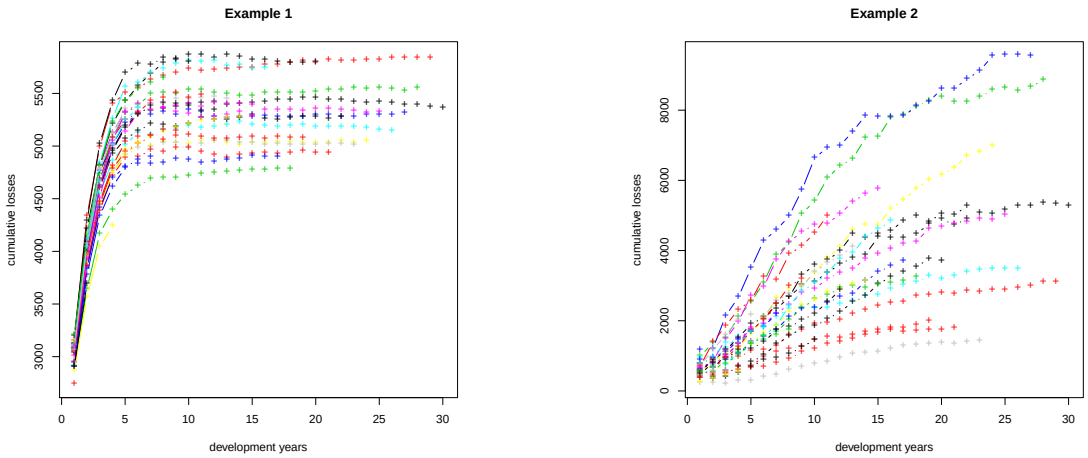



Figure 4: Cumulative losses $C_{i,k}$ against development years $k$ for all underwriting years with short (left) vs. long (right) settlement durations

further increase after a few development years, the right one shows a portfolio with long duration of claims settlement. In the right example we still have to expect losses after 30 development years, so we have to assume ultimate development year $N > n$. On the other hand for Example 1, shown in the left plot of Figure 4, it is obviously that 30 development years are more than enough to model the claims until fully settled.
Under these aspects, we have to adjust the definition of the reserve in case $N \neq n$. By definition of the ultimate development year it follows that the ultimate losses are given by $C_{i,N}$ and the unknown ultimate losses can be estimated by

$$\hat{C}_{i,N} = C_{i,n+1-i} + v_i \sum_{i=n-i+2}^{N} \hat{m}_k \qquad \text{for all } i = 1, \ldots, n \quad \text{with} \quad n - i + 2 \leq N$$

in the additive model or

$$\hat{C}_{i,N} = C_{i,n+1-i} \cdot \prod_{i=n-i+2}^{N} \hat{f}_k \qquad \text{for all } i = 1, \ldots, n \quad \text{with} \quad n - i + 2 \leq N$$

in the multiplicative model. Given the loss history $\mathcal{D}$ the unknown reserve for an underwriting year $i$ is predicted by

$$R_i = \hat{C}_{i,N} - C_{i,n-i+1} \qquad \text{for all } i = 1, \ldots, n \quad \text{with} \quad n - i + 2 \leq N.$$

Note that the parameter estimation is limited to the size $n$ of the given loss triangle $\mathcal{D}$. For $N \leq n$ we can estimate all parameters based on $\mathcal{D}$, but for $N > n$ an extrapolation of parameters is required. The extrapolation depends on the selected loss reserving model. In the additive model we assume the parameters $m_{n+1}, \ldots, m_N$ decrease exponentially to zero. This is estimated by a linear regression of $\ln(m_j)$ on $j$

$$\ln(m_j) = \alpha + \beta \cdot j + \epsilon_j, \qquad j = 1, \ldots, N.$$

The available $n$ estimated parameters $\hat{m}_1, \ldots, \hat{m}_n$ are used to estimate the regression parameters $a$ and $b$. Then the parameters $m_{n+1}, \ldots, m_N$ are estimated by

$$\hat{m}_l = \exp\left\{\hat{a} + \hat{b} \cdot l\right\}, \qquad l = n+1, \ldots, N.$$

Similarly, in the multiplicative model we assume the parameters $f_{n+1}, \ldots, f_N$ decrease exponentially to one. Using a linear regression of $\ln(f_j - 1)$ on $j$

$$\ln(f_j - 1) = \gamma + \delta \cdot j + \epsilon_j, \qquad j = 2, \ldots, N,$$

yields analogously to above estimated tail parameters

$$\hat{f}_l = 1 + \exp\left\{\hat{\gamma} + \hat{\delta} \cdot l\right\}, \qquad l = n+1, \ldots, N,$$

for some estimated regression parameters $\gamma$ and $\delta$. Analogously we can extrapolate the variance parameters $s_k$ and $\sigma_k$ by a linear regression of $\ln(s_k)$ or $\ln(\sigma_k)$ on $k$, since it is assumed that these parameters also decrease exponentially to zero.
In conclusion, our basic loss model is a stochastic model defined by the given loss history $\mathcal{D}$ of a portfolio of basic losses, the selected reserving model (if necessary adjusted because of disturbance in data) and the ultimate development year. Based on this we can estimate model parameters in order to predict the outstanding claims and therefore the required reserve. In the next chapter we are interested in assessing the uncertainty in our prediction.

# Chapter 4

# Quantification of the prediction uncertainty

The modeling process for each basic loss portfolio is divided into two parts, first the selection of the loss model and the prediction of the reserve and then the quantification of the risk that future losses develop differently than predicted. The specification of the loss model and prediction of the reserve has already been described in the previous chapter, hence, we want to consider the uncertainty now. Given a loss model the comprised risk can be assigned to two sources of uncertainty, the reserve risk and the premium risk.
At first, the formal specifications by Solvency II are introduced, which yield that the risk can be assessed by quantifying the prediction uncertainty of the claims development result that we will define in Section 4.2. One possible way to measure the uncertainty in the claims development result of basic losses within an internal model is to construct an empirical distribution for the claims development using the simulation approach presented at the end of this chapter. Before we describe this method, the first two moments required for the simulation are derived.

## 4.1 Solvency II specifications

In Article 101 of the Solvency II Directive (2009) the risk-based capital insurance companies are required to hold is defined as:

> The Solvency Capital Requirement shall be calibrated so as to ensure that all quantifiable risks to which an insurance or reinsurance undertaking is exposed are taken into account. It shall cover existing business, as well as the new business expected to be written over the following 12 months. With respect to existing business, it shall cover only unexpected losses.
> It shall correspond to the Value-at-Risk of the basic own funds of an insurance or reinsurance undertaking subject to a confidence level of 99,5 % over a one-year period. [Article 101.3]

This means, according to Solvency II, insurance companies need to quantify the risk that the next calendar year develops differently as expected. The level is set in order to

ensure that any unexpected losses over a time horizon of one year can be covered with a probability of 99.5%, which limits the ruin probability to less than 0.5%. As mentioned in the introduction, part of the risk we have to quantify are the one-year premium and reserve risk. Within the framework of Solvency II in the Technical Specifications for the third Quantitative Impact Studies QIS3 (CEIOPS, 2007) the reserve risk is described as follows:

> Reserve risk stems from two sources: on the one hand, the absolute level of the claims provisions may be mis-estimated. On the other hand, because of the stochastic nature of future claims payouts, the actual claims will fluctuate around their statistical mean value. [I.3.229]

While the reserve risk refers to the provisions for existing business, the premium risk corresponds to the uncertainty caused by future claims from new business.

> Premium risk is understood to relate to future claims arising during and after the period until the time horizon for the solvency assessment. The risk is that expenses plus the volume of losses (incurred and to be incurred) for these claims (comprising both amounts paid during the period and provisions made at its end) is higher than the premiums received (or if allowance is made elsewhere for the expected profits or losses on the business, that the profitability will be less than expected). [I.3.226]

The premiums and expenses are assumed to be nonrandom, so we consider simplifying the uncertainty in the losses of the next underwriting year.
Overall, we see the calendar year at risk on the next diagonal in the loss development triangle.

## 4.2 The claims development result

Quantifying the risk in unexpected development in claims from existing business and next year's business, is about measuring the uncertainty in the possible change in reserve over one calendar year. Following Wüthrich et al (2009), we denote the difference between the predicted reserves based on the information available in one year and the predicted reserve now as claims development result. Reason for this change are the upcoming losses during the next calendar year, which means that the reserves will have to be recalculated in one year based on the additional information at that point in time.
We assume that all losses are fully settled after $N$ development years and in the interests of simplicity we assume that we have given the loss history of the last $N$ underwriting years, which leads to typical loss development structure. To quantify the uncertainty connected in the re-estimation of the reserves within one accounting year $(N, N+1]$, we differentiate between the information available now at time $t = 0$ using the loss history of the last $N$ underwriting years and one year later at time $t = 1$. Let

$$\mathcal{D}^{(t=0)} = \{C_{i,k} \mid i + k - 1 \leq N\}$$

denote the claims data available at time $t = 0$ and

$$\mathcal{D}^{(t=1)} = \{C_{i,k} \mid i + k - 1 \leq N + 1\}$$

denote the available information one year later. After one year a new loss diagonal is available as illustrated in Table 4.

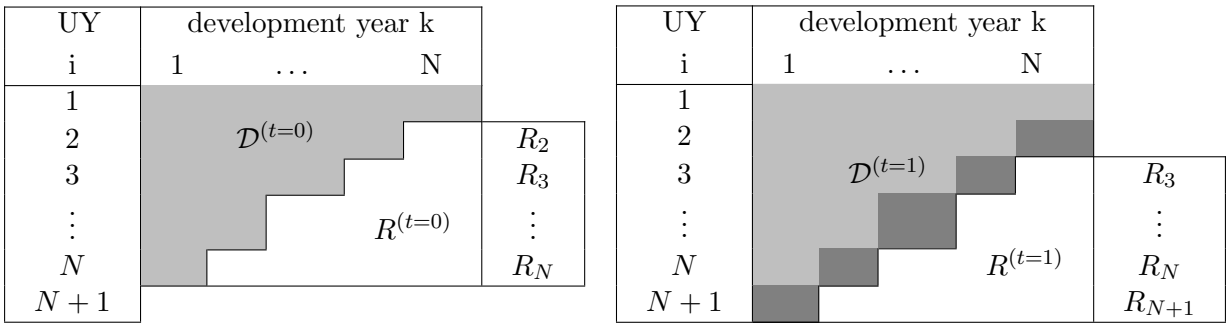



Table 4: Loss development triangle at time $t = 0$ (left) vs. one year later at time $t = 1$ (right)

Accordingly, the loss reserve for underwriting year $i$ for future claims at time $t = 0$ is given by

$$R_i^{(t=0)} = C_{i,N} - C_{i,N-i+1}, \qquad i = 2, \ldots, N, \tag{4.1}$$

and at time $t = 1$ by

$$R_i^{(t=1)} = C_{i,N} - C_{i,N-i+2}, \qquad i = 3, \ldots, N+1. \tag{4.2}$$

Note that the loss reserve of the first underwriting year is zero both at time $t = 0$ and $t = 1$, since it is assumed that all losses are fully developed after $N$ development years. Therefore, also the loss reserve for the second underwriting year is zero at time $t = 1$ because of the additional loss diagonal, but now for the next underwriting year $N + 1$ a loss reserve has to be considered.

**Definition 4.1.** *The true one-year claims development result (CDR) for underwriting year $i = 2, \ldots, N$ is given by*

$$CDR_i^{(t=1)} = E\left[R_i^{(t=0)} \mid D^{(t=0)}\right] - \left(S_{i,N-i+2} + E\left[R_i^{(t=1)} \mid D^{(t=1)}\right]\right),$$

*where $S_{i,N-i+2}$ are the incremental loss amounts between time $t = 0$ and $t = 1$, i.e. in calendar year $N + 1$. For underwriting year $N + 1$ the true one-year CDR is defined as*

$$CDR_{N+1}^{(t=1)} = -\left(S_{N+1,1} + E\left[R_{N+1}^{(t=1)} \mid D^{(t=1)}\right]\right).$$

By the formal characterizations in Section 4.1 the reserve risk refers to the aggregate claims development result

$$\sum_{i=2}^{N} CDR_i^{(t=1)}$$

and the premium risk to

$$CDR_{N+1}^{(t=1)}.$$

An equivalent formulation for the CDR is by (4.1) and (4.2)

$$CDR_i^{(t=1)} = E\left[C_{i,N} \mid D^{(t=0)}\right] - E\left[C_{i,N} \mid D^{(t=1)}\right] \qquad \text{for } i = 2, \dots, N$$

and

$$CDR_{N+1}^{(t=1)} = -E\left[C_{N+1,N} \mid D^{(t=1)}\right].$$

The true CDR is not observable, since the true model parameters, $m_1, \dots, m_N$ or respectively $f_2, \dots, f_N$, are not known. In order to predict the true CDR we therefore have to estimate $E\left[C_{i,N} \mid D^{(t=0)}\right]$ and $E\left[C_{i,N} \mid D^{(t=1)}\right]$ by some $\hat{C}_{i,N}^{(t=0)}$ and $\hat{C}_{i,N}^{(t=1)}$. Similarly, let $\hat{R}_i^{(t=0)}$ and $\hat{R}_i^{(t=1)}$ denote estimates for the expected loss reserve at time $t = 0$ and $t = 1$. Then, the following holds:

**Definition 4.2.** *The observable one-year CDR for underwriting year $i = 2, \dots, N$ is given by*

$$\hat{CDR}_i^{(t=1)} = \hat{R}_i^{(t=0)} - \left(S_{i,N-i+2} + \hat{R}_i^{(t=1)}\right) = \hat{C}_{i,N}^{(t=0)} - \hat{C}_{i,N}^{(t=1)}$$

*and for underwriting year $N + 1$ by*

$$\hat{CDR}_i^{(t=1)} = -\left(S_{N+1,1} + \hat{R}_{N+1}^{(t=1)}\right) = -\hat{C}_{N+1,N}^{(t=1)}.$$

To present this theory and in the naming of this claims development result we still follow Wüthrich et al (2009), but notice that this quantity is actual still not observable, but estimated.
In Section 3.3 we presented unbiased estimators for the cumulative ultimate losses and the reserves in the additive and multiplicative reserving model, which we can use now to estimate the CDR. At time $t = 0$ the loss triangle $\mathcal{D}^{(t=0)}$ is given, so the true model parameters $m_1, \dots, m_N$ and $f_2, \dots, f_N$ at time $t = 0$ can be estimated as outlined in Section 3.3 for $\mathcal{D} = \mathcal{D}^{(t=0)}$ and $n = N$. The estimated parameters at time $t = 0$ are denoted $\hat{m}_1^{(t=0)}, \dots, \hat{m}_N^{(t=0)}$ and $\hat{f}_2^{(t=0)}, \dots, \hat{f}_N^{(t=0)}$, respectively. Given the loss triangle $\mathcal{D}^{(t=0)}$ the expected cumulative loss after $N$ development years for underwriting year $i = 2, \dots, N$ depending on the specified reserving method is by model assumptions (A2) and (M2)

$$E\left[C_{i,N} \mid \mathcal{D}^{(t=0)}\right] = C_{i,N-i+1} + v_i \cdot \sum_{k=N-i+2}^{N} m_k$$

if the additive method is chosen or for the multiplicative method by

$$E\left[C_{i,N} \mid \mathcal{D}^{(t=0)}\right] = C_{i,N-i+1} \cdot \prod_{k=N-i+2}^{N} f_k.$$

In Section 3.5 we mentioned the possibility that the loss model has a transition point $t$ from an additive to multiplicative loss model. In this case the expected ultimate cumulative loss conditional on $\mathcal{D}^{(t=0)}$ is for all underwriting years $i$ with $N - i + 1 \geq t$ equal to the pure multiplicative case, since the additive method is used only until $t$, and otherwise

$$E\left[C_{i,N} \mid \mathcal{D}^{(t=0)}\right] = \left(C_{i,N-i+1} + v_i \cdot \sum_{k=N-i+2}^{t} m_k\right) \cdot \prod_{k=t+1}^{N} f_k, \qquad i = N-t+2, \ldots, N.$$

This can be estimated by using the estimated parameters $\hat{m}_1^{(t=0)}, \ldots, \hat{m}_N^{(t=0)}$ and $\hat{f}_2^{(t=0)}, \ldots, \hat{f}_N^{(t=0)}$ in all three outlined cases. The predicted cumulative loss at time $t = 0$ is denoted by $\hat{C}_{i,N}^{(t=0)}$. The respective loss reserve at time $t = 0$ is, using (4.1), estimated by

$$\hat{R}_i^{(t=0)} = \hat{C}_{i,N}^{(t=0)} - C_{i,N+i-i} \qquad i = 2, \ldots, N.$$

To predict $C_{i,N}$ at time $t = 1$ one diagonal of information (c.f. Table 4) is missing, so we apply a simulation approach to simulate this loss diagonal with mean and variance corresponding to Mack's reserving models.
To simulate the future claims $C_{i,N-i+2}$ we use the expected value and variability at time $t = 0$. Using (3.3) or (3.11), depending on the selected model, the expected losses of the upcoming diagonal at time $t = 0$ are given by

$$E\left[C_{i,N-i+2} \mid \mathcal{D}^{(t=0)}\right] = C_{i,N-i+1} + v_i \cdot m_{N-i+2}$$

for an additive modeled diagonal loss or

$$E\left[C_{i,N-i+2} \mid \mathcal{D}^{(t=0)}\right] = C_{i,N-i+1} \cdot f_{N-i+2}$$

for a loss in the multiplicative model, for all underwriting years $i = 2, \ldots, N$. To predict the first loss of the newest underwriting year $N + 1$ neither a previous loss amount is available nor a multiplicative parameter for the first development is given, so we use always the additive model for the first development year and we have

$$E\left[C_{N+1,1} \mid \mathcal{D}^{(t=0)}\right] = v_{N+1} \cdot m_1.$$

Using now the unbiased estimates $\hat{m}_1^{(t=0)}, \ldots, \hat{m}_N^{(t=0)}$ and $\hat{f}_2^{(t=0)}, \ldots, \hat{f}_N^{(t=0)}$ at time $t = 0$ yields for all underwriting years $i = 2, \ldots, N$ the estimates

$$\hat{E}\left[C_{i,N-i+2} \mid \mathcal{D}^{(t=0)}\right] = \begin{cases} C_{i,N-i+1} + v_i \cdot \hat{m}_{N-i+2}^{(t=0)}, & \text{additive model} \\ C_{i,N-i+1} \cdot \hat{f}_{N-i+2}^{(t=0)}, & \text{multplicative model} \end{cases} \tag{4.3}$$

and for underwriting year $N + 1$

$$\hat{E}\left[C_{N+1,1} \mid \mathcal{D}^{(t=0)}\right] = v_{N+1} \cdot \hat{m}_1^{(t=0)}. \tag{4.4}$$

The estimation of the standard deviation to simulate the loss diagonal is described in the next section.

## 4.3 Conditional mean squared error of prediction

The future losses are random variables, hence, they include some uncertainty. The variability can be described by the mean squared error conditional on the known claims history $\mathcal{D} = \mathcal{D}^{(t=0)}$. To present this we follow beside Wüthrich et al (2009) again Mack (1993, 2002) in this section.

**Theorem 4.3.** *The mean squared error conditional on the known claims history $\mathcal{D} = \mathcal{D}^{(t=0)}$ of the predicted future losses $\hat{C}_{i,N-i+2}$ of the next calendar year for $i = 2, \ldots, N+1$ is given by*

$$\begin{aligned} mse(\hat{C}_{i,N-i+2}) &= E\left[(C_{i,N-i+2} - \hat{C}_{i,N-i+2})^2 \mid \mathcal{D}\right] = \\ &= Var[C_{i,N-i+2} \mid \mathcal{D}] + \left(E[C_{i,N-i+2} \mid \mathcal{D}] - \hat{C}_{i,N-i+2}\right)^2. \end{aligned} \tag{4.5}$$

*Proof.* We follow Mack (2002) and use the general formula

$$E\left[(X - h(Y))^2 \mid Y\right] = Var[X \mid Y] + (E[X \mid Y] - h(Y))^2$$

for $X = C_{i,N-i+2}$, $Y = \mathcal{D}$ and $h(\mathcal{D}) = \hat{C}_{i,N-i+2}$. □

The first term of (4.5) is the variance around the true value for the respective model, while the second term of equation (4.5) describes the difference between estimated value and the unknown true value. The latter can be estimated by

$$\begin{aligned} E\left[\left(E[C_{i,N-i+2} \mid \mathcal{D}] - \hat{C}_{i,N-i+2}\right)^2 \mid \mathcal{D}\right] &= E\left[\left(E[\hat{C}_{i,N-i+2} \mid \mathcal{D}] - \hat{C}_{i,N-i+2}\right)^2 \mid \mathcal{D}\right] \\ &= Var\left[\hat{C}_{i,N+i-2} \mid \mathcal{D}\right], \end{aligned}$$

since $E\left[\hat{C}_{i,N-i+1} \mid \mathcal{D}\right] = E\left[C_{i,N-i+1} \mid \mathcal{D}\right]$ holds by unbiasedness of estimates in both models. Therefore, the mean squared error (4.5) is given by

$$mse(\hat{C}_{i,N-i+2}) = Var[C_{i,N-i+2} \mid \mathcal{D}] + Var\left[\hat{C}_{i,N+i-2} \mid \mathcal{D}\right], \quad i = 2, \ldots, N+1. \tag{4.6}$$

In the **additive model** under the Assumptions 3.1 the second term is given by

$$Var\left[\hat{C}_{i,N-i+2} \mid \mathcal{D}\right] = Var\left[C_{i,N-i+1} + \hat{S}_{i,N-i+2} \mid \mathcal{D}\right] = Var[\hat{S}_{i,N-i+2}], \quad i = 2, \ldots, N+1,$$

since the previous cumulative loss $C_{i,N-i+1}$ is $\mathcal{D}$-measurable and $S_{i,N-i+2}$ is independent of $S_{i,1}, \ldots, S_{i,N-i+1}$ and thus independent of $\mathcal{D}$. Using (3.7) yields

$$Var[\hat{S}_{i,N-i+2}] = Var[v_i \cdot \hat{m}_{N-i+2}] = v_i^2 \cdot Var[\hat{m}_{N-i+2}].$$

Next we consider the first term of (4.6) in the additive model. It holds by equation (3.4)

$$Var\left[C_{i,N-i+2} \mid \mathcal{D}\right] = Var[S_{i,N-i+2}] = v_i \cdot s_{N-i+2}^2.$$

In the **multiplicative case** under the Assumptions 3.2, especially (3.12), the first term of (4.6) is given by

$$Var\left[C_{i,N-i+2} \mid \mathcal{D}\right] = C_{i,N-i+1} \cdot \sigma^2_{N-i+2}.$$

The previous cumulative loss $C_{i,N-i+1}$ is $\mathcal{D}$-measurable and by equation (3.16) the second term of (4.6) in the multiplicative model is given by

$$Var[\hat{C}_{i,N-i+2}] = C^2_{i,N-i+1} \cdot Var[\hat{f}_{N-i+2}].$$

Since the true parameters $s_k$ and $\sigma_k$ are unknown we estimate the conditional mean squared error by using the unbiased estimators presented in Section 3.3. Also the variance of the estimated parameters was determined in that section and an overview of all estimated parameters is given in Table 3. Therefore, the mean squared error for both models can be estimated by

$$\hat{mse}(\hat{C}_{i,N-i+2}) = \begin{cases} v_i \cdot \hat{s}^2_{N-i+1} + v_i^2 \cdot \hat{Var}[\hat{m}_{N-i+2}], & \text{additive,} \\ C_{i,N-i+1} \cdot \hat{\sigma}^2_{N-i+2} + C^2_{i,N-i+1} \cdot \hat{Var}[\hat{f}_{N-i+2} \mid \mathcal{D}], & \text{multiplicative.} \end{cases} \tag{4.7}$$

## 4.4 Stochastic re-reserving

In this section the main steps of the simulation approach re-reserving (Ohlsson and Lauzeningks, 2009; Kraus and Diers, 2010) are presented. In literature the approach is typically used for the one-year reserve risk, but we extend the approach to measure additionally the premium risk. The method consists of three steps, illustrated in Figure 5.

**Step 1: Loss reserve at time t=0**
At first, the best estimate of the cumulative ultimate losses $C_{i,N}$ conditional on the known loss triangle $\mathcal{D}^{(t=0)}$ at time $t = 0$ is calculated for all underwriting years $i = 1, \ldots, N$ associated to the selected model in the first part. This yields predictions $\hat{C}^{(t=0)}_{i,N}$ for the expected ultimate losses and the respective loss reserves $\hat{R}^{(t=0)}_i$ at time $t = 0$ as already described in Section 4.2. This is the best estimate of the outstanding losses at time $t = 0$.

**Step 2: Simulation of loss diagonal**
To predict the loss reserve at time $t = 1$ one diagonal in the loss triangle is missing, so the next step is to simulate the claims between time $t = 0$ and $t = 1$. One possibility to get the simulated diagonal is to simulate from a normal or log-normal distribution with mean given by (4.3) and (4.4) and variance by (4.7). This yields for every simulation run one possible realizations of the cumulative claims $C_{i,N-i+2}$ of the next calendar year for all underwriting years $i = 2, \ldots, N + 1$.

**Step 3: Loss reserve at time $t = 1$ per simulation path**
On basis of the additional loss diagonal of step 2 for each simulation run the reserve at time $t = 1$ can be estimated conditional on the loss triangle $\mathcal{D}^{(t=1)}$ of the related

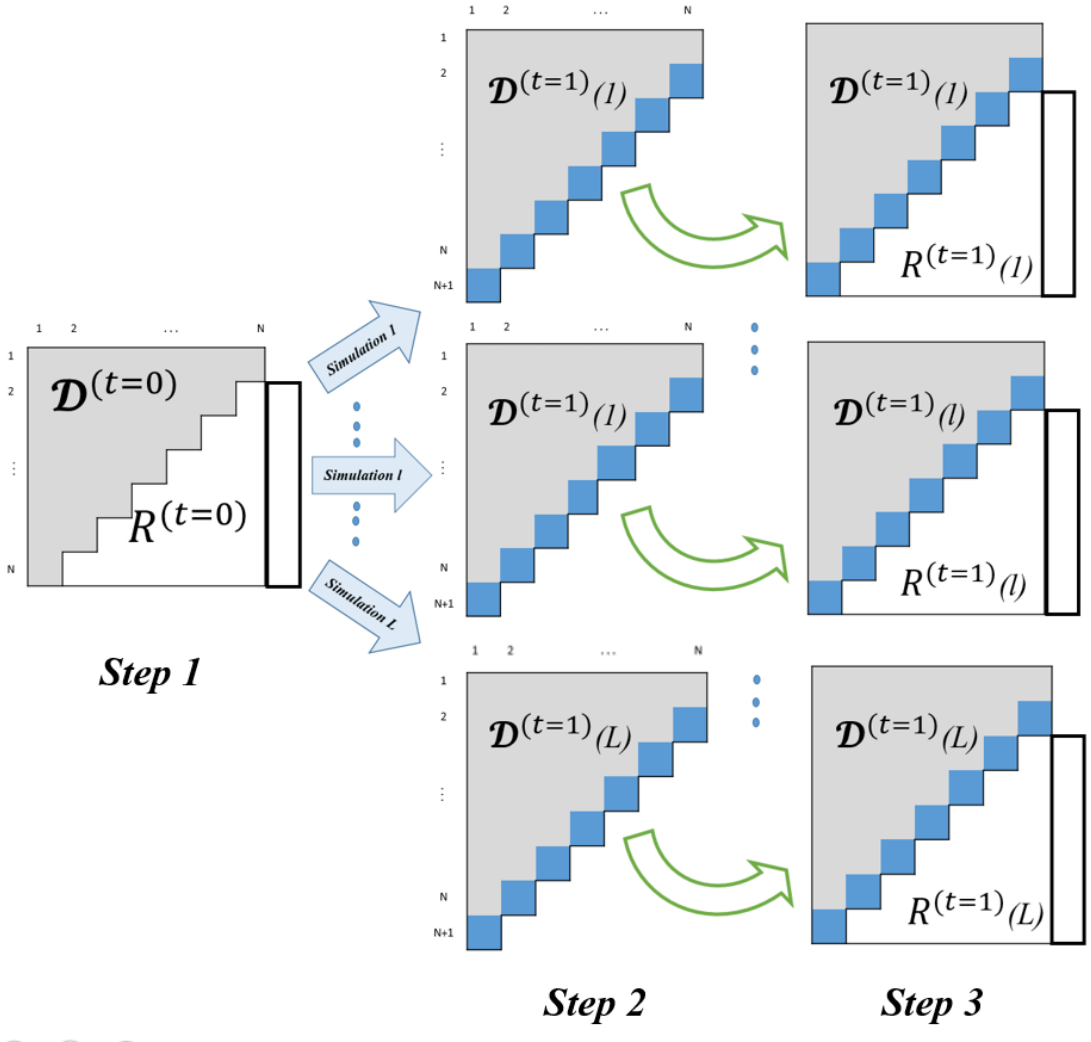



Figure 5: Re-Reserving Approach

simulation run. Similar to the parameter estimates in step 1 new model parameters at time $t = 1$ for each simulation run can be calculated according to Section 3.3 depending on loss triangles $\mathcal{D}^{(t=1)}$ in each simulation path. By this, estimates $\hat{C}_{i,N}^{(t=1)}(l)$ for the predicted cumulative ultimate loss are provided in each simulation. The respective loss reserve is the best estimate for the loss reserve at time $t = 1$ conditional on the loss triangle $\mathcal{D}^{(t=1)}$ for each simulation path.

To be more precise in step 3, let $L$ be the number of simulations and the realizations of simulation run $l$ with $l = 1 \ldots, L$ are denoted by $C_{i,N-i+2}(l)$. The upper triangle $\mathcal{D}^{(t=1)}(l)$ of observed losses in simulation run $l$ is then defined as

$$\mathcal{D}^{(t=1)}(l) \quad := \quad \{C_{i,k}(l) \mid i + k - 1 \leq N + 1\},$$

where

$$C_{i,k}(l) = \begin{cases} C_{i,k}, & i + k - 1 \leq N, \\ C_{i,k}(l), & i + k - 1 = N + 1. \end{cases}$$

This implies

$$\begin{aligned}\mathcal{D}^{(t=1)}(l) &= \{C_{i,k} : i+k-1 \leq N\} \cup \{C_{i,N-i+2}(l) \mid i = 2, \ldots, N+1\} \\ &= \mathcal{D}^{(t=0)} \cup \{C_{i,N-i+2}(l) \mid i = 2, \ldots, N+1\},\end{aligned}$$

which means the loss triangle used to estimate the new reserve is the loss triangle at time $t = 0$ extended by the realizations of the simulated diagonal. Accordingly, we can estimate new model parameters at time $t = 1$ in each simulation run $l$ by applying the introduced formulas from Section 3.3 based on $n = N + 1$ and loss triangle $\mathcal{D} = \mathcal{D}^{(t=1)}(l)$ for all $l = 1, \ldots, L$. The corresponding estimated parameters at $t = 1$ are denoted by $\hat{m}_k^{(t=1)}(l)$ and $\hat{f}_k^{(t=1)}(l)$.
Given the loss triangle $\mathcal{D}^{(t=1)}(l)$ the expected cumulative loss after $N$ development years for underwriting year $i = 3, \ldots, N+1$ depends on the specified reserving method and is

$$E\left[C_{i,N} \mid \mathcal{D}^{(t=1)}\right] = C_{i,N-i+2} + v_i \cdot \sum_{k=N-i+3}^{N} m_k$$

for the additive model or

$$E\left[C_{i,N} \mid \mathcal{D}^{(t=1)}\right] = C_{i,N-i+2} \cdot \prod_{k=N-i+3}^{N} f_k.$$

for the multiplicative model. With a transition point $t$ again for all underwriting years $i$ with development year of last cumulative loss before transition point, i.e. $N + 2 - i < t$, it holds

$$E\left[C_{i,N} \mid \mathcal{D}^{(t=1)}\right] = \left(C_{i,N-i+2} + v_i \cdot \sum_{k=N-i+3}^{t} m_k\right) \cdot \prod_{k=t+1}^{N} f_k.$$

Using the estimated unbiased parameters $\hat{m}_k^{(t=1)}(l)$ and $\hat{f}_k^{(t=1)}(l)$, predictions $\hat{C}_{i,N}^{(t=1)}(l)$ for the cumulative ultimate loss per underwriting year are provided in all simulation runs $l = 1 \ldots, L$. The respective loss reserve at time $t = 1$ is consequently estimated by

$$\hat{R}_i^{(t=1)}(l) = \hat{C}_{i,N}^{(t=1)}(l) - C_{i,N-i+2}(l), \qquad i = 3, \ldots, N+1,$$

where we used (4.2). The approach yields all components to calculate the observable CDR, stated in Definition 4.2, for each simulation. In simulation run $l = 1, \ldots, L$ it is given by

$$\hat{CDR}_i^{(t=1)}(l) = \hat{R}_i^{(t=0)}(l) - \left(S_{i,N-i+2}(l) + \hat{R}_i^{(t=1)}(l)\right) = \hat{C}_{i,N}^{(t=0)}(l) - \hat{C}_{i,N}^{(t=1)}(l)$$

for underwriting years $i = 2, \ldots, N$ and for underwriting year $N + 1$ by

$$\hat{CDR}_{N+1}^{(t=1)}(l) = -\left(S_{N+1,1}(l) + \hat{R}_{N+1}^{(t=1)}\right) = -\hat{C}_{N+1,N}^{(t=1)}(l).$$

The empirical distributions of the aggregate claims development result $\sum_{i=2}^{N} CDR_i^{(t=1)}$ and of the result of the newest underwriting year $CDR_{N+1}$ are used to estimate the probability distributions.

# Chapter 5

# Change point analysis background

Change point analysis is used to identify changes in structure of time series data. Structural changes can arise in a range of application fields, this makes the detection of change points to a highly interesting research area. Since its beginning by Page (1954) change point problems have received a considerable attention in statistical literature resulting in a wide range of problem formulations, methods and algorithms over the years. The interested reader is referred to Truong et al (2018) or Aminikhanghahi and Cook (2016) for more details.
We follow Chen and Gupta (2014) to introduce a general problem formulation of testing for change points at first. Killick et al (2012) propose a solution to this problem with different algorithms for an efficient change point detection. We discuss this method and two algorithms, before we consider a more specific method following Eckley et al (2011) that we will use in the next chapter for the change point detection in claims data. For this we have to discuss regression models with change points, which will complete this chapter.

## 5.1 Problem formulation

Let $Y_1, Y_2, \dots Y_T$ be a sequence of independent random variables with distribution functions $F_1, \dots, F_T$. In general the problem of testing a single change point can be defined as problem of hypothesis testing between null hypothesis of no change

$$\mathcal{H}_0 : F_1 = F_2 = \dots = F_T$$

against the alternative of there is a change point located at $\tau$ so that

$$\mathcal{H}_A : F_1 = \dots = F_\tau \neq F_{\tau+1} = \dots = F_T.$$

If the distributions differ only in parameters $\boldsymbol{\theta}$, but belong to the same family $F(\,\cdot\,)$ it is enough to test

$$\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \dots = \boldsymbol{\theta}_T$$

versus alternative

$$\mathcal{H}_A : \boldsymbol{\theta}_1 = \dots = \boldsymbol{\theta}_\tau \neq \boldsymbol{\theta}_{\tau+1} = \dots = \boldsymbol{\theta}_T.$$

As time series are often affected by more than one structural change, we are more interested in the problem of detecting multiple changes. This includes the problem of finding the optimal number of change points $m$ and their optimal location $\tau_1, \ldots, \tau_m$. For given $\tau_1, \ldots, \tau_m$ we can define the problem again as hypothesis testing between null hypothesis of no change

$$\mathcal{H}_0 : F_1 = F_2 = \ldots = F_T$$

against the alternative of $m$ change points located at $\tau_1, \ldots, \tau_m$ with

$$\begin{aligned} \mathcal{H}_A : F_1 = \ldots = F_{\tau_1} \neq F_{\tau_1+1} = \ldots = F_{\tau_2} &\neq F_{\tau_2+1} \ldots \\ \ldots F_{\tau_m-1} &\neq F_{\tau_m} = \ldots = F_T. \end{aligned}$$

Or in the parametric case we have

$$\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \ldots = \boldsymbol{\theta}_T \tag{5.1}$$

versus alternative

$$\begin{aligned} \mathcal{H}_A : \boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_{\tau_1} \neq \boldsymbol{\theta}_{\tau_1+1} = \ldots = \boldsymbol{\theta}_{\tau_2} &\neq \boldsymbol{\theta}_{\tau_2+1} \ldots \\ \ldots \boldsymbol{\theta}_{\tau_m-1} &\neq \boldsymbol{\theta}_{\tau_m} = \ldots = \boldsymbol{\theta}_T. \end{aligned}$$

## 5.2 Penalized minimization

We follow Killick et al (2012) and Haynes et al (2014). Let $y_1, y_2, \ldots y_T$ be a sequence of independent random variables with $m$ change points located at $\tau_1, \ldots, \tau_m$. The set of change points $\mathcal{T} := \{\boldsymbol{\tau} \mid 0 = \tau_0 < \tau_1 \ldots \tau_m < \tau_{m+1} = T\}$ splits the time series into $m+1$ segments, whereby segment $i$ includes $y_{(\tau_{i-1}+1):\tau_i} := \{y_{\tau_{i-1}+1}, \ldots, y_{\tau_i}\}$. Each segment $i$ has some model parameters $\boldsymbol{\theta}_i$, so we can describe a model $\mathcal{M}_m$ with $m$ change points by a set of parameters

$$\Theta_m = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{m+1}, \tau_1, \ldots, \tau_m\}.$$

One common approach to find the best segmentation is to first define a cost function used to measure the fit of a segment. We denote the cost of a segment $y_{(\tau_{i-1}+1):\tau_i}$ by $\mathcal{C}(y_{(\tau_{i-1}+1):\tau_i})$ and choose the location of change points for a fixed number $m$ by minimizing the sum of costs over the $m+1$ segments

$$\mathcal{Q}_m(y_{1:T}) = \min_{\tau_1,\ldots,\tau_m} \left\{ \sum_{i=1}^{m+1} \left[ \mathcal{C}\left(y_{(\tau_{i-1}+1):\tau_i}\right) \right] \right\}.$$

Since the number of change points is in general unknown we estimate $m$ by solving

$$\min_m \{\mathcal{Q}_m(y_{1:T}) + f(m)\}$$

for some appropriate penalty function $f(m)$, which is increasing with $m$, to avoid overfitting. For a linear penalty function $f(m) = (m+1)\beta$ with some penalty $\beta > 0$, we get the penalized minimization problem

$$\mathcal{Q}(y_{1:T}, \beta) = \min_{m, \tau_{1:m}} \left\{ \sum_{i=1}^{m+1} \left[ \mathcal{C}\left(y_{(\tau_{i-1}+1):\tau_i}\right) + \beta \right] \right\}. \tag{5.2}$$

The choice of model is obviously dependent on the choice of cost function and penalty. Commonly used as cost function is twice the negative maximum log likelihood (Yao (1988), Gupta and Chen (1996)), and the penalty often belongs to an information criteria like Akaike information criterion (AIC; $\beta = 2p$) or the Bayesian information criterion (BIC; $\beta = p \log T$), where $p$ is the number of additional parameters by adding a segment. It is an obvious choice to use the combination of maximum log likelihood as cost function and these penalties, since we have already discussed in Section 2.4 that the information criteria balance model fit and complexity.

## 5.3 Algorithms

Two algorithms are introduced in this section. For that we consider the multiple change point problem again as minimization problem (5.2)

$$\mathcal{Q}(y_{1:T}, \beta) = \min_{m, \tau_{1:m}} \left\{ \sum_{i=1}^{m+1} \left[ \mathcal{C}\left(y_{(\tau_{i-1}+1):\tau_i}\right) + \beta \right] \right\}.$$

with some appropriate penalty $\beta$. The most commonly used method for such a problem is Binary Segmentation proposed by Scott and Knott (1974). Since it is not an exact method, an alternative algorithm will be considered later.

### Binary Segmentation

The reason why Binary Segmentation is this popular is its simplicity and low computational cost. The basic concept is the iterative application of single change point methods on split subsets. This means the first step is to search for a change point in the full time series. If there is at least one point in time $\tau$, which satisfies

$$\mathcal{C}(y_{1:\tau}) + \mathcal{C}(y_{(\tau+1):T}) + \beta < \mathcal{C}(y_{1:T}),$$

then the point with the smallest cost is identified as the first change point, else the method stops and we accept the null hypothesis of no change point. In case a change point $\tau$ is detected the procedure is repeated on the two subsets $y_{1:\tau}$ and $y_{(\tau+1):T}$. In this way the sets are split in smaller subsets until no further change points are detected. But this approach does not necessarily lead to the global minimum of the minimization problem (5.2), this means this is not an exact method (Rohrbeck, 2013).

### Optimal Partitioning

Jackson et al (2005) introduced another method and solve minimization problem (5.2) based on a successive approach. Let $Q(t)$ be the minimization of (5.2) for data $y_{1:t}$ up to point $t$, $1 < t \leq T$, with possible vectors of change points $\mathcal{T}_t := \{\boldsymbol{\tau} \mid \exists m : 0 = \tau_0 < \tau_1 < \ldots < \tau_m < \tau_{m+1} = t\}$ and let $Q(0) = 0$. Then it holds

$$
\begin{aligned}
Q(t) = Q(y_{1:t};\beta) &= \min_{\boldsymbol{\tau}\in\mathcal{T}_t} \left\{ \sum_{i=1}^{m+1} \left[ \mathcal{C}\left(y_{(\tau_{i-1}+1):\tau_i}\right) + \beta \right] \right\} \\
&= \min_{\substack{m \\ \tau_1<\ldots<\tau_{m+1} \\ \tau_{m+1}=t}} \left\{ \sum_{i=1}^{m+1} \left[ \mathcal{C}\left(y_{(\tau_{i-1}+1):\tau_i}\right) + \beta \right] \right\} \\
&= \min_{\substack{m \\ \tau_1<\ldots<\tau_{m+1} \\ \tau_{m+1}=t}} \left\{ \sum_{i=1}^{m} \left[ \mathcal{C}\left(y_{(\tau_{i-1}+1):\tau_i}\right) + \beta \right] + \mathcal{C}(y_{(\tau_m+1):t}) + \beta \right\} \\
&= \min_{s<t} \left\{ \min_{\substack{m \\ \tau_1<\ldots<\tau_m \\ \tau_m=s}} \sum_{i=1}^{m} \left[ \mathcal{C}\left(y_{(\tau_{i-1}+1):\tau_i}\right) + \beta \right] + \mathcal{C}(y_{(s+1):t}) + \beta \right\} \\
&= \min_{s<t} \left\{ \min_{\boldsymbol{\tau}\in\mathcal{T}_s} \sum_{i=1}^{m} \left[ \mathcal{C}\left(y_{(\tau_{i-1}+1):\tau_i}\right) + \beta \right] + \mathcal{C}(y_{(s+1):t}) + \beta \right\} \\
&= \min_{s<t} \left\{ Q(s) + \mathcal{C}(y_{(s+1):t}) + \beta \right\}. \qquad (5.3)
\end{aligned}
$$

The value of $s$ in (5.3) for which $Q(t)$ is minimal is the location of the last change point in segment $y_{1:t}$. The optimal set of change points in $y_{1:T}$ can be calculated recursively. Steps for implementing this method are given in Algorithm 1.

**Algorithm 1:** Optimal Partitioning Method

**input :** A set of data $(y_1, \ldots, y_T)$
A cost function $\mathcal{C}(\cdot)$ dependent on the data
A penalty constant $\beta$
Let $Q(0) = 0$ and $cp(0) = NULL$

**for** $\tau = 1, \ldots, T$ **do**

1. Calculate $Q(\tau) = \min_{0\leq t<\tau}[Q(t) + \mathcal{C}(y_{(t+1):\tau}) + \beta]$.
2. Let $\tau^* = arg\left\{\min_{0\leq t<\tau}[Q(t) + \mathcal{C}(y_{(t+1):\tau}) + \beta]\right\}$.
3. Set $cp(\tau) = [cp(\tau^*), \tau^*]$.

**end**

**output:** The change points recorded in $cp(n)$

## 5.4 Penalized likelihood

A slightly different penalized optimization problem than defined in (5.2), but popular method to detect change points is the penalized likelihood problem (see e.g. Killick et al,

2010) given by

$$\min_{m,\tau_1,\ldots,\tau_m} PL(\mathcal{M}_m).$$

A model $\mathcal{M}_m$ with $m$ changes in some parameters $\boldsymbol{\theta}$ located at $\tau_1, \ldots, \tau_m$, but common variance $\sigma^2$ for all segments is defined by parameter set $\Theta_m = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{m+1}, \sigma^2, \tau_1, \ldots, \tau_m\}$ with

$$p_m = m + 1 + \sum_{j=1}^{m+1} dim(\boldsymbol{\theta}_j)$$

parameters. The penalized likelihood for model $\mathcal{M}_m$ is defined as

$$PL(\mathcal{M}_m) = -2 \log L(\hat{\Theta}_m \mid \boldsymbol{y}) + f(m) \tag{5.4}$$

with maximum value of the likelihood function $L(\hat{\Theta}_m \mid \boldsymbol{y})$ and a penalty function like

$$\begin{aligned} f(m) &= 2p_m \qquad (AIC) \\ f(m) &= p_m \log T \qquad (BIC). \end{aligned}$$

Note that choosing the model that minimizes the penalized likelihood with one of these penalties means choosing the model with the lowest information criteria. It was shown by E. Schwarz (1978) that the AIC does not asymptotically estimate the correct number of parameters, while the Bayesian Information Criterion provides a consistent estimator for the change points as shown by Yao (1988). Therefore, we estimate the number of change points by using the BIC criterion and we finally define our penalized minimization problem as

$$\min_{m,\tau_1,\ldots,\tau_m} PL(\mathcal{M}_m) \quad = \quad \min_{m,\tau_1,\ldots,\tau_m} \left\{ -2 \log L(\hat{\Theta}_m \mid \boldsymbol{y}) + p_m \cdot \log(T) \right\}. \tag{5.5}$$

Testing for change points is then equivalent to finding model $\mathcal{M}_m$ that minimizes the penalized likelihood. If model $\mathcal{M}_0$ without change points and parameters $\boldsymbol{\Theta_0} = \{\boldsymbol{\theta}_0, \sigma^2\}$ has the lowest penalized likelihood, we accept the null hypothesis $\mathcal{H}_0$ of no change point, defined in (5.1). Otherwise, we reject $\mathcal{H}_0$ and determine the number and location of change points by means of model $\mathcal{M}_m$ with the lowest penalized likelihood.

## 5.5 Piecewise regression

In this section we follow Hinkley (1970, 1971) and McGee and Carleton (1970) to present the underlying theory and use the results of Chapter 2 to define the change point problem for a weighted regression model. We start with a single change point before we continue with the formulation of a more general model.
In Section 2.2 we defined the weighted linear regression model in the form

$$Y_i = \psi(i) + \epsilon_i \qquad \text{with} \quad \epsilon_i \sim \mathcal{N}(0, \sigma_w^2 w_i), \qquad i = 1, \ldots, T,$$

for known weights $w_1, \ldots, w_T$, a variance $\sigma^2$ and a linear mean function $\psi(\,\cdot\,)$, which describes the linear relationship between the response and some covariates $x_{i,1}, \ldots, x_{i,k}$ valid for all data points. As already known, this is not true in case of structural changes in the data, so we consider a model which allows for change points. Such models are called piecewise regression models or segmented regression models, since the data is split in linear segments (see e.g. Muggeo, 2003). In case a single change $\tau$ is assumed we define the following model.

**Definition 5.1** (Piecewise regression model)**.** *A piecewise regression model is of the form*

$$Y_i = \begin{cases} \psi_1(i) + \epsilon_i, & i = 1, \ldots, \tau, \\ \psi_2(i) + \epsilon_i, & i = \tau + 1, \ldots, T, \end{cases} \tag{5.6}$$

*where the change point $\tau$ is unknown, $\psi_j(\,\cdot\,)$, $j = 1, 2$, define the mean functions for the two segments and it holds*

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2 w_i), \qquad i = 1, \ldots, T,$$

*for some variance parameter $\sigma^2$.*

In literature mainly three special cases of (5.6) are discussed:

1) Different intercepts $a_1$ and $a_2$:

$$\psi_1(i) = a_1,\ x_i \leq \tau \quad \& \quad \psi_2(i) = a_2,\ x_i > \tau.$$

2) Different slopes and intercepts:

$$\psi_1(i) = a_1 + b_1 x_i,\ x_i \leq \tau \quad \& \quad \psi_2(i) = a_2 + b_2 \cdot (x_i - \tau),\ x_i > \tau.$$

3) Continuous regression with changing slopes $b_1$ and $b_2$:

$$\psi_1(i) = a + b_1(x_i - \tau),\ x_i \leq \tau \quad \& \quad \psi_2(i) = a + b_2(x_i - \tau),\ x_i > \tau.$$

Examples for these special cases are illustrated in Figure 6.
Using this model we can describe a response variable, which shows one abrupt change. Since data is often affected by more than one change, we define a general piecewise regression model with $m$ change points.

**Definition 5.2** (Piecewise regression model with $m$ change points)**.** *A piecewise regression model $\mathcal{M}_m$ with $m$ change points is of the form*

$$Y_i = \psi(i) + \epsilon_i = \begin{cases} \psi_1(i) + \epsilon_i, & i = 1, \ldots, \tau_1, \\ \psi_j(i) + \epsilon_i, & i = \tau_{j-1} + 1, \ldots, \tau_j, \quad j = 2, \ldots, m, \\ \psi_{m+1}(i) + \epsilon_i, & i = \tau_m + 1, \ldots, T, \end{cases} \tag{5.7}$$

*where the change points $\tau_1, \ldots, \tau_m$ are unknown, each $\psi_j(\,\cdot\,)$, $j = 1, \ldots, m+1$, defines the linear mean functions for segment $j$ and*

$$\epsilon_i \sim \mathcal{N}(0, \sigma_m^2 w_i), \qquad i = 1, \ldots, T,$$

*with a common variance parameter $\sigma_m^2$.*

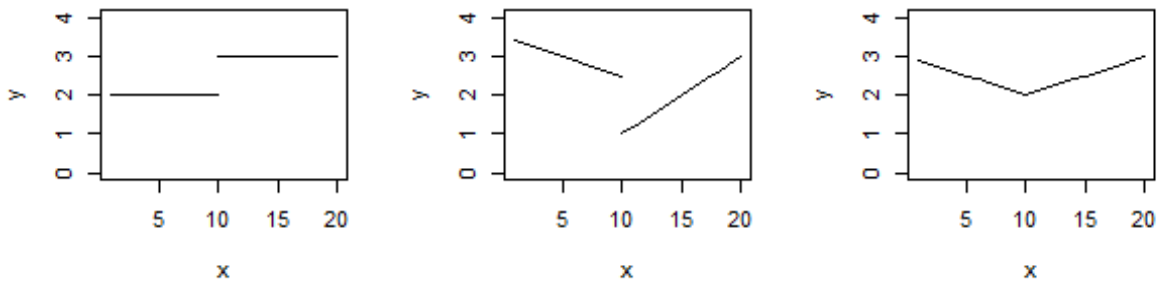



Figure 6: Piecewise regression with single change point in form of special case 1 (left), case 2 (middle) and case 3 (right)

Change points $\tau_1, \ldots, \tau_m$ split the model into $m+1$ segments, where each segment $j$ is described by an individual linear regression with some regression parameters $\boldsymbol{\theta}_j$ that define mean function $\psi_j(\,\cdot\,)$. Therefore, for given change point locations $\tau_1, \ldots, \tau_{m+1}$ we can estimate maximum likelihood estimators for regression parameters $\boldsymbol{\theta}_j$ by fitting a weighted linear regression model with these parameters $\boldsymbol{\theta}_j$ on the data of the corresponding segment $y_{(\tau_{j-1}+1):\tau_j}$. Using the results derived in Section 2.2, especially formula (2.6), yields maximum likelihood estimates $\hat{\boldsymbol{\theta}}_j$ and thus estimated expected values $\hat{\psi}_j(i)$ for all $i = \tau_{j-1}+1, \ldots, \tau_j$ and for all segments $j = 1, \ldots, m+1$.
Rewriting model (5.7) in forms of

$$Y_i = \psi_1(i)\mathbb{1}_{\{1\le i\le\tau\}} + \psi_2(i)\mathbb{1}_{\{\tau_1<i\le\tau_2\}} + \ldots + \psi_{m+1}(i)\mathbb{1}_{\{\tau_m<i\le n\}} + \epsilon_i$$

shows that the whole model is still a weighted regression model. It follows that if we know $\tau_1, \ldots, \tau_m$ we can define

$$\hat{\psi}(\,\cdot\,) := \hat{\psi}_1(\,\cdot\,)\mathbb{1}_{\{1\le i\le\tau\}} + \hat{\psi}_2(\,\cdot\,)\mathbb{1}_{\{\tau_1<i\le\tau_2\}} + \ldots + \hat{\psi}_{m+1}(\,\cdot\,)\mathbb{1}_{\{\tau_m<i\le n\}}$$

and use equation (2.7) to estimate the common variance parameter $\sigma_m^2$. We get

$$\hat{\sigma}_m^2 \;=\; \frac{1}{T}\sum_{i=1}^{T}\frac{1}{w_i}\left(y_i - \hat{\psi}(i)\right)^2 = \frac{1}{T}\sum_{j=1}^{m+1}\sum_{i=\tau_{i-1}+1}^{\tau_i}\frac{1}{w_i}\left(y_i - \hat{\psi}_j(i)\right)^2.$$

Thus, the maximum likelihood estimates $\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_m$ and $\hat{\sigma}_m^2$ depend on the choice of $\tau_1, \ldots, \tau_m$ and $m$ and it remains to estimate the optimal number and location of change points.
We use the penalized likelihood method, which we presented in the previous section, with penalty function corresponding to the BIC to estimate the optimal number $m$ of change points and the optimal locations $\tau_1, \ldots, \tau_m$. Each piecewise regression model in the form (5.7) has a parameter set

$$\Theta_m = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{m+1}, \sigma_m^2, \tau_1, \ldots, \tau_m\}$$

depending on the number of change points $m$. Hence, by (5.5), we have the penalized minimization problem

$$\min_{m,\tau_1,\ldots,\tau_m} PL(\mathcal{M}_m) \quad = \quad \min_{m,\tau_1,\ldots,\tau_m} \left\{ -2\log L(\hat{\Theta}_m \mid \boldsymbol{y}) + dim(\hat{\Theta}_m) \cdot \log(T) \right\}, \qquad (5.8)$$

where, by using (2.15), the maximum likelihood term for this weighted linear model is given by

$$-2\log L(\hat{\Theta}_m \mid \boldsymbol{y}) = T\log(2\pi\hat{\sigma}_m^2) + T + \sum_{i=1}^{T} \log(w_i).$$

The model with change points $\hat{\tau}_1, \ldots, \hat{\tau}_m$ for some $m$ with minimal penalized likelihood yields the estimate for the optimal number and locations of change points $\hat{\tau}_1, \ldots, \hat{\tau}_m$ and accordingly regression parameters $\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{m+1}$ and common variance parameter $\hat{\sigma}_m^2$.
Again testing for change points is equivalent to finding the piecewise regression model $\mathcal{M}_m$ that minimizes the penalized likelihood. In case a model without change points $\mathcal{M}_0$ has minimal likelihood, we do not reject the null hypothesis of no change.
In the next chapter we finally apply this method to loss data to check if the data is disturbed by structural breaks.

# Chapter 6

# Application to loss triangles

In this chapter we fit the presented parts together and use the previous results to apply change point analysis for loss triangles. But in fact, we do not only search for change points, but also for trends. Theoretical foundation for this application are the linear models defined in Chapter 2 and the change point analysis presented in the previous chapter. We apply this to the reserving models discussed in Chapter 3 in order to test for structural changes in the loss development.
As already mentioned in the introduction of this thesis, we present two methods. We start this chapter with a short description of the approach, before we consider candidate models for the first method. This method is outlined in Section 6.3 and defines a way to analyze the loss history separately for each development year. Thereafter, we apply this on an exemplary loss triangle. In Section 6.5 we propose a more extended method to select an overall model. This method can be used to automatize the basic loss model calibration, since it selects and estimates a model based on the loss history under consideration of trends and change points.

## 6.1 Approach

In Section 3.4 we showed that the additive and multiplicative reserving model can be interpreted as weighted regression model in the form of

$$Y_{i,k} = \mu_k + \epsilon_{i,k}, \qquad \epsilon_{i,k} \sim \mathcal{N}\left(0, \sigma_k^2 w_i\right), \tag{6.1}$$

per development year $k = 1, \ldots, N$ in the additive model and $k = 2, \ldots, N$ in the multiplicative model, for all underwriting years $i = 1, \ldots, N$. In the additive model the response variable $Y_{i,k}$ is given by the incremental loss ratios $M_{i,k}$ and the weights are defined as $w_i = \frac{1}{v_i}$ with premiums $v_i$. In the multiplicative model we are interested in responses $F_{i,k}$, which denote the development factors, and weights defined by $w_i = \frac{1}{C_{i,k-1}}$ with cumulative losses $C_{i,k-1}$ from the previous development year $k-1$.
This holds true only under the assumption that for each development year the expected values are constant. We want to test whether the model assumption of equal means is fulfilled for each development year or violated due to a trend or a single or multiple change points. To do this for a given underlying reserving model with respective response and

data, we extend model (6.1) and compare it to other weighted regression models of the form

$$Y_{i,k} = \psi^{(k)}(i) + \epsilon_{i,k}, \qquad \epsilon_{i,k} \sim \mathcal{N}\left(0, \sigma_k^2 w_i\right), \tag{6.2}$$

with some linear mean functions $\psi^{(k)}(\,\cdot\,)$ to describe the behavior of a fixed development year $k$. Then the observations $\boldsymbol{y_k} = \{y_{1,k}, \ldots, y_{N-k+1,k}\}$ of some development year $k$ can be represented as

$$y_{i,k} = \psi^{(k)}(i) + \epsilon_{i,k}, \qquad i = 1, \ldots, N-k+1,$$

for $k = 1, \ldots, N$ in the additive model or $k = 2, \ldots, N$ in the multiplicative model. Because of the triangular form of the given loss data the number of observations declines per development year and obviously it does not make sense to fit too complex models with a lot of parameters to development years with a small number of observations. Furthermore, usually the incremental loss ratios in later development years are quite small and the development factors close to one, since the losses are almost settled, and these development years do not have a substantial impact on the prediction of the reserve compared to early development years. That is why we consider only development years $k$ up to some limit, denoted by $D$, with $D \leq N$.
In the next section we will define candidate models for each development year and fit all presented candidate models to the observed data. Then we use the Bayesian information criterion defined in Section 2.4 to compare our candidate models and evaluate whether trends or structural changes appear in the loss pattern or not.

## 6.2 Candidate models per development year

In this section we consider three different types of models to describe a development year $k = 1, \ldots, D$. All models are weighted regression models of the form of equation (6.2), but differ in the choice of mean function $\psi^{(k)}(\,\cdot\,)$ and therefore also in number of parameters. We denote the number of regression parameters for development year $k$ by $p^{(k)}$.

### Type I: Models with no change point

We consider two models without change point. The first one is our base model and the foundation of the analysis, because we compare all other candidate models to this model.

**Model 1a) Base model**
The base model is an intercept-only model with mean function $\psi^{(k)}(\,\cdot\,)$ of development year $k$ given by

$$\psi^{(k)}(i) = a^{(k)}, \qquad i = 1, \ldots, N.$$

This model corresponds to Mack's (2002) reserving models presented in Section 3.3 with constant means per development year, since in Section 3.4 we have shown for both reserving models that the related maximum likelihood estimators are equal to the there

presented unbiased estimators. All other candidate models are extensions of this model and assume that the data is not constant, but structurally disturbed for some reason.
One reason for a violation in the assumption of a constant level of loss ratios or development factors is that the data comprises a trend, which is covered in the following model.

**Model 1b) Model with additional linear trend**
The mean function changes by an additional trend parameter, which yields

$$\psi^{(k)}(i) = a^{(k)} + b^{(k)} \cdot i, \qquad i = 1, \dots, N.$$

We increased the number of regression parameters $p^{(k)}$ from a single parameter in Model 1a to $p^{(k)} = 2$ in this model, which is a simple linear regression model as given in Definition 2.4 with covariate $x_i = i$. In Section 2.3 we have presented this model together with the intercept-only model that describes Model 1a.
Even if the expected value is not constant in Model 1b, at least both models assume that a single model is valid for all values of development year $k$ and there is no change point. To model a structural break in a development we consider models of type II.

## Type II: Models with a single change point

We consider two different models, which both assume there is a single change point $\tau^{(k)}$, $1 \leq \tau^{(k)} < N$, that splits development year $k$ in two segments.

**Model 2a) Single change in mean**
The mean function

$$\psi^{(k)}(i) = \begin{cases} a_1^{(k)}, & i = 1, \dots, \tau^{(k)}, \\ a_2^{(k)}, & i = \tau^{(k)} + 1, \dots, N, \end{cases}$$

describes a development year with constant mean for all underwriting years up to some year $\tau^{(k)}$ after that the mean changes. Individually both segments $y_{1,k} \dots, y_{\tau^{(k)},k}$ and $y_{\tau^{(k)}+1,k}, \dots, y_{N,k}$ fulfill the assumption of constant means, but not taken as a whole. Within the next model, this does not hold, because it attributes both segments additionally a different trend.

**Model 2b) Single change in intercept and trend**
The model defined by

$$\psi^{(k)}(i) = \begin{cases} a_1^{(k)} + b_1^{(k)} \cdot i, & i = 1, \dots, \tau^{(k)}, \\ a_2^{(k)} + b_2^{(k)} \cdot (i - \tau^{(k)}), & i = \tau^{(k)} + 1, \dots, N, \end{cases}$$

describes a model, that is split in two simple linear regression segments by a change point $\tau^{(k)}$, after which intercept and trend change.
Both models are piecewise linear regression models with one change point as given in Definition 5.1. The number of regression parameters for these two models is $p^{(k)} = p_1^{(k)} + p_2(k)$, where it holds

$$p_j^{(k)} = \begin{cases} 1 & \text{for Model 2a} \\ 2 & \text{for Model 2b} \end{cases} \qquad \text{for } j = 1, 2.$$

The models of type I and type II are illustrated in Figure 7. Each plot shows possible

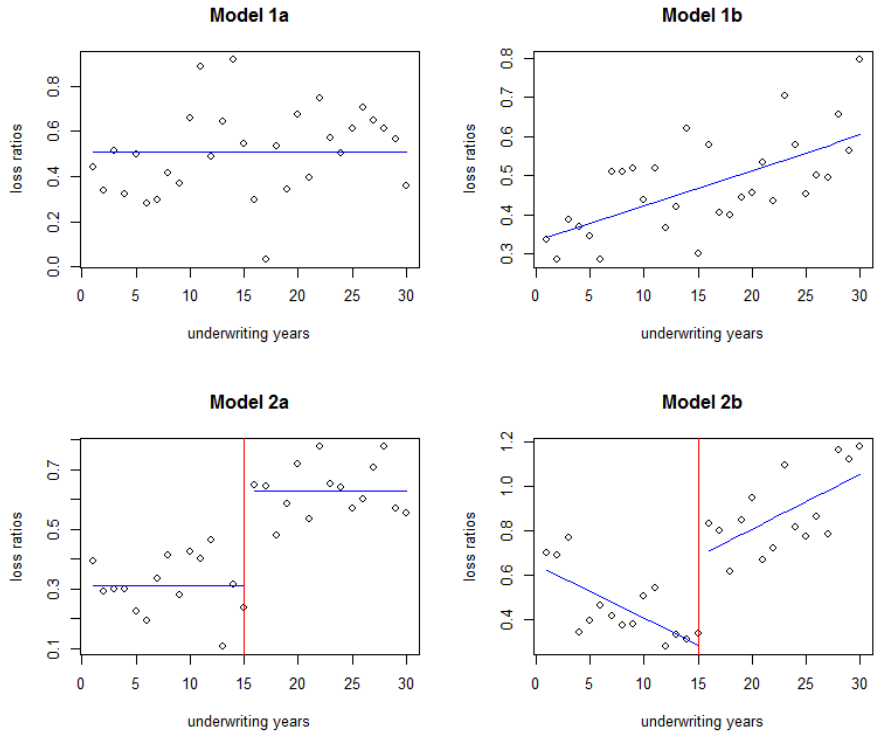



Figure 7: Models of type I (upper plots) and models of type II (lower plots)

observations (in this case loss ratios) for a model of some development year $k$ with 30 observations and the blue line represents the respective mean function the observations are generated from. The upper plots show Model 1a and Model 1b without change point and in the lower plots the red lines highlights the location of the change point that is here defined as $\tau^{(k)} = 15$ for both models of type II with exactly one change point.
Since a development year can be affected by more than one structural break, we need a more general type of models with number of change points $m$ not fixed. This leads to the general piecewise regression model from Definition 5.2.

## Type III: Models with multiple change points

We are interested in models for development year $k$ with $m$ change points located at $\tau_1^{(k)}, \ldots, \tau_m^{(k)}$ for some $m \geq 0$ of the following form:

**Model 3) General model**
The piecewise regression model with mean function

$$\psi^{(k)}(i) = \begin{cases} a_1^{(k)} + b_1^{(k)} \cdot i, & i = 1, \ldots, \tau_1^{(k)}, \\ a_j^{(k)} + b_j^{(k)} \cdot (i - \tau_{j-1}), & i = \tau_{j-1}^{(k)} + 1, \ldots, \tau_j^{(k)}, \quad j = 2, \ldots, m, \\ a_{m+1}^{(k)} + b_{m+1}^{(k)} \cdot (i - \tau_m), & i = \tau_m^{(k)} + 1, \ldots, N, \end{cases}$$

describes a model that splits the development year in $m+1$ individual segments with some regression lines.
We know the number of parameters is penalized, therefore, we want to omit unnecessary parameters in Model 3. If the added penalty for the parameter is higher than the improvement in model fit reduces the BIC, then this parameter is not considered necessary. Thus, we allow for each segment $j$ to either have two parameters $a_j^{(k)}$ and $b_j^{(k)}$ for intercept and slope or only an intercept $a_j^{(k)}$. Starting from second segment, we incorporate also the possibility to avoid a useless new intercept parameter. This is inspired by special case 3 defined in Section 5.5, which assumes a continuous development from previous segment $j-1$ to segment $j$ without jump. In this case segment $j$, for $j > 2$, starts with an intercept equal to the last point of the previous segment, i.e. $a_j^{(k)} = \psi^{(k)}(\tau_{j-1})$, and we do not count $a_j^{(k)}$ as a new parameter. It follows, it is also possible to have either only a new slope parameter $b_j^{(k)}$ for segment $j$ or even no new parameter and the segment is modeled to stay constant on that level $a_j^{(k)}$. This possibility to reduce the number of parameters in Model 3 will be shown in the example in Section 6.4.
To conclude, the number of all regression parameters $p^{(k)}$ for Model 3 is in general given by

$$p^{(k)} = \sum_{j=1}^{m+1} p_j^{(k)} \qquad \text{with} \quad p_1^{(k)} \in \{1,2\} \quad \& \quad p_j^{(k)} \in \{0,1,2\}, \quad j = 2,\dots,m+1,$$

where $p_j^{(k)}$ defines number of regression parameters of segment $j$ for all $j = 1,\dots,m+1$. Note that because the number of parameters per segment and the number of change points is not fixed anymore Model 3 can appear in the form of all models of type I and II in case the optimal number of change points is zero or one.
But due to this variability in number and the locations of change points and number of parameters per segment, we have a high number of possible combinations and we need an efficient way to choose the optimal one. This will be discussed in the next section.
In conclusion, we define the parameter set of development year $k$ for all candidate models in the form

$$\Theta^{(k)} = \{\boldsymbol{\theta}_1^{(k)},\dots,\boldsymbol{\theta}_{m+1}^{(k)}, \sigma_k^2, \tau_1^{(k)},\dots,\tau_m^{(k)}\}, \tag{6.3}$$

where the set of regression parameters of each segment $j$ of development year $k$ is denoted $\boldsymbol{\theta}_j^{(k)}$. If the model has no change point we have only one segment with regression parameters $\boldsymbol{\theta}^{(k)}$ and no $\tau$ in the parameter set, otherwise for some $m \geq 1$ we have $m$ change points and $m+1$ segments with some regression parameters in addition to the variance parameter in the parameter set.
The number of parameters in $\Theta^{(k)}$ for all $m \geq 0$ is given by

$$\dim(\Theta^{(k)}) = 1 + m + p^{(k)},$$

where $p^{(k)}$ is the number of all regression parameters.

## 6.3 Change point analysis per development year

In this section we consider a fixed development year $k$, for some $k = 1, \dots, D$, and the observations $\boldsymbol{y_k} = y_{1,k}, \dots, y_{N-k+1,k}$ of this development year. We want to test if the model assumption of equal means is fulfilled for this development year or not. So we fit the candidate models to the data $\boldsymbol{y_k}$ and compare the BIC of the Model 1a (base model) at first with Model 1b and then with the optimal models of type II and type III.
For a single development year the BIC for all candidate models $\mathcal{M}$ presented in Section 6.2 is in general given by

$$\begin{aligned} BIC(\mathcal{M}) &= -2\log L\left(\hat{\Theta}^{(k)} \mid \boldsymbol{y_k}\right) + dim(\Theta^{(k)}) \cdot \log(N-k+1) \\ &= (N-k+1)\left(\log(2\pi\hat{\sigma}_k^2)+1\right) + \sum_{i=1}^{N-k+1} \log(w_i) + dim(\Theta^{(k)}) \cdot \log(N-k+1) \end{aligned}$$

with maximum likelihood estimate

$$\hat{\sigma}_k^2 = \frac{1}{N-k+1} \sum_{i=1}^{N-k+1} \frac{1}{w_i}(y_i - \hat{\psi}^{(k)}(i))^2$$

for mean function $\psi^{(k)}(\,\cdot\,)$ and parameter set $\Theta^{(k)}$ as defined in (6.3) depending on the candidate model. Here we used Section 2.4, where we derived the BIC for weighted linear regression models. To get the BIC for each candidate model we have to specify the maximum likelihood estimates. In Section 2.3 we presented the maximum likelihood estimation of model parameters for both **models of type I**, which immediately leads the BIC for Model 1a and Model 1b.
The maximum likelihood estimation in piecewise regression models was discussed in Section 5.5, which can be used for models of type II and III. For given locations of change points the regression parameters of each segment can be estimated by fitting the data of this segment to a regression model with these parameters, which yields $\hat{\psi}_j$ for each segment $j$, for $j = 1, \dots, m+1$, and thus maximum likelihood estimate $\hat{\sigma}_k^2$.
For a single change point, i.e. the **models of type II**, we can easily estimate the maximum likelihood of $\tau$ for both models by trying all possible values $\tau = 1, \dots, N-k+1$ and estimate the regression parameters and the corresponding variance parameter based on each $\tau$. Then we can calculate the value of the likelihood function for all sets of estimated parameters and choose the $\tau$ with maximum likelihood. The maximum value of the likelihood function for Model 2a and the one for Model 2b yields the BIC for each model.
For the BIC of the **model of type III** we need to estimate the optimal number and location of change points. As mentioned in previous section, it is not efficient to try all numbers and locations, since the number of change points and parameters per segment is not fixed anymore. Note that since the number of parameters is different for all possible models of this type, it is not sufficient to compare only the likelihood function for different change points, but a penalized minimization approach is required to take the number of parameters into account. So we want to minimize

$$\min_{m,\tau_1,\dots,\tau_m} PL(\mathcal{M}_m) = \min_{m,\tau_1,\dots,\tau_m} \left\{-2\log L\left(\hat{\Theta}^{(k)} \mid \boldsymbol{y_k}\right) + dim(\Theta^{(k)}) \cdot \log(N-k+1)\right\}$$

as discussed in Section 5.4. For an efficient computation we use an algorithm inspired by the Optimal Partitioning method presented in Section 5.3. We adjust the algorithm for our needs, but the concept is to use the successive proceeding from the Optimal Partitioning method and find the optimal location of change points and the optimal parameters per segment by partition of the segment up to a last change point based on the partitions in the previous steps. The corresponding R code to run the approach is shown in the Appendix A.1.
Given the BIC for each model, we compare the BIC of the base model to the BIC of all other models to analyze the differences. We follow Fabozzi et al (2014) and interpret the difference between the BICs of two models as evidence against the model with higher BIC to be the better model. If Model 1b has a lower BIC compared to Model 1a, this implies that this development year has no constant mean but is probably disturbed by a trend. Similar in case the BIC of Model 2a or Model 2b is clearly lower than the BIC of Model 1a, then there is evidence that the model assumption is violated in this development year because of a change point. Model 3 has always lowest BIC by definition, so in case Model 3 is not equal to Model 1a, then there is evidence that the development year is disturbed by a trend, change point or both. But we are interested in the significance of this evidence, so we denote the absolute difference between the BICs of two models $\Delta$ and use the principles derived by Kass and Raftery (1995), which yield the following interpretations for the model with higher BIC:

| | | |
|---|---|---|
| $\mathbf{\Delta < 2}$ | : | Evidence against model with higher BIC is not significant. |
| $\mathbf{2 \leq \Delta < 6}$ | : | There is a positive evidence against model with higher BIC. |
| $\mathbf{6 \leq \Delta < 10}$ | : | The evidence against model with higher BIC is strong. |
| $\mathbf{10 \leq \Delta}$ | : | The evidence against model with higher BIC is decisive. |

By this we can judge the evidence against the original reserving model. The reason why we do not only compare Model 3 and Model 1a but also Model 1b and models with exactly one change point is that it can happen that Model 3 becomes very complex because the number of change points and trends is not limited. In terms of practical application, one of the other models would be preferred in this case. If we compare the BIC of Model 3 with the other models, we can evaluate whether another model fits almost as well but is not that complex.

## 6.4 Exemplary application

We demonstrate the approach presented in the previous section by means of an example based on the cumulative loss triangle given in Appendix A.3. For the analysis we use the R functions presented in Appendix A.1 to fit the models and calculate the BIC for each candidate model. We show the method on the basis of the multiplicative reserving model as underlying reserving model, thus, we are interested in development factors as response and the cumulative losses from the respective previous development year as weights. As already mentioned, the development factors start from the second development year and we present the results from analyzing development years two, three and four.

The **second development year** has the following observed development factors:

```
 [1] 1.169324 1.162750 1.161218 1.177176 1.103187 1.148052 1.158462 1.146301 1.117710 1.111361 1.158134
[12] 1.091129 1.061845 1.092284 1.121587 1.089020 1.113378 1.082967 1.103472 1.130769 1.206529 1.156928
[23] 1.149189 1.107265 1.204623 1.142124 1.202818 1.225892 1.190426 1.226929 1.290268 1.236598 1.302107
[34] 1.257289 1.264768 1.116647 1.161820 1.169067 1.160897 1.136097 1.121968 1.087960 1.167328 1.159322
[45] 1.182102 1.187929 1.133476 1.147208 1.168649
```

Fitting the data to Model 1a and Model 1b yields the resulting models of type I illustrated in Figure 8 together with the determined BIC. The left plot in the figure shows the

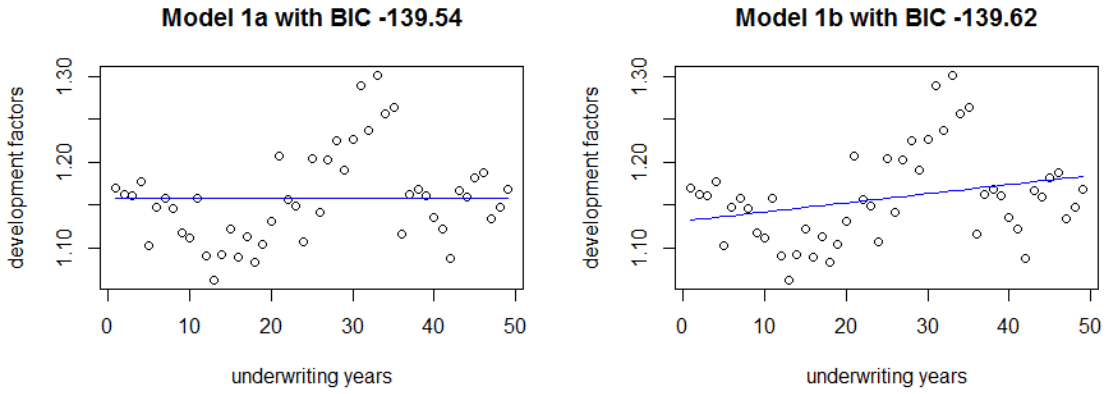



Figure 8: Model 1a (left) and Model 1b (right) for observed development factors of DY2

base model, which corresponds to the original multiplicative reserving model assuming a constant development factor for development year 2. Allowing for an additional trend parameter leads to Model 1b (right plot). The figure shows that the model with trend has minimal lower BIC than the base model, but on this basis we do not reject the base model, since the evidence against Model 1a is not significant.
Next step is to test for one change point, thus, we fit the data to both models of type II with the result given by Figure 9. Comparing the information criteria of these two models

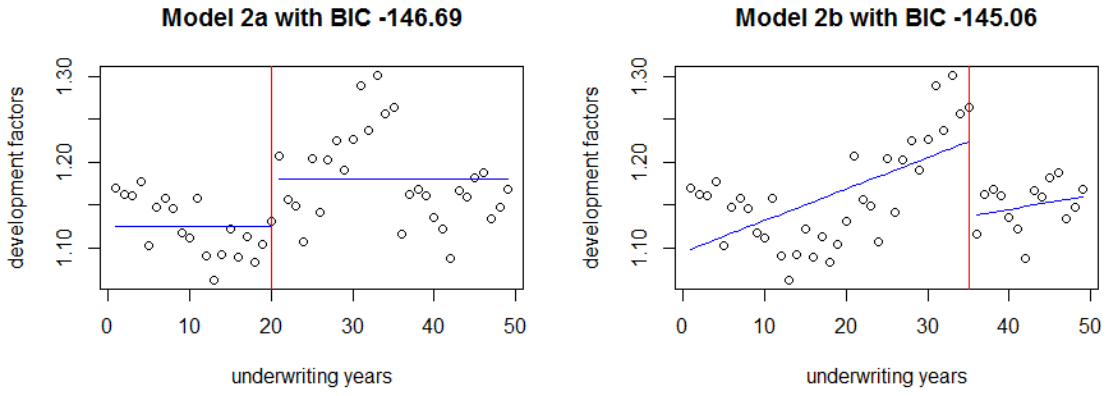



Figure 9: Model 2a (left) and Model 2b (right) for observed development factors of DY2

to Model 1a indicates that there is at least one change point in data. The difference for both models is too high to ignore the evidence against the base model. But the fact that both models propose a different change point hints that probably one change point is not enough, thus we fit the data to Model 3.
Allowing for each number of change points and testing for the optimal number and location leads to the model presented in Figure 10. Model 3 has two change points and the evidence that this model fits clearly better and against Model 1a is decisive. Obviously we have to reject the hypothesis that development year 2 has development factors with a constant level for all underwriting years, since our proposed method revealed that the data is disturbed by structural changes.

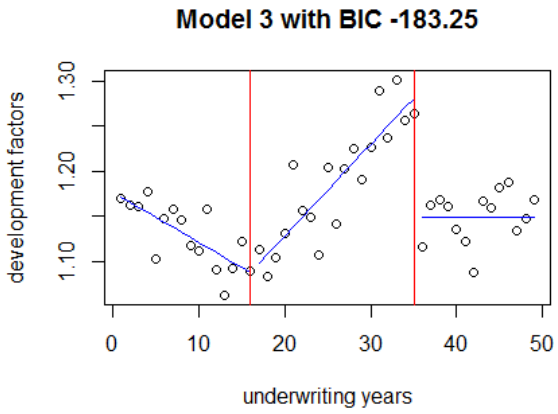



Figure 10: Model 3 for observed development factors of DY2

Note that Model 3 for this data is a good example for what we explained in Section 6.2 with regard to avoiding unnecessary parameters. The second segment starts at the last level of the previous segment and we do not need a new intercept parameter. Also for the third segment we need only one new parameter, since this segment has obviously a new level, but an additional trend parameter is not worth the increasing complexity.
We continue with the **third development year** and apply the same approach on the observations of development year 3, which are given by:

```
 [1] 1.129547 1.141429 1.092039 1.090876 1.118347 1.066648 1.087056 1.141425 1.114531 1.073510 1.059459
[12] 1.096958 1.175040 1.090072 1.092764 1.074150 1.082954 1.134907 1.106821 1.098110 1.182680 1.229721
[23] 1.209252 1.185986 1.196175 1.236189 1.161573 1.147865 1.210819 1.208338 1.213113 1.253227 1.182577
[34] 1.165724 1.247450 1.150170 1.202611 1.189428 1.216641 1.196323 1.226974 1.159509 1.257347 1.197520
[45] 1.209851 1.225442 1.221560 1.223073
```

All candidate models with the respective information criteria are concluded in Figure 11. Again we can draw conclusions from analyzing all candidate models. Firstly, adding a trend parameter obviously improves the model fit a lot. At this point of time before analyzing the other models we would already reject the hypothesis that the third development year has a constant expected development factor for all underwriting years.
But considering Model 2a and Model 2b reveals that a change point is even more suitable to describe the increase from older underwriting years to newer ones. Now both models of type II propose the same change point and the difference to the BIC of Model 1a is

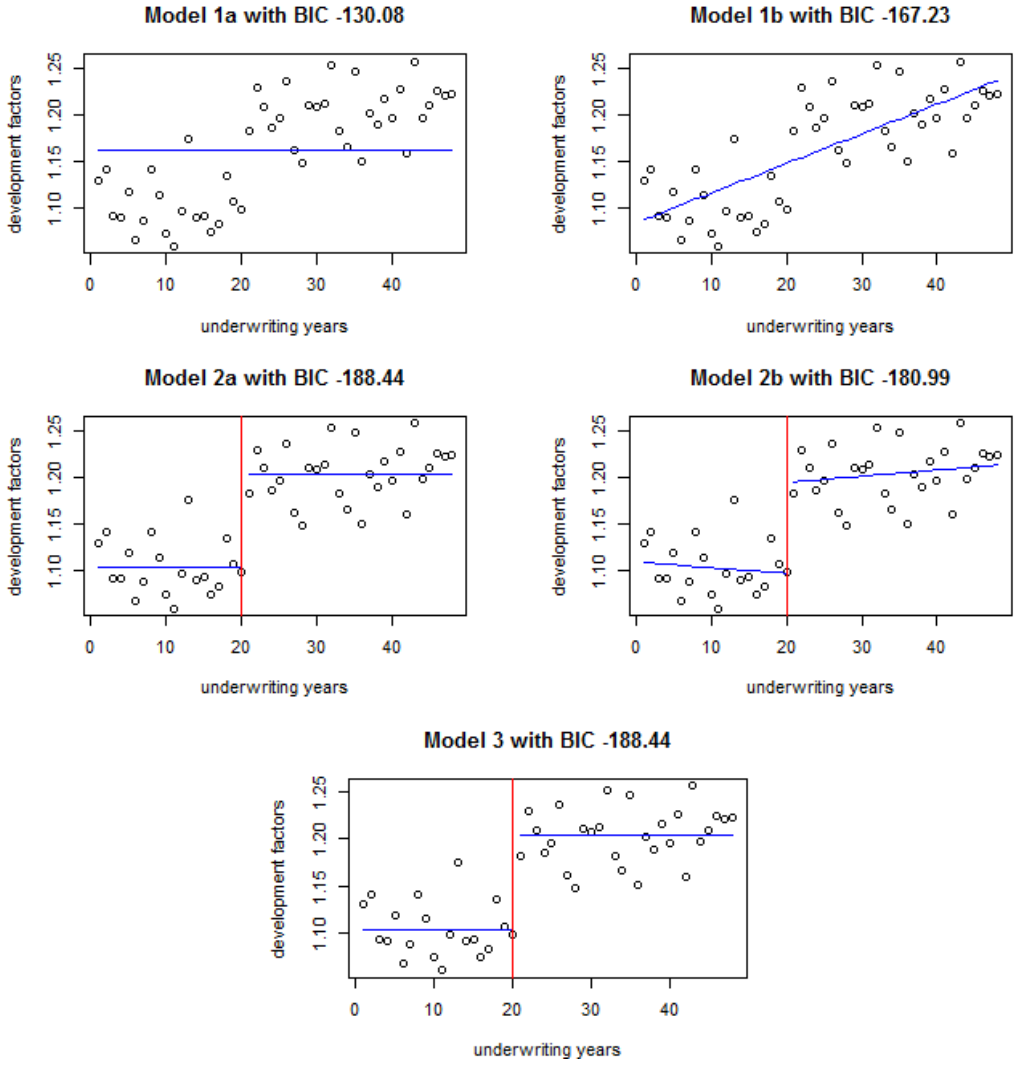



Figure 11: All candidate models based on the observed development factors of DY3

very large, but the fact that Model 2a has a lower BIC than Model 2b shows that it is not necessary to add a trend parameter to both segments.
Model 3 reveals that more change points do not improve the result anymore, instead Model 2a is already the optimal model to describe development year 3. Therefore, we still have to reject the assumption that the development factors of development year 3 have the same mean, but now we evaluated that it is because of a change point. The evidence against Model 1a is decisive and our method manifests that development year 3 is split in two segments with each a constant mean.
Lastly, we apply the approach on the **forth development year** with observations

```
 [1] 1.098521 1.088052 1.110494 1.125790 1.091408 1.097900 1.092555 1.106848 1.095334 1.104255 1.093042
[12] 1.100402 1.118174 1.082582 1.113441 1.113822 1.109559 1.103727 1.094070 1.100368 1.098434 1.116399
[23] 1.098194 1.121952 1.096454 1.128102 1.091105 1.124826 1.116646 1.110916 1.105111 1.094164 1.086173
[34] 1.109784 1.120197 1.103618 1.108863 1.101601 1.102480 1.078373 1.118762 1.094677 1.098704 1.098172
[45] 1.119247 1.112747 1.107858
```

We present the result of fitting these observations to each candidate model again in form

of the plotted models with each BIC. For development year 4 this is given in Figure 12. This figure shows that in case of development year 4 there is no evidence at all against Model 1a, since Model 1b as well as the models of type II show a higher BIC than the base model and Model 3 is equivalent to Model 1a. Thus, we verified that the data of this development year is not disturbed by trends or change points and the original reserving model is the optimal model to describe this development year.

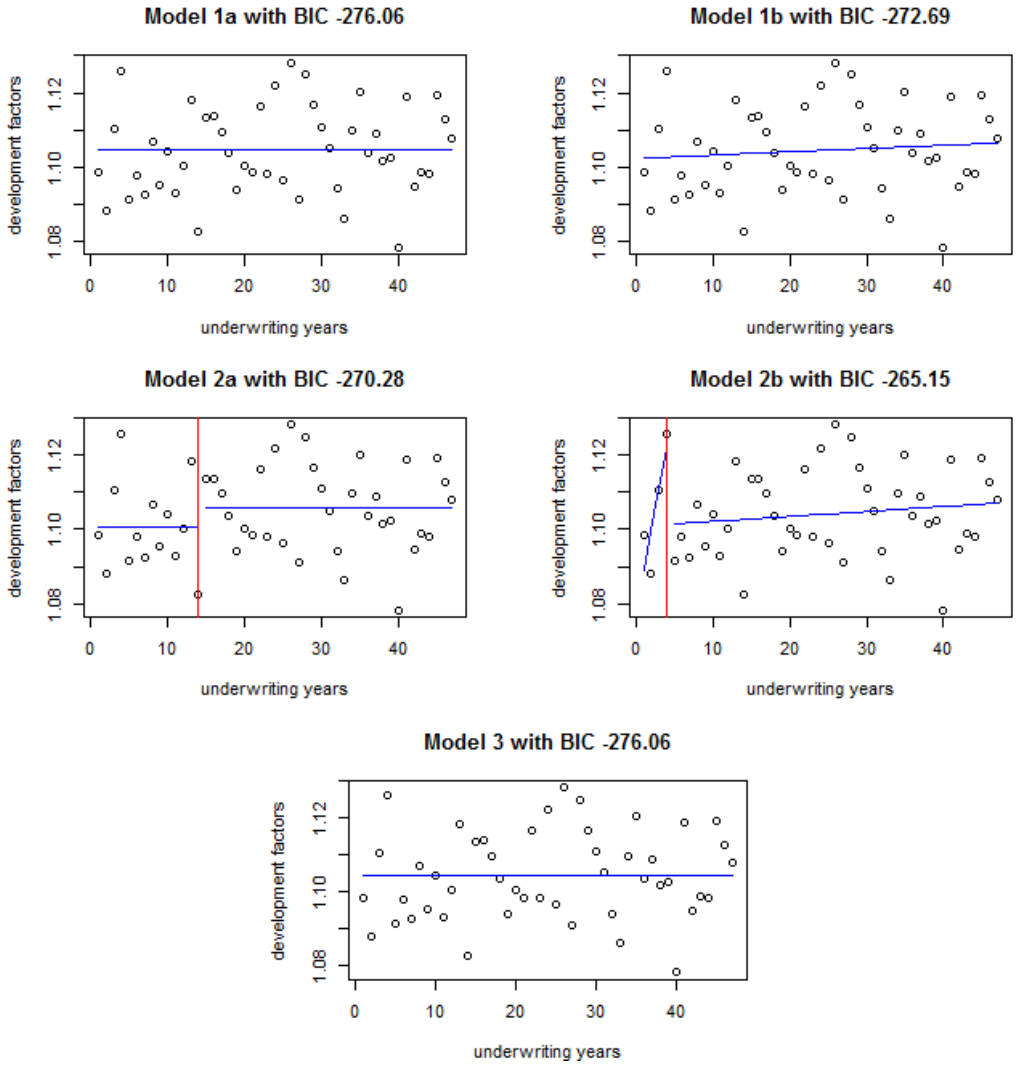



Figure 12: All candidate models based on the observed development factors of DY4

## 6.5 Automatized model selection

The change point analysis per development year introduced in Section 6.3 and discussed in the previous section by using an example is a helpful tool to analyze the data and determine whether the data is disturbed by unwanted effects like trends or change points. In

practice, we expect that change points do not appear individually for some development years, but ranging through the total claims history. That is why it is useful to consider all development years $k$ up to the chosen limit $D$ and test if there are common change points to obtain an overall model. In this section we present an approach for that and use again the penalized likelihood approach discussed in Section 5.4 to decide on the best model.
In Section 3.5 we stated two different scenarios how change points can show up in the data. The first one has change points in underwriting years and the second one has changes in calendar years. We formulate one model for each case.

**Model A) Change points in underwriting years**
The model with $m$ change points in the loss development located at some underwriting years $\tau_1, \dots, \tau_m$ is for all development years $k$ defined as

$$Y_{i,k} = \psi^{(k)}(i) + \epsilon_{i,k} \qquad \text{with} \qquad \epsilon_i \sim \mathcal{N}\left(0, \sigma_k^2 w_i\right), \qquad i,k = 1, \dots, N,$$

where the mean functions of all development years $k$ are in the form

$$\psi^{(k)}(i) = \begin{cases} \psi_1^{(k)}(i), & i \leq \tau_1, \\ \psi_j^{(k)}(i), & \tau_{j-1} < i \leq \tau_j, \quad j = 2, \dots, m, \\ \psi_{m+1}^{(k)}(i), & \tau_m < i \leq N. \end{cases}$$

Similar we can define a model with change points in calendar years.

**Model B) Change points in calendar years**
The model with $m$ change points in the loss development located at some calendar years $\tau_1, \dots, \tau_m$ is defined as

$$Y_{i,k} = \psi^{(k)}(i) + \epsilon_{i,k} \qquad \text{with} \qquad \epsilon_i \sim \mathcal{N}\left(0, \sigma_k^2 w_i\right), \qquad i,k = 1, \dots, N,$$

but now the change points effect each development year $k$ in the following way:

$$\psi^{(k)}(i) = \begin{cases} \psi_1^{(k)}(i), & i + k - 1 \leq \tau_1, \\ \psi_j^{(k)}(i), & \tau_{j-1} < i + k - 1 \leq \tau_j, \quad j = 2, \dots, m, \\ \psi_{m+1}^{(k)}(i), & \tau_m < i + k - 1 \leq N. \end{cases}$$

Both models are illustrated in Figure 13. In contrast to change point analysis per development year we assume that if there are change points $\tau_1, \dots, \tau_m$, then these change points concern all development years and not that each development year has individual change points $\tau_1^{(k)}, \dots, \tau_m^{(k)}$.
For both models we use a penalized likelihood approach to test for the respective types of change points and find the optimal number and location of change points, in case there are change points. For given observations

$$\{y_{i,k} \mid k = 1, \dots, D, i = 1, \dots, N - k + 1\} = \{\boldsymbol{y_k} \mid k = 1, \dots, D\}$$

and cost function defined as the sum of the maximum likelihood of all considered development years, we get for a model $\mathcal{M}_m$ in form of Model A or Model B in both cases the

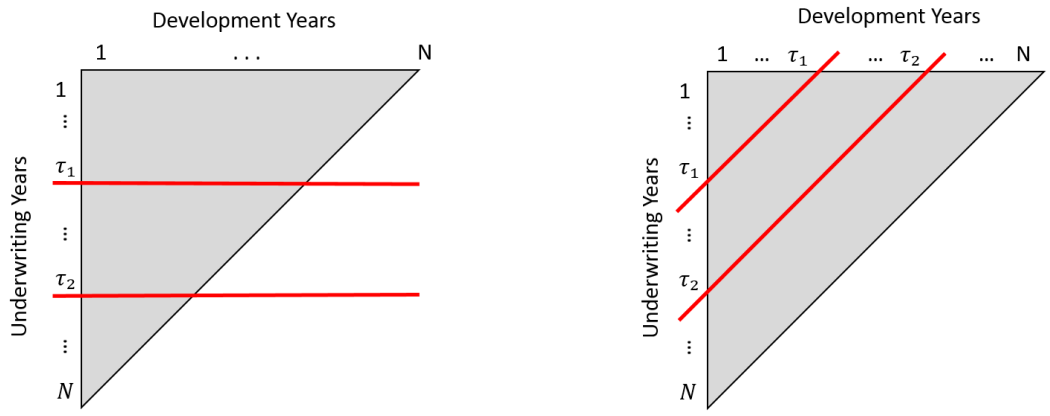



Figure 13: Model A (left) and Model B (right) with two change points

penalized minimization problem

$$\min_{m,\tau_1,\ldots,\tau_m} PL(\mathcal{M}_m) \quad = \quad \min_{m,\tau_1,\ldots,\tau_m} \left\{ \sum_{k=1}^{D} -2 \log L \left( \hat{\boldsymbol{\theta}}^{(k)}, \sigma_k^2 \mid \boldsymbol{y_k} \right) + dim(\Theta) \cdot \log(T) \right\},$$

where $T$ is the number of observations with $T = \sum_{k=1}^{D}(N-k+1)$ and the parameter set for both models is

$$\Theta = \{\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(D)}, \tau_1, \ldots, \tau_{m+1}, \sigma_1^2, \ldots, \sigma_D^2\}$$

for some $m \geq 0$, where $\{\tau_1, \ldots, \tau_{m+1}\} = \emptyset$ for $m = 0$.
For same values of change points $\tau_1, \ldots, \tau_m$, Model A and Model B differ in the location of these change points in the development years, since for Model B these belong to calendar years and for Model A to underwriting years. If for example Model A and Model B both have change point $\tau = 10$, then in Model A this change point is in all development years underwriting year 10, which yields first segments from the first to the tenth element in the loss triangle for all development years. Model B has in this example calendar year 10 as change point, i.e. in the first development year the first segment consists again of the first ten elements, but in the second development year calendar year 10 corresponds to underwriting year 9 and the first segment has nine elements, in the third development year the first segment has eight elements and so on. Different segments per development year lead to different regression parameters and thus different variance parameters for Model A and Model B. But the approach to find the optimal locations of change points in both models is the same. We determined the maximum likelihood term per development year in the previous section, which yields

$$PL(\mathcal{M}_m) = \sum_{k=1}^{D} \left[ (N-k+1) \left( \log(2\pi\hat{\sigma}_k^2) + 1 \right) + \sum_{i=1}^{N-k+1} \log(w_i) \right] + dim(\Theta) \cdot \log(T).$$

Again there is a high number of possible models and we need an efficient way to calculate the optimal number and location of change points for Model A and Model B. We adjust

the algorithm used for change point analysis per development year in order to find an appropriate number and location of change points in this minimization problem. The R code for this method is presented in Appendix A.2. We will evaluate the performance of this method for both models in the simulation study in the next chapter.
The model selection then works as follows. Running the penalized likelihood approach to test for one of both models yields a winner model with a number and locations of change points, which minimize the penalized likelihood. This is our final model. In case we want to test for both types of change points we run the approach once for Model A and once for Model B and that results in two winner models. We trust the model with lowest penalized likelihood of both models and select this one as the final model.
If the final model with minimal penalized likelihood is a model $\mathcal{M}_0$ without change point, we verified that there are no change points in the data. In case the optimal model has no change point and each mean function per development year is constant, i.e the regression parameters per development year consist only of an intercept parameter, this corresponds to the original reserving model. Otherwise, if the optimal model contains additionally some slope parameters, we have to assume trends are contained in some development years and we should not trust the original reserving model. Similarly, in the case the final selected model is a model with one or more change points, we have to reject the hypothesis that there are no change points in the claims data and the means are equal per development year.
If we reject the original reserving model, because we assume that the model assumptions are violated, then we should consider these effects for an adequate prediction of the future losses. That is why we adjust the reserving model and use the final model, which is an extension of the original one by additional parameters, to predict the values of the lower triangle

$$\{Y_{i,k} \mid i + k - 1 > N\}.$$

The final model is defined by the regression parameters with some maximum likelihood estimates $\hat{\boldsymbol{\theta}}^{(k)}$, $k = 1, \ldots, D$, which we can use to predict the lower triangle up to the chosen limit $D < N$. Since we assume that all losses almost settled after development year $D$ and thus later development years have no essential impact on the prediction, for simplicity we use the estimated model parameters from the original reserving model for the prediction of the remaining development years $D + 1, \ldots, N$.
But for all development years $k = 1, \ldots, D$, we use the estimated regression parameters, which specify the mean function $\psi^{(k)}$ for each development year. This means we predict the future development by

$$\hat{Y}_{i,k} = \hat{\psi}^{(k)}(i) \qquad \text{for all } k = 1, \ldots, D \quad \text{and for all } i + k - 1 > N.$$

In the simulation study we evaluate the performance of the change point detection approach on the one hand and one the other hand we use this prediction method to examine the accuracy of estimates and compare it to the original model.

# Chapter 7

# Simulation study

In this chapter we provide a simulation study to explore the performance of the change point and trend detection method proposed in Section 6.5. We generate data sets for different scenarios and apply the method for each data set. By this we can estimate the number and location of change points, if there are some, and compare them to the true ones, which we generated in the simulation process. Furthermore, we can use the model, which is selected by this procedure, to predict the reserve and compare it to the true reserve. To quantify the impact of neglected trends and change points on the accuracy of the prediction in the original reserving method, we also predict the reserve by using the original method without adjustments.
First, we give an overview of the simulation setup and then we explain the scenarios with the choice of the parameters in detail. After that, we present the results for each scenario and conclude the chapter with a discussion of these results.

## 7.1 Overview of simulation setup

In this study we focus on the additive model as original reserving model and therefore we use loss ratios as data basis to analyze the loss development and predict the reserve. We consider different scenarios and simulate 100 data sets per scenario. The data sets contain $N$ underwriting years with premiums and a $N \times N$-matrix of loss ratios $M_{i,k}$, which define the loss ratios of all underwriting years $i$ and development years $k$, $i, k = 1, \ldots, N$. The loss ratios are generated by

$$M_{i,k} \sim \mathcal{N}\left(\mu_{i,k}, \sigma_{i,k}^2\right), \qquad i, k = 1, \ldots, N, \tag{7.1}$$

where $\mu_{i,k}$ are chosen scenario specific and discussed in detail in the next section. While the upper triangle of loss ratios

$$\{M_{i,k} \mid i + k - 1 \leq N\}$$

defines the observed loss ratios, which we will use for model selection and prediction, the lower triangle

$$\{M_{i,k} \mid i + k - 1 > N\}$$

is considered to be unknown and represents the true future loss development, which needs to be predicted. But before we analyze the prediction, we want to examine the accuracy of the change point detection of our approach. We are interested in the number and location of identified change points in each simulated data set. For each scenario we investigate whether each true change point was detected or not and the number of wrongly detected change points. In a scenario without true change points the detection is considered correct, when no change point is falsely identified.
Thereafter, we verify the precision of the reserve prediction by using the selected model, which allows for trends and change points, and the impact of this adjustment compared to the original additive reserving method. To make the absolute reserves and the derivations between true and predicted reserve comparable, we fix the premiums for all simulated data sets and all scenarios by some $v_1, \dots, v_N$. The reserve, as specified in Definition 3.2, is given by

$$R = \sum_{i=2}^{N} R_i \qquad \text{with} \qquad R_i = \sum_{k=N-i+2}^{N} S_{i,k}, \qquad i = 2, \dots, N,$$

where $S_{i,k}$ denote the incremental losses. By definition of loss ratios $M_{i,k}$, the incremental losses are given by $S_{i,k} = v_i \cdot M_{i,k}$. It follows that for simulated data set $l$, for $l = 1, \dots, 100$, we have the true reserve

$$R(l) = \sum_{i=1}^{N} R_i(l) \qquad \text{with} \qquad R_i(l) = \sum_{k=N-i+2}^{N} S_{i,k}(l) = v_i \sum_{k=N-i+2}^{N} M_{i,k}(l), \quad i = 2, \dots, N,$$

where $M_{i,k}(l)$ denote the realized loss ratios of simulation $l$. We compare the true reserve to the predicted reserve with the proposed model of the presented model selection approach from previous chapter. For each simulated data set $l$, the so predicted reserve is denoted by $\hat{R}^*(l)$ and it holds

$$\hat{R}_i^*(l) = \sum_{k=N-i+2}^{N} \hat{S}_{i,k}(l) = v_i \sum_{k=N-i+2}^{N} \hat{M}_{i,k}(l), \quad i = 2, \dots, N,$$

where we predict $M_{i,k}$ as discussed in Section 6.5 by

$$\hat{M}_{i,k} = \hat{\psi}^{(k)}(i)$$

for some mean function of development year $k = 1, \dots, N$ based on observations

$$\{M_{i,k}(l) \mid i + k - 1 \leq N\}.$$

Additionally, we estimate the reserve in the original additive reserving model as presented in Section 3.3, where we introduced this loss reserving model, based on the upper triangle

$$\{S_{i,k}(l) \mid i + k - 1 > N\}$$

with realized incremental losses $S_{i,k}(l) = v_i \cdot M_{i,k}(l)$. We denote the predicted reserve in the original additive reserving model $\hat{R}^0$. To compare the results we consider for the

original and the new prediction model the deviation between the true and the predicted reserve.
The deviation between the true reserve and the predicted reserve in simulation run $l$ is given by

$$R(l) - \hat{R}(l), \qquad l = 1, \dots, N, \qquad \text{for } \hat{R} = \begin{cases} \hat{R}^0, & \text{original model,} \\ \hat{R}^*, & \text{new model.} \end{cases}$$

We use these deviations to analyze both predictions of a scenario and calculate the average absolute deviation

$$\frac{1}{100} \sum_{l=1}^{100} \left| R(l) - \hat{R}(l) \right|$$

and the mean squared error

$$\frac{1}{100} \sum_{l=1}^{100} \left( R(l) - \hat{R}(l) \right)^2$$

in each scenario for both models. Furthermore, we consider the mean relative error

$$\frac{1}{100} \sum_{l=1}^{100} \frac{\left| R(l) - \hat{R}(l) \right|}{R(l)}$$

in both cases as evidence for or against the accuracy of the estimated reserve.
Section 7.3 presents these results for all scenarios that we discuss in the next section extensively.

## 7.2 Scenario design

In this section we finalize the simulation setup and describe the design of the scenarios in detail. This simulation study comprises seven different scenarios, where we generate the loss ratios from. We discuss the choice of reasonable values for the expected values $\mu_{i,k}$ and variance parameters $\sigma^2_{i,k}$ in (7.1) for each scenario.
We set the number of underwriting years $N$ to 30 and consider the following **scenarios**:

**(S1)** No change point and all development years have constant means

**(S2)** No change point, but there are trends in all development years

**(S3)** One change point in underwriting year $\tau = 15$ with constant means before and after the change point (Model A with one change point)

**(S4)** Same scenario as S3, but we increase the overall volatility parameters (Model A with one change point)

**(S5)** Two change points located in underwriting years $\tau_1 = 10$ and $\tau_2 = 20$ and additional trends in development years (Model A with two change points)

**(S6)** One change point in calendar year $\tau = 15$ with constant means before and after the change point (Model B with one change point)

**(S7)** One change point in calendar year $\tau = 20$ with constant means before the change point and trends after the change point (Model B with one change point)

with 100 replicates in each scenario.
In **scenario (S1)** all development years have a constant mean, this means it holds

$$\mu_{i,k}^{S1} = \mu^{(k)} \qquad \text{for all} \quad i, k = 1, \ldots, 30$$

for some $\mu^{(1)}, \ldots, \mu^{(30)}$. Since it is assumed that the loss ratios decrease exponentially to zero, we define

$$\mu^{(k)} = \exp\{-0.8 - 0.4 \cdot k\}, \qquad k = 1, \ldots, 30, \tag{7.2}$$

which yields the mean loss ratios per development year shown in Figure 14.

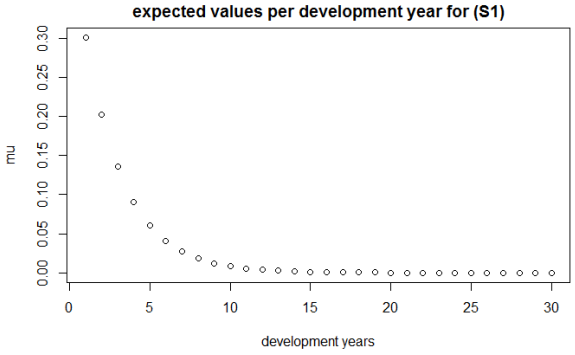



Figure 14: Expected loss ratios $\mu_{i,k}^{S1} = \mu^{(k)}$ for all underwriting years $i$ of development years $k$ in scenario (S1)

We define the limit of development years that we consider for trend and change point detection by $D = 10$, since the losses are almost settled at that point. To define the variances $\sigma_{i,k}^2$ we have to define the premiums $v_1, \ldots, v_{30}$ at first, since we assume heteroscedasticity in the original and the extended model with variances in the form

$$\sigma_{i,k}^2 = \frac{\sigma_k^2}{v_i}, \qquad i, k = 1, \ldots, 30.$$

As mentioned before, for reasons of comparability we fix the premiums for all data sets. Typically, the premiums of a portfolio are different for all underwriting years, but of similar size. We use the randomly generated premiums, which are all more or less 10 million per underwriting year, shown in Figure 15. We do not fix equal premiums for all underwriting

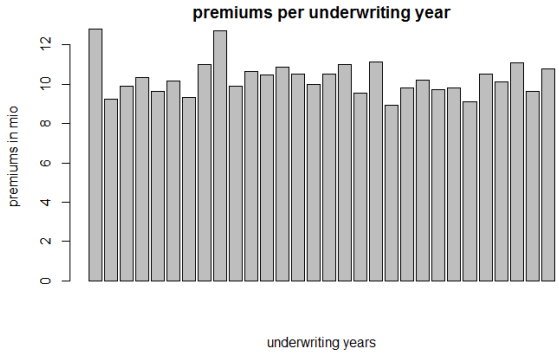



Figure 15: Simulated premiums per underwriting year

years to allow for the different variances depending on the premiums as weights. But the volume adjusted variance parameter $\sigma_k^2$ is assumed to be constant per development year. We fix this parameter per development year so that all loss ratios $M_{1,k}, \dots, M_{N,k}$ have an average variation coefficient of 0.3, which yields for variance parameters $\sigma_{i,k}^2$

$$\sigma_{i,k} = 0.3\ \mu^{(k)}\ \frac{\frac{1}{\sqrt{v_i}}}{\frac{1}{30}\sum_{j=1}^{30}\frac{1}{\sqrt{v_j}}},$$

where we used the mean parameters $\mu^{(k)}$, defined in (7.2). We fix these variances for all scenarios except of Scenario 4, where we consider the impact of higher variances on the accuracy of the change point detection.
For **scenario (S2)** we use the previously defined loss ratios $\mu^{(k)}$ from (S1) as starting point for each development year, but add a linear increasing trend of 2% per underwriting year. This means we define the mean values in (S2) as

$$\mu_{i,k}^{S2} = \mu^{(k)} + 0.02 \cdot \mu^{(k)} \cdot i, \qquad i = 1, \dots, 30,$$

which is illustrated in Figure 16.

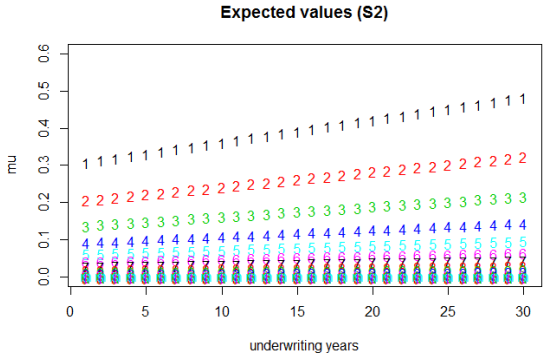



Figure 16: Expected loss ratios $\mu_{i,k}^{S2}$ against underwriting years $i$ per development year $k$ in scenario (S2)

To generate the data sets for **scenario (S3)** we split the data in two blocks and define the expected values for (S3) by

$$\mu_{i,k}^{S3} = \begin{cases} \mu_1^{(k)} = \mu^{(k)} & \text{for all } i = 1, \dots, 15, \\ \mu_2^{(k)} = \exp\{-0.1 - 0.4 \cdot k\} & \text{for all } i = 16, \dots, 30, \end{cases} \qquad k = 1, \dots, 30,$$

where we used the expected values $\mu^{(k)}$ of scenario (S1) as loss ratios for the upper block and defined the loss ratios for the lower block again as exponentially decreasing. The expected loss ratios for both are plotted in Figure 17, which shows that we set the loss ratios in the lower block higher compared to the loss ratios in older underwriting years.

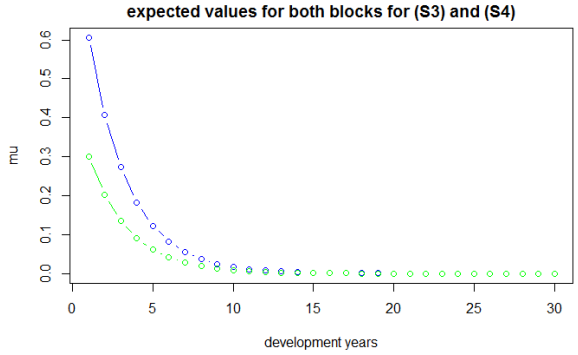



Figure 17: Expected loss ratios $\mu_1^{(k)}$ for the upper (green) and $\mu_2^{(k)}$ for the lower (blue) block of all development years in scenario (S3) and (S4)

For **scenario (S4)** we use the same expected values as defined in (S3), but double the variances compared to the other scenarios. This means we set

$$\mu_{i,k}^{S4} = \mu_{i,k}^{S3} \qquad \text{and} \qquad {\sigma_{i,k}^*}^2 = 2 \cdot \sigma_{i,k}^2$$

in order to investigate the impact of increasing volatility on the performance of the change point detection and the accuracy of the predicted reserve.
**Scenario (S5)** is the most complex scenario with two change points and trends.We split the data sets for this scenario in three blocks and define

$$\mu_{i,k}^{S5} = \begin{cases} \mu^{(k)} & \text{for all } i = 1, \dots, 10 \\ \mu_2^{(k)} - 0.07 \cdot \mu_2^{(k)} \cdot i & \text{for all } i = 11, \dots, 20 \\ \mu_{20,k}^{S5} & \text{for all } i = 21, \dots, 30 \end{cases} \qquad \text{for } k = 1, \dots, 30,$$

which is illustrated in Figure 18. Here we used again the mean loss ratios $\mu^{(k)}$ from scenario (S1) as mean of the first segment of each development year. For the second segments we used the means $\mu_2^{(k)}$, defined in scenario (S3), but only as intercept parameter and added a strong decreasing trend. After the second change point the mean values for the third segments of all development years remain constant at the reached level in the previous segments. By this scenario we want to evaluate, whether complex scenarios with

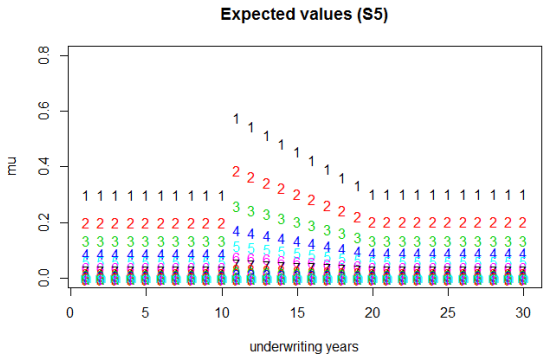



Figure 18: Expected loss ratios $\mu_{i,k}^{S5}$ against underwriting years $i$ per development year $k$ in scenario (S5)

many parameters can be identified, especially under consideration of the low number of observations.
Scenarios (S3)-(S5) are three scenarios with change points in underwriting years, this means they belong to Model A as defined in Section 6.5. To evaluate also the performance of the method for detection of change points in calendar years we consider two more scenarios of Model B.
For **scenario (S6)** we use the same values for the location of the change point and the means before and after the change points as defined in (S3), but now the loss development is separated by the change point $\tau = 15$ in calendar year (instead of underwriting year). That means, we define

$$\mu_{i,k}^{S6} = \begin{cases} \mu^{(k)} & \text{for all } i \leq 15 - k + 1, \\ \mu_2^{(k)} & \text{for all } i > 15 - k + 1, \end{cases} \qquad k = 1, \ldots, 30,$$

which is illustrated in Figure 19.

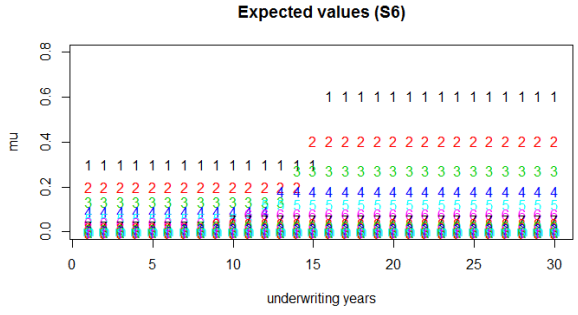



Figure 19: Expected loss ratios $\mu_{i,k}^{S6}$ against underwriting years $i$ per development year $k$ in scenario (S6)

Analogously, we define **scenario (S7)** as a scenario of Model B with one change point in a calendar year but now located at $\tau = 20$ and trend after this change points. We set

$$\mu_{i,k}^{S6} = \begin{cases} \mu^{(k)} & \text{for all } i \leq 20 - k + 1, \\ \mu_2^{(k)} + 0.02 \cdot \mu_2^{(k)} & \text{for all } i > 20 - k + 1, \end{cases} \qquad k = 1, \ldots, 30,$$

where we used again the values $\mu^{(k)}$ and $\mu_2^{(k)}$ from scenario (S3), but added an increasing trend after change point in calendar year $\tau = 20$. The expected values of (S7) are shown in Figure 20.

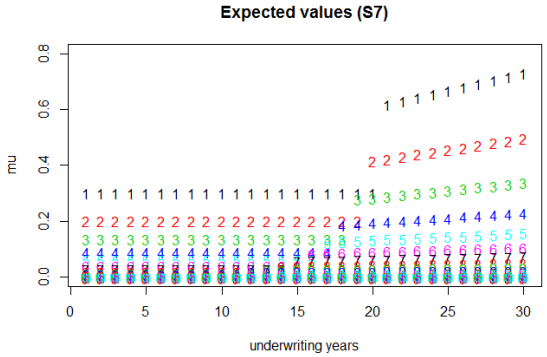



Figure 20: Expected loss ratios $\mu_{i,k}^{S7}$ against underwriting years $i$ per development year $k$ in scenario (S7)

Now we can generate data sets for each scenario and run the method defined in Section 6.5 to test for structural breaks in underwriting years and in calendar years by using the R functions presented in Appendix A.2 and predict the reserve under consideration of change points and trends. More details are given in Appendix B.1, there we present the generated premiums and the setup of the simulation study in R and also the results, which are described in the next section.

## 7.3 Simulation results

This section presents the results obtained in the simulation study for the discussed scenarios (S1)-(S7). For each scenario we analyze at first the number and locations of the detected change points and compare them to the true change points of each scenario. We count a detected change point in one of the simulated data sets as *exact detected true change point* in case the underlying scenario has exactly this change point. The remaining true change points with a detected change point in a range of the true change point $\pm 5$ are considered as at least *nearly detected*. For a data set with a true change point that is neither exactly nor nearly detected, we mark the true change point as *not detected*. In the case of scenario (S5) we analyze the two true change points separately. To evaluate the accuracy of the change point detection per scenario we have to identify also the number of detected change points, which are no true change points. Such a change point is a

*falsely detected change point*. The results of the change point detection for all scenarios are shown in Table 5.

The table shows that in both scenarios (S1) and (S2) without any true change point

| | Scenarios | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| change points | (S1) | (S2) | (S3) | (S4) | (S5) | | (S6) | (S7) |
| true change point in scenario | no | no | one | one | $1^{st}$ | $2^{nd}$ | one | one |
| total number exact detected | 0 | 0 | 100 | 100 | 100 | 24 | 100 | 100 |
| total number nearly detected | 0 | 0 | 0 | 0 | 0 | 76 | 0 | 0 |
| total number not detected | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| total number falsely detected | 0 | 0 | 11 | 18 | 60 | | 14 | 18 |

Table 5: Summary of detected change points for all scenarios with 100 data sets

actually also no change point was falsely identified as change point. In scenario (S3) with a single change point we detected all true change points exactly, but identified 11 times additional to the true change point falsely some point as change point. The number of falsely detected changes point increases in scenario (S4) to 18. As expected, Scenario 6 and Scenario 7 show a similar picture to Scenario 3 and Scenario 4, which shows that the method works for the detection of changes in underwriting years and calendar years. In scenario (S5) we have two true change points, the first one was exactly detected in all simulated data sets. The second change point was also identified in each case but in the majority of cases only nearly, furthermore the number of falsely detected change points is 60 in this scenario. We will discuss reasons for this in the next section, but to conclude all true change points were detected.

Probably even more interesting is the accuracy of the predicted reserve especially compared to the accuracy of the original reserving model. To measure the accuracy of estimates we use the average absolute deviation, the mean squared error and the mean relative error, which we defined in Section 7.1. We compare these statistical values for the predicted reserve based on the new model to the reserve predicted with the original model. The results are presented in Table 6. As expected the table shows that for scenario (S1) the predicted reserves are for both models very similar and close to the true reserve, but for all other scenarios we have large differences in the predicted reserve and therefore in deviation to the true reserve. While the predicted reserve taking the detected change points and trends into account is in every scenario quite precise with a low absolute and relative deviation to the true reserve, the original reserving method appears as not appropriate to predict the reserve for scenarios (S2)-(S7). In scenario (S2), which comprises only an increasing trend, the original reserving method neglects this and underestimates the reserve by more than 20% or in volume terms in average about 6.5 million based on an average true reserve of 30.1 million (see Appendix B.2) in this scenario. Scenarios (S3) and (S4) tell the same story in case of a change point, more than 12 million difference in average to a true reserve

**(I) Deviations between true and predicted reserve with the new model**

| | Scenarios | | | | | | |
|---|---|---|---|---|---|---|---|
| deviation | (S1) | (S2) | (S3) | (S4) | (S5) | (S6) | (S7) |
| average absolute in mio | 0.85 | 1.20 | 1.95 | 1.92 | 1.90 | 1.31 | 4.73 |
| mean squared in billion | 109 | 238 | 526 | 667 | 355 | 287 | 2571 |
| mean relative | 4.4% | 3.9% | 5.0% | 5.0% | 7.8% | 3.4% | 9.7% |

**(II) Deviations between true and predicted reserve with the original model**

| | Scenarios | | | | | | |
|---|---|---|---|---|---|---|---|
| deviation | (S1) | (S2) | (S3) | (S4) | (S5) | (S6) | (S7) |
| average absolute in mio | 0.80 | 6.54 | 12.77 | 12.58 | 3.01 | 8.24 | 20.4 |
| mean squared in billion | 102 | 4392 | 16467 | 16073 | 978 | 6946 | 41830 |
| mean relative | 4.1% | 21.6% | 32.9% | 32.5% | 15.4% | 21.3% | 42.0% |

Table 6: Comparison of prediction accuracy between new (I) and original (II) model

of about 38 million shows that the predicted reserve is not accurate. The result expressed as a percentage is an underestimation of the reserve by more than 32%. In contrast, the proposed model achieves more precise predictions, we have average absolute deviations of a few percentage points. Under consideration of the random deviation of the true reserve, which makes an exact prediction impossible, the predicted reserve by the new method proves as very accurate. Furthermore the new method shows random deviations in both directions, while the original method shows for all scenarios (S2)-(S7) a systematic over- or underestimation depending on the set-up of the scenario. Even for the complex scenario (S5) with two change points and a trend, the new method performed better compared to the original method. The results of scenario (S6) and scenario (S7) show that also in case of change points in calendar year we achieve with the new method more accurate predictions than the original method. Also the mean squared errors confirm the complete picture of clearly better prediction results in case of trends and change points compared to the original reserving model.

## 7.4 Discussion

By these results we can draw conclusions with regard to the accuracy of the detected change points on the one hand and the precision of the predicted reserve on the other

hand.
With the proposed method we were able to identify each true change point in all scenarios, both in underwriting and in calendar years. For scenarios (S3), (S4), (S6) and (S7) and the first change point of scenario (S5) this detection was totally exact, only the second true change point in (S5) is harder to detect exact. The reason for that is pretty clear, when the transition from one segment to the next is without a jump from one level to another but instead continuous, it is more difficult to identify the exact point, where the change is located. Nevertheless, we were able to identify the second true change point in this scenario at least nearly and in most data sets the true change point was only one point away. The fact, that there are some wrongly identified change points shows that the change point detection is sensitive and tends to slightly overfit the number of change points. But in both scenarios without change point no point was falsely identified as change point, which verifies that this overestimation of change points seems to happen only in case the data is actually disturbed by one or more change points. Scenario (S4) indicates that by increasing the common variance also the sensitivity increases, which is quite obvious. We should keep in mind that we do not have many observations per development year and change points divide these observations into even smaller connected sets, thus, it is clear that a precise estimation is increasingly complex with increasing parameters, which is again confirmed by scenario (S5) and scenario (S7). Therefore, the more structural changes and the higher the volatility for a low number of observations, the more sensitive the change point detection. It follows, in case more than one change point is detected it can be helpful to analyze how significant the change points are. For that one can use the change point analysis per development year presented in Section 6.3. If for most development years a model with exactly these change points is the model with lowest BIC and the difference to the model with only one change point is clearly, then this is an explicit evidence that all change points are actually true change points. But if we do not trust a change point, the selected model can be easily adjusted by removing this change point and using the maximum likelihood estimates in the model without this change point for prediction.
The comparison of the accuracy of the predicted reserves points out two aspects. The first one is that the proposed model proves as an appropriate way to predict the reserve, the predicted reserve is accurate in all tested scenarios. Only scenarios S5 and S7 differ from the 3-5 percent of other scenarios with slightly higher variance. The reason for the higher absolute deviation in both scenarios is that the last point is later than in the other scenarios and thus, the last segment on which the prediction is based is shorter. Again a smaller number of observations makes it more difficult to make a precise prediction, but compared to the original method the prediction accuracy is still clearly better.
In contrast, the original reserving method seems to be not sufficient to predict the reserve in case the data basis is disturbed by change points or trends. Scenarios (S2)-(S7) show that neglecting trends and structural changes in the loss data leads to bad estimations for the reserve. The relative error in the original model is at least double up to almost seven times as high as in the new model for these scenarios. Only if no trends and change points appear, which is rather untypical in practice, the classic model assumptions are fulfilled and the original model should be used. In case the data shows only change points but no trends the original reserving method can be used by ignoring the data before the

last change occurred, but even in this case a method is required to detect these changes. If trends are contained, the original method is not applicable at all. An increasing trend will always lead to an underestimation and a decreasing trend to an overestimation of the reserve, when the model is not adjusted.
Note that the results and the conclusions depend strong on the assumed scenarios. This includes the assumption of a normal distribution. We chose the normal distribution for two main reasons. The first one is because the maximum likelihood estimates correspond to the estimates in the original reserving models, which provide the basis for the proposed extension. Secondly, we need a distribution, which allows for negative values, since we discussed in Section 3.1, that there can be negative incremental losses. Using a method which is restricted to a positive response is therefore often not applicable in practice. But since typically log-normal or gamma distribution is used to model loss data, we investigated the impact on the outcome when using a different distribution assumption. For this we used the expected values and variances of scenario (S3), but generated the loss ratios from a log-normal distribution. We applied the proposed new method without any adjustment, i.e. we used a normal distribution to fit the data. The results as well as the code used to generate the data sets and results are given in detail in Appendix B.2. These results show that the proposed method is efficient even if the observed data is not normally distributed. We detected all true change points exactly and the predicted reserve is again very accurate with an average discrepancy of only 4.7% to the true reserve, while the original method gives no accurate estimates in case the data is disturbed by a change point. This certifies that the proposed method is robust even if the assumption of a normal distribution is not fulfilled.
To conclude, the proposed method solves two issues of the original method, which can lead to an inadequate estimation of the reserve at once, namely trends and change points. No separate analysis for both is required and we have shown in the simulation study that using the model, which includes the detected change points and trends, appears as an improved method to predict the reserve accurately.

# Conclusion

In this thesis we provided an approach to detect changes and abrupt breaks in claims data of non-life business portfolios, which form the basis for the assessment of the claims reserve and in an internal model also for the underwriting risk resulting from fluctuations in the claims settlement.
We modified two established reserving models by integrating trends and change points based on the theory of linear models and change point analysis. After describing the necessary framework and theoretical basics, we proposed a statistical method to investigate the loss history for structural changes. Then, we adapted this method to combine more development years in order to automatize the analyses and model selection, which we evaluated by means of a simulation study. By application of the presented method we detected all change points and using the selected model, which allows for trends and change points, for prediction of the reserve demonstrated precise prediction results. On the other hand it does not appear appropriate to predict the reserve in the presence of trends and change points by the original models without any adjustment.
The promising results provide opportunity for further research. In future work, it would be interesting to evaluate in detail the impact of change points and trends on the assessment of the premium and reserve risk. As we explained in Chapter 4, in an internal model the original reserving models are also used for the quantification of the premium risk and reserve risk. Obviously the uncertainty of the prediction cannot be correctly measured in case the past claims history is not modeled adequately, thus, the results of this thesis apply equally for risk modeling purposes as for the reserving. We expect similar results, but this should also be quantified in the same manner as for the reserve based on a simulation study but by assessing the fluctuation of the claims settlement.
Furthermore, the robustness of the methodology can be further extended with regard to practical application. Currently the data should not contain obvious outliers, since these can distort the results. Therefore, instead of cleaning up the data to avoid outliers before application of our proposed method, this could be easily integrated in the method to further advance automatization.
Already from the current results we can draw important conclusions. We believe that it is essential to analyze the claims history precisely and ensure that in case there are any structural breaks, they will be detected. Besides change points we should also pay attention to trends, because they can have a strong leverage on the future claims development. In practice, standard models without trend consideration are widely used for reserving and risk modeling purposes and if the data is analyzed for change points then on the basis of expert judgment. The presented approach is an objective analytical method to identify both change points and trends. This can improve analyses, model decisions and

results and furthermore, it can help to justify these decisions to regulatory supervision. As we have seen, only when the observed claims development is precisely analyzed, its structural behavior understood and appropriately modeled, it is possible to use the past claims to predict the future claims development accurately and thus the claims reserve and the risk capital. Keeping in mind, the enormous relevance of an adequate valuation of reserves and risk capital for insurance companies, the practical benefit of an application based on the ideas of this thesis becomes evident.

# Appendix A

# R Code

This appendix provides R code for the change point analysis per development year approach proposed in Section 6.3 and the R Code for the automatized model selection approach presented in Section 6.5, which is used for the simulation study in Chapter 7. We use no available package for change point detection, only the package pracma (2017) for some calculations.

We consider a data set with some $N \times N$-matrix **losses**, which defines the incremental loss triangle, and a $N$-dimensional vector **premium** for the premiums of all underwriting years. We can define the response and the weights depending on the chosen original reserving model by the following commands.

```
#calculate observed loss ratios and development factors
  M<-losses/premium
  F<-cbind( apply(losses[(1:(N-1)),1:2],1,sum)/losses[1:(N-1),1], sapply(3:N, function(k){apply(losses
       [1:(N-1),1:k],1,sum)/apply(losses[,1:(k-1)],1,sum)}))

#choose type="add" for the additive reserving model and type="mult" for the multiplicative reserving
     model
  if(type=="add"){
    Y=M
    W=matrix(premium,N,N)
  } else {
    Y=F
    W=cbind(losses[1:(N-1),1], sapply(2:(N-2), function(k){c(apply(losses[1:(N-k),1:k],1,sum),rep(NA,k
         -1))}))
    N=N-1
}
```

Note that we shifted all development years in the multiplicative model by $-1$ in order to start from DY 1. We will need this later.

## A.1 Analysis per development year

In this section we present the commands and defined functions to perform change point analysis per development year. We consider a fixed development year **k** and start with models of type I. We fit the data of this development year to Model 1a and Model 1b to calculate the BIC for both.

```
y=Y[1:(N-k+1),k]
w=W[1:(N-k+1),k]
n=length(y)

#fit models
    model1a=lm(y~1, weights=w)
```

```
    model1b=lm(y~c(1:n), weights= w)

#calculate sum of squared residuals
    SS1a=sum(w*(model1a$residuals)^2)
    SS1b=sum(w*(model1b$residuals)^2)

#calculate BIC
    Q1a=2*log(n)+n*(log(2*pi*SS1a/n)+1)-sum(log(w))
    Q1b=3*log(n)+n*(log(2*pi*SS1b/n)+1)-sum(log(w))
```

$Q1a$ is the BIC of Model 1a and $Q1b$ is the BIC of Model 1b. We compare the BIC of all models to the BIC $Q1a$ of Model 1a and analyze the difference between the BICs to decide if there is evidence against the base model and how strong it is. For a graphical illustration of Model 1a vs. Model 1b we use:

```
mu1a=predict(model1a)
mu1b=predict(model1b)

par(mfrow=c(1,2))
plot(y,main=paste("Model 1a with BIC", round(Q1a,2)))
lines(c(1:n),mu1a,col="blue")
plot(y,main=paste("Model 1b with BIC", round(Q1b,2)))
lines(c(1:n),mu1b,col="blue")
```

Next step is to find the optimal location of the change point $\tau$ for Model 2a and Model 2b of type II. We choose for each model the one with the smallest sum of squared residuals, since this yields the maximum possible likelihood for both models.

```
SS2a<-SS2b<-rep(NA,n-1)
#test all tau , but we want a minimum segment length of 3
  for(tau in 3:(n-3)){
  model2a.left=lm(y[1:tau]~1,weights=w[1:tau])
  model2a.right=lm(y[(tau+1):n]~1,weights=w[(tau+1):n])
  SS2a[tau]=sum(w[1:tau]*model2a.left$residuals^2)+sum(w[(tau+1):n]*model2a.right$residuals^2)
  model2b.left=lm(y[1:tau]~c(1:tau),weights=w[1:tau])
  model2b.right=lm(y[(tau+1):n]~c(1:(n-tau)),weights=w[(tau+1):n])
  SS2b[tau]=sum(w[1:tau]*model2b.left$residuals^2)+sum(w[(tau+1):n]*model2b.right$residuals^2)  }

#choose tau with minimal sum of squared residuals
tau.a=which.min(SS2a)
tau.b=which.min(SS2b)

#calculate BIC
Q2a=4*log(n)+n*(log(2*pi*SS2a[tau.a]/n)+1)-sum(log(w))
Q2b=6*log(n)+n*(log(2*pi*SS2b[tau.b]/n)+1)-sum(log(w))
```

The graphics for models 2a and models 2b are generated by

```
mu.a=c(predict(lm(y[1:tau.a]~1,weights=w[1:tau.a])),predict(lm(y[(tau.a+1):n]~1,weights=w[(tau.a+1):n])))
mu.b=c(predict(lm(y[1:tau.b]~c(1:tau.b),weights=w[1:tau.b])),predict(lm(y[(tau.b+1):n]~c(1:(n-tau.b)),weights=w[(tau.b
    +1):n])))

par(mfrow=c(1,2))
plot(y,main=paste("Model 2a with BIC", round(Q2a,2)))
lines(c(1:tau.a),mu.a[1:tau.a],col="blue")
lines(c((tau.a+1):n),mu.a[(tau.a+1):n],col="blue")
plot(y,main=paste("Model 2b with BIC", round(Q2b,2)))
lines(c(1:tau.b),mu.b[1:tau.b],col="blue")
lines(c((tau.b+1):n),mu.b[(tau.b+1):n],col="blue")
```

For models for type III we run a penalized likelihood approach to find the optimal number and location of change points and use the following:

```
  cp=replicate((n+1),list(0))
  model=replicate(n,list(SS=0, p=0, parameter=c(0,0,0,0,0), Q=NA))
#successive approach
  for(tau in 3:(N-k-2)){              # minimum segment length of 3
    index=c(1,which((c(1:tau)<=(tau-2))&(c(1:tau)>3)))
    interim.models<-sapply(index,function(t){ fit.DY(y,t,tau,w,previous.model=model[,t]) })
    problem<-sapply(1:length(index),function(t){ interim.models[,t]$Q})
    tau.star=which.min(problem)
    model[,tau+1]=interim.models[,tau.star]
    cp[[(tau+1)]]=unique(c(cp[[index[tau.star]]],tau)) }
#choose model with minimal BIC
q<-which.min(model[4,])
Q=model[4,q]$Q
cpts<-cp[[q]]
 #check if Model 1a or Model 1b with m=0 has lower cost, then this is optimal
    if(min(Q1a,Q1b)<=Q){
        Q$Q=min(Q1a,Q1b)
        cpts=NA
    if(Q1a<=Q1b){
         mu=rep(model1a$coefficients,n)
```

```
        } else{ mu=model1b$coefficients[1]+c(1:n)*model1b$coefficients[2]}
  }else{
    parameter=model[3,(cpts+1)]
    l=length(cpts)
    mu=numeric(0)
    for( j in 2:l){ mu= c(mu, (parameter[[j]][1]+parameter[[j]][2]*c(1:(cpts[j]-cpts[j-1])))) }
    mu=c(mu, (parameter[[l]][4]+parameter[[l]][5]*c(1:(n-cpts[l]))))
    cpts=cpts[which(cpts!=0)]
  }
```

Here we used the function fit.DY, which fits the best model with last change point tau based on the previous calculated models with last change point at t before tau. This function is defined as:

```
fit.DY<-function(y,t,tau,w,previous.model){
  n=length(y)

#try all possible models for segment t:tau
  model.a<-lm(y[t:tau]~c(1:(tau-t+1))+1,weights=w[t:tau]) #two new parameters: slope and intercept
  model.b<-lm(y[t:tau]~1,weights=w[t:tau])  #one new parameter: just intercept, no slope
  model.c<-lm(I(y[t:tau]-previous.model[[3]][3] - previous.model[[3]][2]*c(1:(tau-t+1)))~c(1:(tau-t+1))
      -1,weights=w[t:tau])  #one new parameter: slope change
  model.d<-lm(I(y[t:tau]-previous.model[[3]][3])~0,weights=w[t:tau]) #no new parameter: model stays at
      latest level
  models=list(model.a,model.b,model.c,model.d)
  SS=sapply(1:4, function(l){sum(w[t:tau]*(models[[l]]$residuals)^2)})

#now models for tau+1:n
  model2.a<-lm(y[(tau+1):n]~c(1:(n-tau))+1,weights=w[(tau+1):n]) #two new parameters: slope and
      intercept
  model2.b<-lm(y[(tau+1):n]~1,weights=w[(tau+1):n])  #one new parameter: just intercept, no slope
#the other two models depend on the previous fitted segment
  SS.right=c(sum(w[(tau+1):n]*(model2.a$residuals)^2),sum(w[(tau+1):n]*(model2.b$residuals)^2))
  p=c(2,1,1,0,2,1)
  a=c(coef(model.a)[1],coef(model.b)[1],previous.model[[3]][3],previous.model[[3]][3])
  b=c(coef(model.a)[2],0,previous.model[[3]][2]+coef(model.c)[1],0)
  L=4
  if(t==1){ L=2}  #if t=1 this is the first segment and we allow only for model a or b
  Q=matrix(NA,L,4)
  a.right=cbind(rep(coef(model2.a)[1],L), rep(coef(model2.b)[1],L),matrix(NA,L,2))
  b.right=cbind(rep(coef(model2.a)[2],L), rep(0,L),matrix(NA,L,2))
  new.level=a+(tau-t+1)*b #latest level of segment t:tau depending for all possible models
  SS.rest<-matrix(NA,L,2)

#now we calculate the cost Q for each combination of first and second model
  for(l in 1:L){
    for(q in 1:2){
      Q[l,q]=n*(log(2*pi*(previous.model[[1]]+SS[l]+SS.right[q])/(n))+1)-sum(log(w))+ (previous.model
          [[2]]+p[l]+p[(4+q)]+2)*log(n)     }
    model2.c<-lm(I(y[(tau+1):n]-new.level[l]-b[l]*c(1:(n-tau)))~c(1:(n-tau))-1,weights=w[(tau+1):n])
    model2.d<-lm(I(y[(tau+1):n]-new.level[l])~0,weights=w[(tau+1):n])
    a.right[l,3:4]=c(new.level[l],new.level[l])
    b.right[l,3:4]=c(b[l]+coef(model2.c)[1],0)
    SS.rest[l,]= c(sum(w[(tau+1):n]*(model2.c$residuals)^2),sum(w[(tau+1):n]*(model2.d$residuals)^2))
    Q[l,3]=n*(log(2*pi*(previous.model[[1]]+SS[l]+SS.rest[l,1])/(n))+1)-sum(log(w))+ (previous.model
        [[2]]+p[l]+1+2)*log(n)
    Q[l,4]=n*(log(2*pi*(previous.model[[1]]+SS[l]+SS.rest[l,2])/(n))+1)-sum(log(w))+ (previous.model
        [[2]]+p[l]+2)*log(n)   }
  best.l=which(Q==min(Q),arr.ind=TRUE)[1,1]
  best.r=which(Q==min(Q),arr.ind=TRUE)[1,2]
  return(list(SS=(previous.model[[1]]+ SS[best.l]), p=(previous.model[[2]]+p[best.l]+1), parameter=c(a
      [best.l], b[best.l], new.level[best.l], a.right[best.l,best.r] , b.right[best.l,best.r]), Q=
      min(Q)))
}
```

The final $Q$ is the BIC of the optimal model with change points located at *cpts*. We can plot the final model with lowest BIC by:

```
plot(y,main=paste("Model 3 with BIC", round(Q,2)),xlab="Underwriting Years",ylab="Loss Ratios",pch=16)
    if(any(is.na(cpts))){
      lines(c(1:n),mu,col="blue")
    }else{
    lines(c(1:cpts[1]),mu[1:cpts[1]],col="blue")
    if(length(cpts)>1){
      for(l in 2:length(cpts)){
        lines(c((cpts[l-1]+1):cpts[l]),mu[(cpts[l-1]+1):cpts[l]],col="blue") }}
    lines(c((cpts[length(cpts)]+1):n), mu[(cpts[length(cpts)]+1):n],col="blue")
    abline(v=cpts+0.5,col="red") }
```

## A.2 Automatized model selection

In this section we present the approach to find the optimal number and location of change points for more development years, at first located at the same underwriting years (Model A) and then in the same calendar years (Model B). The limit $D$ is denoted $DYs$ in the code below. We need three slightly different functions for Model A and Model B. The first one starts with $m = 0$, i.e. fits the best model per development year without change points but allows for trends and is equal for Model A and Model B.

```
typeI<-function(Y,W,N,DYs,U){                    # U is the total number of observations
#fit model with and without for all DYs
  model1a=sapply(1:DYs, function(k){lm(Y[1:(N-k+1),k]~1, weights=W[1:(N-k+1),k])})
  model1b=sapply(1:DYs, function(k){lm(Y[1:(N-k+1),k]~c(1:(N-k+1)), weights=W[1:(N-k+1),k])})
#calculate sum of squared residuals
  SS1a=sapply(1:DYs, function(k){sum(W[1:(N-k+1),k]*(model1a[,k]$residuals)^2)})
  SS1b=sapply(1:DYs, function(k){sum(W[1:(N-k+1),k]*(model1b[,k]$residuals)^2)})
#calculate cost per development year
  Q1a=sapply(1:DYs, function(k){2*log(U)+(N-k+1)*log(2*pi*SS1a[k]/(N-k+1))})
  Q1b=sapply(1:DYs, function(k){2*log(U)+(N-k+1)*log(2*pi*SS1b[k]/(N-k+1))})
#choose minimal per development year
  mu=sapply(1:DYs, function(k){if(Q1a[k]<=Q1b[k]){
    rep(model1a[,k]$coefficients,N)
  }else{model1b[,k]$coefficients[1]+model1b[,k]$coefficients[2]*c(1:N) }})
  Q=sapply(1:DYs, function(k){if(Q1a[k]<=Q1b[k]){
    Q1a[k]
  }else{ Q1b[k] }})
  p=rep(2,DYs)
  p[which(Q==Q1b)]=3
  parameter=sapply(1:DYs,function(k){if(Q1a[k]<=Q1b[k]){
    c(model1a[,k]$coefficients[1],0)
  }else{ c(model1b[,k]$coefficients[1],model1b[,k]$coefficients[2])}})
  return(list(Q.total=sum(Q),Q=Q,mu=mu, parameter=parameter, p=p))
}
```

The selected $\mu$ in by the function are the predicted values per development year in case the selected model without change points has minimal cost. We need the function later for the main function, where all possible models with $m > 0$ are tested. There we need again a function that fits the optimal model to the last segments based on the previous models. This is given for Model A by the following function, which is an extension of the function *fit.DY()* presented in the previous section of this appendix.

```
fit<-function(y,t,tau,w,previous.model,N,DYs,U,typI){
  nSS<-nQ<-rep(0,DYs)
  np<-rep(0,(DYs+1))
  nparameter=matrix(0,5,DYs)
  K=DYs
  if((N-tau+1)<=DYs){
    for(j in (N-tau+1):DYs){
      nSS[j]=previous.model$SS[j]
      np[j]=previous.model$p[j]
      nparameter[,j]=previous.model$parameter[,j]
      nQ[j]=previous.model$Q[j]
      if(t==1){
        nQ[j]=typI$Q[j]
        np[j]=typI$p[j]
        nparameter[,j]=c(typI$parameter[,j],NA, NA,NA)
      }
    }
    K=N-tau
  }
  model.a<-sapply(1:K, function(k){lm(Y[t:tau,k]~c(1:(tau-t+1))+1, weights=W[t:tau,k])})
  model.b<-sapply(1:K, function(k){lm(Y[t:tau,k]~1, weights=W[t:tau,k])})
  model.c<-sapply(1:K, function(k){lm(I(Y[t:tau,k]-previous.model$parameter[3,k]-c(1:(tau-t+1))*
      previous.model$parameter[2,k])~c(1:(tau-t+1))-1, weights=W[t:tau,k])})
  model.d<-sapply(1:K, function(k){lm(I(Y[t:tau,k]-previous.model$parameter[3,k])~0, weights=W[t:tau,k
      ])})
  model.e<-sapply(1:K, function(k){lm(I(Y[t:tau,k]-previous.model$parameter[3,k]-c(1:(tau-t+1))*
      previous.model$parameter[2,k])~0, weights=W[t:tau,k])})
  models=list(model.a,model.b, model.c,model.d, model.e)
  SS.left=sapply(1:5, function(l){sapply(1:K, function(k){sum(W[t:tau,k]*(models[[l]][,k]$residuals)^2)
      })})
  p=c(2,1,1,0,0,2,1)
  model2.a<-sapply(1:K, function(k){lm(Y[(tau+1):(N-k+1),k]~c(1:(N-k+1-tau))+1, weights=W[(tau+1):(N-k
      +1),k])})
  model2.b<-sapply(1:K, function(k){lm(Y[(tau+1):(N-k+1),k]~1, weights=W[(tau+1):(N-k+1),k])})
  a<-sapply(1:K, function(k){c(model.a[,k]$coefficients[1],model.b[,k]$coefficients[1],previous.model$
      parameter[3,k],previous.model$parameter[3,k],previous.model$parameter[3,k])})
  b<-sapply(1:K, function(k){c(model.a[,k]$coefficients[2],0,previous.model$parameter[2,k]+model.c[,k]$
      coefficients[1],0,previous.model$parameter[2,k])})
```

```
  SS.right=cbind(sapply(1:K, function(k){sum(W[(tau+1):(N-k+1),k]*(model2.a[,k]$residuals)^2)}),sapply
      (1:K, function(k){sum(W[(tau+1):(N-k+1),k]*(model2.b[,k]$residuals)^2)}))
  L=5
  if(t==1){ L=2}
  Q=array(NA,dim=c(L,5,K))
  a.r=sapply(1:K, function(k){c(model2.a[,k]$coefficients[1],model2.b[,k]$coefficients[1],0,0,0)})
  b.r=sapply(1:K, function(k){c(model2.a[,k]$coefficients[2],0,0,0,0)})
  a.right=replicate(L,a.r)
  b.right=replicate(L,b.r)
  SS.rest<-array(NA,dim=c(L,3,K))
  new.level=sapply(1:L, function(l){a[l,]+(tau-t+1)*b[l,]})
  for(l in 1:L){
    for(q in 1:2){
      Q[l,q,]=c(N:(N-K+1))*log(2*pi*(previous.model$SS[1:K]+SS.left[,l]+SS.right[,q])/c(N:(N-K+1)))+ (
          previous.model$p[1:K]+p[l]+p[(5+q)])*log(U)
    }
    mu2=sapply(1:K, function(k){ mu2=a[l,k]+c(1:(N-k+1-tau))*b[l,k]})
    model2.c<-sapply(1:K, function(k){lm(I(Y[(tau+1):(N-k+1),k]-new.level[k,l]-c(1:(N-k+1-tau))*b[l,k])
        ~c(1:(N-k+1-tau))-1, weights=W[(tau+1):(N-k+1),k])})
    model2.d<-sapply(1:K, function(k){lm(I(Y[(tau+1):(N-k+1),k]-new.level[k,l])~0, weights=W[(tau+1):(N
        -k+1),k])})
    model2.e<-sapply(1:K, function(k){lm(I(Y[(tau+1):(N-k+1),k]-new.level[k,l]-c(1:(N-k+1-tau))*b[l,k])
        ~0, weights=W[(tau+1):(N-k+1),k])})
    a.right[3,,l]<-new.level[,l]
    a.right[4,,l]<-new.level[,l]
    a.right[5,,l]<-new.level[,l]
    b.right[3,,l]=b[l,]+sapply(1:K, function(k){model2.c[,k]$coefficients[1]})
    b.right[4,,l]=rep(0,K)
    b.right[5,,l]=b[l,]
    SS.rest[l,,]= rbind(sapply(1:K, function(k){sum(W[(tau+1):(N-k+1),k]*(model2.c[,k]$residuals)^2)}),
        sapply(1:K, function(k){sum(W[(tau+1):(N-k+1),k]*(model2.d[,k]$residuals)^2)}),sapply(1:K,
        function(k){sum(W[(tau+1):(N-k+1),k]*(model2.e[,k]$residuals)^2)}))
    Q[l,3,]=c(N:(N-K+1))*log(2*pi*(previous.model$SS[1:K]+SS.left[,l]+SS.rest[l,1,])/c(N:(N-K+1)))+ (
        previous.model$p[1:K]+p[l]+1)*log(U)
    Q[l,4,]=c(N:(N-K+1))*log(2*pi*(previous.model$SS[1:K]+SS.left[,l]+SS.rest[l,2,])/c(N:(N-K+1)))+ (
        previous.model$p[1:K]+p[l])*log(U)
    Q[l,5,]=c(N:(N-K+1))*log(2*pi*(previous.model$SS[1:K]+SS.left[,l]+SS.rest[l,3,])/c(N:(N-K+1)))+ (
        previous.model$p[1:K]+p[l])*log(U)
  }
  for(k in 1:K){
    nQ[k]=min(Q[,,k])
    best.l=which(Q[,,k]==min(Q[,,k]),arr.ind=TRUE)[1,1]
    best.r=which(Q[,,k]==min(Q[,,k]),arr.ind=TRUE)[1,2]
    nSS[k]=previous.model$SS[k]+SS.left[k, best.l]
    np[k]=previous.model$p[k]+p[best.l]
    nparameter[,k]=c(a[best.l,k ], b[best.l,k ], new.level[k,best.l], a.right[best.r,k,best.l] , b.
        right[best.r,k,best.l])
  }
  Q.total=sum(nQ)+(previous.model$p[(DYs+1)]+1)*log(U)
  np[length(np)]=previous.model$p[(DYs+1)]+1
  return( list(SS=nSS, p=np, parameter=nparameter, Q=nQ, Q.total=Q.total))
}
```

We need this function again in the main function *CP*, which estimates the optimal number and location of change points for Model A.

```
CP<-function(Y,W,DYs,N){
  E=fliplr(triu(ones(N),0))[,1:DYs]
#total number of observations
  U=sum(E)
  cp=replicate((N+1),list(0))
#use function typeI to get the optimal model without change point
  typI=typeI(Y,W,N,DYs,U)
  model=replicate(N,list(SS=rep(0,DYs), p=rep(0,DYs+1), parameter=matrix(0,5,DYs), Q=rep(NA,DYs), Q.
      total=NA))
  model[,1]$p=c(rep(1,DYs),0)

  for(tau in 3:(N-3)){
    index=c(1,which((c(1:tau)<=(tau-2))&(c(1:tau)>3)))
    interim.models<-sapply(index,function(t){ fit(Y,t,tau,W,previous.model=model[,t],N,DYs,U,typI) })
    problem<-sapply(1:length(index),function(t){ interim.models[,t]$Q.total})
    tau.star=which.min(problem)
    model[,(tau+1)]=interim.models[,tau.star]
    cp[[tau+1]]=unique(c(cp[[index[tau.star]]],tau))
  }
  q<-which.min(model[5,])
  Q.total=model[5,q]$Q
  cpts<-cp[[q]]
  if(typI$Q.total<=Q.total){
    Q.total=typI$Q.total
    cpts=NA
    mu=typI$mu
  } else{
    parameter=model[3,(cpts+1)]
    l=length(cpts)
    mu=rep(numeric(0),DYs)
    for( j in 2:l){ mu= rbind(mu, sapply(1:DYs, function(k){ parameter[[j]][1,k]+parameter[[j]][2,k]*c
        (1:(cpts[j]-cpts[j-1]))})) }
    mu= rbind(mu, sapply(1:DYs, function(k){parameter[[1]][4,k]+parameter[[l]][5,k]*c(1:(N-cpts[l]))}))
    cpts=cpts[which(cpts!=0)]
  }
```

```
  return(list(Q.total=Q.total,Q=model[4,q], cpts=cpts,mu=mu))
}
```

For Model B the main function and the fit function work similarly to detect trends and change points in calendar years. The fit function (denoted *fit.CY*) is given by:

```
fit.CY<-function(y,t,tau,w,previous.model,N,DYs,U,typI){
  nSS<-nQ<-rep(0,DYs)
  np<-rep(0,(DYs+1))
  nparameter=matrix(0,5,DYs)
  nmu=previous.model$mu
  K=DYs
  if(DYs>tau){
    for(j in (tau+1):DYs){
      nQ[j]=typI$Q[j]
      np[j]=typI$p[j]
      nmu[,j]=typI$mu[,j]
      nparameter[,j]=c(0,0,0, typI$parameter[,j])
    }
    K=tau
  }
  model.a<-sapply(1:K, function(k){lm(Y[max(t-k+1,1):(tau-k+1),k]~c(1:((tau-k+1)-max(t-k+1,1)+1))+1,
      weights=W[max(t-k+1,1):(tau-k+1),k])})
  model.b<-sapply(1:K, function(k){lm(Y[max(t-k+1,1):(tau-k+1),k]~1, weights=W[max(t-k+1,1):(tau-k+1),k
      ])})
  model.c<-sapply(1:K, function(k){lm(I(Y[max(t-k+1,1):(tau-k+1),k]-previous.model$parameter[3,k]-c
      (1:((tau-k+1)-max(t-k+1,1)+1))*previous.model$parameter[2,k])~c(1:((tau-k+1)-max(t-k+1,1)+1))-1,
       weights=W[max(t-k+1,1):(tau-k+1),k])})
  model.d<-sapply(1:K, function(k){lm(I(Y[max(t-k+1,1):(tau-k+1),k]-previous.model$parameter[3,k])~0,
      weights=W[max(t-k+1,1):(tau-k+1),k])})
  model.e<-sapply(1:K, function(k){lm(I(Y[max(t-k+1,1):(tau-k+1),k]-previous.model$parameter[3,k]-c
      (1:((tau-k+1)-max(t-k+1,1)+1))*previous.model$parameter[2,k])~0, weights=W[max(t-k+1,1):(tau-k
      +1),k])})
  models=list(model.a,model.b, model.c,model.d, model.e)
  SS.left=sapply(1:5, function(l){sapply(1:K, function(k){sum(W[max(t-k+1,1):(tau-k+1),k]*(models[[l
      ]][,k]$residuals)^2)})})
  p=c(2,1,1,0,0,2,1)
  x2=sapply(1:K,function(k){c(1:(N-tau+k-1))})
  model2.a<-sapply(1:K, function(k){lm(Y[(tau-k+2):N,k]~x2[[k]]+1, weights=W[(tau-k+2):N,k])})
  model2.b<-sapply(1:K, function(k){lm(Y[(tau-k+2):N,k]~1, weights=W[(tau-k+2):N,k])})
  a<-sapply(1:K, function(k){c(model.a[,k]$coefficients[1],model.b[,k]$coefficients[1],previous.model$
      parameter[3,k],previous.model$parameter[3,k],previous.model$parameter[3,k])})
  b<-sapply(1:K, function(k){c(model.a[,k]$coefficients[2],0,previous.model$parameter[2,k]+model.c[,k]$
      coefficients[1],0,previous.model$parameter[2,k])})
  SS.right=cbind(sapply(1:K, function(k){sum(W[(tau-k+2):(N-k+1),k]*(model2.a[[k]]$residuals)^2)}),
      sapply(1:K, function(k){sum(W[(tau-k+2):(N-k+1),k]*(model2.b[[k]]$residuals)^2)}))
  L=5
  if(t==1){ L=2}
  Q=array(NA,dim=c(L,5,K))
  a.r=sapply(1:K, function(k){c(model2.a[[k]]$coefficients[1],model2.b[[k]]$coefficients[1],0,0,0)})
  b.r=sapply(1:K, function(k){c(model2.a[[k]]$coefficients[2],0,0,0,0)})
  a.right=replicate(L,a.r)
  b.right=replicate(L,b.r)
  new.level=sapply(1:L, function(l){sapply(1:K, function(k){a[l,k]+((tau-t+1)-max(t-k+1,1)+1)*ifelse(is
      .na(b[l,k]),0,b[l,k])})})
  SS.rest<-array(NA,dim=c(L,3,K))
  model2.c<-sapply(1:L, function(l){sapply(1:K, function(k){lm(I(Y[(tau-k+2):N,k]-new.level[k,l]-x2[[k
      ]]*ifelse(is.na(b[l,k]),0,b[l,k]))~x2[[k]]-1, weights=W[(tau-k+2):N,k])})})
  model2.d<-sapply(1:L, function(l){sapply(1:K, function(k){lm(I(Y[(tau-k+2):N,k]-new.level[k,l])~0,
      weights=W[(tau-k+2):N,k])})})
  model2.e<-sapply(1:L, function(l){sapply(1:K, function(k){lm(I(Y[(tau-k+2):N,k]-new.level[k,l]-x2[[k
      ]]*ifelse(is.na(b[l,k]),0,b[l,k]))~0, weights=W[(tau-k+2):N,k])})})
  for(l in 1:L){
    for(q in 1:2){
      Q[l,q,]=c(N:(N-K+1))*log(2*pi*(previous.model$SS[1:K]+SS.left[,l]+SS.right[,q])/c(N:(N-K+1)))+ (
          previous.model$p[1:K]+p[l]+p[(5+q)])*log(U)
    }
    SS.rest[l,,]= rbind(sapply(1:K, function(k){sum(W[(tau-k+2):(N-k+1),k]*(model2.c[,l][[k]]$residuals
        )^2)}),sapply(1:K, function(k){sum(W[(tau-k+2):(N-k+1),k]*(model2.d[,l][[k]]$residuals)^2)}),
        sapply(1:K, function(k){sum(W[(tau-k+2):(N-k+1),k]*(model2.e[,l][[k]]$residuals)^2)}))
    a.right[3,,l]<-new.level[,l]
    a.right[4,,l]<-new.level[,l]
    a.right[5,,l]<-new.level[,l]
    b.right[3,,l]=ifelse(is.na(b[l,k]),0,b[l,k])+sapply(1:K, function(k){model2.c[[k]]$coefficients
        [1]})
    b.right[4,,l]=rep(0,K)
    b.right[5,,l]=ifelse(is.na(b[l,k]),0,b[l,k])
    Q[l,3,]=c(N:(N-K+1))*log(2*pi*(previous.model$SS[1:K]+SS.left[,l]+SS.rest[l,1,])/c(N:(N-K+1)))+ (
        previous.model$p[1:K]+p[l]+1)*log(U)
    Q[l,4,]=c(N:(N-K+1))*log(2*pi*(previous.model$SS[1:K]+SS.left[,l]+SS.rest[l,2,])/c(N:(N-K+1)))+ (
        previous.model$p[1:K]+p[l])*log(U)
    Q[l,5,]=c(N:(N-K+1))*log(2*pi*(previous.model$SS[1:K]+SS.left[,l]+SS.rest[l,3,])/c(N:(N-K+1)))+ (
        previous.model$p[1:K]+p[l])*log(U)
  }
  for(k in 1:K){
    L.k=L
    if((t-k+1)<=1){L.k=2}
    nQ[k]=min(Q[,,k])
    best.l=which(Q[,,k]==min(Q[1:L.k,,k]),arr.ind=TRUE)[1,1]
    best.r=which(Q[,,k]==min(Q[1:L.k,,k]),arr.ind=TRUE)[1,2]
    nSS[k]=previous.model$SS[k]+SS.left[k, best.l]
```

```
    np[k]=previous.model$p[k]+p[best.l]
    nmu[max(t-k+1,1):(tau-k+1),k]=models[[best.l]][,k]$fitted.values+rbind(rep(0,((tau-k+1)-max(t-k
        +1,1)+1)),rep(0,((tau-k+1)-max(t-k+1,1)+1)), previous.model$parameter[3,k]+c(1:((tau-k+1)-max(
        t-k+1,1)+1)), rep(previous.model$parameter[3,k],((tau-k+1)-max(t-k+1,1)+1)), previous.model$
        parameter[3,k]+c(1:((tau-k+1)-max(t-k+1,1)+1)))[best.l,]
    nmu[(tau-k+2):N,k]=rbind(predict(model2.a[[k]],data.frame(x2[[k]])),predict(model2.b[[k]],data.
        frame(x2[[k]])),predict(model2.c[,best.l][[k]],data.frame(x2[[k]]))+new.level[k,best.l]+x2[[k
        ]]*ifelse(is.na(b[best.l,k]),0,b[best.l,k]),predict(model2.d[,best.l][[k]],data.frame(x2[[k]])
        )+new.level[k,best.l],predict(model2.e[,best.l][[k]],data.frame(x2[[k]]))+new.level[k,best.l]+
        x2[[k]]*ifelse(is.na(b[best.l,k]),0,b[best.l,k]))[best.r,]
    nparameter[,k]=c(a[best.l,k], b[best.l,k], new.level[k,best.l], a.right[best.r,k,best.l], b.
        right[best.r,k,best.l])
  }
  Q.total=sum(nQ)+(previous.model$p[(DYs+1)]+1)*log(U)
  np[length(np)]=previous.model$p[(DYs+1)]+1
  return(list(SS=nSS, p=np, parameter=nparameter, Q=nQ, Q.total=Q.total, mu=nmu, cp=c(previous.model$
      cp,tau)))
}
```

Using the *fit.CY* function we can run the successive approach to get the optimal number and location of change points in calendar years by function *CP.CY*:

```
CP.CY<-function(Y,W,DYs,N){
  E=fliplr(triu(ones(N),0))[,1:DYs]
  U=sum(E)
  cp=replicate((N+1),list(0))
  typI=typeI(Y,W,N,DYs,U)
  model=replicate(N,list(SS=rep(0,DYs), p=rep(0,DYs+1), parameter=matrix(0,5,DYs), Q=rep(NA,DYs), Q.
      total=NA,mu=matrix(0,N,DYs),cp=numeric(0)))
  model[,1]$p=c(rep(1,DYs),0)
  for(tau in 3:(N-3)){
    index=c(1,which((c(1:tau)<=(tau-2))&(c(1:tau)>3)))
    interim.models<-sapply(index,function(t){ fit.CY(Y,t,tau,W,previous.model=model[,t],N,DYs,U,typI)
        })
    problem<-sapply(1:length(index),function(t){ interim.models[,t]$Q.total})
    tau.star=which.min(problem)
    model[,(tau+1)]=interim.models[,tau.star]
  }
  q<-which.min(model[5,4:(N-2)])+3
  Q.total=model[5,q]$Q.total
  if(typI$Q.total<=Q.total){
    Q.total=typI$Q.total
    cpts=NA
    mu=typI$mu
  } else{
    mu=model[6,q]$mu
    cpts=model[7,q]$cp
  }
  return(list(Q.total=Q.total,Q=model[4,q], cpts=cpts,mu=mu))
}
```

## A.3 Example

The first development years of the cumulative loss triangle used in the example of Section 6.4 are presented in the following table.

| UY | DY 1 | DY 2 | DY 3 | DY 4 | DY 5 | DY 6 | DY 7 | DY 8 | DY 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1970 | 844370.00 | 987342.17 | 1115249.05 | 1225124.76 | 1321455.66 | 1360466.81 | 1377140.47 | 1381460.78 | 1383066.07 |
| 1971 | 1020430.00 | 1186505.06 | 1354310.94 | 1473560.69 | 1592463.74 | 1635972.00 | 1654515.86 | 1661871.63 | 1664328.32 |
| 1972 | 931780.00 | 1081999.39 | 1181585.17 | 1312143.71 | 1414116.34 | 1461575.67 | 1475056.51 | 1481876.07 | 1484234.44 |
| 1973 | 1141040.00 | 1343205.36 | 1465270.90 | 1649588.00 | 1816154.27 | 1871573.26 | 1889668.85 | 1897057.42 | 1899462.85 |
| 1974 | 962380.00 | 1061685.56 | 1187332.95 | 1295865.26 | 1401378.40 | 1445569.98 | 1461420.93 | 1467129.62 | 1469442.63 |
| 1975 | 1017480.00 | 1168119.79 | 1245972.55 | 1367953.36 | 1498247.81 | 1542641.29 | 1561670.20 | 1568137.49 | 1570102.29 |
| 1976 | 822560.00 | 952904.14 | 1035859.77 | 1131733.48 | 1228164.84 | 1263913.31 | 1276373.92 | 1282075.44 | 1284216.50 |
| 1977 | 1040670.00 | 1192921.13 | 1361630.19 | 1507117.80 | 1656212.73 | 1721199.26 | 1744326.79 | 1751170.07 | 1754245.51 |
| 1978 | 1010360.00 | 1129288.98 | 1258627.70 | 1378617.22 | 1478982.52 | 1524101.74 | 1541837.31 | 1548424.85 | 1551165.41 |
| 1979 | 1174750.00 | 1305571.55 | 1401544.50 | 1547662.82 | 1694672.32 | 1730788.79 | 1747441.24 | 1754158.31 | 1757244.68 |
| 1980 | 983530.00 | 1139059.61 | 1206787.38 | 1319069.76 | 1422383.06 | 1467418.66 | 1483447.11 | 1490087.34 | 1492128.21 |
| 1981 | 943080.00 | 1029022.37 | 1128794.02 | 1242127.70 | 1350384.56 | 1396870.47 | 1413805.27 | 1419513.18 | 1422167.46 |
| 1982 | 987140.00 | 1048189.42 | 1231664.58 | 1377215.56 | 1506924.25 | 1550473.64 | 1570027.28 | 1575730.95 | 1578162.13 |
| 1983 | 1102400.00 | 1204134.29 | 1312593.08 | 1420989.40 | 1555102.04 | 1599451.53 | 1618672.90 | 1624062.91 | 1626574.91 |
| 1984 | 962910.00 | 1079986.92 | 1180170.55 | 1314050.21 | 1412572.32 | 1459900.77 | 1477553.34 | 1482956.98 | 1485485.37 |
| 1985 | 923650.00 | 1005873.41 | 1080458.59 | 1203438.66 | 1316116.48 | 1349141.40 | 1361325.91 | 1367857.93 | 1369615.16 |
| 1986 | 1139310.00 | 1268483.24 | 1373708.99 | 1524211.04 | 1631192.95 | 1673233.20 | 1692975.03 | 1700082.36 | 1702506.74 |
| 1987 | 893020.00 | 967111.57 | 1097581.60 | 1211430.20 | 1329570.67 | 1383838.77 | 1400250.72 | 1405523.05 | 1407551.78 |
| 1988 | 1139040.00 | 1256898.70 | 1391162.31 | 1522028.28 | 1644433.99 | 1693414.66 | 1710403.56 | 1717215.23 | 1720457.52 |
| 1989 | 960560.00 | 1086171.80 | 1192736.14 | 1312449.09 | 1421212.69 | 1468297.33 | 1484853.97 | 1490061.93 | 1492759.25 |
| 1990 | 1165380.00 | 1406064.86 | 1662925.24 | 1826613.30 | 1969219.18 | 2028715.64 | 2050679.30 | 2058302.86 | 2061135.52 |
| 1991 | 1109300.00 | 1283380.73 | 1578199.80 | 1761901.04 | 1902498.20 | 1958843.20 | 1981475.48 | 1989548.43 | 1992082.11 |
| 1992 | 835410.00 | 960044.28 | 1160935.19 | 1274931.86 | 1374144.09 | 1428306.41 | 1446211.15 | 1452295.59 | 1454542.40 |
| 1993 | 1120480.00 | 1240667.94 | 1471414.75 | 1650856.86 | 1767058.64 | 1819130.60 | 1841628.82 | 1850217.91 | 1853370.63 |
| 1994 | 915090.00 | 1102338.88 | 1318589.70 | 1445772.59 | 1568207.90 | 1612075.71 | 1628629.19 | 1635707.09 | 1638141.03 |
| 1995 | 1094660.00 | 1250237.83 | 1545530.41 | 1743515.86 | 1885579.52 | 1950462.76 | 1975501.65 | 1983240.20 | 1986441.95 |
| 1996 | 960820.00 | 1155692.00 | 1342420.50 | 1464721.56 | 1586207.33 | 1637929.11 | 1655989.04 | 1661480.40 | 1663834.01 |
| 1997 | 843340.00 | 1033843.47 | 1186712.41 | 1334845.26 | 1437454.98 | 1476800.15 | 1492639.92 | 1497335.51 | 1499718.44 |
| 1998 | 1051830.00 | 1252125.94 | 1516098.43 | 1692944.49 | 1839699.62 | 1891316.85 | 1910925.62 | 1919169.01 | 1922088.04 |
| 1999 | 882700.00 | 1083010.24 | 1308642.34 | 1453791.41 | 1573903.55 | 1612901.61 | 1631682.95 | 1637863.24 | 1639839.14 |
| 2000 | 1205050.00 | 1554837.26 | 1886193.97 | 2084453.84 | 2263210.29 | 2319671.80 | 2347924.28 | 2357726.77 | 2361147.35 |
| 2001 | 1028100.00 | 1271346.28 | 1593285.40 | 1743314.79 | 1873671.25 | 1943170.28 | 1964132.73 | 1972140.48 | 1975585.87 |
| 2002 | 928580.00 | 1209110.73 | 1429866.35 | 1553082.62 | 1668930.39 | 1720203.80 | 1740329.93 | 1748440.96 | 1751187.12 |
| 2003 | 976410.00 | 1227629.91 | 1431077.59 | 1588186.54 | 1719959.81 | 1769461.86 | 1792701.46 | 1800535.88 | 1803260.07 |
| 2004 | 936950.00 | 1185024.65 | 1478259.02 | 1655940.78 | 1796342.80 | 1855430.80 | 1876727.22 | 1884780.03 | 1888487.17 |
| 2005 | 1127190.00 | 1258673.38 | 1447688.46 | 1597695.52 | 1724389.31 | 1780276.60 | 1801263.89 | 1807289.83 | 1809813.11 |
| 2006 | 998920.00 | 1160564.96 | 1395708.51 | 1547649.55 | 1653807.69 | 1701516.79 | 1719420.34 | 1728145.76 | 1730962.92 |
| 2007 | 1175940.00 | 1374752.60 | 1635169.28 | 1801304.21 | 1949935.63 | 2003479.78 | 2026148.07 | 2035470.20 | 2038551.02 |
| 2008 | 986690.00 | 1145445.26 | 1393596.05 | 1536411.44 | 1646373.87 | 1694197.43 | 1712292.96 | 1720099.66 | 1722943.93 |
| 2009 | 1064300.00 | 1209147.91 | 1446530.90 | 1559899.20 | 1688487.62 | 1745843.53 | 1763447.28 | 1769937.36 | 1772362.96 |
| 2010 | 919720.00 | 1031896.16 | 1266109.50 | 1416475.68 | 1539040.06 | 1579342.22 | 1597309.80 | 1603544.67 | 1606180.43 |
| 2011 | 1096340.00 | 1192774.08 | 1383032.40 | 1513973.26 | 1637073.24 | 1689078.76 | 1707764.97 | 1714876.18 | 1717545.07 |
| 2012 | 925700.00 | 1080595.60 | 1358683.54 | 1492790.66 | 1594496.69 | 1644539.22 | 1664256.08 | 1671556.41 | |
| 2013 | 711370.00 | 824706.80 | 987603.07 | 1084557.65 | 1172176.29 | 1204969.43 | 1218199.95 | | |
| 2014 | 957590.00 | 1131968.62 | 1369513.10 | 1532823.62 | 1651109.90 | 1704578.70 | | | |
| 2015 | 923550.00 | 1097112.27 | 1344447.89 | 1496029.87 | 1608155.38 | | | | |
| 2016 | 1128330.00 | 1278934.98 | 1562296.37 | 1730802.97 | | | | | |
| 2017 | 997050.00 | 1143823.78 | 1398980.17 | | | | | | |
| 2018 | 821720.00 | 960302.33 | | | | | | | |
| 2019 | 1047250.00 | | | | | | | | |

Table 7: Example

# Appendix B

# Addition to simulation study

In this appendix we present the parameters we used to generate the scenarios (S1) - (S7) and additionally the results for another scenario.

## B.1 Set-up and output

The premiums are randomly randomly generated by

```
premium=round(1000000*rnorm(n,10,1))
```

and are given by

```
> premium
 [1] 12820951  9226667  9911129 10362665  9622959 10153757  9334240 10993251 12723273  9913918 10650710
[12] 10464656 10865958 10538631 10006405 10500871 11018792  9569703 11120863  8925830  9798526 10205751
[23]  9706809  9815132  9125897 10540689 10117661 11112323  9640390 10767499
```

The mean values and sigmas are generated by

```
mu=exp(-0.8-0.4*c(1:n))
mu2=exp(-0.1-0.4*c(1:n)))
sigma=0.3*(n*mu)/sum(1/sqrt(premium))

#expected values per scenario
EW=array(NA, dim=c(n,n,7))
EW[,,1]=repmat(mu,n,1)
EW[,,2]=sapply(1:n, function(k){mu[k]+c(1:n)*0.02*mu[k]})
EW[,,3]=rbind(repmat(mu,(n/2),1), repmat(mu2,(n/2),1))
EW[,,4]=rbind(repmat(mu,(n/2),1), repmat(mu2,(n/2),1))
EW[,,5]=rbind(repmat(mu,10,1),sapply(1:n, function(k){mu2[k]-c(1:10)*0.07*mu2[k]}), repmat((mu2-10*0.07
    *mu2),10,1))
EW[,,6]=sapply(1:n, function(k){c(rep(mu[k], max(15-k+1,0)), rep(mu2[k], min((n-(15-k+1)),n))) })
EW[,,7]=sapply(1:n, function(k){c(rep(mu[k], max(20-k+1,0)), mu2[k]+c(1:min((n-20+k-1),n))*0.02*mu2[k])
    })
se=repmat(sigma,7,1)
se[4,]=sqrt(2)*se[4,]
```

We used the following code to generate data sets, test for change points and predict the reserve:

```
L=100
W=matrix(premium,n,n)
SimulationResults=replicate(7,list(M=array(NA, dim=c(n,n,L)), cp=matrix(0,5,L), R.true=rep(NA,L), R.hat
    =rep(NA,L), R0.hat=rep(NA,L)))
DYs=10

for(s in 1:7){
M=array(NA, dim=c(n,n,L))
for(i in 1:n){
  for(k in 1:n){
    M[i,k,]=   rnorm(L,EW[i,k,s],se[s,k]/sqrt(premium[i]))
  }}
R.hat<-R0.hat<-R.true<-rep(NA,L)
cp<-matrix(0,6,L)
for(j in 1:L){
  Y=M[,,j]
 if(s<=5){   Model=CP(Y,W,DYs,n)
```

```
 } else{ Model=CP.CY(Y,W,DYs,n)}
  m=Model$mu
  cpts=Model$cpts
  if(any(!is.na(cpts))){
    for(l in 1: length(cpts))
    cp[l,j]=cpts[l]
  }
  mu0.hat<-apply(Y*W*upT,2,sum)/apply(W*upT,2,sum)
  mu.hat=cbind(m, repmat(mu0.hat[(DYs+1):n],n,1))
  lossesthat=lowT*mu.hat*W
  losses0hat=lowT*mu0.hat*W
  R.true[j]=sum(lowT*Y*W)
  R.hat[j]=sum(lossesthat)
  R0.hat[j]=sum(losses0hat)
}
SimulationResults[,s]=list(M=M, cp=cp, R.true=R.true, R.hat=R.hat, R0.hat=R0.hat)
}
```

The results are analyzed by the following commands:

```
cpts.detected=sapply(1:5, function(s){length(which(SimulationResults[2,s]$cp[1:5,]>0))})
true=c(0,0,sapply(3:4, function(s){length(which(SimulationResults[2,s]$cp[,]==15))}),length(which(
    SimulationResults[2,5]$cp[,]==10)),length(which(SimulationResults[2,5]$cp[,]==20)),length(which(
    SimulationResults[2,6]$cp[,]==15)),length(which(SimulationResults[2,7]$cp[,]==20)))
detected=rep(NA,100)
for(l in 1:100){
  if(!any(SimulationResults[2,5]$cp[1:5,l]==20))
    detected[l]=SimulationResults[2,5]$cp[min(which((abs(SimulationResults[2,5]$cp[1:5,l]-20)==min(abs(
        SimulationResults[2,5]$cp[1:5,l]-20))&(SimulationResults[2,5]$cp[1:5,l]!=10))))),l]
}
nearly=sum(!is.na(detected)&(detected<=25)&(detected>=15))
min(detected,na.rm=TRUE)
max(detected,na.rm=TRUE)

wrongly.detected=cpts.detected-c(true[1:4],sum(true[5:6])+nearly,true[7:8])
mean.absolute.error=sapply(1:7, function(s){mean(abs(SimulationResults[3,s]$R.true-SimulationResults[4,
    s]$R.hat))})
mean.squared.absolute.error=sapply(1:7, function(s){mean((SimulationResults[3,s]$R.true-
    SimulationResults[4,s]$R.hat)^2)})
mean.relative.error=sapply(1:7, function(s){mean(abs(SimulationResults[3,s]$R.true-SimulationResults[4,
    s]$R.hat)/SimulationResults[3,s]$R.true)})

mean.absolute.error0=sapply(1:7, function(s){mean(abs(SimulationResults[3,s]$R.true-SimulationResults
    [5,s]$R0.hat))})
mean.squared.absolute.error0=sapply(1:7, function(s){mean((SimulationResults[3,s]$R.true-
    SimulationResults[5,s]$R0.hat)^2)})
mean.relative.error0=sapply(17, function(s){mean(abs(SimulationResults[3,s]$R.true-SimulationResults[5,
    s]$R0.hat)/SimulationResults[3,s]$R.true)})

average.true.reserve=sapply(1:7, function(s){mean(SimulationResults[3,s]$R.true)})

Overview=t(data.frame(mean.absolute.error,mean.squared.absolute.error,mean.relative.error, mean.
    absolute.error0, mean.squared.absolute.error0, mean.relative.error0))
colnames(Overview)=c("S1","S2","S3","S4","S5", "S6", "S7")
```

This yields the following results:

```
> average.true.reserve/1000000
[1] 19.21734 30.09068 38.70915 38.53945 19.44387 38.66612 48.61927
> mean.absolute.error/1000000
[1] 0.851246 1.202557 1.951265 1.925225 1.523650 1.312879 4.732396
> mean.absolute.error0/1000000
[1]  0.8018979  6.5400562 12.7713190 12.5801290  2.9305673  8.2476909 20.4144506
> mean.squared.absolute.error/10000000000
[1]  108.5097  238.8031  525.5819  667.1223  354.5872  286.8041 2571.2717
> mean.squared.absolute.error0/10000000000
[1]   101.9750  4392.2211 16466.7350 16073.3327   977.9889  6946.2440 41829.5707
> mean.relative.error*100
[1] 4.366898 3.943553 4.982589 4.976084 7.824046 3.398808 9.702227
> mean.relative.error0*100
[1]  4.094885 21.640087 32.929542 32.538903 15.396266 21.259453 41.954862
```

Or summarized in a Table:

| | S1 | S2 | S3 | S4 | S5 | S6 | S7 |
|---|---|---|---|---|---|---|---|
| mean.absolute.error | 8.512460e+05 | 1.202557e+06 | 1.951265e+06 | 1.925225e+06 | 1.523650e+06 | 1.312879e+06 | 4.732396e+06 |
| mean.squared.absolute.error | 1.085097e+12 | 2.388031e+12 | 5.255819e+12 | 6.671223e+12 | 3.545872e+12 | 2.868041e+12 | 2.571272e+13 |
| mean.relative.error | 4.366898e-02 | 3.943553e-02 | 4.982589e-02 | 4.976084e-02 | 7.824046e-02 | 3.398808e-02 | 9.702227e-02 |
| mean.absolute.error0 | 8.018979e+05 | 6.540056e+06 | 1.277132e+07 | 1.258013e+07 | 2.930567e+06 | 8.247691e+06 | 2.041445e+07 |
| mean.squared.absolute.error0 | 1.019750e+12 | 4.392221e+13 | 1.646673e+14 | 1.607333e+14 | 9.779889e+12 | 6.946244e+13 | 4.182957e+14 |
| mean.relative.error0 | 4.094885e-02 | 2.164009e-01 | 3.292954e-01 | 3.253890e-01 | 1.539627e-01 | 2.125945e-01 | 4.195486e-01 |

## B.2 Effect of misspecification of distribution

This scenario is generated by using the expected values and variances of scenario (S3), but by using a log-normal distribution.

```
M=array(NA, dim=c(n,n,L))
for(i in 1:n){
  for(k in 1:n){
    M[i,k,]=   rlnorm(L,log(EW[i,k,3]^2/sqrt(EW[i,k,3]^2+se[3,k]^2/premium[i])),sqrt(log(se[3,k]^2/(
        premium[i]*EW[i,k,3]^2)+1)))
  }}
```

The results are produced analogously to above.

```
R.hat<-R0.hat<-R.true<-rep(NA,L)
cp<-matrix(0,6,L)
DYs=10
for(j in 1:L){
  Y=M[,,j]
  Model=CP(Y,W,DYs,n)
  m=Model$mu
  cpts=Model$cpts
  if(any(!is.na(cpts))){
    for(l in 1: length(cpts))
      cp[l,j]=cpts[l]
  }
  mu0.hat<-apply(Y*W*upT,2,sum)/apply(W*upT,2,sum)
  mu.hat=cbind(m, repmat(mu0.hat[(DYs+1):n],n,1))
  lossesthat=lowT*mu.hat*W
  losses0hat=lowT*mu0.hat*W
  R.true[j]=sum(lowT*Y*W)
  R.hat[j]=sum(lossesthat)
  R0.hat[j]=sum(losses0hat)
}

S8=list(M=M, cp=cp, R.true=R.true, R.hat=R.hat, R0.hat=R0.hat)
```

This yields the following results

```
> length(which(S8$cp[,]==15))
[1] 100
> length(which(S8$cp[,]>0))-100
[1] 25
> (mean.absolute.error=mean(abs(S8$R.true-S8$R.hat)))
[1] 1818823
> (mean.absolute.error0=mean(abs(S8$R.true-S8$R0.hat)))
[1] 11304868
> (mean.squared.absolute.error=mean((S8$R.true-S8$R.hat)^2))
[1] 4.597557e+12
> (mean.squared.absolute.error0=mean((S8$R.true-S8$R0.hat)^2))
[1] 1.289314e+14
> (mean.relative.error=mean(abs(S8$R.true-S8$R.hat)/S8$R.true))
[1] 0.04743045
> (mean.relative.error0=mean(abs(S8$R.true-S8$R0.hat)/S8$R.true))
[1] 0.297722
```

This means we detected all true change points and the predicted reserve by the new method is again very close to the true reserve, while the original method shows a large deviation.

# References


Akaike H (1973) Information Theory and an Extension of the Maximum Likelihood Principle, Springer New York, New York, NY, pp 199–213

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19(6):716–723, DOI 10.1109/TAC.1974.1100705, URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1100705

Aminikhanghahi S, Cook D (2016) A Survey of Methods for Time Series Change Point Detection. Knowledge and Information Systems 51, DOI 10.1007/s10115-016-0987-z

Borchers HW (2017) R package "pracma."

Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin) (2016) Solvency II URL https://www.bafin.de/EN/Aufsicht/VersichererPensionsfonds/Aufsichtsregime/SolvencyII/solvency_II_artikel_en.html

CEIOPS (2007) QIS3 Technical Specifications PART I: INSTRUCTIONS. https://eiopa.europa.eu/Publications/QIS/QIS3TechnicalSpecificationsPart1.PDF, [Online; accessed 20-July-2019]

Chen J, Gupta A (2014) Parametric statistical change point analysis: With applications to genetics, medicine, and finance. Birkhauser Boston, DOI 10.1007/978-0-8176-4801-5

Cipra T (2010) Financial and Insurance Formulas. Springer Science & Business Media, DOI 10.1007/978-3-7908-2593-0

Dimitrijevic D (2015) The detection and prevention of manipulations in the balance sheet and the cash flow statement. Ekonomski horizonti 17:137–153, DOI 10.5937/ekonhor1502137d

Ding J, Tarokh V, Yang Y (2018) Model Selection Techniques – An Overview. IEEE Signal Processing Magazine 35, DOI 10.1109/MSP.2018.2867638

E Schwarz G (1978) Estimating the Dimension of a Model. The Annals of Statistics 6, DOI 10.1214/aos/1176344136

Eckley IA, Fearnhead P, Killick R (2011) Analysis of changepoint models. Bayesian Time Series Models pp 205–224

England P, Verrall R (2002) Stochastic Claims Reserving in General Insurance. British Actuarial Journal 8(3):443–518, DOI 10.1017/S1357321700003809

European Commission (2009) Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II) URL https://www.eur-lex.europa.eu

Fabozzi F, Focardi S, Rachev T, Arshanapalli B (2014) Appendix E: Model Selection Criterion: AIC and BIC, Wiley, pp 399–403. DOI 10.1002/9781118856406.app5

Fahrmeir L, Kneib T, Lang S (2007) Regression: Modelle, Methoden und Anwendungen. Springer, DOI 10.1007/978-3-540-33933-5

Forte S, Ialenti M, Pirra M (2012) Claims reserving uncertainty in the development of internal risk models. Mathematical and Statistical Methods for Actuarial Sciences and Finance pp 203–210, DOI 10.1007/978-88-470-2342-0_24

Gupta A, Chen J (1996) Detecting Changes of Mean in Multidimensional Normal Sequences with Applications to Literature and Geology. Computational Statistics 11

Haynes K, A Eckley I, Fearnhead P (2014) Efficient penalty search for multiple changepoint problems. arXiv preprint arXiv:14123617

Hinkley D (1970) Inference About the Change-Point in a Sequence of Random Variables. Biometrika 57, DOI 10.1093/biomet/57.1.1

Hinkley D (1971) Inference in Two-Phase Regression. Journal of The American Statistical Association - J AMER STATIST ASSN 66:736–743, DOI 10.1080/01621459.1971.10482337

Institute of Actuaries (1997) Claims Reserving Manual. URL https://www.actuaries.org.uk/system/files/field/document/Claims%20Reserving%20Manual%20V1%20complete.pdf

Jackson B, Scargle J, Barnes D, Arabhi S, Alt A, Gioumousis P, Gwin E, Sangtrakulcharoen P, Tan L, Tsai T (2005) An algorithm for optimal partitioning of data on an interval. Signal Processing Letters, IEEE 12:105 – 108, DOI 10.1109/LSP.2001.838216

Kass RE, Raftery AE (1995) Bayes Factors. Journal of the American Statistical Association 90(430):773–795, DOI 10.1080/01621459.1995.10476572, URL https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572

Killick R, Eckley IA, Ewans K, Jonathan P (2010) Detection of changes in variance of oceanographic time-series using changepoint analysis. Ocean Engineering 37(13):1120–1126

Killick R, Fearnhead P, Eckley I (2012) Optimal Detection of Changepoints With a Linear Computational Cost. Journal of the American Statistical Association 107:1590–1598, DOI 10.1080/01621459.2012.737745

Klenke A (2013) Wahrscheinlichkeitstheorie. Springer Berlin Heidelberg

Kraus C, Diers D (2010) Das stochastische Re-Reserving. Zeitschrift für die gesamte Versicherungswissenschaft 99, DOI 10.1007/s12297-009-0071-9

Mack T (1993) Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates. ASTIN Bulletin 23(2):213–225, DOI 10.2143/AST.23.2.2005092

Mack T (2002) Schadenversicherungsmathematik. Schriftenreihe angewandte Versicherungsmathematik, Verlag Versicherungswirtschaft, URL https://books.google.de/books?id=Z6bJygAACAAJ

McGee VE, Carleton WT (1970) Piecewise Regression. Journal of the American Statistical Association 65(331):1109–1124, URL http://www.jstor.org/stable/2284278

Merz M, Wüthrich M (2008) Modelling the Claims Development Result for Solvency Purposes. Casualty Actuarial Society E-Forum

Muggeo V (2003) Estimating Regression Models with Unknown Break-Points. Statistics in medicine 22:3055–71, DOI 10.1002/sim.1545

Ohlsson E, Lauzeningks J (2009) The one-year non-life insurance risk. Insurance: Mathematics and Economics 45:203–208, DOI 10.1016/j.insmatheco.2009.06.001

Page E (1954) Continuous Inspection Schemes. Biometrika 41:100, DOI 10.2307/2333009

Quarg G, Mack T (2004) Munich chain ladder. Blätter der DGVFM 26(4):597–630, DOI 10.1007/BF02808969, URL https://doi.org/10.1007/BF02808969

R Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org

Rao C, Toutenbourg H (1999) Linear models. Least squares and alternatives. 2nd ed. DOI 10.1007/b98889

Rohrbeck C (2013) Detection of changes in variance using binary segmentation and optimal partitioning

Schmidt KD (2006) Methods and models of loss reserving based on run-off triangles: a unifying survey. Techn. Univ., Inst. für Mathematische Stochastik

Scott A, Knott M (1974) A Cluster Analysis Method for Grouping Means in the Analysis of Variance. Biometrics 30, DOI 10.2307/2529204

Truong C, Oudre L, Vayatis N (2018) A review of change point detection methods. ArXiv abs/1801.00718

Wüthrich M, Merz M (2008) Stochastic Claims Reserving Methods in Insurance. Wiley

Wüthrich M, Merz M, Lysenko N (2009) Uncertainty of the claims development result in the chain ladder method. Scandinavian Actuarial Journal 2009:63–84, DOI 10.1080/03461230801979732

Yao YC (1988) Estimating the Number of Change-Points Via Schwarz' Criterion. Statistics & Probability Letters 6:181–189, DOI 10.1016/0167-7152(88)90118-6