UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MODÉLISATION DES RÉSERVES EN ASSURANCE I.A.R.D. – UTILISATION DE MÉTHODES ROBUSTES

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

ZAHRA GHASEMIVANANI

JANVIER 2018

ACKNOWLEDGEMENTS

Writing this note of thanks is the finishing touch on my dissertation. It has been a period of learning for me, not only in the scientific area, but also on a personal level. Writing this dissertation has had a big impact on me. I would like to reflect on the people who have supported and helped me so much throughout this period.

First, I would like to thank my supervisor Professor Mathieu Pigeon, I want to thank you for your excellent cooperation and for all of the opportunities I was given to conduct my research and further my thesis. In addition, I would like to thank all my tutors, for their valuable guidance. You definitely provided me with the tools that I needed to choose the right direction and successfully complete my thesis. Moreover, I want to thank Mrs Gisèle Legault who helps me to solve the software and hardware issues I have had.

I would also like to thank my mother for her support and sympathetic ear. You are always there for me. Finally, there are my friends. We were not only able to support each other by deliberating over our problems and findings, but also happily by talking about things other than just our papers.

Thank you very much, everyone!

CONTENTS

LIST OF TABLES

LIST OF FIGURES

RÉSUMÉ

Une police d'assurance reçoit une prime initiale et elle paiera les réclamations futures probables. En conséquence, l'assureur ne connaît pas tous les coûts futurs, ainsi que le calendrier des paiements. Par conséquent, l'un des plus gros passifs de l'assureur est mettre un capital de côté afin de couvrir tous ces paiements futurs. Parmi toutes les méthodes de réservation existantes en assurance générale, nous considérons deux modèles dans ce travail: le modèle stochastique de la Chain-Ladder (ou le modèle de Mack Chain-Ladder) et le modèle linéaire généralisé (ou GLM) pour les réserves. La première est la technique la plus utilisée pour calculer la réserve actuarielle et l'autre s'applique à un large groupe de sinistres qui appartenant à la famille exponentielle. L'existence d'un outlier (ou des valeurs aberrantes) dans l'ensemble de données créera une réserve sous-estimée ou surestimée pour les deux méthodes. Cette erreur de prédiction peut être grande et il est essentiel de modifier ces méthodes de manière à ce qu'elles deviennent moins sensibles et plus précises, même en présence de valeurs aberrantes. Dans cette étude, nous présentons le cadre général pour des statistiques robustes, ainsi que nous expliquons les approches basiques dans la réserve de perte telles que le modèle stochastique de la chaîne et le GLM pour les réserves. Ensuite, nous étudions des versions robustes de ces modèles avant de passer par un exemple basé sur un ensemble de données réelles.

Mots-clés: Statistiques robustes, provisionnement des sinistres, assurance générale, modèle Mack Chain-Ladder, modèle linéaire généralisé (GLM).

ABSTRACT

Insurance companies receive an upfront premium and will pay future probable claims. As a result, the insurer knows neither the future costs nor the payment schedule. One of the largest liabilities of the insurer is therefore to put capital aside in order to cover all these future payments.

Among all existing reserving methods in general insurance, we consider two models in this work: the stochastic Chain-Ladder model (or Mack Chain-Ladder model) and the generalized linear model (or GLM) for reserves. The former is the most widely used technique to calculate the actuarial reserve and the latter is applicable for a wide range of claims which belong to the exponential family. The existence of an outliers in the dataset will result in an underestimated or overestimated reserve for both methods. This prediction error may be large and it is essential to modify these methods in a way that they become less sensitive and more accurate, even in presence of outliers.

In this study, we introduce the general framework for robust statistics, as well as explain basic approaches in loss reserving such as the stochastic Chain-Ladder model and the GLM for reserves. We then study robust versions of these models before going through an example based on a real data set.

Keywords: Robust statistics, loss reserving, general insurance, Mack Chain-Ladder model, Generalized linear model(GLM).

INTRODUCTION

The evaluation of future profit or loss becomes more and more important every day. The market is so competitive and only insurance companies with an accurate plan can survive. At each moment, an insurance company needs to put aside a sufficient capital amount in order to be able to pay all future liabilities generated by the contracts that have been sold to the clients. This capital forms the *reserve* (or provision) of the non-life company. Periodically, insurance regulators require an evaluation on this reserve in order to control the financial solvency of the company and to protect policyholders.

For a very long time, all calculations for evaluating loss reserve in insurance companies were done using simple deterministic algorithms. But since the early 1990s, actuaries have started to propose more sophisticated stochastic methods to manage the solvency of the company.

Two of the most important methods for estimating the outstanding reserves are the stochastic Chain-Ladder model (or Mack's model) and the generalized linear model (or GLM) for reserves. The ultimate goal of these methods is to accurately predict the amount of reserve but, as we will see, both methods are highly dependent on the data. In fact, the stochastic Chain-Ladder model uses the claims history in order to predict future developments and the GLM for reserves uses claims history in order to estimate unknown parameters. Both approaches can quickly become inaccurate in case of existence of outliers.

An outlier is a sample value that is significantly different when compared with the majority of the sample. It is not necessarily a wrong value, but it should always be checked for a transcription error. The existence of an outlier could have a large effect on the estimation procedure as well as on the total reserve amount. It could cause reserve underestimation or overestimation and problems in the solvency on the insurance company. This is a well-known issue in several areas and many robust statistical methods have been developed to be less sensitive to the outliers (see Wüthrich et Merz (2008)).

Since 1953, when Box first gave the word "robust" its statistical meaning, this field has evolved considerably, although the concept already existed for several decades (see Stigler (1972) for more details).

Several persons consider that the fundamental work in robust statistics was done in the 1960s, and the early 1970s by Tukey (1960, 1962), Huber (1964, 1967) and Frank Hampel (1971a, 1974). One of the possible reasons for this significant development was the accessibility of modern and fast computers. From the early 1980s, robust statistics have experienced considerable growth and many of the most influential books in this field, such as Huber (1981), Hampel *et al.* (1986a), Rousseeuw et Leroy (1986) and Staudte et Sheather (1990), were written during this period. Several researches have appeared in recent years on applying robust analysis to loss modeling and to reserve evaluation. For example, see Gather et Schultze (1999), Wüthrich et Merz (2008) , Cantoni et Ronchetti (1999) and Verdonck *et al.* (2009).

Based on Maronna *et al.* (2006) and Hampel *et al.* (1986b), we introduce "robust statistics" and all related concepts such as M-estimator in the first chapter. In the second chapter, we present two of the most important reserving methods in non-life insurance: the stochastic Chain-Ladder model (based on Wüthrich et Merz (2008)) the family of generalized linear models (GLM) for reserves. In this chapter, we also study how sensitive are these methods to outliers. As a solution of high dependency of reserve to outlier data, we introduce a "robustified" version of both models in Chapter 3. The robust stochastic Chain-Ladder model is based on Verdonck *et al.* (2009) and Verdonck et Debruyne (2011), while the robust GLM is based on Cantoni et Ronchetti (1999). Finally, we go through a real study case in the Chapter 4.

CHAPTER I

ROBUST STATISTICS

In this chapter we introduce robust statistics and we present some measures of robustness such as the empirical influence function, the influence function and the breakdown point.
Robust means model's, test's or system's ability to perform effectively and without failure while its assumptions or variables altered or violated. For statistics, a robust model can still provides insight to a problem despite having its assumptions altered. It is expected the robust statistics gives more reliable results than classic statistics.

1.1 Introduction and Motivation

Statistics play an important role in reducing and organizing information provided in a dataset. For instance, the sample mean summarizes the "central tendency" of a sample to a single value. Robust statistics go beyond that by offering good performance even if some standard assumptions such as normality, linearity or independence are not thoroughly verified. For example, robust methods will remain valid for samples drawn from a wide range of probability distributions and in particular from a non-normal distribution.

Outliers can occur in a purely random manner in any distribution, but they should always be checked for transcription errors, measurement errors, etc. They can ruin standard statistical methods because in the presence of outliers, most of the statistics often used in practice become useless and traditional estimators, such as maximum likelihood estimators or moment-based estimators, have poor performance (see Maronna *et al.* (2006)). Thankfully, many robust statistical methods have been developed to be less sensitive to the outliers.

Let $x = \{x_1, x_2, ..., x_n\}$ be a set of observed values. The sample mean $\bar{x}$ and the sample standard deviation s are defined by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

The sample mean is just the arithmetic average of the data and, as such one might expect that, it provides a good estimate of the "center" or "location" of the data. Likewise, one might expect that the sample standard deviation would provide a good estimate of the "dispersion" of the data. Now we shall investigate how much influence a single outlier can have on these estimates often used in practice.

Example 1.1. *We consider a portfolio with* 24 *payments in thousands of dollars, in ascending order:*

2.20	2.20	2.40	2.40	2.50	2.70	2.80	2.90
3.03	3.03	3.10	3.37	3.40	3.40	3.40	3.50
3.60	3.70	3.70	3.70	3.70	3.77	5.28	28.95.

The maximum sample value, 28.95, *is meaningfully larger than the remaining* 23 *observations in the sample and the difference with its previous value* (5.28) *is very large. This makes us think that it could be an outlier probably caused by a* *misplaced decimal point. The original value could have been* 2.895.

We calculate the sample mean and the standard deviation which are $\bar{x} = 4.28$ *and* $s = 5.3$. *As we can see in the sample, there are only two observations above the mean which is an evidence for us to realize that the mean is not between the bulk of the data and it is not a good estimation of the central tendency of the data. After deleting the suspected outlier and re-calculating the sample mean and the standard deviation, we obtain* $\bar{x} = 3.21$ *and* $s = 0.69$. *The new mean is between the eleventh and the twelfth number, which is a more reasonable estimate of the center of the data. Also, the standard deviation is over seven times smaller than it was in the presence of an outlier.*

To complete this introductory example, we are interested in the effect of any outlier on these two statistics. We assume that the value of the outlier is replaced by a random variable X *which can take any values between* $-\infty$ *and* $+\infty$. *Consequently, the resulting sample mean will take values between* $-\infty$ *to* $+\infty$. *Thus, we can say that a single outlier has an unbounded effect on the sample mean. Similarly, the standard deviation will take a positive value up to* $+\infty$ *and we can conclude that a single outlier has an unbounded effect on the standard deviation.*

The above example shows that deleting outliers is a simple way to handle the problem but it may cause some problems:

- deletion is a subjective decision because the analyst has to decide when an observation is outlier enough to be removed from the sample;
- by removing some "extreme" points in the sample, there is a risk of underestimating (or overestimating) location and/or dispersion; and

- eliminating some points from the sample may affect some properties of statistics (unbiasness, etc.).

As shown in the previous example, the sample mean and the standard deviation are not the best statistics, since they are highly influenced by outlirs. Fortunately, there are other options for estimating the central tendency and the dispersion of the data, for instance the *median* and the *median absolute deviation.* We rank the elements of a random sample $(x_1, x_2, \ldots, x_n)$ in ascending order

$$\mathbf{x} = x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)},$$

and the sample median is defined by

$$\text{median}(\mathbf{x}) = \begin{cases} x_{(m)}, & \text{if } n = 2m-1 \\ \dfrac{x_{(m)} + x_{(m+1)}}{2}, & \text{if } n = 2m. \end{cases}$$

In the introductory example, the sample median for the original sample is 3.37, while the sample median without the outlier is 3.37. Moreover, when the largest value in the sample (28.95) is replaced by a random value in $(-\infty, +\infty)$, the sample median does not change. As the median is not influenced by the presence of an outlier, we conclude it is a robust alternative to the sample mean. We say that the median is resistant to outliers whereas the mean is not. In fact, the median can tolerate a proportion of outliers up to 50 % before it becomes arbitrarily large and useless. Formally, we could say that its *breakdown point* is 50 % whereas the breakdown point for the mean is 0 %.

A robust alternative to the standard deviation is the median absolute deviation about the median (MAD) which is defined by

$$\begin{aligned}\text{MAD}(\mathbf{x}) &= \text{MAD}(x_1, x_2, \ldots, x_n) \\ &= \text{Median}\{|\mathbf{x} - \text{Median}(\mathbf{x})|\}.\end{aligned}$$

In order to compare the median absolute deviation about the median to standard deviation, we define a "normalized" version of the MAD (MADN) as

$$\text{MADN}(\mathbf{x}) = \frac{\text{MAD}(\mathbf{x})}{0.6745},$$

where 0.6745 is the median absolute deviation about the median of a standard normal random variable. Indeed, we have

$$\Pr[|Z| \leq y] = \frac{1}{2},$$

where y is the unknown value and $Z \sim \text{Normal}(0, 1)$. Therefore we have

$$\Phi(y) - \Phi(-y) = \frac{1}{2}.$$

Since $\Phi(-y) = 1 - \Phi(y)$, we have

$$y = \phi^{-1}\left(\frac{3}{4}\right) = 0.6745.$$

In the Example 1.1, the normalized version of the MAD is 0.53 for the complete sample and it decreases to 0.5 after deleting the outlier (28.95). It pointed out that the (normalized version of the) median absolute deviation about the median is not influenced very much by the presence of an outlier and provides a robust alternative to the standard deviation. For the Example 1.1, we present the full results in Table 1.1.

1.2 Measuring Robustness

The basic tools to describe and measure robustness are the empirical influence function, the influence function and the breakdown point which are defined in

Table 1.1 Statistics in presence and absence of an outlier for Example 1.1

statistics	$\bar{x}$	s	Median (**x**)	MADN (**x**)
in presence of outlier	4.28	5.30	3.37	0.53
in absence of outlier	3.20	0.68	3.37	0.50

subsections 1.2.1, 1.2.2 and 1.2.3. These approaches are part of the field called *quantitative robustness* in the literature (see Hampel *et al.* (1986a) and Maronna *et al.* (2006)) because they really allow to measure the robustness of a statistic. Conversely, we introduce in subsection 1.2.4 a different way to describe the robustness of a statistic called *qualitative robustness*. In the following chapters of this document, we mainly focus on the empirical influence function and the influence function in order to measure the robustness of our estimators. Finally, mathematical definitions are gathered in Appendix A.

1.2.1 Empirical Influence Function

The empirical influence function is a measure of how a statistic is related to (or depends) a single point of the sample. This measure does not depend on the model since it simply relies on re-calculating the estimator with a different sample.

In order to formalize the concept, we need[1] a probability space $\{\Omega, \mathcal{A}, \mathbb{P}\}$ and two measure spaces $\{\chi, \Sigma\}$ and $\{\Gamma, S\}$. We consider a vector of independent and identically distributed (iid) random variables $X_1, \dots, X_n$, $n \in \mathbb{N}$, where

$$X_i : \{\Omega, \mathcal{A}\} \to \{\chi, \Sigma\}.$$

A sample from these random variables is denoted by $\mathbf{x} = \{x_1, \dots, x_n\}$. Finally,

[1] See Appendix A for the definitions of Ω, $\mathcal{A}$, etc.

we define a statistic

$$T_n : \{\chi^n, \Sigma^n\} \to \{\Gamma, S\}.$$

For the i^{th} observation in the sample, the empirical influence function (EIF) of the statistic T_n is given by

$$\begin{aligned} \text{EIF}_i : \chi &\to \mathbb{R} \\ x &\mapsto n\left(T_n(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n) - T_n(\mathbf{x})\right). \end{aligned} \tag{1.2.1}$$

The empirical influence function measures the amount of change in the statistic if we replace the i^{th} observation by an arbitrary value x.

Example 1.2. *We want to study the effect of a change of one of the observations in the sample of the Example 1.1 on both the sample mean and the sample median by using the empirical influence function (note that* 28.95 *is present in the sample).*

As a first step, we have randomly selected the observation x_{19} and changed it from 3.7 *to* 6.78*. The new sample mean is $\bar{x} = 4.40$ and for the sample median, the new value is* 3.38*. The value of the empirical influence function for the sample sample mean is*

$$\begin{aligned} EIF_{19}(6.78) &= 24\left(T_{24}(x_1, \ldots, x_{18}, 6.78, x_{20}, \ldots, x_{24}) - T_{24}(\mathbf{x})\right) \\ &= 24(4.40 - 4.28) \\ &= 2.88 \end{aligned}$$

and for the sample median is

$$\begin{aligned} EIF_{19}(6.78) &= 24\left(T_{24}(x_1, \ldots, x_{18}, 6.78, x_{20}, \ldots, x_{24}) - T_{24}(\mathbf{x})\right) \\ &= 24(3.38 - 3.38) \\ &= 0. \end{aligned}$$

This last value shows that the sample median is not affected when we replace the observation x_{19} by a different value.

As a second step, we evaluate the empirical influence function for an arbitrary value x and we create the plot presented in the Figure 1.1. This graph shows that the EIF for the sample mean is unbounded, that is to say the impact of an outlier on this statistic can be huge.

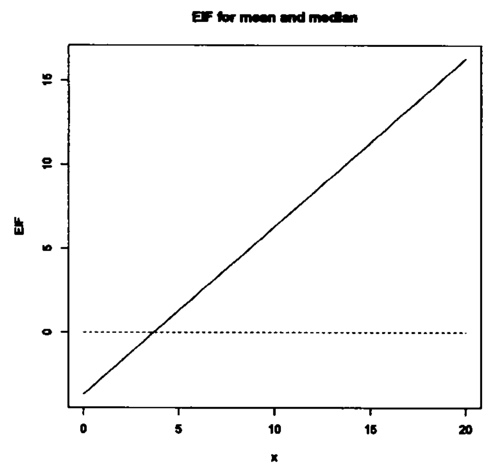

Figure 1.1 Empirical influence function for the sample mean (red line) and the sample median (blue line)

1.2.2 Influence Function

The influence function (IF) is a measure of the dependence between an estimator and the model's distribution. It does not just rely on the sample data; instead, it considers the effect of a slight change of the distribution on an estimator. In order to define the influence function in mathematical format we need to define some concepts.

Functionals

Let F_θ be the parametric cumulative distribution function (cdf) and f_θ be the corresponding probability density function (pdf). We need to find an estimator for the parameter θ based on a dataset. For a random sample $(X_1, \ldots, X_n)$, we define the empirical distribution function

$$G_n = G_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \leq x),$$

where $\mathbf{1}(A)$ is the indicator function of the event A. One should note that the empirical distribution function does not take in to account the order of the observations.

An estimator of θ is

$$\widehat{\theta} = T_n = T_n(X_1, \ldots, X_n) = T_n(G_n).$$

It means that (1) $\widehat{\theta}$ can be seen as the sequence of statistics $\{T_n; n \geq 1\}$, one for each sample size n, and (2) $\widehat{\theta}$ does not depend on the order of the observations. Estimators are *functionals* or can be replaced by functionals. It means it exists a functional $T : \Omega_T \to \mathbb{R}$, where Ω_T stands for the domain of T, such that

$$T_n(x_1, \ldots, x_n) \xrightarrow[n\to\infty]{} T(G) \tag{1.2.2}$$

in probability when the observation are independent and identically distributed according to the distribution G in Ω_T.

Moreover, we assume that the functional T is Fisher consistent, that is to say that at the true model F, the sequence $\{T_n; n \geq 1\}$ measures asymptotically the correct quantity.

$$T(F_\theta) = \theta \ \ \forall\, \theta \text{ in } \Theta, \tag{1.2.3}$$

where Θ is the parameter space of dimension $p \in \mathbb{N}$.

Gâteaux-derivative

The influence function is the limit of the Gâteaux derivative with respect to a smooth deviation, as the deviation approaches a point mass. The Gâteaux derivative limit provides a way to calculate the solution to the functional equation, where the influence function appears as the limit of the Gâteaux derivative. In this way the Gâteaux derivative limit provides a way to circumvent the need to try to guess the solution of the functional equation.

We say that the functional $T(G)$ is Gâteaux-differentiable at the distribution F in Ω_T if it exists a real function a such that

$$\lim_{t\to 0}\left(\frac{T((1-t)F+tG)-T(F)}{t}\right)=\int a(x)\,dG(x), \qquad \forall G\in\Omega_T. \tag{1.2.4}$$

The Gâteaux-derivative generalizes the directional derivative. In particular, if we choose $G=\Delta_x$, the empirical distribution which gives mass 1 to x, we obtain

$$\begin{aligned}\lim_{t\to 0}\left(\frac{T((1-t)F+t\Delta_x)-T(F)}{t}\right) &= \int a(y)\,d\Delta_x(y)\\ &= \mathrm{a}(x).\end{aligned} \tag{1.2.5}$$

Here $T(G)$ is the asymptotic value of the estimator sequence $\{T_n; n\geq 1\}$ as mentioned above.

Back to the influence function

Remind that we aim at evaluating what will happen when the data do not follow the model F exactly but they follow a different distribution G. To do that, we want to evaluate the one-sided directional derivative of T at F, in the direction of $G-F$, i.e.

$$\lim_{t\to 0}\left(\frac{T((1-t)F+tG)-T(F)}{t}\right). \tag{1.2.6}$$

This last expression is true for all distribution $G\in\Omega_T$. In particular, if we want to evaluate the contamination effect of a point x on the estimator, we may assume that $G = \Delta_x$. Then, the influence function is defined by

$$IF(x;T,F) = \lim_{t\to 0}\left(\frac{T((1-t)F + t\Delta_x) - T(F)}{t}\right). \tag{1.2.7}$$

If in Equation (1.2.5) we put $G = F$, we obtain

$$\lim_{t\to 0}\left(\frac{T((1-t)F + tF) - T(F)}{t}\right) = \int a(y)\, dF(y). \tag{1.2.8}$$

By putting together Equation (1.2.5) and Equation (1.2.8), we obtain

$$\int IF(y;T,F)\, dF(y) = 0.$$

In order to clarify the concept of influence function and the notation introduced before, we consider the influence function of the arithmetic mean. Let the sample space be $\chi = \mathbb{R}$ and the parameter space be $\theta = \mathbb{R}$. We assume that the model's distribution is the standard Normal with probability density function

$$\phi(x) = (2\pi)^{-(1/2)} \exp^{-(1/2)x^2}, \qquad x \in \mathbb{R},$$

the true value of the parameter is $\theta_0 = 0$ and thus $F_{\theta_0} = \Phi$. The arithmetic mean is

$$T_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the corresponding functional for all probability measures with existing first moment is

$$T(G) = \int u\, dG(u).$$

From Equations (1.2.3), we conclude that T is Fisher consistent

$$\begin{aligned} T(F_\theta) &= \int u\, dF_\Theta(u) \\ &= \int u(2\pi)^{-(1/2)} \exp^{-(1/2)u^2}\, du \\ &= 0. \end{aligned}$$

0 is the value of the parameter Θ.

By using Equation (1.2.8), it follows that

$$\begin{aligned} IF(x;T,F) &= \left(\lim_{t\to 0}\frac{\int u\,d[(1-t)\Phi + t\Delta_x](u) - \int u\,d\Phi(u)}{t}\right) \\ &= \lim_{t\to 0}\left(\frac{(1-t)\int u\,d\Phi(u) + t\int u\,d\Delta_x(u) - \int u\,d\Phi(u)}{t}\right) \\ &= \lim_{t\to 0}\left(\frac{0+tx-0}{t}\right) \end{aligned} \tag{1.2.9}$$

because $\int u\,d\Phi(u) = 0$. Therefore

$$IF(x;T,F) = x. \tag{1.2.10}$$

Moreover, we assume that the asymptotic normality assumption is verified (see Hampel *et al.* (1986a))

$$f_G\left(\frac{T_n - T(G)}{\sqrt{n}}\right) \xrightarrow[n\to\infty]{weakly} \mathrm{N}\left(0, V(T,G)\right), \tag{1.2.11}$$

where $V(T,G)$ is the asymptotic variance given by

$$V(T,F) = \int IF(x;T,F)^2\,dF(x).$$

we will define $IF(x;T,F)$ in equation (1.2.7).

Example 1.3. *Suppose that the theoretical distribution of the data in Example 1.1 is the standard Normal. We want to evaluate the influence function of* 2 *estimators for the location parameter* μ*: the empirical mean* $\bar{x}$ *and the median* $median(x)$.

Obviously the empirical distribution is different since mean and variance are not 0 *and* 1 *respectively. In order to calculate the influence function for the functional* $T_n(G) = \bar{x}$*, we use Equation* (1.2.10) *which is the arithmetic mean's influence function for the standard Normal distribution. For the median, we have*

$$T_n(G) = \arg\min\left(\sum_{i=1}^{24}|x_i - \mu|\right).$$

It is easy to show that $\hat{\mu} = median(x) = 3.38$ *in our example.*

In Subsection 1.3, we show that the influence function of the median in this model is proportional to $sign\,(x_i - \hat{\mu})$. *In this example,*

$$sign\,(x_i - median(x)) = sign\,(x_i - 3.38)\,.$$

The Figure 1.2 illustrates both influence functions for some values of x. *As it has been illustrated in the figure, the influence function for the mean seems to be unbounded and the influence function for median is bounded.*

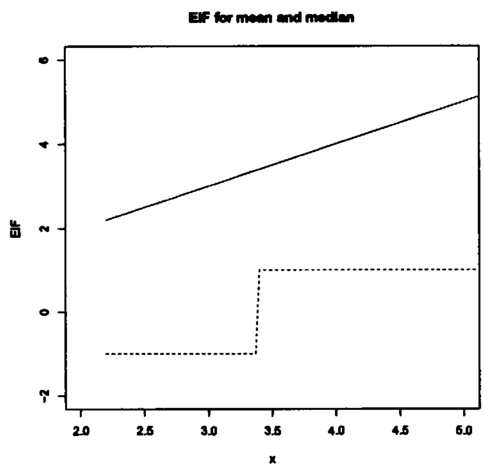

Figure 1.2 Influence Function for the arithmetic mean (red line) and the median (blue line)

1.2.3 Breakdown Point

The influence function, which describes the infinitesimal stability of an estimator, is a very useful tool for measuring the robustness but it has one major limitation: it is a local concept. Thus, it must be complemented by at least one more global measure to assess the reliability of an estimator. This measure must show up to what "distance" from the model distribution the estimator still gives some relevant information. The *breakdown point* of an estimator is the proportion of incorrect observations an estimator can handle before giving an incorrect result. The contamination should not be able to drive the estimator $\hat{\theta}$ to infinity, or to the boundary of the parameter space Θ when it is not empty, in order for the estimator $\hat{\theta}$ to give some information about θ.

The breakdown point ε^* of the sequence of estimators $\{T_n; n \geq 1\}$ at the distribution F is defined below. The evaluation of parameter is Θ_∞ as x goes to ∞ (x is a sample observation).

Definition 1.1. *The asymptotic contamination breakdown point of the estimator $\hat{\theta}$ at the distribution F, denoted by $\varepsilon^*(\hat{\theta}, F)$, is the largest $\varepsilon^* \in [0,1)$ such that for $\varepsilon^* < \varepsilon$, $\hat{\theta}_\infty((1-\varepsilon)F + \varepsilon G)$ as a function of G remains bounded, and also bounded away from the boundary of Θ.*

We interpret the above definition in the following way: there exists a bounded and closed set $K \subset \Theta$ such that $K \cap \mathrm{D}_\Theta = \emptyset$ (where D_Θ denotes the boundary of the parameter space Θ) and

$$\hat{\theta}_\infty((1-\varepsilon)F + \varepsilon G) \in K, \qquad \forall \varepsilon < \varepsilon^* \text{ and } \forall G. \tag{1.2.12}$$

As we illustrate in Example 1.1, changing (or removing) one observation in the sample has a large impact on the estimated value of the mean. If we replace the larger value in the sample by an arbitrary large value, we obtain an arbitrary large value for the empirical mean. Therefore, we conclude that the breakdown point for the sample mean is 0. Conversely, the sample median showed some resistance to the modification of an observation. We conclude its breakdown point is larger that 0. Actually, it is possible to show that the breakdown point of the sample median in this example is 50 %, which means it will be resistant up to 50 % of contamination in the sample (see Maronna *et al.* (2006)).

The breakdown point is the smallest fraction of gross errors which can carry the statistic over all bounds. Somewhat more precisely, it is the distance from the model distribution beyond which the statistic becomes totally unreliable and uninformative. Finally, it measures directly the global reliability of a statistic.

1.2.4 Qualitative Robustness

We turn next to the definition of a global measure of the robustness of a statistic. This is, in some sense, a *qualitative* measure of the robustness of an estimator. By qualitative, we mean a dichotomous measure (yes or no) of the robustness as opposed to a *quantitative* measure that evaluates the level of robustness of a statistic.

In order to define the concept of qualitative robustness, we remind the well-known definition of the continuity of a real-valued function.

Definition 1.2. *Let I be an interval in $\mathbb{R}$ and $f : I \to \mathbb{R}$ be a function. f is a* continuous function *at $\alpha \in I$, if $\forall \epsilon > 0$ there exists $\delta > 0$ such that $\forall x \in I$, we have*

$$|x - \alpha| < \delta \Rightarrow |f(x) - f(\alpha)| < \epsilon.$$

Now, we generalize this definition to the case where we have a mapping between two metric space.

Definition 1.3. *Let (E, d) and (F, h) be 2 metric spaces and define the mapping $f : E \to F$. This mapping is continuous at $\alpha \in E$ if, $\forall \epsilon > 0$ there exists $\delta > 0$ such that $\forall x \in E$, we have*

$$d(x, a) < \delta \Rightarrow h(f(x), f(\alpha)) < \epsilon.$$

This continuity makes it possible to evaluate the qualitative robustness of a statistic. The original definition of this measure was given by Hampel (1971b).

Definition 1.4. *We consider a space* $(\Omega, \mathcal{A})$*,* $\mathcal{P}(\Omega)$ *the set of all probability measures which can be define on this space and a statistic* T_n*. For two distributions* $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(\Omega)$*,* T_n *is qualitatively robust at* $\mathbb{P}$ *if* $\forall \epsilon > 0$*, there is* $\delta > 0$ *such that*

$$\pi(\mathbb{P}, \mathbb{Q}) \leq \delta \Rightarrow \pi(\mathcal{L}_{\mathbf{P}}(T_n), \mathcal{L}_{\mathbf{Q}}(T_n)) \leq \epsilon,$$

when $n \to \infty$*.* $\mathcal{L}_{\mathbf{P}}(T_n)$ *is the cumulative distribution function of the statistics* T_n *calculated from a sample of size* n *of a random variable with distribution* $\mathbb{P}$*. Finally,* π *is an appropriate measure of the distance between two distributions.*

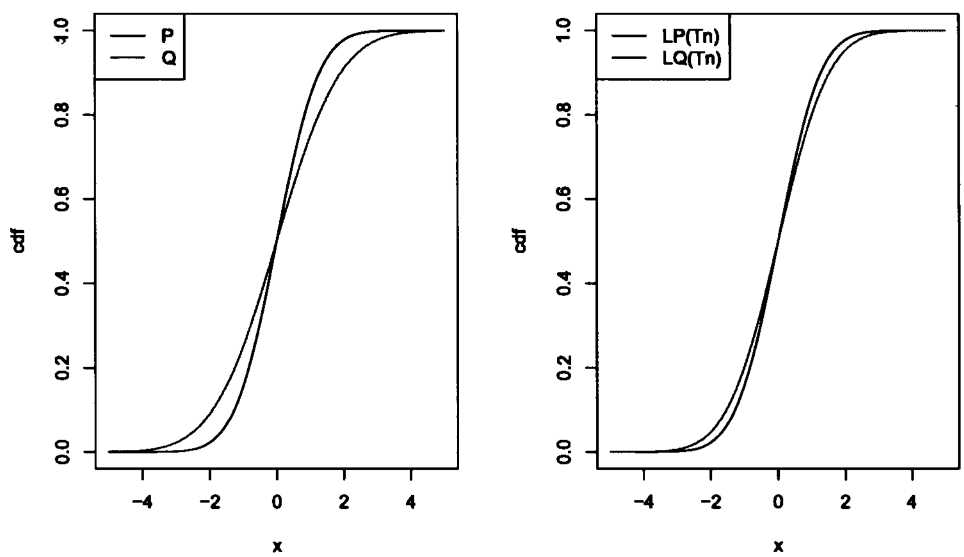

Figure 1.3 Qualitative Robustness

In Figure 1.3, we show on the left-hand side graph two theoretical cumulative distribution functions (for two random variables P and Q) which are close. On the right-hand side, we illustrate the empirical cumulative distribution function of the statistic T_n based on both, the distribution F_P and F_Q. Since the statistic T_n is robust, these two distributions are very close too.

This approach has many practical limitations, mainly because of its dichotomy. Thus, we need a more detailed measure which can be used in practice. In the following sections we will introduce a "quantitative" definition of robustness.

1.3 M-Estimators

M-estimators are a very general class of estimators obtained by evaluating the minimum of sums of functions of data. By Maronna *et al.* (2006), as special cases of M-estimators we can point out

- least-square estimators (LSE) which are defined as the minimum of the sum of squared residuals;
- maximum likelihood estimators (MLE) which are obtained by finding the maximum of the likelihood function with respect to a (vector-valued) parameter θ on the parameter space Θ.

To illustrate the concept, we consider a random sample of size n from a parametric distribution f_θ and a maximum likelihood estimator $T_n = T_n(X_1, X_2, \dots, X_n)$ which belongs to the class of M-estimator. The MLE is obtained by maximizing the likelihood function

$$\widehat{\theta} = T_n = \arg\max_{T_n \in \Theta} \prod_{i=1}^{n} f_{T_n}(X_i),$$

or, equivalently, by minimizing the negative log-likelihood function

$$\widehat{\theta} = T_n = \arg\min_{T_n \in \Theta} \sum_{i=1}^{n} - \ln\left(f_{T_n}(X_i)\right).$$

The idea behind an M-estimator is to generalize this last expression to

$$\widehat{\theta} = T_n = \arg\min_{T_n \in \Theta} \sum_{i=1}^{n} \rho\left(X_i, T_n\right),$$

where ρ is a function.

To formalize the concept, we consider, as in Subsection 1.2.1, two measure spaces $\{\chi, \Sigma\}$ and $\{\Theta \in \mathbb{R}^d, S\}$ where d is a positive integer and $\theta \in \Theta$ is a vector of unknown parameters. An M-estimator T is defined through a measurable function

$$\rho : \chi \times \Theta \to \mathbb{R}.$$

which maps a probability distribution F to the value $T(F) \in \Theta$ that minimizes

$$\int_{\chi} \rho(x, \theta)\, dF(x).$$

We assume that the function ρ has a derivative $\frac{\partial}{\partial \theta}\rho\left(x, \theta\right) = \psi\left(x, \theta\right)$. Thus, an M-estimator can also be defined as the solution of

$$\sum_{i=1}^{n} \psi(X_i, T_n) = 0 \tag{1.3.1}$$

or, equivalently, as the solution of

$$\int_{\chi} \psi(y, T(G))\, dG(y) = 0. \tag{1.3.2}$$

Now, we are interested in finding a general expression for the influence function of an M-estimator. Let $G = (1-t)F + t\Delta_x$ in Equation (1.3.2)

$$\begin{aligned}
&\int \psi\left(y, T((1-t)F + t\Delta_x)\right)\, d\left((1-t)F + t\Delta_x\right)(y) = 0 \\
&\Rightarrow (1-t)\int \psi\left(y, T((1-t)F + t\Delta_x)\right)\, dF(y) \\
&\quad + t\int \psi\left(y, T((1-t)F + t\Delta_x)\right)\, d\Delta_x(y) = 0 \\
&\Rightarrow (1-t)\int \psi\left(y, T((1-t)F + t\Delta_x)\right)\, dF(y) \\
&\quad + t\psi\left(x, T((1-t)F + t\Delta_x)\right) = 0.
\end{aligned} \tag{1.3.3}$$

Now we take the derivative of Equation (1.3.3) with respect to t and we should remember that T is a function of t. Therefore.

$$
\begin{aligned}
&\frac{\partial}{\partial t}\left(\int (1-t)\psi(y,T((1-t)F+t\Delta_x))\,dF(y) + t\psi(x,T((1-t)F+t\Delta_x))\right) = 0\\
&\quad \Rightarrow -\int \psi(y,T((1-t)F+t\Delta_x))\,dF(y)\\
&\quad + (1-t)\int \left(\frac{\partial\psi(y,T((1-t)F+t\Delta_x))}{\partial T((1-t)F+t\Delta_x)}\right)\left(\frac{\partial T((1-t)F+t\Delta_x)}{\partial t}\right) dF(y)\\
&\quad + \psi(x,T((1-t)F+t\Delta_x)\\
&\quad + t\left(\frac{\partial\psi(x,T((1-t)F+t\Delta_x))}{\partial T((1-t)F+t\Delta_x)}\right)\left(\frac{\partial T((1-t)F+t\Delta_x)}{\partial t}\right) = 0\\
&\quad \Rightarrow \int \psi(y,T((1-t)F+t\Delta_x))d(\Delta_x - F)(y)\\
&\quad + (1-t)\int \left(\frac{\partial\psi(y,T((1-t)F+t\Delta_x))}{\partial T((1-t)F+t\Delta_x)}\right)\left(\frac{\partial T((1-t)F+t\Delta_x)}{\partial t}\right) dF(y)\\
&\quad + t\left(\frac{\partial\psi(x,T((1-t)F+t\Delta_x))}{\partial T((1-t)F+t\Delta_x)}\right)\left(\frac{\partial T((1-t)F+t\Delta_x)}{\partial t}\right) = 0 \qquad (1.3.4)
\end{aligned}
$$

Now, we calculate the limit of Equation (1.3.4) when $t \to 0$

$$
\begin{aligned}
&\lim_{t\to 0}\left(\int \psi(y,T((1-t)F+t\Delta_x)\ d(\Delta_x - F)(y)\right.\\
&\quad + (1-t)\int \left(\frac{\partial\psi(y,T((1-t)F+t\Delta_x))}{\partial T((1-t)F+t\Delta_x)}\right)\left(\frac{\partial T((1-t)F+t\Delta_x)}{\partial t}\right) dF(y)\\
&\quad + \left.\left(\frac{\partial\psi(x,T((1-t)F+t\Delta_x))}{\partial T((1-t)F+t\Delta_x)}\right)\left(\frac{\partial T((1-t)F+t\Delta_x)}{\partial t}\right)\right)\\
&\quad = -\int \psi(y,T(F))\,dF(y)\\
&\quad + \psi(x,T(F))\left(\frac{\partial T((1-t)F+t\Delta_x)}{\partial t}\right)\Bigg|_{t=0} dF(y) + 0 = 0, \qquad (1.3.5)
\end{aligned}
$$

where we assume that the order of derivative and integration can be switched. Making use of Equation (1.2.8) and Equation (1.3.2), we obtain

$$
\begin{aligned}
\psi(x,T(F)) + \left(\frac{\partial\psi(y,T((1-t)F+t\Delta_x))}{\partial T((1-t)F+t\Delta_x)}\right)\Bigg|_{T((1-t)F+t\Delta_x)=T(F)} dF(y)\\
\times IF(x;T,F) = 0.
\end{aligned}
$$

Therefore, the influence function for a M-estimator is

$$IF(x;T,F) = \frac{\psi(x,T(F))}{\left(\frac{\partial\psi(y,T((1-t)F+t\Delta_x))}{-\partial T((1-t)F+t\Delta_x)}\right)\Big|_{T((1-t)F+t\Delta_x)=T(F)} dF(y)}. \tag{1.3.6}$$

For a maximum likelihood estimator, the ψ function is given by

$$\psi(x,T) = -\frac{d}{dT}\ln\left(f_T(x)\right)$$

and let $\widetilde{T}$ be the maximum likelihood estimator. Then, the influence function can be written as

$$\begin{aligned} IF(x;T,F_{\widetilde{T}}) &= \frac{\left(-\frac{d}{dT}\ln\left(f_T(x)\right)\right)\big|_{T=\widetilde{T}}}{\left(-\int \frac{d^2}{dT^2}\ln\left(f_T(y)\right)\right)\big|_{T=T(F_{\widetilde{T}})} dF_{\widetilde{T}}(y)} \\ &= \frac{\left(-\frac{d}{dT}\ln\left(f_T(x)\right)\right)\big|_{T=\widetilde{T}}}{\left(\int\left(\frac{d}{dT}\ln\left(f_T(y)\right)\right)\big|_{T=T(F_{\widetilde{T}})}\right)^2 dF_{\widetilde{T}}(y)}. \end{aligned} \tag{1.3.7}$$

To conclude this chapter, we look at some special cases of influence function for maximum likelihood estimators.

Example 1.4. *We assume a location normal model*

$$F(x-\mu) = \Phi(x-\mu)$$

where the likelihood function is

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp -0.5(x_i-\mu)^2$$

and the log-likelihood function is

$$l = \sum_{i=1}^{n} \ln\left(\sqrt{2\pi}\right) - 0.5(x_i-\mu)^2.$$

The derivative of the log-likelihood function with respect to the parameter is

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{n}(x_i-\mu),$$

then we put the derivative equal to zero to find the estimator

$$\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\Rightarrow \hat{\mu} = \bar{x}.$$

Let $\psi(x) = x$ *and the influence function is*

$$IF(x; \bar{x}, F_{\bar{x}}) = \frac{x}{\int y^2\, dF_{\bar{x}}(y)} = \frac{x}{1} = x,$$

since

$$\int y^2\, dF_{\bar{x}}(y) = \int y^2 f_{\bar{x}}\, dy = \mathrm{Var}[Y] - \mathrm{E}[Y]^2 = 1 - 0 = 1.$$

Example 1.5. *We assume a location logistic model*

$$F(x-\mu) = \frac{1}{1+\exp\{-(x-\mu)\}}$$

where the density function is

$$f(x-\mu) = \frac{dF(x-\mu)}{d(x-\mu)} = \frac{\exp\{-(x-\mu)\}}{(1+\exp\{-(x-\mu)\})^2}.$$

The likelihood function is

$$L = \prod_{i=1}^{n} \frac{\exp\{-(x_i-\mu)\}}{(1+\exp\{-(x_i-\mu)\})^2}$$

and the log-likelihood function is

$$l = \sum_{i=1}^{n} -(x_i - \mu) - 2\ln(1+\exp^{-(x_i-\mu)}).$$

The derivative of the log-likelihood function with respect to parameter (μ) *is*

$$\frac{\partial l}{\partial(x_i-\mu)} = \sum_{i=1}^{n} -1 + 2\left(\frac{\exp^{-(x_i-\mu)}}{(1+\exp^{-(x_i-\mu)})}\right)$$

then we put the derivative equal to zero to find the estimator

$$\sum_{i=1}^{n} 2\left(\frac{\exp^{-(x_i-\mu)}}{(1+\exp^{-(x_i-\mu)})}\right) - 1 = 0$$

$$\Rightarrow \sum_{i=1}^{n} 2F(x_i - \mu) - 1 = 0,$$

where there is no explicit solution for $\hat{\mu}$. *Let* $\psi(x) = 2F(x) - 1$ *and the influence function is*

$$IF(x; \bar{x}, F_{\hat{\mu}}) \propto \psi(x).$$

Example 1.6. *We assume a location Laplace model where the probability density functin is*

$$f(x - \mu) = 0.5 \ \exp\{-|x - \mu|\}.$$

The likelihood function is

$$L = \prod_{i=1}^{n} 0.5 \exp\{-|x_i - \mu|\}$$

and the log-likelihood function is

$$l = \sum_{i=1}^{n} - \ln(2) - |x_i - \mu|$$

$$\hat{\mu} = \arg\min\left(\sum_{i=1}^{n} |x_i - \mu|\right)$$

and

$$\hat{\mu} = median(x).$$

The derivative of log-likelihood function with respect to the parameter is

$$\frac{\partial l}{\partial(x_i - \mu)} = \begin{cases} +1 & x_i > \mu \\ -1 & x_i < \mu \end{cases}$$

$$= sign(x - \mu).$$

Therefore the influence function is

$$IF(x;\bar{x},F_{median(x)}) = \frac{\left(-\frac{d}{dT}\ln f_T(x)\right)\Big|_{T=\tilde{T}}}{\int \left(\left(\frac{d}{dT}\ln f_T(y)\right)\Big|_{T=T(F_{\tilde{T}})}\right)^2 dF_{\tilde{T}}(y)} \propto \psi(x)$$

since the integration in denominator is a constant.

CHAPTER II

CLASSICAL MODEL FOR RESERVE IN GENERAL INSURANCE

2.1 Introduction and Motivation

In Figure 2.1, we illustrate the typical development of a claim in non-life insurance. In many cases, it is not possible for the insurer to settle a claim immediately after its occurrence. The main reasons for such a delay are:

- Reporting delay: in each type of line of business there is an acceptable period in which the insured can notify the insurer of the claim occurrence. For example, in liability insurance, reporting a claim can take years. In Table 2.1, we report some descriptive statistics about the rate of payment in various lines of business, while n is the accident year (see Denuit et Charpentier (2004)).

- Claim verification: an insurance company may want to make all necessary verifications about the accuracy and the authenticity of the claim. Obviously, it takes some time before the settlement to do this process.

- Reopening of a case: in the case of gathering new information about an old case, it may be necessary to reopen the case and update the claim amount.

At each moment, insurance companies need to have the sufficient capital aside in

Table 2.1 Rate of payment in various lines of business.

	n	$n+1$	$n+2$	$n+3$	$n+4$
Home	55%	90%	94%	95%	96%
Car	55%	79%	84%	90%	99%
Bodily injury	13%	38%	50%	65%	72%
Public liability	10%	25%	35%	40%	45%

order to be able to pay their future liabilities which are generated by the contracts that have been sold to the clients. This capital forms the *reserve* (or provision) of the non-life company. Periodically, insurance regulators require an evaluation on this reserve in order to control the financial solvency of the company and to protect policyholders. The core task of a reserving actuary is to calculate this reserve and to determine some characteristics of its distribution.

On the evaluation date, claims are classified in different groups according to their stage of development. These groups are illustrated on Figure 2.1. When a claim has been reported before the evaluation date but has not been completely paid, we classify the claim as *reported but not settled* (RBNS). When the evaluation date is after the occurrence of the claim but before the reporting date, we classify the claim as *incurred but not reported* (IBNR). Finally, when the evaluation date is after the occurrence date but the claim has not started to be paid, the claim is *reported but not (yet) paid* (RBNP). Often, this last category is grouped with the RBNS class, as illustrated on Figure 2.1. In general insurance, it is not possible to have an accurate financial condition without an accurate unpaid claim evaluation. As noted, there are some elements that needed to be evaluated in order to estimate the total reserve amount: a provision for RBNS claims, a provision for RBNP

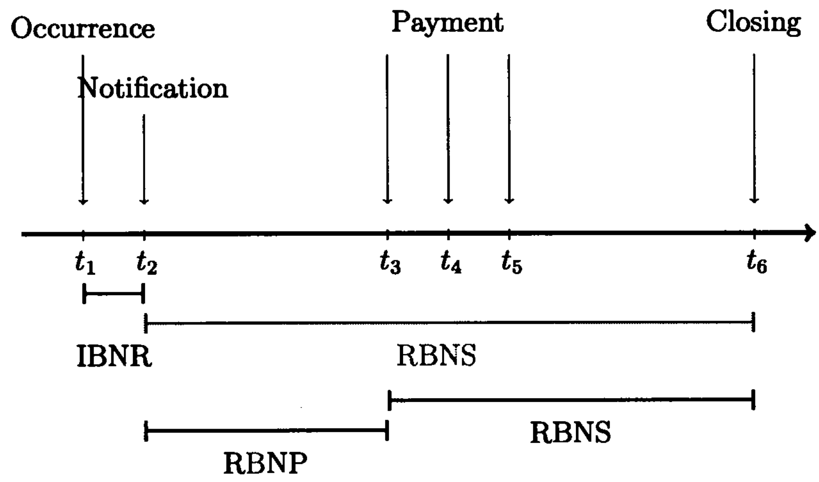

Figure 2.1 Claim development.

claims, a provision for IBNR claims, an estimation for reopened claim, etc.

2.2 Collective and Individual Approaches

Existing models for loss reserving can be divided into two main groups based on the granularity of the database: one can construct models for reserve by using aggregated data, i.e., data summarized by occurrence period, by development period, etc., or by using very detailed database, i.e., by using individual covariates such as information on the claim, information on the policyholder, etc.

2.2.1 Collective Approach

For nearly 40 years, the lack of detailed and accurate databases has forced the insurance companies to develop stochastic models for aggregated information. Reserving actuaries summarize all the information about claim payments and incurred losses in a development triangle, or *run-off triangle*. This structure shows the development of claims over time. It has an incremental and a cumulative form.

In Table 2.2, there is a toy example of a run-off triangle for occurrence years 2010 to 2012 and annual reporting periods. A run-off triangle can be read from a variety

Table 2.2 Cumulative run-off triangle (in millions of dollars).

Accident year	12 months	24 months	36 months
2010	100	150	170
2011	110	160	
2012	120		

of perspectives:

- each row represents one accident year, so in our example, the first row contains claims with occurrence year 2010, the second row contains claim payments with occurrence in 2011 and so on;
- each column represents age or maturity date: in the Table 2.2, the first column contains claims payments after 12 months, the second column contains the cumulative claims payments after 24 months and so on; and
- each diagonal represents one calendar year. In our example, the first diagonal (upper left corner of the triangle) is a portrait of the situation at the end of year 2010 and so on.

More formally, a collective model is constructed for a run-off triangle with occurrence periods $i = 1, 2, \ldots, I$ and development periods $j = 1, 2, \ldots, J$ (without loss of generality we assume that $J = I$ in the following)

$$\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1(I-1)} & C_{1I} \\ C_{21} & C_{22} & \cdots & C_{2(I-1)} & \\ \vdots & \ddots & & & \\ C_{I1} & & & & \end{bmatrix},$$

where C_{ij} represents the total cumulative amount of claims occurred in the i^{th} period and paid up to period j. In order to calculate the reserve we need to predict the lower triangle in the above matrix

$$\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1(I-1)} & C_{1I} \\ C_{21} & C_{22} & \cdots & C_{2(I-1)} & \widehat{C}_{2I} \\ C_{31} & C_{32} & \cdots & \widehat{C}_{3(I-1)} & \widehat{C}_{3I} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{I1} & \widehat{C}_{I2} & \cdots & \widehat{C}_{I(I-1)} & \widehat{C}_{II} \end{bmatrix}.$$

Then, the total amount of the reserve is predicted by

$$\widehat{R} = \sum_{t=1}^{I} \widehat{C}_{tI} - \sum_{t=1}^{I} C_{t(I-t+1)}$$

There exist similar models for incremental run-off triangles (see Hertig (1985), Renshaw et Verrall (1998a), England et Verrall (2002), Taylor (2000)).

Collective models have been studied since the early 80s. An interested reader can consult Wüthrich et Merz (2008) for an almost complete overview. Among the collective models, the stochastic Chain-Ladder model (or Mack's model) has a special position because it produces the same reserve estimate as the Chain-Ladder algorithm and it does not make assumption about the underlying distribution of a claim amount (see Carrato *et al.* (1999)). Also, the mean square error of the predicted reserve amount can be calculated with an analytic formula. It will be presented in Section 2.4.

2.2.2 Individual Approach

For individual (or micro-level) approaches, a detailed and accurate database is essential to predict the claim development and to estimate the parameters of the model. In recent decades, the availability of detailed information on each reported claim and the development of strong computational tools allow creative researchers to propose individual approaches for evaluating loss reserves. Among these researchers Arjas (1989) and Norberg (1999) proposed an individual stochastic structure within a continuous time framework. From this base, several other models have been developed, e.g., Arjas et Haastrup (1996), Larsen (2007), Zhao *et al.* (2009), Zhao et Zhou (2010) and Antonio et Plat (2014). In a discrete time framework, Pigeon *et al.* (2013) and Pigeon *et al.* (2014) proposed an individual model based on the Chain-Ladder structure.

This family of approaches offers a better performance when compared to collective models, allows for individual predictions and establishes a time/payment schedule. Individual approach will not be studied in this research.

2.3 Generalized Linear Models

Generalized linear models (GLM) are a very popular class of approaches used to specify and to quantify the relation between a response variable and some covariates. Generalized linear models can model both individual and collective loss reserving approaches. The first researchers who proposed loss reserving models based on generalized linear models were Renshaw et Verrall (1998a) and Renshaw et Verrall (1998b).

Generalized linear models differ from linear models in three important points:

- the distribution of the response variable is a member of the exponential family (since the Normal distribution is a member of this family, the linear model is a special case of GLM but it is no longer the only option);

- the relation between the expected value of the response variable and the covariates (or independent variables) is not necessary linear: a link function describes the relation between those two components; and
- the variance of the response variable is not necessarily a constant (as in the basic linear model) and can vary according to the expected value of the dependent variable.

More details can be found in Appendix B and in McCullagh et Nelder (1989).

The role of generalized linear models in general insurance is important because the Normal distribution is rarely a valid option to model loss distributions. Moreover, the relation between the expected value of a claim and its characteristics is mostly multiplicative and hardly additive.

Definition 2.1. *Let X_{ij}, where $X_{ij} = C_{ij} - C_{i(j-1)}$, $2 \leq j \leq I$ and $X_{i1} = C_{i1}$, $1 \leq i \leq I$ be incremental data. A generalized linear model for loss reserving is defined by*

i. incremental data X_{ij} for different occurrence years ($i_1 \neq i_2$) and/or different development years ($j_1 \neq j_2$) are independent;

ii. the probability density function of X_{ij} is

$$f(x;\theta_{ij},\phi_{ij}) = c(x;\phi_{ij}) \exp\left(\frac{x\theta_{ij} - a(\theta_{ij})}{\phi_{ij}}\right),$$

where θ_{ij} is the canonical parameter, ϕ_{ij} is the dispersion parameter and $a()$ and $c()$ are two functions.

Based on Definition 2.1, we obtain

$$\mathrm{E}[X_{ij}] = \mu_{ij} = a'(\theta_{ij})$$
$$\mathrm{Var}[X_{ij}] = \phi_{ij}\mathrm{V}[x_{ij}] = \phi_{ij}a''(\theta_{ij}),$$

where $\phi_{ij} > 0$ and V[.] is a function called *variance function.*

The model above has $I \times J$ unknown parameters. They need to be estimated using information in the run-off triangle which contains less than $I \times J$ observations. Thus, it is inevitable to add more structure (or constraints) in the model in order to decrease the number of unknown parameters.

We assume a multiplicative structure such as

$$\mathrm{E}[X_{ij}] = \alpha_i\psi_j,$$

where α_i stands for the accident year effect and ψ_j stands for the development period effect. Consequently, there remain $I+J$ unknown parameters in the model: $\alpha_1, \dots, \alpha_I$ and $\psi_1, \dots, \psi_J$.

By selecting the logarithmic link function, the model is straightforward and we obtain

$$\ln(\mathrm{E}[X_{ij}]) = \ln(\mu_{ij}) = \ln(\alpha_i) + \ln(\psi_j)\,.$$

To make the model identifiable, we add an additional constraint on the model: $\sum_{j=1}^{J}\psi_j = 1$ or $\alpha_1 = 1$. In the following, we choose $\alpha_1 = 1$ which leads to $\ln(\alpha_1) = 0$. Then, the number of parameters to estimate is $I + J - 1$.

In order to write the model in the generalized linear model framework, we group the parameters in a vector as below

$$\boldsymbol{\beta} = \left[\ln(\alpha_2) \quad \cdots \quad \ln(\alpha_I) \quad \ln(\psi_1) \quad \ln(\psi_2) \quad \cdots \quad \ln(\psi_J)\right]'$$

and we define design matrices as

$$Z_{1j} = \begin{bmatrix} 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots \end{bmatrix}$$
$$Z_{ij} = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots \end{bmatrix}$$

where 1 are in position $i-1$ position $I+j-1$. Hence, the linear predictor of the model is

$$\eta_{ij} = \ln\left(\mathrm{E}[X_{ij}]\right) = \ln\left(\mu_{ij}\right) = Z_{ij}\boldsymbol{\beta}. \tag{2.3.1}$$

Therefore the expected claim payment for cell (i, j) is

$$\mathrm{E}[X_{ij}] = \exp\{Z_{ij}\boldsymbol{\beta}\} = \exp\{\ln(\alpha_i) + \ln(\psi_j)\},$$

where $\alpha_1 = 1$.

All parameters can be estimated by a maximum likelihood approach which is available in virtually all statistical software. The estimation of $\boldsymbol{\beta}$ by maximum likelihood method is

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\ln(\alpha_2)} & \cdots & \widehat{\ln(\psi_J)} \end{bmatrix}',$$

and therefore $\widehat{X_{ij}}$ evaluated by

$$\widehat{X_{ij}} = \exp\{Z_{ij}\widehat{\boldsymbol{\beta}}\}.$$

In order to evaluate the risk of our reserve estimates, we need to calculate the mean square error of prediction as defined in the Theorem 2.3.1.

Theorem 2.3.1. *The mean square error of prediction (MSEP) for the total paid amount* $\sum_i \widehat{C}_{i,J}$ *is*

$$MSEP_{\sum_i C_{iJ}}\left(\sum_{i=1}^{I} \widehat{C}_{iJ}\right) = \sum_{i+j>I} \phi_{ij} \mathrm{V}\left[\widehat{X}_{ij}\right] + \sum_{i+j>I, n+m>I} \widehat{X}_{ij}\widehat{X}_{nm} Z_{ij} H(\hat{\beta})^{-1} Z'_{mn}.$$

Proof. Given

$$\mathcal{D}_T = \{C_{ij} : i + j \leq T + 1, 1 \leq j \leq T\},$$

the conditional mean square error of prediction (MSEP) for the total paid amount $\sum_i \widehat{C}_{i,J}$, is

$$\begin{aligned}
\mathrm{MSEP}_{\sum_i C_{iJ}|\mathcal{D}_I}\left(\sum_{i=1}^{I} \widehat{C}_{iJ}\right) &= \mathrm{E}\left[\left(\sum_{i=1}^{I} \widehat{C}_{iJ} - \sum_{i=1}^{I} C_{iJ}\right)^2 |\mathcal{D}_I\right] \\
&= \mathrm{E}\left[\left(\sum_{i+j>I} \widehat{X}_{ij} - \sum_{i+j>I} X_{ij}\right)^2 |\mathcal{D}_I\right] \\
&= \mathrm{Var}\left[\sum_{i+j>I} X_{ij}|\mathcal{D}_I\right] + \left(\sum_{i+j>I} \left(\widehat{X}_{ij} - \mathrm{E}[X_{ij}|\mathcal{D}_I]\right)|\mathcal{D}_I\right)^2 \\
&= \underbrace{\mathrm{Var}\left[\sum_{i+j>I} X_{ij}\right]}_{\text{stoc. error}} + \underbrace{\left(\sum_{i+j>I} \left(\widehat{X}_{ij} - \mathrm{E}[X_{ij}]\right)\right)^2}_{\text{est. error}},
\end{aligned}$$

where for $i + j > I$, X_{ij} and $\mathcal{D}_I$ are independent. By using the between-cell independence, we have

$$\begin{aligned}
\mathrm{MSEP}_{\sum_i C_{iJ}|\mathcal{D}_I}\left(\sum_{i=1}^{I} \widehat{C}_{iJ}\right) &= \sum_{i+j>I} \mathrm{Var}[X_{ij}] + \left(\sum_{i+j>I} \left(\widehat{X}_{ij} - \mathrm{E}[X_{ij}]\right)\right)^2 \\
&= \sum_{i+j>I} \phi_{ij}\mathrm{V}[X_{ij}] + \left(\sum_{i+j>I} \left(\widehat{X}_{ij} - \mathrm{E}[X_{ij}]\right)\right)^2.
\end{aligned}$$

We calculate the stochastic error and the estimation error separately and then we sum these up to obtain the (conditional) mean square error of prediction for the total paid amount $\sum_i \widehat{C}_{i,J}$. The difficult part is calculating the estimation error.

The unconditional MSEP is given by:

$$\text{MSEP}_{\sum_i C_{iJ}}\left(\sum_{i=1}^{I}\widehat{C}_{iJ}\right) = \text{E}\left[\text{MSEP}_{\sum_i C_{iJ}|\mathcal{D}_I}\left(\sum_{i=1}^{I}\widehat{C}_{iJ}\right)\right]$$
$$= \sum_{i+j>I}\phi_{ij}\text{V}[X_{ij}] + \text{E}\left[\left(\sum_{i+j>I}\left(\widehat{X}_{ij} - \text{E}[X_{ij}]\right)\right)^2\right]$$

Now we first estimate the last part

$$\text{E}\left[\left(\sum_{i+j>I}\left(\widehat{X}_{ij} - \text{E}[X_{ij}]\right)\right)^2\right] = \sum_{i+j>I,m+n>I}\text{E}\left[\left(\widehat{X}_{ij} - \text{E}[X_{ij}]\right)\left(\widehat{X}_{mn} - \text{E}[X_{mn}]\right)\right] \tag{2.3.2}$$

since $\hat{X}_{ij}$ is not an unbiased estimator for $E[X_{ij}]$ then (2.3.2) may have bias. The quadratic term in (2.3.2) will be approximated by

$$\begin{aligned}
\text{Var}\left[\widehat{X}_{ij}\right] &= \text{Var}\left[\exp\{Z_{ij}\hat{\beta}\}\right] \\
&= \exp\{2Z_{ij}\beta\}\text{Var}\left[\exp\{Z_{ij}\hat{\beta} - Z_{ij}\beta\}\right] \\
&\approx \exp\{2Z_{ij}\beta\}\text{Var}\left[Z_{ij}\hat{\beta}\right] \\
&= X_{ij}^2 Z_{ij}\text{Cov}\left[\hat{\beta},\hat{\beta}\right]Z'_{ij}
\end{aligned}$$

we have used the linearization $\exp(z) \approx 1+z$ for $z \approx 0$, in third to fourth equality. The cross term in equation (2.3.2) will be approximated by

$$\begin{aligned}
\text{Cov}\left[\widehat{X}_{ij},\widehat{X}_{mn}\right] &\approx \exp\{Z_{ij}\beta + Z_{nm}\beta\}\text{Cov}\left[Z_{ij}\hat{\beta} + Z_{nm}\hat{\beta}\right] \\
&= X_{ij}X_{nm}Z_{ij}\text{Cov}\left[\hat{\beta},\hat{\beta}\right]Z'_{mn}
\end{aligned}$$

Here we need to calculate $\text{Cov}\left[\hat{\beta},\hat{\beta}\right]$. We know $Z_{ij}^{(k)}$ is the k^{th} coordinate of the design matrix then

$$\widehat{\mathrm{Cov}\left[\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\beta}}\right]}$$
$$= \left(\left(\sum_{i+j\leq I} \mathrm{Var}\left[\widehat{X}_{ij}\right]^{-1} Z_{ij}^{(k)} Z_{ij}^{(l)}\right)_{k,l=1,2,\ldots,I+J-1}\right)^{-1}$$
$$= H(\hat{\beta})^{-1}.$$

Then

$$\mathrm{MSEP}_{\sum_i C_{iJ}}\left(\sum_{i=1}^{I}\widehat{C}_{iJ}\right) = \sum_{i+j>I}\phi_{ij}\mathrm{V}\left[\widehat{X}_{ij}\right] + \sum_{i+j>I,n+m>I}\widehat{X}_{ij}\widehat{X}_{nm}Z_{ij}H(\hat{\beta})^{-1}Z'_{mn}.$$

□

2.3.1 The (Over-dispersed) Poisson Model for Loss Reserving

If we choose $\mathrm{V}[\mu_{ij}] = (1)\mu_{ij}$ (recall that $\mu_{ij} = \mathrm{E}[X_{ij}]$) and a logarithmic link function in the generalized linear model framework (other choices possible for the link function. Here in order to respect multiplicative affect we choose logarithmic link), we obtain a Poisson model for loss reserving. It means that incremental payments X_{ij} are independent random variables following Poisson distributions. Since the constraint has no effect on the total reserve amount, we can choose $\sum_{j=1}^{J}\psi_j = 1$ or $\alpha_1 = 1$ and adapt the Z_{ij} based on our choice. In the following, we use $\sum_{j=1}^{J}\psi_j = 1$ to simplify the presentation.

In the Poisson model, parameters β_t, $1 \leq t \leq I+J-1$ are estimated by maximum likelihood method as below:

$$
\begin{aligned}
L(x;\boldsymbol{\alpha},\boldsymbol{\psi},\boldsymbol{\beta}) &= \prod_{i+j\leq I} \frac{\exp\{-\alpha_i\psi_j\}(\alpha_i\psi_j)^{x_{ij}}}{x_{ij}!} \\
\ell(x;\boldsymbol{\alpha},\boldsymbol{\psi},\boldsymbol{\beta}) &= \sum_{i+j\leq I} (-\alpha_i\psi_j) + x_{ij}\ln(\alpha_i\psi_j) - \ln(x_{ij}!) \\
\ell(x;\boldsymbol{\alpha},\boldsymbol{\psi},\boldsymbol{\beta}) &= \sum_{i+j\leq I} -e^{Z_{ij}\boldsymbol{\beta}} + x_{ij}Z_{ij}\boldsymbol{\beta} - \ln(x_{ij}!) \\
\frac{\partial\ell(x;\boldsymbol{\alpha},\boldsymbol{\psi},\boldsymbol{\beta})}{\partial\beta_t} &= \sum_{i+j\leq I} (x_{ij} - e^{Z_{ij}\boldsymbol{\beta}})Z_{ijt} = 0, \quad t = 1,\ldots,I+J+1, \qquad (2.3.3)
\end{aligned}
$$

where Z_{ijt} is the t^{th} element in the matrix Z_{ij} and, as we saw in the previous section, $\ln(\alpha_i\psi_j) = Z_{ij}\boldsymbol{\beta}$. The solution of the system of equations given by (2.3.3) gives us estimated values for parameters $\boldsymbol{\beta}$.

Here we calculate the MSEP for the total paid amount $\sum_i C_{iJ}$ with a Poisson distribution. By Theorem 2.3.1, the MSEP for a member of the exponential family is

$$
\begin{aligned}
\mathrm{MSEP}_{\sum_i C_{iJ}}\left(\sum_{i=1}^{I}\widehat{C}_{iJ}\right) &= \sum_{i+j>I} \phi_{ij}\mathrm{V}[\widehat{x}_{ij}] \\
&\quad + \sum_{i+j>I,n+m>I} \widehat{x}_{ij}\widehat{x}_{nm}Z_{ij}H(\hat{\beta})^{-1}Z'_{mn}.
\end{aligned}
$$

For a Poisson random variable, we have $\phi_{ij} = 1$ and $V(\widehat{X}_{ij}) = \phi_{ij}V(\widehat{x}_{ij}) = V(\mu_{ij}) = \mu_{ij}$. Then

$$
\begin{aligned}
H(\hat{\beta})^{-1} &= \left(\left(\sum_{i+j\leq I} \mathrm{Var}\left[\widehat{X}_{ij}\right]^{-1} Z_{ij}^{(k)}Z_{ij}^{(l)}\right)_{k,l=1,2,\ldots,I+J-1}\right)^{-1} \\
&= \left(\left(\sum_{i+j\leq I} \mu_{ij}^{-1}Z_{ij}^{(k)}Z_{ij}^{(l)}\right)_{k,l=1,2,\ldots,I+J-1}\right)^{-1}.
\end{aligned}
$$

Therefore

$$
\mathrm{MSEP}_{\sum_i C_{iJ}}\left(\sum_{i=1}^{I}\widehat{C}_{iJ}\right) = \sum_{i+j>I}\mu_{ij} + \sum_{i+j>I,n+m>I} \widehat{X}_{ij}\widehat{X}_{nm}Z_{ij}H(\hat{\beta})^{-1}Z'_{mn}.
$$

Poisson model has its limitations. In fact, it rarely happens that the expected value and the variance of an incremental payment in a cell (i,j) have the same value as assumed in the Poisson model. When the observed variance is higher than the observed expected value of the model, over-dispersion occurs. In other words, $\mathrm{Var}[Y_{ij}] > \mathrm{E}[Y_{ij}]$ since we know that in the Poisson model, the theoretical variance is assumed to be equal to theoretical expected value. Over-dispersion is one of the most common issues in a dataset and it causes failure of the mean-variance relation.

Suppose $Y_1, Y_2, ..., Y_n$ are independently and identically distributed random variables following a Poisson distribution with expected value θ, which is estimated by $\bar{Y} = \sum Y_i/n$. Based on the theoretical distribution, we have $\mathrm{E}[Y_i] = \theta$ and $\mathrm{Var}[Y_i] = \theta$, while based on the empirical distribution, we may observe $\mathrm{Var}[Y_i] = \phi\theta = \phi V(\mu_i)$ where $\phi > 1$.

In the over-dispersion (or quasi-) Poisson model, parameters β_t, $1 \le t \le I+J-1$ are estimated by maximum likelihood method. β_t is the solution of system in Equation (2.3.3).
Recall that we assume a logarithmic link function: $Y_i = \exp(X_i'\boldsymbol{\beta})$. We can estimate ϕ by using the Pearson chi-squared statistics

$$X^2 = \sum_{i=1}^{n} \frac{(Y_i - \widehat{\mu}_i)^2}{V(\widehat{\mu}_i)},$$

where we assume, for simplicity, that n is the total number of observations. We define the scaled Pearson chi-squared statistic as

$$X_s^2 = \frac{X^2}{\phi}.$$

and $X_s^2 \sim \mathcal{X}_{n-p}^2$ where p is the number of unknown parameters.

As we know, the expected value of $\mathcal{X}_{n-p}^2$ is $n-p$ and we use the approximation

$\mathcal{X}_s^2 \approx n - p$

$$\hat{\phi} = \frac{X^2}{n-p}.$$

To calculate the MSEP for the total paid amount $\sum_i C_{iJ}$, we use again the Theorem 2.3.1 where we use the following definitions

$$V(\widehat{X}_{ij}) = \phi_{ij} V(\widehat{x}_{ij}) = \phi_{ij}\mu_{ij}$$

$$H(\widehat{\beta})^{-1} = \left(\left(\sum_{i+j\leq I} \mathrm{Var}\left[\widehat{X}_{ij}\right]^{-1} Z_{ij}^{(k)} Z_{ij}^{(l)}\right)_{k,l=1,2,\ldots,I+J-1}\right)^{-1}$$

$$= \left(\left(\sum_{i+j\leq I} \phi_{ij}\mu_{ij}^{-1} Z_{ij}^{(k)} Z_{ij}^{(l)}\right)_{k,l=1,2,\ldots,I+J-1}\right)^{-1}.$$

Therefore, we have

$$\mathrm{MSEP}_{\sum_i C_{iJ}}\left(\sum_{i=1}^{I} \widehat{C}_{iJ}\right) = \sum_{i+j>I} \widehat{\phi}_{ij}\mu_{ij} + \sum_{i+j>I, n+m>I} \widehat{X}_{ij}\widehat{X}_{nm} Z_{ij} H(\widehat{\beta})^{-1} Z'_{mn}.$$

Example 2.1. *The run-off triangle presented in Table 2.3 shows the cumulative dataset used in Taylor et Ashe (1983). As a first step, we evaluate the total*

Table 2.3 Cumulative run-off triangle from Taylor et Ashe (1983).

i	1	2	3	4	5	6	7	8	9	10
1	357 848	1 124 788	1 735 330	2 218 270	2 745 596	3 319 994	3 466 336	3 606 286	3 833 515	3 901 463
2	352 118	1 236 139	2 170 033	3 353 322	3 799 067	4 120 063	4 647 867	4 914 039	5 339 085	
3	290 507	1 292 306	2 218 525	3 235 179	3 985 995	4 132 918	4 628 910	4 909 315		
4	310 608	1 418 858	2 195 047	3 757 447	4 029 929	4 381 982	4 588 268			
5	443 160	1 136 350	2 128 333	2 897 821	3 402 672	3 873 311				
6	396 132	1 333 217	2 180 715	2 985 752	3 691 712					
7	440 832	1 288 463	2 419 861	3 483 130						
8	359 480	1 421 128	2 864 498							
9	376 686	1 363 294								
10	344 014									

reserve amount obtained by using the classical Poisson model for loss reserving. Predicted values for each occurrence year and each payment period are presented in Table 2.4. The predicted total amount for the reserve, which is the sum of all predicted claims, is 18 680 856$ *and the mean square error of prediction is* 12 843.46. *Now if we multiply the incremental value* $x_{2,1}$ *by* 10 *and reevaluate the*

Table 2.4 Predicted incremental triangle obtained from the classical Poisson model for loss reserving.

i	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	856 804
3	0	0	0	0	0	0	0	0	1 018 834	897 410
4	0	0	0	0	0	0	0	1 310 258	1 089 616	959 756
5	0	0	0	0	0	0	605 548	725 788	603 569	531 636
6	0	0	0	0	0	383 287	424 501	508 792	423 113	372 687
7	0	0	0	0	334 148	351 548	389 349	466 660	388 076	341 826
8	0	0	0	247 190	226 674	238 477	264 121	316 566	263 257	231 882
9	0	0	375 833	370 179	339 456	357 132	395 534	474 073	394 241	347 255
10	0	94 634	93 678	92 268	84 611	89 016	98 588	118 164	98 266	86 555

amount of reserve we will obtain completely different predictions. These values are presented in Table 2.5. The predicted total reserve amount is now 13 064 239$ *and the mean square error of prediction is* 8 310.85.

The significant difference between the predicted total reserve amount before and after introducing an outlier shows that the Poisson model for loss reserving is not robust. Since the dispersion parameter ϕ *does not intervene in predictions, results for both Poisson model and over-dispersed Poisson model are identical.*

Table 2.5 Predicted incremental triangle in presence of an outlier, completed with classical GLM method.

i	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	438 841
3	0	0	0	0	0	0	0	0	778 247	446 899
4	0	0	0	0	0	0	0	1 082 322	809 160	464 651
5	0	0	0	0	0	0	516 774	585 569	437 780	251 390
6	0	0	0	0	0	325 808	353 007	400 001	299 047	171 724
7	0	0	0	0	278 776	289 155	313 295	355 002	265 405	152 405
8	0	0	0	197 960	179 141	185 810	201 322	228 123	170 548	97 935
9	0	0	273 943	267 074	241 685	250 683	271 611	307 769	230 093	132 128
10	0	150 805	91 872	89 568	81 054	84 071	91 090	103 216	77 166	44 312

2.4 Stochastic Chain-Ladder Model

The Chain-Ladder algorithm is a distribution free, purely computational, method for evaluating the reserve. It gives an estimated value for the expected ultimate cumulative claim amount but it says nothing about how good is this estimator, about the variability of the reserve amount, about the complete predictive distribution, etc.

The Mack's model is a stochastic version of the Chain-Ladder algorithm, and it provides more information about the variance. (see England et Verrall (2002) and Wright (1990)). Suppose we have the information up to time T, then we define

$$\mathcal{D}_T = \{C_{ij} : i + j \leq T + 1, 1 \leq j \leq T\},$$

which is simply the upper left triangle (run-off triangle), and

$$\mathcal{B}_k = \{C_{ij} : i + j \leq I + 1, 1 \leq j \leq k\} \subseteq \mathcal{D}_I,$$

with $\mathcal{B}_I = \mathcal{D}_I$ (see Figure 2.2 and Figure 2.3).

Definition 2.2. *The stochastic Chain-Ladder model is based on the following hypotheses:*

(CL1) *cumulative payments for different occurrence periods are independent, i.e.,* $\left(C_{ij}\right)_{j=1,\dots,I}$ *is independent of* $\left(C_{i'j}\right)_{j=1,\dots,I}$ *for* $i \neq i'$*; and*

(CL2) *we have the following structure for the conditional first moment of* C_{ij} *(knowing the past* $C_{i1},\dots,C_{i(j-1)}$*)*

$$\mathrm{E}\left[C_{ij}|C_{i1},\dots,C_{i(j-1)}\right] = \lambda_{j-1}C_{i(j-1)}, \qquad j = 2,\dots,I.$$

Under hypotheses **(CL1-CL2)**, we have

$$\mathrm{E}[C_{iI}|\mathcal{D}_I] = \mathrm{E}\left[C_{iI}|C_{i(I-i+1)}\right] = \lambda_{I-1}\lambda_{I-2} \times \cdots \times \lambda_{I-i+1}C_{i(I-i+1)}, \;\; 2 \leq i \leq I. \tag{2.4.1}$$

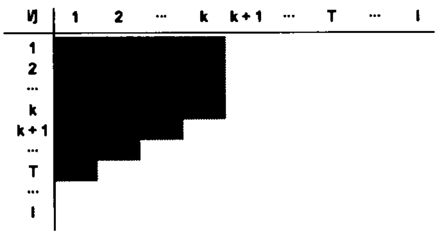

Figure 2.2 Set $\mathcal{B}_k$.

Equation (2.4.1) provides an algorithm to predict the ultimate payment knowing the information up to I. Conditional on the values of development factors (λ_j), the best estimate for the reserve amount for occurrence period i at time I is

$$(R_i|\mathcal{D}_I) = \mathrm{E}[C_{iI}|\mathcal{D}_I] - C_{i(I-i)} = C_{i(I-i)}(\lambda_{I-i} \times \cdots \times \lambda_{I-1} - 1).$$

In practice, development factors are unknown and have to be estimated. Under the assumptions **(CL1-CL2)**, the standard estimators for development factors are

$$\widehat{\lambda}_j = \frac{\sum_{i=1}^{I-j} C_{i(j+1)}}{\sum_{i=1}^{I-j} C_{ij}}, \qquad j = 2, \ldots, I-1. \tag{2.4.2}$$

Based on hypotheses in Definition 2.2, we can prove that (see Wüthrich et Merz (2008) for the detailed proof)

- given $\mathcal{B}_j$, $\widehat{\lambda}_j$ is an unbiased estimator of λ_j;

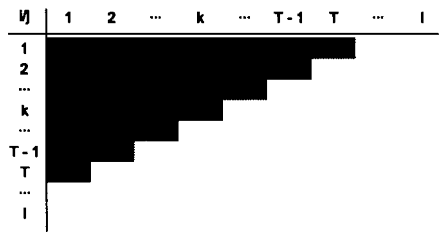

Figure 2.3 Set $\mathcal{D}_T$.

- $\widehat{\lambda}_j$ is an unbiased estimator of λ_j;
- $\widehat{\lambda}_0, \ldots, \widehat{\lambda}_j$ are non-correlated estimators;
- given $C_{i(I-i)}$, $\widehat{C}_{iI}$ is an unbiased estimator of $\mathrm{E}[C_{iI}|\mathcal{D}_I] = \mathrm{E}\big[C_{iI}|C_{i(I-i)}\big]$; and
- $\widehat{C}_{iI}$ is an unbiased estimator of $\mathrm{E}[C_{iI}]$.

The advantage of the stochastic version of the Chain-Ladder model is to enable the evaluation of the prediction error. In order to be able to calculate this prediction error, we need to modify the second hypothesis in Definition 2.2 as follows

(CL2b) cumulative claim amounts C_{ij}, $j = 1, 2, \ldots$ form a Markov chain. More precisely, there exist strictly positive factors $\lambda_1, \ldots, \lambda_{I-1}$ and variance parameters $\sigma_1^2, \ldots, \sigma_{I-1}^2$ such that for $1 \le i \le I$ and $2 \le j \le I$,

$$\begin{aligned} \mathrm{E}\left[C_{ij} | C_{i(j-1)}\right] &= \lambda_{(j-1)} C_{i(j-1)} \\ \mathrm{Var}\left[C_{ij} | C_{i(j-1)}\right] &= C_{i(j-1)} \sigma_{j-1}^2. \end{aligned}$$

Variance parameters are estimated by

$$\widehat{\sigma}_j^2 = \frac{1}{I-j-1} \sum_{i=0}^{I-j-1} C_{ij} \left(\frac{C_{i(j-1)}}{C_{ij}} - \widehat{\lambda}_j \right)^2, \qquad j = 1, \ldots, I. \tag{2.4.3}$$

Under the first hypothesis in Definition 2.2 and Equation (2.4.3), we have (see Buchwalder *et al.* (2006) and Barnett et Zehnwirth (2000))

- given $\mathcal{B}_j$, $\widehat{\sigma}_j$ is an unbiased estimator of σ_j; and
- $\widehat{\sigma}_j$ is an unbiased estimator of σ_j.

When there is not enough data ($I \le J$), $\widehat{\sigma}_{(J-1)}$ is obtained by extrapolation:

$$\frac{\widehat{\sigma}_{J-4}^2}{\widehat{\sigma}_{J-3}^2} = \frac{\widehat{\sigma}_{J-3}^2}{\widehat{\sigma}_{J-2}^2}.$$

To measure the quality of the claim reserve estimator, we want to consider the second moment. Hence, we need to calculate the mean square error of the prediction (MSEP).

Definition 2.3. *Let X be a random variable and $\mathcal{D}$ be a set of observations. Moreover, we assume that $\widehat{X}$ is a $\mathcal{D}$-measurable estimate for* $\mathrm{E}[X|\mathcal{D}]$. *The conditional mean square error of prediction for the predictor $\widehat{X}$ of X is*

$$\begin{aligned} MSEP_{X|\mathcal{D}}(\widehat{X}) &= \mathrm{E}\left[(\widehat{X} - X)^2 | \mathcal{D}\right] \\ &= \mathrm{Var}[X|\mathcal{D}] + (\widehat{X} - \mathrm{E}[X|\mathcal{D}])^2. \end{aligned}$$

The conditional mean square error of prediction is a measure of the quality of a predictor for the ultimate claim. Though in many cases, calculation of prediction distribution is not possible and numerical methods such as bootstrap or MCMC can be used to simulate the predictive distribution.

In the decomposition of the mean square error of prediction, the first term on the right hand side is the stochastic error which is the error of the stochastic model and can not be eliminated. The second term on the right hand side is the estimation error which shows the uncertainty in the parameters estimation and conditional expectation.

For the mean square error of prediction of the total paid amount for occurrence year i, we have

$$\mathrm{MSEP}_{C_{iI}|\mathcal{D}_j}(\widehat{C}_{iI}) = \mathrm{Var}[C_{iI}|\mathcal{D}_j] + (\widehat{C}_{iI} - \mathrm{E}[C_{iI}|\mathcal{D}_j])^2.$$

Therefore to calculate the mean square error of prediction, we should calculate both errors separately:

$$\begin{aligned}
\mathrm{Var}[C_{iI}|\mathcal{D}_j] &= \mathrm{Var}\left[C_{iI}|C_{i(I-i)}\right] \\
&= \mathrm{E}\left[\mathrm{Var}\left[C_{iI}|C_{i(I-1)}\right]|C_{i(I-i)}\right] + \mathrm{Var}\left[\mathrm{E}\left[C_{iI}|C_{i(I-1)}\right]|C_{i(I-i)}\right] \\
&= \sigma_{I-1}^2 C_{i(I-i)} \prod_{m=I-i}^{I-2} \lambda_m + \lambda_{I-1}^2 \mathrm{Var}\left[C_{i(I-1)}|C_{i(I-i)}\right].
\end{aligned}$$

After the iteration of this equation, we get

$$\begin{aligned}
\mathrm{Var}\left[C_{iI}|C_{i(I-i)}\right] &= C_{i(I-i)} \sum_{j=I-i}^{I-1} \prod_{n=j+1}^{I-1} \lambda_n^2 \sigma_j^2 \prod_{m=I-i}^{j-1} \lambda_m \\
&= \sum_{j=I-i}^{I-1} \prod_{n=j+1}^{I-1} \lambda_n^2 \sigma_j^2 \mathrm{E}\left[C_{ij}|C_{i(I-i)}\right] \\
&= \left(\mathrm{E}\left[C_{iI}|C_{i(I-i)}\right]\right)^2 \sum_{j=I-i}^{I-1} \frac{\sigma_j^2/\lambda_j^2}{\mathrm{E}\left[C_{ij}|C_{i(I-i)}\right]}.
\end{aligned}$$

This is the stochastic error of prediction of the total reserve amount for one occurrence year in the Mack's model.

Now we have the estimation of conditional estimation error as below:

$$\begin{aligned}(\widehat{C}_{iI} - \mathrm{E}[C_{iI}|\mathcal{D}_{\mathcal{I}}])^2 &= C^2_{i(I-i)}(\widehat{\lambda}_{I-i} \times \ldots \times \widehat{\lambda}_{I-1} - \lambda_{I-i} \times \ldots \times \lambda_{I-1})^2 \\ &= C^2_{i(I-i)}(\prod_{j=I-i}^{I-1} \widehat{\lambda}^2_j + \prod_{j=I-i}^{I-1} \lambda^2_j - 2\prod_{j=I-i}^{I-1} \widehat{\lambda}_j\lambda_j).\end{aligned}$$

This will give us an estimate for the accuracy of the factor estimate $\widehat{\lambda}_j$. This equation can not been calculated since $\lambda_{I-i}, \ldots, \lambda_{I-1}$ are unknown. Although $\widehat{\lambda}_{I-i}, \ldots, \widehat{\lambda}_{I-1}$ are known. We will determine the behaviour of $\widehat{\lambda}$ around λ and calculate the fluctuation amount in a fixed accident year. Then, the main problem is calculating the square of the estimated factor. We follow the conditional approach (see Wüthrich et Merz (2008)), in which the information in hand is

$$\mathcal{D}^o_{I-i} = \{C_{kj} \in D_I; j > I - i\} \subseteq D_I.$$

Definition 2.4.

i. hypothesis (CL1) in Definition 2.2 is valid here;

ii. there are constants $\lambda_j > 0$, $\sigma_j > 0$ and random variables $\epsilon_{i(j+1)}$ such that, for $i \in \{1, \ldots, I\}$ and $j \in \{1, \ldots, I-1\}$, we have

$$C_{i(j+1)} = \lambda_j C_{ij} + \sigma_j \sqrt{C_{ij}}\epsilon_{i(j+1)}$$

where $\mathcal{B}_1 = \{C_{ij} : i + j \leq I + 1, j = 1\}$, $\mathrm{E}\big[\epsilon_{i(j+1)}|\mathcal{B}_1\big] = 0$, $\mathrm{E}\Big[\epsilon^2_{i(j+1)}|\mathcal{B}_1\Big] = 1$ and $\mathrm{E}\big[C_{i(j+1)} > 0|\mathcal{B}_1\big] = 1$.

This condition on $\mathcal{B}_1$ is a guarantee that the cumulative amount will be positive. The reason is $\mathcal{B}_1$ has the first column has all positive claim amount and all other

columns are based on the first column. Then, we will have this condition on proceed calculations.
we re-sample the observations which will generate the estimated development factors given $\mathcal{D}_I$. Then by conditioning on the $\mathcal{D}_I$, we fix the denominator on the λ_j and the only variable will be $\sum C_{i(j+1)}$ in the nominator. Here we generate new observations $\tilde{C}_{i(j+1)}$ as:

$$\tilde{C}_{i(j+1)} = \lambda_j C_{ij} + \sigma_j \sqrt{C_{ij}} \tilde{\epsilon}_{i(j+1)},$$

and, by using these new observations, we estimate the new development factors:

$$\begin{aligned}\widehat{\lambda}_j &= \frac{\sum_{i=1}^{I-j-1} \tilde{C}_{i(j+1)}}{\sum_{i=1}^{I-j-1} C_{ij}} \\ &= \lambda_j + \frac{\sigma_j}{\sum_{i=1}^{I-j-1} C_{ij}} \sum_{i=0}^{I-j-1} \sqrt{C_{ij}} \tilde{\epsilon}_{i(j+1)}.\end{aligned}$$

It is necessary to know $\tilde{C}_{i(j+1)}$ and $\widehat{\lambda}_j$ are no longer constant values. The distribution of $\widehat{\lambda}_j$ given $\mathcal{B}_j$ is the same as λ_j.

Theorem 2.4.1. *Under the conditional probability measure* $\Pr[.|\mathcal{D}_{\mathcal{I}}]$, *we have*

i. $\mathrm{E}\left[\widehat{\lambda}_j\right] = \lambda_j$, *and*

ii. $\mathrm{E}\left[(\widehat{\lambda}_j)^2\right] = \lambda_j^2 + \frac{\sigma_j^2}{\sum_{i=1}^{I-j-1} C_{ij}}$.

Proof:

i.

$$\begin{aligned}
\mathrm{E}\left[\widehat{\lambda}_j\right] &= \mathrm{E}\left[\lambda_j + \frac{\sigma_j}{\sum_{i=1}^{I-j-1} C_{ij}} \sum_{i=1}^{I-j-1} \sqrt{C_{ij}}\tilde{\epsilon}_{i(j+1)}\right] \\
&= \mathrm{E}[\lambda_j] + \mathrm{E}\left[\frac{\sigma_j}{\sum_{i=1}^{I-j-1} C_{ij}} \sum_{i=1}^{I-j-1} \sqrt{C_{ij}}\tilde{\epsilon}_{i(j+1)}\right] \\
&= \lambda_j + 0 = \lambda_j.
\end{aligned}$$

ii.

$$\begin{aligned}
\mathrm{E}\left[\widehat{\lambda}_j^2\right] &= \mathrm{E}\left[(\lambda_j + \frac{\sigma_j}{\sum\limits_{i=1}^{I-j-1} C_{ij}} \sum_{i=1}^{I-j-1} \sqrt{C_{ij}}\tilde{\epsilon}_{i(j+1)})^2\right] \\
&= \mathrm{E}[\lambda_j^2] + \mathrm{E}\left[(\frac{\sigma_j}{\sum\limits_{i=1}^{I-j-1} C_{ij}} \sum_{i=1}^{I-j-1} \sqrt{C_{ij}}\tilde{\epsilon}_{i(j+1)})^2\right] \\
&+ 2\mathrm{E}\left[(\lambda_j)(\frac{\sigma_j}{\sum_{i=1}^{I-j-1} C_{ij}} \sum_{i=1}^{I-j-1} \sqrt{C_{ij}}\tilde{\epsilon}_{i(j+1)})\right] \\
&= \lambda_j^2 + \frac{\mathrm{E}\left[\sum\limits_{i=1}^{I-j-1} \sigma_j^2 C_{ij}\tilde{\epsilon}_{i(j+1)}^2\right] + 0}{\left(\sum\limits_{i=1}^{I-j-1} C_{ij}\right)^2} + 0 \\
&= \lambda_j^2 + \frac{\sigma_j^2}{\sum\limits_{i=1}^{I-j-1} C_{ij}}.
\end{aligned}$$

Therefore, the estimation error is

$$\begin{aligned}
\mathrm{E}\Big[\widehat{C}_{iI} - \mathrm{E}[C_{iI}|\mathcal{D}_\mathcal{I}]\Big]^2 &= C^2_{i(I-i)}\mathrm{E}\left[\left(\prod_{j=I-i}^{I-1}\widehat{\lambda}_j - \prod_{j=I-i}^{I-1}\lambda_j\right)^2\right] \\
&= C^2_{i(I-i)}\mathrm{Var}\left[\prod_{j=I-i}^{I-1}\widehat{\lambda}_j\right] \\
&= C^2_{i(I-i)}\left(\mathrm{E}\left[\prod_{j=I-i}^{I-1}\widehat{\lambda}_j\right]^2 - \mathrm{E}\left[\prod_{j=I-i}^{I-1}\widehat{\lambda}_j\right]^2\right) \\
&= C^2_{i(I-i)}\left(\prod_{j=I-i}^{I-1}\mathrm{E}\Big[\widehat{\lambda}_j\Big]^2 - \prod_{j=I-i}^{I-1}\widehat{\lambda}^2_j\right) \\
&= C^2_{i(I-i)}\left(\prod_{j=I-i}^{I-1}\left(\lambda^2_j + \frac{\sigma^2_j}{\sum_{i=1}^{I-j-1} C_{ij}}\right) - \prod_{j=I-i}^{I-1}\widehat{\lambda}^2_j\right),
\end{aligned}$$

and the mean square error of prediction, which is the sum of the process error and the estimation error, is:

$$\begin{aligned}
\widehat{\mathrm{MSEP}}_{C_{iI}|\mathcal{D}_j}(\widehat{C}_{iI}) &= \mathrm{E}\big[C_{iI}|C_{i(I-i)}\big]^2\sum_{j=I-i}^{I-1}\frac{\frac{\sigma^2_j}{\lambda^2_j}}{\mathrm{E}\big[C_{ij}|C_{i(I-i)}\big]} \qquad (2.4.4) \\
&+ C^2_{i(I-i)}\left(\prod_{j=I-i}^{I-1}\left(\lambda^2_j + \frac{\sigma^2_j}{\sum_{i=1}^{I-j-1} C_{ij}}\right) - \prod_{j=I-i}^{I-1}\widehat{\lambda}^2_j\right),
\end{aligned}$$

where we replace the unknown values by Equations (2.4.2) and (2.4.3)

$$\widehat{MSEP}_{C_{iI}|\mathcal{D}_j}(\widehat{C}_{iI}) = (C_{iI})^2\left(\sum_{j=I-i}^{I-1}\frac{\widehat{\sigma}^2_j/\widehat{\lambda}^2_j}{\widehat{C}_{iI}} + \prod_{j=I-i}^{I-1}\left(\frac{\widehat{\sigma}^2_j/\widehat{\lambda}^2_j}{\widehat{C}_{iI}} + 1\right) - 1\right).$$

For the total mean square error, we sum up mean square error of rows and we take in to account the correlation between rows

$$\widehat{MSEP}_{\sum_i C_{iI}}\left(\sum_{i=1}^{I}\widehat{C}_{iI}\right)=\sum_{i=1}^{I}\widehat{MSEP}_{C_{iI}|\mathcal{D}_j}(\widehat{C}_{iI})$$

$$+2\sum_{1\leq i<k\leq I}C_{i(I-i)}\hat{C}_{k(I-i)}\left(\prod_{j=I-i}^{I-1}\left(\widehat{\lambda}_j^2+\frac{\widehat{\sigma}_j^2}{\sum_{i=0}^{I-j-1}\widehat{C}_{ij}}\right)-\prod_{j=I-i}^{I-1}\widehat{\lambda}_j^2\right).$$

For detailed information, see Merz et Wüthrich (2010).

Example 2.2. *We use the same database (see Taylor et Ashe (1983)) which has been presented in Table 2.3. As a first step, we evaluate the total reserve amount with the stochastic Chain-Ladder model. The full matrix (observed and predicted*

Table 2.6 Run-off triangle (observed and predicted cumulative paid amounts) with the stochastic Chain-Ladder model.

i	1	2	3	4	5	6	7	8	9	10
1	357 848	1 124 788	1 735 330	2 218 270	2 745 596	3 319 994	3 466 336	3 606 286	3 833 515	3 901 463
2	352 118	1 236 139	2 170 033	3 353 322	3 799 067	4 120 063	4 647 867	4 914 039	5 339 085	5 433 719
3	290 507	1 292 306	2 218 525	3 235 179	3 985 995	4 132 918	4 628 910	4 909 315	5 285 148	5 378 826
4	310 608	1 418 858	2 195 047	3 757 447	4 029 929	4 381 982	4 588 268	4 835 458	5 205 637	5 297 906
5	443 160	1 136 350	2 128 333	2 897 821	3 402 672	3 873 311	4 207 459	4 434 133	4 773 589	4 858 200
6	396 132	1 333 217	2 180 715	2 985 752	3 691 712	4 074 999	4 426 546	4 665 023	5 022 155	5 111 171
7	440 832	1 288 463	2 419 861	3 483 130	4 088 678	4 513 179	4 902 528	5 166 649	5 562 182	5 660 771
8	359 480	1 421 128	2 864 498	4 174 756	4 900 545	5 409 337	5 875 997	6 192 562	6 666 635	6 784 799
9	376 686	1 363 294	2 382 128	3 471 744	4 075 313	4 498 426	4 886 502	5 149 760	5 544 000	5 642 266
10	344 014	1 200 818	2 098 228	3 057 984	3 589 620	3 962 307	4 304 132	4 536 015	4 883 270	4 969 825

cumulative paid amounts) is presented in Table 2.6. The total reserve amount is 18 680 856$ *and the mean square error of prediction is* 2 441 364.

As a second step we are going to manipulate one observation and calculate the total reserve amount again. We multiply the $X_{2,1}$ *by* 10 *and we reevaluate the total reserve amount. The new full matrix is presented in Table 2.7. The total amount of the reserve is now* 13 064 239$ *and the mean square error of prediction* *is* 3 108 085. *The significant difference between amounts shows that the stochastic*

Table 2.7 Run-off triangle in presence of an outlier with the stochastic Chain-Ladder model.

i	1	2	3	4	5	6	7	8	9	10
1	357 848	1 124 788	1 735 330	2 218 270	2 745 596	3 319 994	3 466 336	3 606 286	3 833 515	3 901 463
2	3 521 180	4 405 201	5 339 095	6 522 384	6 968 129	7 289 125	7 816 929	8 083 101	8 508 147	8 658 952
3	290 507	1 292 306	2 218 525	3 235 179	3 985 995	4 132 918	4 628 910	4 909 315	5 183 258	5 275 130
4	310 608	1 418 858	2 195 047	3 757 447	4 029 929	4 381 982	4 588 268	4 786 228	5 053 302	5 142 871
5	443 160	1 136 350	2 128 333	2 897 821	3 402 672	3 873 311	4 152 087	4 331 228	4 572 913	4 653 967
6	396 132	1 333 217	2 180 715	2 985 752	3 691 712	4 017 520	4 306 675	4 492 486	4 743 169	4 827 240
7	440 832	1 288 463	2 419 861	3 483 130	3 999 904	4 352 911	4 666 205	4 867 528	5 139 139	5 230 229
8	359 480	1 421 128	2 864 498	3 946 820	4 532 389	4 932 390	5 287 392	5 515 515	5 823 284	5 926 500
9	376 686	1 363 294	2 141 541	2 950 702	3 388 482	3 687 528	3 952 933	4 123 481	4 353 574	4 430 740
10	344 014	782 855	1 229 755	1 694 405	1 945 795	2 117 519	2 269 925	2 367 860	2 499 988	2 544 300

Chain-Ladder method in presence of an outlier gives not a very accurate result for the amount of reserve.

2.5 Influence Function

2.5.1 Influence Function for the Generalized Linear Model for Reserves

In this section we consider the influence function for the parameters' estimators $\widehat{\boldsymbol{\beta}}$ in the GLM for reserves. The non-robust version of $\widehat{\boldsymbol{\beta}}$ is a maximum likelihood estimator which is the solution of the following equation:

$$\sum_{i=1}^{I}\sum_{j=1}^{I-i-1}\frac{(x_{ij}-\mu_{ij})}{\mathrm{V}[\mu_{ij}]}\mu'_{ij}=0 \tag{2.5.1}$$

For more details refer to Appendix B.3. We have seen the definition of μ_{ij} and $\mathrm{V}[\mu_{ij}]$ in section 2.3 . Also, $\mu'_{ij}=\frac{\partial}{\partial\beta}\mu_{ij}$. As illustrated in Equation (1.3.6), the influence function for an M-estimators is proportional to $\psi(x_{ij},\mu_{ij})$. Therefore the influence function is proportional to $\left(\left(x_{ij}-\mu_{ij}\right)/\mathrm{V}[\mu_{ij}]\right)\mu'_{ij}$ and it is unbounded:

large deviations of the x_{ij} from its mean will have a large effect of the estimator. This fact has been illustrated in Example 2.1 where an over-dispersion Poisson model has been applied to a dataset.

2.5.2 Influence Function for the Stochastic Chain-Ladder Model

As we mentioned before, the Mack's model is a stochastic version of the Chain-Ladder model. In order to evaluate the behaviour of this model, we need to calculate its influence function. Since the stochastic Chain-Ladder model is a distribution free model, we need to make some assumptions before proceeding:

i. incremental amounts X_{ij} has a multiplicative structure $X_{ij} = \alpha_i \psi_j$; and

ii. X_{ij} are independent random variables with a distribution belonging to the exponential family.

We assume a logarithmic link function and we obtain

$$\mathrm{E}[X_{ij}] = \exp\left(\ln(\alpha_i) + \ln(\psi_j)\right)$$

If we assume that $X_{ij} \sim \text{Poisson}(\alpha_i \psi_j)$, then the evaluated reserve will be identical to the one obtained from the stochastic Chain-Ladder model. We impose the condition $\sum_{j=1}^{I} \psi_j = 1$ and the α_i and ψ_j are obtained by maximum likelihood method. This corresponds to a multiplicative generalized linear model with (over-dispersed) Poisson errors and logarithmic link function. The parameters will be the same as the estimated parameters by the stochastic Chain-Ladder model.

In order to estimate parameters, the equality of this logarithmic generalized linear model with the marginal totals method, has been used (see Kaas *et al.* (2009)). The sum of the i^{th} row in the run-off triangle is

$$RS_i = \sum_{j=1}^{I-i+1} X_{ij},$$

which is equal to $\sum_{j=1}^{I-i+1} \widehat{\alpha}_i \widehat{\psi}_j$, and sum of the j^{th} column is

$$CS_j = \sum_{i=1}^{I-j+1} X_{ij},$$

which is equal to $\sum_{i=1}^{I-j+1} \widehat{\alpha}_i \widehat{\psi}_j$. In the first row of the run-off triangle, we have

$$RS_1 = \sum_{j=1}^{I} \widehat{\alpha}_1 \widehat{\psi}_j = \widehat{\alpha}_1,$$

since $\sum_{j=1}^{I} \hat{\psi}_j = 1$. In the last column, we have only one observation, so

$$\begin{aligned} CS_I &= \widehat{\alpha}_1 \widehat{\psi}_I \\ \rightarrow \widehat{\psi}_I &= \frac{CS_I}{\widehat{\alpha}_1}. \end{aligned}$$

Afterwards, we evaluate

$$\widehat{\alpha}_2 = \frac{RS_2}{1 - \widehat{\psi}_I}$$

and

$$\widehat{\psi}_{I-1} = \frac{CS_{I-1}}{\widehat{\alpha}_1 + \widehat{\alpha}_2}.$$

By repeating these steps, we find estimate all parameters and we obtain the general form

$$\begin{aligned} \widehat{\alpha}_l &= \frac{RS_l}{1 - \sum_{j=I-l+2}^{I} \widehat{\psi}_j} \\ \widehat{\psi}_{I-l+1} &= \frac{CS_{I-l+1}}{\sum_{i=1}^{l} \widehat{\alpha}_i}, \end{aligned}$$

for $l = 1, \ldots, I$. Now, we can see these estimators as functional in order to obtain their influence functions.

Theorem 2.5.1. *For any distribution F and $\{X_{11}, \ldots, X_{I1}\} \sim F$, the estimators of the stochastic Chain-Ladder model have the following functional representation*

$$T_{\alpha_l}(F) = \frac{\mathrm{E}\left[\sum_{j=1}^{I-l+1} X_{lj}\right]_F}{1 - \sum_{j=I-l+2}^{I} T_{\psi_j}(F)}, \qquad \forall l = 1, \ldots, I$$

and

$$T_{\psi_{I-l+1}}(F) = \frac{\mathrm{E}\left[\sum_{i=1}^{I} X_{i(I-l+1)}\right]_F}{\sum_{i=1}^{l} T_{\alpha_i}(F)}, \qquad \forall l = 1, \ldots, I.$$

while $\mathrm{E}[\cdot]_F$ is the expected value of a random variable following distribution F. Moreover, these functionals are Fisher consistent.

Proof. We show these functionals are related to the stochastic Chain-Ladder's estimators. Suppose $\{x_{11}, \ldots, x_{I1}\}$ is a random sample with empirical distribution F_n. We know $\mathrm{E}[X_{ij}]_{F_n} = x_{ij}$ and we assume empty summation is equal to zero, therefore

$$\begin{aligned}
T_{\alpha_1}(F_n) &= \mathrm{E}\left[\sum_{j=1}^{I} X_{1j}\right]_{F_n} \\
&= \sum_{j=1}^{I} \mathrm{E}[X_{1j}]_{F_n} \\
&= \sum_{j=1}^{I} x_{1j} \\
&= \widehat{\alpha}_1 \\
T_{\beta_I}(F_n) &= \frac{\mathrm{E}[X_{1I}]_{F_n}}{T_{\alpha_1}(F_n)} \\
&= \frac{\mathrm{E}[X_{1I}]_{F_n}}{\widehat{\alpha}_1} \\
&= \frac{x_{1I}}{\widehat{\alpha}_1} \\
&= \widehat{\psi}_I
\end{aligned}$$

We assume this is valid for $T_{\alpha_1}(F_n), \ldots, T_{\alpha_{l-1}}(F_n)$ and $T_{\psi_I}(F_n), \ldots, T_{\psi_{I-l+2}}(F_n)$ where $l \neq I$. Then, we have

$$
\begin{aligned}
T_{\alpha_l}(F_n) &= \frac{\mathrm{E}\Big[\sum_{j=1}^{I-l+1} X_{lj}\Big]_{F_n}}{1-\sum_{j=I-l+2}^{I} T_{\psi}(F_n)} \\
&= \frac{\sum_{j=1}^{I-l+1} \mathrm{E}[X_{lj}]_{F_n}}{1-\sum_{j=I-l+2}^{I} \widehat{\psi}_j} \\
&= \frac{\sum_{j=1}^{I-l+1} x_{lj}}{1-\sum_{j=I-l+2}^{I} \widehat{\psi}_j} \\
&= \widehat{\alpha}_l
\end{aligned}
$$

$$
\begin{aligned}
T_{\beta_{I-l+1}}(F_n) &= \frac{\mathrm{E}\Big[\sum_{i=1}^{l} X_{i,I-l+1}\Big]_{F_n}}{\sum_{i=1}^{l} T_{\alpha_i}(F_n)} \\
&= \frac{\sum_{i=1}^{l} \mathrm{E}[X_{i,I-l+1}]_{F_n}}{\sum_{i=1}^{l} \widehat{\alpha}_i} \\
&= \frac{\sum_{i=1}^{l} x_{i,I-l+1}}{\sum_{i=1}^{l} \widehat{\alpha}_i} \\
&= \widehat{\psi}_{I-l+1}.
\end{aligned}
$$

By using the same induction principle, we prove the Fisher consistency of these functionals. Suppose $X_{ij} \sim F_{\alpha,\psi}$, a parametric distribution.Then,

$$\begin{aligned}
T_{\alpha_1}(F_{\alpha,\psi}) &= \mathrm{E}\left[\sum_{j=1}^{I} X_{1j}\right]_{F_{\alpha,\psi}} \\
&= \sum_{j=1}^{I} \mathrm{E}[X_{1j}]_{F_{\alpha,\psi}} \\
&= \sum_{j=1}^{I} \alpha_1 \psi_j \\
&= \alpha_1 \\
T_{\psi_I}(F_{\alpha,\psi}) &= \frac{\mathrm{E}[X_{1I}]_{F_{\alpha,\psi}}}{T_{\alpha_1}(F_{\alpha,\psi})} \\
&= \frac{\alpha_1 \psi_I}{\alpha_1} \\
&= \psi_I .
\end{aligned}$$

We assume now this is valid for $T_{\alpha_1}(F_n), \ldots, T_{\alpha_{l-1}}(F_n)$ and $T_{\psi_I}(F_n), \ldots, T_{\psi_{I-l+2}}(F_n)$ where $l \neq I$. Thus,

$$\begin{aligned}
T_{\alpha_l}(F_{\alpha,\psi}) &= \frac{\mathrm{E}\left[\sum_{j=1}^{I-l+1} X_{lj}\right]_{F_{\alpha,\psi}}}{1 - \sum_{j=I-l+2}^{l} T_{\psi_j}(F_n)} \\
&= \frac{\sum_{j=1}^{I-l+1} \mathrm{E}[X_{lj}]_{F_{\alpha,\psi}}}{1 - \sum_{j=I-l+2}^{l} \psi_j} \\
&= \frac{\sum_{j=1}^{I-l+1} \alpha_l \psi_j}{1 - \sum_{j=I-l+2}^{l} \psi_j} \\
&= \alpha_l \frac{\sum_{j=1}^{I-l+1} \psi_j}{\sum_{j=1}^{I-l+1} \psi_j} \\
&= \alpha_l
\end{aligned}$$

and

$$
\begin{aligned}
T_{\psi_{I-l+1}}(F_{\alpha,\psi}) &= \frac{\mathrm{E}\left[\sum_{i=1}^{l} X_{i,I-l+1}\right]_{F_{\alpha,\psi}}}{T_{\alpha_i}(F_n)} \\
&= \frac{\sum_{i=1}^{l} \mathrm{E}[X_{i,I-l+1}]_{F_{\alpha,\beta}}}{\sum_{i=1}^{l} \alpha_i} \\
&= \frac{\sum_{i=1}^{l} \alpha_i \psi I - l + 1}{\sum_{i=1}^{l} \alpha_i} \\
&= \psi_{I-l+1} \frac{\sum_{i=1}^{l} \alpha_i}{\sum_{i=1}^{l} \alpha_i} \\
&= \psi_{I-l+1}.
\end{aligned}
$$

□

We use these functionals to define the following functionals which correspond to the evaluation of future claims X_{ij}

$$
\begin{aligned}
T_{\alpha_i\psi_j} &= T_{\alpha_i} T_{\psi_j} \qquad \forall i = 1, \ldots, \qquad \forall j = 1, \ldots, I \\
T_{\sum_{j=I-i+2}^{I} \alpha_i \psi_j} &= \sum_{j=I-i+2}^{I} T_{\alpha_i \psi_j} \qquad \forall i = 1, \ldots, I \\
T_{\sum_{i=2}^{I} \sum_{j=I-i+2}^{I} \alpha_i \psi_j} &= \sum_{i=2}^{I} \sum_{j=I-i+2}^{I} T_{\alpha_i \psi_j}.
\end{aligned}
$$

Since we know that $\mathrm{E}[X_{ij}] = \alpha_i \psi_j$, then

$$
\mathrm{E}[R_i] = \sum_{j=I-i+2}^{I} \alpha_i \psi_j
$$

and

$$
\mathrm{E}[R] = \sum_{i=2}^{I} \sum_{j=I-i+2}^{I} \alpha_i \psi_j.
$$

Here we are not interested in calculating the amount of reserve but we want to evaluate the influence function. For this purpose, we suppose $\{X_{11}, \ldots, X_{I1}\} \sim F$ and F_{ij} is the distribution of X_{ij} which is Poisson$(\alpha_i \psi_j)$.

To define a contaminated dataset, we assume that a small fraction $\epsilon > 0$ of the data is replaced by the value z and the remaining fraction $(1 - \epsilon)$ comes from the original distribution F.

Theorem 2.5.2. *Let $\epsilon > 0$ and $0 \leq p \leq q \leq I$ and define the distribution as below*

$$F_{p,q,\epsilon,x} = \begin{cases} X_{ij} \sim F_{ij} & \forall (i,j) \neq (p,q) \\ X_{p,q} \sim (1-\epsilon)F_{p,q} + \epsilon \Delta_z, & \textit{elsewhere.} \end{cases} \tag{2.5.2}$$

Then, the influence functions of the functionals $T_{\alpha_1}, \ldots, T_{\alpha_I}$ and $T_{\psi_1}, \ldots, T_{\psi_I}$ are

$$\begin{aligned} IF([z,p,q];T_{\alpha_l},F) &= \lim_{\epsilon \to 0} \frac{T_{\alpha_l}(F_{p,q,\epsilon,z}) - T_{\alpha_l}(F)}{\epsilon} \\ &= \begin{cases} \frac{\alpha_l [\sum_{i=I-l+2}^{I} IF([z,p,q];T_{\psi_i},F)]}{1-\sum_{i=I-l+2}^{I} \psi_i} & p \neq l \\ \frac{z-\alpha_l \psi_q + \alpha_l [\sum_{i=I-l+2}^{I} IF([z,l,q];T_{\psi_i},F)]}{1-\sum_{i=I-l+2}^{I} \psi_i} & p = l \end{cases} \end{aligned}$$

$$IF([z,p,q];T_{\psi_{I-l+1}},F) = \lim_{\epsilon \to 0} \frac{T_{\psi_{I-l+1}}(F_{p,q,\epsilon,z}) - T_{\psi_{I-l+1}}(F)}{\epsilon} \tag{2.5.3}$$

$$= \begin{cases} -\frac{\psi_{I-l+1} [\sum_{i=I}^{I} IF([z,p,q];T_{\alpha_i},F)]}{\sum_{i=1}^{l} \alpha_i} & q \neq I-l+1 \\ \frac{z-\alpha_p \psi_{I-l+1} - \psi_{I-l+1} [\sum_{i=l}^{l} IF([z,p,I-l+1];T_{\alpha_i},F)]}{\sum_{i=1}^{l} \alpha_i} & q = I-l+1 \end{cases}$$

Proof. First, we need to define the estimators' functional forms corresponding to the contaminated distribution $F_{p,q,\epsilon,z}$ which have been introduced in Equation (2.5.2). The functional form of $\widehat{\alpha}_1$ is

$$T_{\alpha_1}(F_{p,q,\epsilon,z}) = \begin{cases} \mathrm{E}\Big[\sum_{j=1,j\neq q}^{I} X_{1j} + (1-\epsilon)X_{1q} + \epsilon z\Big]_F & p=1 \\ \mathrm{E}\Big[\sum_{j=1}^{I} X_{1j}\Big]_F & p\neq 1 \end{cases}$$

$$= \begin{cases} T_{\alpha_1}(F) - \epsilon\mathrm{E}[X_{1q}]_F + \epsilon z & p=1 \\ T_{\alpha_1}(F) & p\neq 1 \end{cases}$$

$$= \begin{cases} T_{\alpha_1}(F) - \epsilon T_{\alpha_1}(F)T_{\psi_i}(F) + \epsilon z & p=1 \\ T_{\alpha_1}(F) & p\neq 1 \end{cases}$$

and the functional form of $\widehat{\psi_I}$ is

$$T_{\beta_I}(F_{p,q,\epsilon,z}) = \begin{cases} \frac{\mathrm{E}[(1-\epsilon)X_{1I}+\epsilon z]_F}{T_{\alpha_1}(F_{p,q,\epsilon,z})}, & p=1, q=I \\ \frac{\mathrm{E}[X_{1I}]_F}{T_{\alpha_1}(F_{p,q,\epsilon,z})}, & p=1, q\neq I \\ \frac{\mathrm{E}[X_{1I}]_F}{T_{\alpha_1}(F_{p,q,\epsilon,z})}, & p\neq 1 \end{cases}$$

$$= \begin{cases} \frac{(1-\epsilon)\mathrm{E}[X_{1I}]_F+\epsilon z}{T_{\alpha_1}(F)+\epsilon(z-T_{\alpha_l}(F)T_{\psi_q}(F))} & p=1, q=I \\ \frac{\mathrm{E}[X_{1I}]_F}{T_{\alpha_1}(F)+\epsilon(z-T_{\alpha_1}(F)T_{\psi_q}(F))} & p=1, q\neq I \\ \frac{\mathrm{E}[X_{1I}]_F}{T_{\alpha_1}(F)} & p\neq 1 \end{cases}$$

$$= \begin{cases} \frac{(1-\epsilon)T_{\alpha_1}(F)T_{\psi_I}(F)+\epsilon z}{T_{\alpha_1}(F)+\epsilon(z-T_{\alpha_1}(F)T_{\psi_q}(F))} & p=1, q=I \\ \frac{T_{\alpha_1}(F)T_{\psi_I}(F)}{T_{\alpha_1}(F)+\epsilon(z-T_{\alpha_1}(F)T_{\psi_q}(F))} & p=1, q\neq I \\ T_{\psi_I}(F) & p\neq 1. \end{cases}$$

We suppose it is valid for $T_{\alpha_1}(F_{p,q,\epsilon,z}), \ldots, T_{\alpha_{l-1}}(F_{p,q,\epsilon,z})$ and for $T_{\psi_{I-l+2}}(F_{p,q,\epsilon,z})$

$, \ldots, T_{\psi_I}(F_{p,q,\epsilon,z})$, where $l < I$. Then we find $T_{\alpha_l}(F_{p,q,\epsilon,z})$

$$T_{\alpha_l}(F_{p,q,\epsilon,z}) = \begin{cases} \frac{\mathrm{E}\left[\sum_{j=1,j\neq q}^{I-l+1} X_{lj}+(1-\epsilon)X_{lq}+\epsilon z\right]_F}{1-\sum_{j=I-l+2}^{I} T_{\psi_I}(F_{p,q,\epsilon,z})} & p=l \\ \frac{\mathrm{E}\left[\sum_{j=1}^{I-l+1} X_{lj}\right]_F}{1-\sum_{j=I-l+2}^{I} T_{\psi_j}(F_{p,q,\epsilon,z})} & p\neq l \end{cases}$$

$$= \begin{cases} \frac{\sum_{j=1,j\neq q}^{I-l+1} \mathrm{E}\left[X_{lj}\right]_F+(1-\epsilon)\mathrm{E}\left[X_{lq}\right]_F+\epsilon z}{1-\sum_{j=I-l+2}^{I} T_{\psi_j}(F_{p,q,\epsilon,z})} & p=l \\ \frac{\sum_{j=1}^{I-l+1} \mathrm{E}\left[X_{lj}\right]}{1-\sum_{j=I-l+2}^{I} T_{\psi_j}(F_{p,q,\epsilon,z})} & p\neq l \end{cases}$$

$$= \begin{cases} \frac{\sum_{j=1,j\neq q}^{I-l+1} T_{\alpha_l}(F)T_{\psi_j}(F)+(1-\epsilon)T_{\alpha_l}(F)T_{\psi_q}(F)+\epsilon z}{1-\sum_{j=I-l+2}^{I} T_{\psi_j}(F_{p,q,\epsilon,z})} & p=l \\ \frac{\sum_{j=1}^{I-l+1} T_{\alpha_l}(F)T_{\psi_j}(F)}{1-\sum_{j=I-l+2}^{I} T_{\psi_j}(F_{p,q,\epsilon,z})} & p\neq l \end{cases}$$

$$= \begin{cases} \frac{\sum_{j=1,j\neq q}^{I-l+1} T_{\alpha_l}(F)T_{\psi_j}(F)+\epsilon(z-T_{\alpha_l}(F)T_{\psi_q}(F))}{1-\sum_{j=I-l+2}^{I} T_{\psi_j}(F_{p,q,\epsilon,z})} & p=l \\ \frac{\sum_{j=1}^{I-l+1} T_{\alpha_l}(F)T_{\psi_j}(F)}{1-\sum_{j=I-l+2}^{I} T_{\psi_j}(F_{p,q,\epsilon,z})} & p\neq l \end{cases}$$

and $T_{\psi_{I-l+1}}(F_{p,q,\epsilon,z})$

$$T_{\psi_{I-l+1}}(F_{p,q,\epsilon,z}) = \begin{cases} \frac{\mathrm{E}\left[\sum_{i=1,i\neq p}^{l} X_{i,I-l+1}+(1-\epsilon)X_{p,I-l+1}+\epsilon z\right]_F}{\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})} & q=I-l+1 \\ \frac{\mathrm{E}\left[\sum_{i=1}^{l} X_{i,I-l+1}\right]_F}{\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})} & q\neq I-l+1 \end{cases}$$

$$= \begin{cases} \frac{\sum_{i=1,i\neq p}^{l} \mathrm{E}\left[X_{i,I-l+1}\right]_F+(1-\epsilon)\mathrm{E}\left[X_{p,I-l+1}\right]_F+\epsilon z}{\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})} & q=I-l+1 \\ \frac{\sum_{i=1}^{l} \mathrm{E}\left[X_{i,I-l+1}\right]_F}{\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})} & q\neq I-l+1 \end{cases}$$

$$= \begin{cases} \frac{\sum_{i=1,i\neq p}^{l} T_{\alpha_i}(F)T_{\psi_{I-l+1}}(F)+(1-\epsilon)T_{\alpha_p}(F)T_{\psi_{I-l+1}}(F)+\epsilon z}{\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})} & q=I-l+1 \\ \frac{\sum_{i=1}^{l} T_{\alpha_i}(F)T_{\psi_{I-l+1}}(F)}{\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})} & q\neq I-l+1 \end{cases}$$

$$= \begin{cases} \frac{\sum_{i=1}^{l} T_{\alpha_i}(F)T_{\psi_{I-l+1}}(F)+\epsilon(z-T_{\alpha_p}(F)T_{\psi_{I-l+1}}(F))}{\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})} & q=I-l+1 \\ \frac{\sum_{i=1}^{l} T_{\alpha_i}(F)T_{\psi_{I-l+1}}(F)}{\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})} & q\neq I-l+1. \end{cases}$$

Second, we calculate the influence function. It is enough to take the derivative of the functionals with respect to ϵ. We calculate the influence function for T_{α_1} and

T_{ψ_I} then calculate the influence function for the general case of T_{α_l} and $T_{\psi_{I-l+1}}$:

$$\begin{aligned} IF([z,1,q];T_{\alpha_1},F) =& \frac{d}{d\epsilon}T_{\alpha_1}(F_{p,q,\epsilon,z})|_{\epsilon=0} \\ =& \frac{d}{d\epsilon}(T_{\alpha_1}(F) - \epsilon T_{\alpha_1}(F)T_{\psi_q}(F) + \epsilon z) \\ =& z - T_{\alpha_1}(F)T_{\psi_q}(F) \end{aligned}$$

$$\begin{aligned} IF([z,p,q];T_{\alpha_1},F) =& \frac{d}{d\epsilon}T_{\alpha_1}(F) \\ =& 0 \qquad p > 1. \end{aligned}$$

It is the influence function for T_{α_1}, then with the same approach, the influence function for T_{ψ_I} is

$$IF([z,1,q];T_{\psi_I},F) = \frac{d}{d\epsilon}T_{\psi_I}(F_{p,q,\epsilon,z})|_{\epsilon=0}$$

$$= \begin{cases} \frac{-T_{\alpha_l}(F)T_{\psi_I}(F)(z-T_{\alpha_l}(F)T_{\psi_q}(F))}{[T_{\alpha_l}(F)+\epsilon(z-T_{\alpha_l}(F)T_{\psi_q}(F))]^2}|_{\epsilon=0} & q \neq n \\ \frac{(-T_{\alpha_l}(F)T_{\psi_I}(F)+z)[T_{\alpha_l}(F)+\epsilon(z-T_{\alpha_l}(F)T_{\psi_I}(F))]-(z-T_{\alpha_l}(F)T_{\psi_I}(F))[(1-\epsilon)T_{\alpha_l}(F)T_{\psi_I}(F)+\epsilon z]}{[T_{\alpha_l}(F)+\epsilon(z-T_{\alpha_l}(F)T_{\psi_I}(F))]^2}|_{\epsilon=0} & q = n \end{cases}$$

$$= \begin{cases} \frac{-T_{\psi_I}(F)(z-T_{\alpha_l}(F)T_{\psi_q}(F))}{T_{\alpha_l}(F)} & q \neq n \\ \frac{(1-T_{\psi_I}(F))(z-T_{\alpha_l}(F)T_{\psi_I}(F))}{T_{\alpha_l}(F)} & q = n \end{cases}$$

$$IF([z,p,q];T_{\psi_I},F) = 0 \qquad p > 1.$$

It is obvious that the influence function for these two functionals are unbounded.

Here we calculate the influence function for general cases

$$IF([z,p,q];T_{\alpha_l},F) = \frac{d}{d\epsilon}T_{\alpha_l}(F_{p,q,\epsilon,z})|_{\epsilon=0}$$

$$= \begin{cases} \frac{d}{d\epsilon}\frac{\sum_{j=1}^{I-l+1}T_{\alpha_l}(F)T_{\psi_j}(F)+\epsilon(z-T_{\alpha_l}(F)T_{\psi_q}(F))}{1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F_{p,q,\epsilon,z})}|_{\epsilon=0} & p=l \\ \frac{d}{d\epsilon}\frac{\sum_{j=1}^{I-l+1}T_{\alpha_l}(F)T_{\psi_j}(F)}{1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F_{p,q,\epsilon,z})}|_{\epsilon=0} & p\neq l \end{cases}$$

$$= \begin{cases} \frac{(z-T_{\alpha_l}(F)T_{\psi_q}(F))[1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F_{p,q,\epsilon,z})]}{[1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F_{p,q,\epsilon,z})]^2} \\ \quad +\frac{\sum_{j=I-l+2}^{I}IF([p,q];T_{\psi_j},F)[\sum_{j=1}^{I-l+1}T_{\alpha_l}(F)T_{\psi_j}(F)+\epsilon(z-T_{\alpha_l}(F)T_{\psi_q}(F))]}{[1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F_{p,q,\epsilon,z})]^2}|_{\epsilon=0} & p=l \\ \frac{\sum_{j=I-l+2}^{I}IF([p,q];T_{\psi_j},F)[\sum_{j=1}^{I-l+1}T_{\alpha_l}(F)T_{\psi_j}(F)]}{[1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F_{p,q,\epsilon,z})]^2}|_{\epsilon=0} & p\neq l \end{cases}$$

$$= \begin{cases} \frac{(z-T_{\alpha_l}(F)T_{\psi_q}(F))[1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F)]+\sum_{j=I-l+2}^{I}IF([p,q];T_{\psi_j},F)[\sum_{j=1}^{I-l+1}T_{\alpha_l}(F)T_{\psi_j}(F)]}{[1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F)]^2} & p=l \\ \frac{\sum_{j=I-l+2}^{I}IF([p,q];T_{\psi_j},F)[\sum_{j=1}^{I-l+1}T_{\alpha_l}(F)T_{\psi_j}(F)]}{[1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F)]^2} & p\neq l \end{cases}$$

$$= \begin{cases} \frac{[1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F)][(z-T_{\alpha_l}(F)T_{\psi_q}(F))+T_{\alpha_l}(F)\sum_{j=I-l+2}^{I}IF([p,q];T_{\psi_j},F)]}{[1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F)]^2} & p=l \\ \frac{T_{\alpha_l}(F)\sum_{j=I-l+2}^{I}IF([p,q];T_{\psi_j},F)}{1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F)} & p\neq l \end{cases}$$

$$= \begin{cases} \frac{(z-T_{\alpha_l}(F)T_{\psi_q}(F))+T_{\alpha_l}(F)\sum_{j=I-l+2}^{I}IF([p,q];T_{\psi_j},F)}{1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F)} & p=l \\ \frac{T_{\alpha_l}(F)\sum_{j=I-l+2}^{I}IF([p,q];T_{\psi_j},F)}{1-\sum_{j=I-l+2}^{I}T_{\psi_j}(F)} & p\neq l \end{cases}$$

and

$$IF([z,p,q];T_{\psi_{I-l+1}},F) = \frac{d}{d\epsilon}T_{\psi_{I-l+1}}(F_{p,q,\epsilon,z})|_{\epsilon=0}$$

$$= \begin{cases} \frac{d}{d\epsilon}\frac{\sum_{i=1}^{l} T_{\alpha_i}(F)T_{\psi_{j-l+1}}(F)+\epsilon(z-T_{\alpha_p}(F)T_{\psi_{I-l+1}}(F))}{\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})}|_{\epsilon=0} & q = I-l+1 \\ \frac{d}{d\epsilon}\frac{\sum_{i=1}^{l} T_{\alpha_i}(F)T_{\psi_{j-l+1}}(F)}{\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})}|_{\epsilon=0} & q \neq I-l+1 \end{cases}$$

$$= \begin{cases} \frac{(z-T_{\alpha_p}(F)T_{\psi_{I-l+1}}(F))[\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})]}{[\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})]^2} \\ \quad + \frac{\sum_{i=1}^{l} IF([z,p,q];T_{\alpha_i},F)[\sum_{i=1}^{l} T_{\alpha_i}(F)T_{\psi_{I-l+1}}(F)+\epsilon(z-T_{\alpha_p}(F)T_{\psi_{I-l+1}}(F))]}{[\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})]^2}|_{\epsilon=0} \\ & q = I-l+1 \\ \frac{\sum_{i=1}^{l} IF([z,p,q];T_{\alpha_i},F)[\sum_{i=1}^{l} T_{\alpha_i}(F)T_{\psi_{I-l+1}}(F)]}{[\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})]^2}|_{\epsilon=0} & q \neq I-l+1 \end{cases}$$

$$= \begin{cases} \frac{(z-T_{\alpha_p}(F)T_{\psi_{I-l+1}}(F))[\sum_{i=1}^{l} T_{\alpha_i}(P)]-\sum_{i=1}^{l} IF([z,p,q];T_{\alpha_i},F)[\sum_{i=1}^{l} T_{\alpha_i}(F)T_{\psi_{I-l+1}}(F)]}{[\sum_{i=1}^{l} T_{\alpha_i}(F)]^2} \\ & q = I-l+1 \\ \frac{\sum_{i=1}^{l} IF([z,p,q];T_{\alpha_i},F)[\sum_{i=1}^{l} T_{\alpha_i}(F)T_{\psi_{I-l+1}}(F)]}{[\sum_{i=1}^{l} T_{\alpha_i}(F_{p,q,\epsilon,z})]^2} & q \neq I-l+1 \end{cases}$$

$$= \begin{cases} \frac{\sum_{i=1}^{l} T_{\alpha_i}(F)[(z-T_{\alpha_p}(F)T_{\psi_{I-l+1}}(F))-T_{\psi_{I-l+1}}(F)\sum_{i=1}^{l} IF([z,p,q];T_{\alpha_i},F]}{[\sum_{i=1}^{l} T_{\alpha_i}(F)]^2} & q = I-l+1 \\ \frac{T_{\psi_{I-l+1}}(F)\sum_{i=1}^{l} IF([z,p,q];T_{\alpha_i},F)}{\sum_{i=1}^{l} T_{\alpha_i}(F)} & q \neq I-l+1 \end{cases}$$

$$= \begin{cases} \frac{(z-T_{\alpha_p}(F)T_{\psi_{I-l+1}}(F))-T_{\psi_{I-l+1}}(F)\sum_{i=1}^{l} IF([z,p,q];T_{\psi_i},F)}{\sum_{i=1}^{l} T_{\psi_i}(F)} & q = I-l+1 \\ \frac{T_{\psi_{I-l+1}}(F)\sum_{i=1}^{l} IF([z,p,q];T_{\alpha_i},F)}{\sum_{i=1}^{l} T_{\alpha_i}(F)} & q \neq I-l+1. \end{cases}$$

□

As a result, we see that influence functions for general cases of T_{α_l} and $T_{\beta_{I-l+1}}$ are unbounded. Thus, the stochastic Chain-Ladder model is not robust and a small number of outlier(s) can have large effects on results. In the next chapter, we present a robust version of this model.

CHAPTER III

ROBUST PROCEDURES IN LOSS RESERVING MODELS

3.1 Robust Generalized Linear Models

The general framework of generalized linear models (GLM) for loss reserving has been discussed in Section 2.3. In this section, we introduce a robust version of this model. To simplify the presentation and avoid unnecessary double subscripts, we replace x_{11} by x_1, x_{12} by x_2 and so on.

We suppose that incremental paid amounts X_i, $i = 1, 2, \ldots, n = I + I$ belong to the exponential family such that $\mathrm{E}[X_i] = \mu_i$, $\mathrm{Var}[X_i] = \mathrm{V}[\mu_i]$ and (see Equation (2.3.1))

$$\eta_i = \ln(\mathrm{E}[X_i]) = \ln(\mu_i) = z_i\boldsymbol{\beta}.$$

The non-robust estimator of $\boldsymbol{\beta}$, which is the solution of Equation (2.5.1), has been studied in the previous chapter. We remind that this maximum likelihood estimator belongs to the general family of M-estimator. Thus, as showed in Equation (1.3.7), the influence function is given by the general expression

$$IF(x; \psi, F) = M(\psi, F)^{-1}\psi(x, \mu),$$

with

$$M(\psi, F) = -\mathrm{E}\left[\frac{\partial}{\partial \beta}\psi(X, \mu)\right].$$

Moreover, the estimator has an asymptotic Normal distribution with asymptotic variance

$$\Omega = M(\psi, F)^{-1}Q(\psi, F)M(\psi, F),$$

where

$$Q(\psi, F) = \mathrm{E}\left[\psi(X, \mu)\psi(X, \mu)^T\right].$$

See Cantoni et Ronchetti (1999) for more details. As suggested by Cantoni et Ronchetti (1999), we assume in this chapter that $\psi(x, \mu)$ can be rewritten as $\psi(x, \mu) = \nu(x, \mu)\omega(z)\mu' - a(\boldsymbol{\beta})$, where

$$a(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[\nu(x_i, \mu_i)]\,\omega(z_i)\mu_i'$$

and $\mu_i = g^{-1}(z_i\boldsymbol{\beta})$ with $g(\cdot)$ a logarithmic link function. The term $a(\boldsymbol{\beta})$ makes sure that the estimator is Fisher consistent (proof is available in Cantoni et Ronchetti (1999), p. 19). Therefore, the influence function of the estimator of $\boldsymbol{\beta}$ is

$$IF(x; \psi, F) = \frac{\nu(x, \mu)\omega(z)\mu' - a(\boldsymbol{\beta})}{-\mathrm{E}\left[\frac{\partial}{\partial\boldsymbol{\beta}}(\nu(X, \mu)\omega(z)\mu' - a(\boldsymbol{\beta}))\right]}.$$

This particular form for the influence function allows us to bound separately the impact of outliers in response variables (x_i) and in covariates (z_i). The bounded function ψ will result in a bounded influence function and thus the choice of ψ plays an important role.

3.1.1 The Structure of the Robust Poisson Model for Reserves

In order to model reserves, we consider a specific case of the robust generalized linear model. We assume that $X_i \sim$ Poisson with $\mathrm{E}[X_i] = \mathrm{V}[X_i] = \mu_i$. As suggested by Cantoni et Ronchetti (1999), a simple choice for the weight function $\nu(\cdot)$ is

$$\nu(x_i, \mu_i) = \psi_c(r_i)\frac{1}{\mathrm{V}[\mu_i]^{1/2}},$$

with

$$r_i = \frac{x_i - \mu_i}{\mathrm{V}[\mu_i]^{1/2}}$$

and $\psi_c(r_i)$ is the Huber function defined as

$$\psi_c(r_i) = \begin{cases} r & |r| \leq c, \\ (c)\ \mathrm{sign}(r) & |r| > c, \end{cases} \tag{3.1.1}$$

where c is a tuning constant which guarantees a level of efficiency of reserve. From empirical evidence (see Verdonck et Debruyne (2011)), the default value for this constant is $c = 1.345$.

For the weight function $\omega(\cdot)$, one possible choice is

$$\omega(z_i) = \sqrt{1 - h_i},$$

where h_i is the i^{th} diagonal element of the hat matrix $H = Z(Z^T Z)^{-1} Z^T$. According to Cantoni et Ronchetti (1999), this specific choice for $\omega(\cdot)$ is not optimal, mainly because H does not have high breakdown properties.

In the loss reserving process with the (robust) Poisson distribution, the model is based on a row effect and on a column effect (see Equation (2.3.1)). Therefore, it is really unlikely to have one or more outlier(s) in the Z matrix. Thus, we can choose the same weight for all observations or $\omega(z_i) = 1$.

Closer examination by Verdonck et Debruyne (2011) has showed that the default value $c = 1.345$ is usually too low. They suggest a new approach to define this tuning constant: first adjusting the robust GLM with $c = 1.345$ and calculating residuals, then setting the new value of the constant c^* to be the 75% quantile of the residuals' distribution and finally adjusting again the robust GLM with $c = c^*$. The selection of a higher quantile will result in a higher efficiency but will also result in a lower breakdown value. The 75% quantile is an optimal option which keeps a good balance between efficiency and robustness.

There is a package (named **robustbase**) in R to estimate the robust GLM for the Poisson and the Binomial distribution robust model (see Maechler et al. (2016) for more details).

3.1.2 The Influence Function of the Robust Poisson Model for Reserves

We calculate the influence function for the model introduced below. In order to evaluate the function $\psi(x, \mu) = \nu(x, \mu)\omega(Z)\mu' - a(\boldsymbol{\beta})$, we need to find

$$
\begin{aligned}
a(\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[\psi_c(r_i)]\,\omega(z_i)\frac{1}{\sqrt{\mathrm{V}[\mu_i]}}\mu_i' \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[\psi_c(r_i)]\,(1)\frac{1}{\sqrt{\mu_i}}\mu_i'.
\end{aligned}
$$

Since the function $\psi_c(r_i)$ takes different values in the two intervals $|r| \leq c$ and $|r| > c$ (see Equation (3.1.1)), it is necessary to open the summation

$$
\begin{aligned}
&|r_i| < c \\
&\Rightarrow -c \leq \frac{j - \mu_i}{\sqrt{\mathrm{V}[\mu_i]}} \leq c \\
&\Rightarrow -c\sqrt{\mathrm{V}[\mu_i]} + \mu_i \leq j \leq c\sqrt{\mathrm{V}[\mu_i]} + \mu_i.
\end{aligned}
$$

Let $j_1 = \lfloor \mu_i - c\sqrt{\mathrm{V}[\mu_i]} \rfloor$ and $j_2 = \lfloor \mu_i + c\sqrt{\mathrm{V}[\mu_i]} \rfloor$, then

$$\begin{aligned}
\mathrm{E}[\psi_c(r_i)] &= \mathrm{E}\left[\psi_c\left(\frac{X_i - \mu_i}{\sqrt{\mathrm{V}[\mu_i]}}\right)\right] \\
&= \sum_{j=-\infty}^{\infty} \psi_c\left(\frac{j-\mu_i}{\sqrt{\mathrm{V}[\mu_i]}}\right) \Pr[X_i = j]\,\mathbb{I}_{[0,\infty]}(j) \\
&= \sum_{j=0}^{\infty} \psi_c\left(\frac{j-\mu_i}{\sqrt{\mathrm{V}[\mu_i]}}\right) \Pr[X_i = j]\,\mathbb{I}_{[0,\infty]}(j) \\
&= c\Pr[X_i \le j_1]\,(-1) + c\Pr[X_i \ge j_2 + 1]\,(+1) \\
&\quad + \sum_{j=j_1+1}^{j_2} \left(\frac{j-\mu_i}{\sqrt{\mathrm{V}[\mu_i]}}\right) \Pr[X_i = j] \\
&= c\left(\Pr[X_i \ge j_2+1] - \Pr[X_i \le j_1]\right) \\
&\quad + \frac{1}{\sqrt{\mathrm{V}[\mu_i]}}\left(\sum_{j=j_1+1}^{j_2} j\Pr[X_i = j] - \sum_{j=j_1+1}^{j_2} \mu_i \Pr[X_i = j]\right) \\
&= c\left(\Pr[X_i \ge j_2+1] - \Pr[X_i \le j_1]\right) \\
&\quad + \frac{1}{\sqrt{\mathrm{V}[\mu_i]}}\left(\mu_i\left(\Pr[X_i \le j_2 - 1] - \Pr[X_i \le j_1 - 1]\right)\right) \\
&\quad - \frac{1}{\sqrt{\mathrm{V}[\mu_i]}}\left(\mu_i\left(\Pr[X_i \le j_2] - \Pr[X_i \le j_1]\right)\right) \\
&= c\left(\Pr[X_i \ge j_2+1] - \Pr[X_i \le j_1]\right) \\
&\quad + \frac{\mu_i}{\sqrt{\mathrm{V}[\mu_i]}}\left(\Pr[X_i = j_1] - \Pr[X_i = j_2]\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
a(\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[\psi_c(r_i)]\,\frac{1}{\sqrt{\mu_i}}\mu_i' \\
&= \frac{1}{n}\sum_{i=1}^{n} c\left(\Pr[X_i \ge j_2+1] - \Pr[X_i \le j_1]\right) z_i \mu_i^{1/2} \\
&\quad + \left(\Pr[X_i = j_1] - \Pr[X_i = j_2]\right) z_i \mu_i.
\end{aligned}$$

Thus, the ψ function is

$$\begin{aligned}\psi(x,\mu) &= \nu(x,\mu)\omega(z)\mu' - a(\boldsymbol{\beta}) \\ &= \begin{cases} \left(\frac{x_i-\mu_i}{\mu_i}\right) z_i\mu_i - a(\boldsymbol{\beta}) & j_1 < x_i < j_2 \\ \left(\frac{(c)\text{sign}(r_i)}{\mu_i^{1/2}}\right) z_i\mu_i - a(\boldsymbol{\beta}) & x_i < j_1 \text{ or } x_i > j_2. \end{cases}\end{aligned}$$

Since both cases are bounded, we conclude that the influence function is bounded for the Poisson model with Huber function.

Example 3.1. *The run-off triangle presented in Table 3.1 shows the cumulative dataset used in Taylor et Ashe (1983). As a first step, we evaluate the total reserve*

Table 3.1 Cumulative run-off triangle from Taylor et Ashe (1983).

i	1	2	3	4	5	6	7	8	9	10
1	357 848	1 124 788	1 735 330	2 218 270	2 745 596	3 319 994	3 466 336	3 606 286	3 833 515	3 901 463
2	352 118	1 236 139	2 170 033	3 353 322	3 799 067	4 120 063	4 647 867	4 914 039	5 339 085	
3	290 507	1 292 306	2 218 525	3 235 179	3 985 995	4 132 918	4 628 910	4 909 315		
4	310 608	1 418 858	2 195 047	3 757 447	4 029 929	4 381 982	4 588 268			
5	443 160	1 136 350	2 128 333	2 897 821	3 402 672	3 873 311				
6	396 132	1 333 217	2 180 715	2 985 752	3 691 712					
7	440 832	1 288 463	2 419 861	3 483 130						
8	359 480	1 421 128	2 864 498							
9	376 686	1 363 294								
10	344 014									

amount with the robust GLM for reserves. Predicted values for each occurrence year and each payment period are presented in Table 3.2. The expected total amount for the reserve, which is the sum of all expected claims, is 18 839 333\$.

As a second step, we introduce an outlier in the second row by multiplying the value in the cell (2, 1) *by* 10. *We reevaluate the total reserve amount with the robust GLM method. Predicted values for each occurrence year and each payment*

Table 3.2 Predicted incremental triangle.

i	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	878 152
3	0	0	0	0	0	0	0	0	1 010 600	900 139
4	0	0	0	0	0	0	0	1 258 065	1 066 615	950 031
5	0	0	0	0	0	0	608 186	717 086	607 961	541 510
6	0	0	0	0	0	380 249	415 017	489 329	414 864	369 518
7	0	0	0	0	340 042	363 615	396 862	467 923	396 716	353 354
8	0	0	0	258 700	242 192	258 982	282 662	333 275	282 557	251 673
9	0	0	387 833	372 590	348 814	372 995	407 100	479 995	406 950	362 470
10	0	94 083	95 220	91 478	85 640	91 577	99 951	117 848	99 914	88 993

period are presented in Table 3.3. The total amount of reserve is 13 536 364$. *As expected, we obtain completely different predictions.*

We have considered a decimal error mistake (by multiplying a value by 10*) for various cells. As we can see in the Table 3.4, the effect of an outlier is not the same in the predicted reserve depending on its position in the run-off triangle. In some cases, it even make the robust GLM reserve greater than the classical GLM reserve. As expected, when there is an outlier in the last row, the robust method has no impact on the predicted reserve.*

As we considered in Section 3.1.1, the defaul value of c based on Verdonck et Debruyne (2011) is equal to 1.345*. In the Table 3.4 you can find some example of reserve amount by Robust GLM method while the default* $c = 1.345$.

The modified option for c based on Verdonck et Debruyne (2011) has been studied

Table 3.3 Predicted incremental triangle in presence of an outlier.

i	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	478 054
3	0	0	0	0	0	0	0	0	795 062	476 027
4	0	0	0	0	0	0	0	1 102 697	828 248	495 896
5	0	0	0	0	0	0	546 073	623 169	468 069	280 247
6	0	0	0	0	0	335 571	367 270	419 122	314 807	188 484
7	0	0	0	0	285 649	299 933	328 266	374 611	281 374	168 467
8	0	0	0	187 585	170 464	178 988	195 896	223 553	167 913	100 534
9	0	0	283 180	266 618	242 284	254 400	278 431	317 741	238 659	142 892
10	0	146 415	95 647	90 053	81 834	85 926	94 043	107 320	80 609	48 263

Table 3.4 Predicted reserves ($c = 1.345$).

outlier	Classic GLM	Robust GLM
NO	18 839 333	18 839 333
$(1,3)$	15 813 130	15 908 930
$(2,3)$	16 044 692	15 934 573
$(3,1)$	14 594 660	15 119 785
$(4,3)$	19 166 515	18 992 655
$(6,5)$	27 889 017	26 949 513
$(10,1)$	60 313 152	61 101 893

in section (3.1.1). In the first run $c = 1.345$ *and in the second run* c *is the* 75^{th} *quantile of pearson residuals of the previous run. The degree of improvement in accuracy of evaluated reserve presented in the Table 3.5.*

Table 3.5 Estimated reserve with modified c.

outlier	Robust GLM with modified c	Robust GLM with fix c
$(2,1)$	18 177 617	13 536 364
$(1,3)$	18 603 226	15 908 930
$(2,3)$	18 794 322	15 934 573
$(3,1)$	18 070 067	15 119 785
$(4,3)$	19 261 166	18 992 655
$(6,5)$	18 905 464	26 949 513
$(10,1)$	60 555 812	61 101 893

As we see in Table 3.5, the reserve amount with modified c in presence of an outlier is less volatile than with a fix c. It shows modified c is prefered.

3.2 Robust Version of the Stochastic Chain-Ladder Model

The amount of reserve directly influences the insurance company solvency and the profit amount. We have shown, in Section 2.4, that outliers highly influence the total amount of the reserve in the stochastic version of the Chain-Ladder model, or Mack's model. Based on Verdonck *et al.* (2009), we describe in this section a method for robustifying this model.

This method would be able to firstly recognize outliers in the run-off triangle and secondly smooths the detected outliers. It smooths the run-off triangle in such a way that the outstanding claim reserve is very close to the outstanding claim reserve without outliers.

3.2.1 Robust development factor

In the stochastic version of the Chain-Ladder model, outliers have a large influence of the total amount of the reserve because this model is constructed from cumulative amounts. Thus, only one outlier in a column may impact all other columns. For example, an outlier in the first column affects all other columns and affects all development factors. The general idea to robustify the Mack's model is to consider incremental data, as in the GLM approach, instead of cumulative data. This will decrease the impact of outliers and by using incremental data, an outlier can affect at most two development factors.

As shown in Equation (2.4.2), a development factor is the ratio of the sum of cumulative paid amounts for one development period to the sum of cumulative paid amounts for the previous development period. To robustify a development factor, we could replace this ratio by

$$\widehat{\lambda}_j = \text{mean}\left(\frac{X_{ij}}{X_{i(j-1)}}\right)_{i=1,2,\ldots,I-i+1,\; j=2,\ldots,I}.$$

In Equation (1.2.10), we demonstrate that mean is not an appropriate choice in presence of outliers since it is very sensitive to them. However, we show that median is a better choice and thus, we may replace the mean by the median in order to obtain a robust estimator for development factor

$$\widehat{\lambda}_j = \text{median}\left(\frac{X_{ij}}{X_{i(j-1)}}\right)_{i=1,2,\ldots,I-i+1,\; j=2,\ldots,I}. \tag{3.2.1}$$

Obviously, we can not control the impact of outlier in the last two columns ($I-1$ and I) by this method since there is only one row remaining and the mean and median of two observations are equal. Thus, it is important to be careful in the construction of a robust version of the Mack's model. In the next section, we present the model introduced in Verdonck *et al.* (2009).

3.2.2 The Structure of the Robust Chain-Ladder Model

The model has been initially presented in Renshaw et Verrall (1998a) and Verdonck *et al.* (2009). For better understanding of this subject, an example will immediately calculate the algorithme step by step.

We assume that each X_{ij} follows an over-dispersed Poisson distribution with

$$\mathrm{E}[X_{ij}] = m_{ij}$$

and

$$\mathrm{Var}[X_{ij}] = \phi m_{ij},$$

where X_{ij} is, as usual, the incremental paid amount for occurrence period i and development period j, ϕ is the dipersion parameter and m_{ij} will be precisely defined in the following. We estimate ϕ as below

$$\widehat{\phi} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{I-i+1} \frac{(x_{ij} - m_{ij})^2}{m_{ij}}}{n - (I + I - 1)}.$$

In the first step we calculate development factors using Equation (3.2.1) with cumulative data. In spite of the fact that development factors based on incremental data are more robust, this approach, as explained by Verdonck *et al.* (2009), will work with cumulative data. Then, we construct the fitted run-off triangle in the following way:

- the fitted cumulative paid to date diagonal, $\widehat{C}_{i(I-i+1)}$, $i = 1, 2, \ldots, I$, is the same as the actual cumulative paid to date diagonal, $C_{i(I-i+1)}$, $i = 1, 2, \ldots, I$, so $\widehat{C}_{i(I-i+1)} = C_{i(I-i+1)}$, $i = 1, 2, \ldots, I$, is the fitted diagonal;
- then we calculate backward the upper left part of the fitted triangle by dividing the cumulative data by the previously calculated development factors; and

- finally, by subtracting two consecutive payments, we find incremental payments, $m_{ij} = \widehat{C}_{i(j+1)} - \widehat{C}_{ij}$, $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, I$.

By differencing cumulative fitted values, the two corner points $(I, 1)$ and $(1, I)$ are the same as observed incremental data.

The Pearson residuals are defined as:

$$r_{ij} = \frac{X_{ij} - m_{ij}}{\sqrt{\phi m_{ij}}}. \tag{3.2.2}$$

In order to detect outliers, we construct a classical Tukey boxplot based on Pearson residuals

$$[Q_1 - 3IQR,\ Q_3 + 3IQR],$$

where Q_1 and Q_3 are residual's first and third quartiles. Outliers are most likely to be outside this interval and by comparing the residuals of the first column, r_{i1}, $i = 1, \ldots, I$, to this interval we detect the outliers in the first column. We will do the same for the second column. If we detect an outlier, we adjust it in the first column as follows: we check the residual in the second column in the same row as the detected oulier value; if this value is also an outlier, we change the cumulative payment by the median of cumulative payments in the first column, however if the residual in the second column is not an outlier, we calculate the payment in the first column by dividing the second column cumulative payment by the first development factor. More formally, suppose r_{k1} is an outlier. Then if r_{k2} is also an outlier,

$$C_{k1} = \text{median}\left(C_{i1} | i = 1, 2, \ldots, I\right),$$

and if it is not,

$$C_{k1} = \frac{X_{k2}}{\text{median}\left(\frac{X_{i2}}{X_{i1}} | i = 1, 2, \ldots, I - 1\right)}.$$

After making the first column outlier free, we investigate other columns, but we do not use the residuals to find the outliers in other columns because they are based on the development factors which have been calculated by using cumulative data. Therefore, we use the first column incremental outlier free data and calculate the development factors as

$$\widehat{\lambda}_j^1 = \text{median}\left(\frac{X_{ij}}{X_{i1}}|i = 1, 2, \ldots, I - j + 1\right), \quad 2 \leq j \leq I.$$

Then, the future claims will be

$$\widehat{X}_{j,I-i+2}^1 = X_{j1}\widehat{\lambda}_{I-i+2}^1, \quad 1 \leq j \leq I, \quad 2 \leq i \leq I.$$

After fitting the upper triangle by multiplying first column incremental claims by development factors, we calculate again the Pearson residuals, but instead of m_{ij}, we use fitted incremental claims $\widehat{X}_{ij}^1$:

$$r_{ij}^1 = \frac{X_{ij} - \widehat{X}_{ij}^1}{\sqrt{\phi\widehat{X}_{ij}^1}}.$$

Residuals, except residuals of the first column, will be examined by using the boxplot as mentioned before but based on new r_{ij}^1. The residuals detected as outlier are replaced by the median of all residuals except residuals of the first column. For instance, suppose r_{kl}^1 is an outlier, then it will be replaced by

$$r_{kl}^1 = \text{median}\left(r_{ij}^1|i = 1, 2, \ldots, I - 1; \ j = 2, \ldots, I - i + 1\right).$$

At last, we back transform outlier-free residuals to incremental data

$$X_{ij}^r = r_{ij}^1\sqrt{\phi\widehat{X}_{ij}^1} + \widehat{X}_{ij}^1.$$

Data are robustified, then we can apply the usual stochastic Chain-Ladder model to evaluate the reserve.

In this method, the two corner points X_{I1} and X_{1I} can not be detected if they are outlier. X_{I1} is the first observation in the last claim year and it is the only observation in this year. By comparing this value with the median of the rest of data in the first column, we realise if it is an outlier or not. If there is not much difference between these two values, it is not an outlier but in case of huge difference, we replace X_{I1} by

$$\text{median}\,(X_{1j}|j = 1, 2, \ldots, I-1)\,.$$

The X_{1I} case is more difficult to handle because it is the only observation in the I^{th} development year. The proposed solution is an extrapolation of the last development factor by curve estimation based on previous development factors. If the extrapolated claim is close to the actual claim amount, it is not an outlier but if the two values are very different X_{1I} is an outlier and should be replaced by the extrapolated amount.

Example 3.2. *In order to clarify the algorithm of the robust Clain-Ladder model, the process will be presented step by step with a toy example based on the run-off triangle illustrated in Table 3.6*

Table 3.6 Cumulative run-off triangle (in millions of dollars).

Accident year	12 months	24 months	36 months
2010	100	150	170
2011	110	160	
2012	120		

- *Step 1. We compute development factors, we calculate fitted values and dispersion parameter and we evaluate Pearson residuals:*

$$\widehat{\lambda}_2 = median\left(\frac{C_{12}}{C_{11}}, \frac{C_{22}}{C_{21}}\right) = median\left(\frac{150}{100}, \frac{160}{110}\right) = 1.477$$
$$\widehat{\lambda}_3 = median\left(\frac{C_{13}}{C_{12}}\right) = median\left(\frac{170}{150}\right) = 1.1333.$$

The fitted run-off triangle based on the computed development factors is presented in Table 3.7, where results are obtained as follow:

Table 3.7 Evaluated run-off triangle (in millions of dollars).

Accident year	12 months	24 months	36 months
2010	101.54	150	170
2011	108.31	160	
2012	120		

$$\widehat{X}_{13} = 170 = X_{13} \qquad \widehat{X}_{12} = 150 = \frac{170}{1.1333} \qquad \widehat{X}_{11} = 101.54 = \frac{150}{1.477}$$
$$\widehat{X}_{22} = 160 = X_{22} \qquad \widehat{X}_{21} = 108.31 = \frac{160}{1.477}$$
$$\widehat{X}_{31} = 120 = X_{31}.$$

Then we calculate the dispersion parameter as

$$\begin{aligned}\widehat{\phi} &= \frac{\sum_{i=1}^{I}\sum_{j=1}^{I-i+1}\frac{(x_{ij}-m_{ij})^2}{m_{ij}}}{n-(I+I-1)} \\ &= \frac{\sum_{i=1}^{3}\sum_{j=1}^{3-i+1}\frac{(x_{ij}-m_{ij})^2}{m_{ij}}}{6-(3+3-1)} \\ &= 0.023310023 + 0 + 0 + 0.026442308 + 0 + 0 = 0.049752331.\end{aligned}$$

Finally, Pearson residuals based on

$$r_{ij} = \frac{X_{ij} - m_{ij}}{\sqrt{\phi m_{ij}}}$$

are presented in Table 3.8.

- *Step 2. We detect and modify potential outlier(s) in the first row. The classical boxplot interval is* $[-3.693\ 3.727]$. *If the residuals in the first column*

Table 3.8 Pearson residulals.

i/j	1	2	3
1	-0.68	0.99	0.00
2	0.73	−1.06	
3	0.00		

is outside of this interval, it means that observation is an outlier and the correction of the first column is based on second column. Suppose r_{k1} is an outlier, then:

- *if r_{k2} is not an outlier $C_{k1} = \frac{X_{k2}}{median\left(\frac{X_{i2}}{X_{i1}}|i=1,2,3\right)}$*
- *if r_{k2} is an outlier $C_{k1} = median\,(C_{i1}|i = 1, 2, 3)$.*

In this example, all residuals of the first column are inside the boxplot interval therefore there is no change in the first column.

- *Step 3. we compute modified development factors*

$$\widehat{\lambda}_j^1 = median\left(\frac{X_{ij}}{X_{i1}}|i = 1, \ldots, 3-j+1\right), \quad j = 2, 3.$$

Then, we calculate fitted incremental claims based on the first column (outlier-free)

$$\widehat{X}_{ij}^1 = X_{i1}\widehat{\lambda}_j^1, \quad i = 1, 2, \quad j = 2, 3.$$

Incremental claims in second and third columns are presented in Table 3.9.

Then, we detect outlier(s) in other columns (except the first one):

$$r_{ij}^1 = \frac{X_{ij} - \widehat{X}_{ij}^1}{\sqrt{\phi \widehat{X}_{ij}^1}}.$$

Table 3.9 Incremental claims

i/j	2	3
1	47.7	20.0
2	52.5	

Table 3.10 Pearson residuals

i/j	2	3
1	1.47	0.00
2	−1.55	

The residuals are presented in Table 3.10. Once again, the same rule as we used for the first column. The boxplot interval is [−5.31, 5.27]. *By comparing the residuals we conclude there is no outlier(s) in the second and the third column.*

- *Step 4. We can finally apply the classical Chain-Ladder model.*

Here we wish to show the effects of robustifying Mack Chain-Ladder with a real example:

Example 3.3. *We consider again the cumulative run-off triangle illustrated in Table 3.1. As a first step, we evaluate the total reserve amount with the robust version of the Mack's model presented in the previous subsection. The full run-off triangle, occurred and predicted amounts, are presented in Table 3.14. Obviously, this is identical to the full run-off triangle obtained with the traditional, non-robust, Mack's model. The total reserve amount is* 18 680 856$.

Table 3.11 Occurred and predicted cumulative amounts.

i	1	2	3	4	5	6	7	8	9	10
1	357 848	1 124 788	1 735 330	2 218 270	2 745 596	3 319 994	3 466 336	3 606 286	3 833 515	3 901 463
2	352 118	1 236 139	2 170 033	3 353 322	3 799 067	4 120 063	4 647 867	4 914 039	5 339 085	5 433 719
3	290 507	1 292 306	2 218 525	3 235 179	3 985 995	4 132 918	4 628 910	4 909 315	5 285 148	5 378 826
4	310 608	1 418 858	2 195 047	3 757 447	4 029 929	4 381 982	4 588 268	4 835 458	5 205 637	5 297 906
5	443 160	1 136 350	2 128 333	2 897 821	3 402 672	3 873 311	4 207 459	4 434 133	4 773 589	4 858 200
6	396 132	1 333 217	2 180 715	2 985 752	3 691 712	4 074 999	4 426 546	4 665 023	5 022 155	5 111 171
7	440 832	1 288 463	2 419 861	3 483 130	4 088 678	4 513 179	4 902 528	5 166 649	5 562 182	5 660 771
8	359 480	1 421 128	2 864 498	4 174 756	4 900 545	5 409 337	5 875 997	6 192 562	6 666 635	6 784 799
9	376 686	1 363 294	2 382 128	3 471 744	4 075 313	4 498 426	4 886 502	5 149 760	5 544 000	5 642 266
10	344 014	1 200 818	2 098 228	3 057 984	3 589 620	3 962 307	4 304 132	4 536 015	4 883 270	4 969 825

As a second step, we introduce an outlier in the second row by multiplying the value in the cell $(2, 1)$ *by* 10. *We reevaluate the total amount of reserve and the new full run-off triangle is presented in Table 3.15. Total reserve amount is* 18 619 218\$. *As expected, we obtain completely different predictions.*

Table 3.12 Full matrix in presence of an outlier.

i	1	2	3	4	5	6	7	8	9	10
1	357 848	1 124 788	1 735 330	2 218 270	2 745 596	3 319 994	3 466 336	3 606 286	3 833 515	3 901 463
2	373 700	1 257 721	2 191 615	3 374 904	3 820 649	4 141 645	4 669 449	4 935 621	5 360 667	5 455 684
3	290 507	1 292 306	2 218 525	3 235 179	3 985 995	4 132 918	4 628 910	4 909 315	5 284 199	5 377 860
4	310 608	1 418 858	2 195 047	3 757 447	4 029 929	4 381 982	4 588 268	4 835 040	5 204 252	5 296 496
5	443 160	1 136 350	2 128 333	2 897 821	3 402 672	3 873 311	4 207 008	4 433 274	4 771 807	4 856 386
6	396 132	1 333 217	2 180 715	2 985 752	3 691 712	4 074 539	4 425 572	4 663 593	5 019 713	5 108 687
7	440 832	1 288 463	2 419 861	3 483 130	4 087 970	4 511 889	4 900 601	5 164 171	5 558 516	5 657 039
8	359 480	1 421 128	2 864 498	4 172 880	4 897 494	5 405 359	5 871 046	6 186 810	6 659 246	6 777 279
9	376 686	1 363 294	2 379 988	3 467 065	4 069 116	4 491 079	4 877 998	5 140 353	5 532 880	5 630 948
10	344 014	1 195 296	2 086 703	3 039 820	3 567 681	3 937 645	4 276 885	4 506 910	4 851 065	4 937 049

In order to study the effect of outlier(s) in different accident year and payment *period, we have had consider the decimal error mistake (by multiplying x_{ij} by 10) for various cells.*

As we can see in Table 3.16, the effect of an outlier on the value of predicted reserve depends on the position of outlier in the run-off triangle. In some cases existing outlier make the Robust Mack Chain-Ladder reserve become greater than the Classical Mack Chain-Ladder reserve. But, it can be the opposite.

The volatility of reserve in presence of an outlier in Robust Mack Chain-Ladder

Table 3.13 Predicted reserve.

Outlier	Classic CL	Robust MCL
NO	18 680 856	18 680 856
(1, 3)	15 813 130	15 908 930
(2, 3)	16 044 692	15 934 573
(3, 1)	14 594 660	15 119 785
(4, 3)	19 166 515	18 992 655
(6, 5)	27 889 017	26 949 513
(10, 1)	60 313 152	61 101 893

method is less than classic Mack Chain-Ladder. We conclude the Robust Mack chain ladder is a less sensitive method in presence of outlier(s) then it is more suitable method. The reserve amount with this method can be less or more than the reserve which is evaluated by classical Mack Chain-Ladder. As expected when there is an outlier in the last raw, robust method has no impact on predicted reserve. Since in the last raw there is just one element, the robust method has no effect on it.

3.2.3 The Influence Function of the Robust Chain-Ladder Model for Reserves

Instead of calculating the theoretical influence function (IF), we calculate the empirical influence function (EIF).

The reason behind not using the IF is if we replace the classical parameter estimates in the IF, it will fail to detect outlier(s) due to masking effects and if we replace the robust parameter estimates in IF, it will have a small IF because robust method down-weights the outliers.

The calculation of EIF is complicated then here we just provide the general formula based on Verdonck et Debruyne (2011) which does not provide proof and details.

First, the unknown distribution function F has been replaced by the empirical distribution function F_n (see EIF in Section 1.2.1). Then, the stochastic version of robust parameter estimates $\widehat{\alpha}_i^s$ and $\widehat{\beta}_j^s$ has been used in the empirical influence function of the (non-robust) Chain-Ladder model. At every observation (p, q) in the upper part of the run-off triangle, we assume that $z = X_{pq}$ because we are considering the EIF of the observation. Thus, we have

$$EIF_{\alpha_l}(X_{pq}) = \begin{cases} \dfrac{\hat{\alpha}_l^s(F_n)\left(\sum_{i=I-l+2}^{I} EIF_{\hat{\beta}_i^s(X_{pq})}\right)}{1-\sum_{i=I-l+2}^{I}\hat{\beta}_i^s(F_n)} & p \neq l \\ \\ \dfrac{X_{ij} - \hat{\alpha}_l^s(F_n)\hat{\beta}_q^s(F_n) + \hat{\alpha}_l^s(F_n)\sum_{i=I-l+2}^{I} EIF_{\hat{\beta}_i^s(X_{pq})}}{1-\sum_{i=I-l+2}^{I}\hat{\beta}_i^s(F_n)} & p = l \end{cases}$$

and

$$EIF_{\beta_{J-l+1}}(X_{pq}) = \begin{cases} \dfrac{-\hat{\beta}_{J-l+1}^s(F_n)\sum_{i=1}^{l} EIF_{\hat{\alpha}_i^s(X_{pq})}}{\sum_{i=1}^{l}\hat{\alpha}_i^s(F_n)} & q \neq J-l+1 \\ \\ \dfrac{X_{ij} - \hat{\alpha}_p^s(F_n)\hat{\beta}_{J-l+1}^s(F_n) - \hat{\beta}_{J-l+1}^s(F_n)\sum_{i=1}^{I} EIF_{\hat{\alpha}_i^s(X_{pq})}}{\sum_{i=1}^{l}\hat{\alpha}_i^s(F_n)} & \\ & q = J-l+1. \end{cases}$$

By these equations, we measure the influence of claims X_{pq} in the run-off triangle on the values of $\widehat{\alpha}_1^s, \ldots, \widehat{\alpha}_I^s$ and $\widehat{\beta}_1^s, \ldots, \widehat{\beta}_J^s$.
In this thesis we just use the below example to compare Classic Mack Chain Ladder and Robust Mack Chain Ladder.

Example 3.4. *We consider again the cumulative run-off triangle illustrated in Table 3.1. As a first step, we evaluate the total reserve amount with the robust version of the Mack's model presented in the previous subsection. The full run-off triangle, occurred and predicted amounts, are presented in Table 3.14. Obviously, this is identical to the full run-off triangle obtained with the traditional, non-robust, Mack's model. The total reserve amount is* 18 680 856$.

Table 3.14 Occurred and predicted cumulative amounts.

i	1	2	3	4	5	6	7	8	9	10
1	357 848	1 124 788	1 735 330	2 218 270	2 745 596	3 319 994	3 466 336	3 606 286	3 833 515	3 901 463
2	352 118	1 236 139	2 170 033	3 353 322	3 799 067	4 120 063	4 647 867	4 914 039	5 339 085	5 433 719
3	290 507	1 292 306	2 218 525	3 235 179	3 985 995	4 132 918	4 628 910	4 909 315	5 285 148	5 378 826
4	310 608	1 418 858	2 195 047	3 757 447	4 029 929	4 381 982	4 588 268	4 835 458	5 205 637	5 297 906
5	443 160	1 136 350	2 128 333	2 897 821	3 402 672	3 873 311	4 207 459	4 434 133	4 773 589	4 858 200
6	396 132	1 333 217	2 180 715	2 985 752	3 691 712	4 074 999	4 426 546	4 665 023	5 022 155	5 111 171
7	440 832	1 288 463	2 419 861	3 483 130	4 088 678	4 513 179	4 902 528	5 166 649	5 562 182	5 660 771
8	359 480	1 421 128	2 864 498	4 174 756	4 900 545	5 409 337	5 875 997	6 192 562	6 666 635	6 784 799
9	376 686	1 363 294	2 382 128	3 471 744	4 075 313	4 498 426	4 886 502	5 149 760	5 544 000	5 642 266
10	344 014	1 200 818	2 098 228	3 057 984	3 589 620	3 962 307	4 304 132	4 536 015	4 883 270	4 969 825

As a second step, we introduce an outlier in the second row by multiplying the value in the cell (2, 1) *by* 10. *We reevaluate the total amount of reserve and the new full run-off triangle is presented in Table 3.15. Total reserve amount is* 18 619 218$. *As expected, we obtain completely different predictions.*

In order to study the effect of outlier(s) in different accident year and payment period, we have had consider the decimal error mistake (by multiplying x_{ij} *by* 10*)*

Table 3.15 Full matrix in presence of an outlier.

i	1	2	3	4	5	6	7	8	9	10
1	357 848	1 124 788	1 735 330	2 218 270	2 745 596	3 319 994	3 466 336	3 606 286	3 833 515	3 901 463
2	373 700	1 257 721	2 191 615	3 374 904	3 820 649	4 141 645	4 669 449	4 935 621	5 360 667	5 455 684
3	290 507	1 292 306	2 218 525	3 235 179	3 985 995	4 132 918	4 628 910	4 909 315	5 284 199	5 377 860
4	310 608	1 418 858	2 195 047	3 757 447	4 029 929	4 381 982	4 588 268	4 835 040	5 204 252	5 296 496
5	443 160	1 136 350	2 128 333	2 897 821	3 402 672	3 873 311	4 207 008	4 433 274	4 771 807	4 856 386
6	396 132	1 333 217	2 180 715	2 985 752	3 691 712	4 074 539	4 425 572	4 663 593	5 019 713	5 108 687
7	440 832	1 288 463	2 419 861	3 483 130	4 087 970	4 511 889	4 900 601	5 164 171	5 558 516	5 657 039
8	359 480	1 421 128	2 864 498	4 172 880	4 897 494	5 405 359	5 871 046	6 186 810	6 659 246	6 777 279
9	376 686	1 363 294	2 379 988	3 467 065	4 069 116	4 491 079	4 877 998	5 140 353	5 532 880	5 630 948
10	344 014	1 195 296	2 086 703	3 039 820	3 567 681	3 937 645	4 276 885	4 506 910	4 851 065	4 937 049

for various cells.

As we can see in Table 3.16, the effect of an outlier on the value of predicted reserve depends on the position of outlier in the run-off triangle. In some cases existing outlier make the Robust Mack Chain-Ladder reserve become greater than the Classical Mack Chain-Ladder reserve. But, it can be the opposite.

The volatility of reserve in presence of an outlier in Robust Mack Chain-Ladder method is less than classic Mack Chain-Ladder. We conclude the Robust Mack chain ladder is a less sensitive method in presence of outlier(s) then it is more suitable method. The reserve amount with this method can be less or more than the reserve which is evaluated by classical Mack Chain-Ladder. As expected when there is an outlier in the last raw, robust method has no impact on predicted reserve. Since in the last raw there is just one element, the robust method has no effect on it.

Table 3.16 Predicted reserve.

Outlier	Classic CL	Robust MCL
NO	18 680 856	18 680 856
(1, 3)	15 813 130	15 908 930
(2, 3)	16 044 692	15 934 573
(3, 1)	14 594 660	15 119 785
(4, 3)	19 166 515	18 992 655
(6, 5)	27 889 017	26 949 513
(10, 1)	60 313 152	61 101 893

CHAPTER IV

CASE STUDY

In this chapter we analyze a real dataset from a French-German general insurance company. The original dataset covers occurrence years from 1999 to 2010 (partial) and development years from 1999 to 2011. In our example, we restrict ourselves to years 1999 to 2008 to calibrate various classical and robust models, and we compare our prediction results with observed payments in 2009 and 2010. In Section 4.1, we consider the generalized linear model for reserves and in Section 4.2, we present results with the stochastic Chain-Ladder model. Finally, we compare predictive distributions in Section 4.3

4.1 Generalized Linear Model

From the dataset, we construct the incremental run-off triangle presented in Table 4.1. In this table, development years 2009 and 2010 has been indicated.

Here is the cumulative run-off triangle in Table 4.2. As we see in the table, development of the claim's payment is not totally complete. The table shows most part of the claims has been paid in first nine years. Then we conclude the claims which have occurred from 1999 to 2002 already have almost paid but the claims which has been occurred from 2003 to 2008 are still open and will develop.

Table 4.1 Incremental run-off triangle.

i	1	2	3	4	5	6	7	8	9	10	11	12
1	224 029	426 494	191 046	362 554	45 718	49 109	42 549	76 976	5 871	8 353	**1 629**	**11 852**
2	233 083	362 137	250 703	169 742	247 624	36 337	135 285	54 530	127 599	**22 423**	**2 901**	
3	272 653	380 774	230 831	152 406	126 293	182 292	157 453	102 873	**96 262**	**12 468**		
4	270 892	421 649	259 148	266 317	237 599	241 236	79 175	**27 494**	**44 302**			
5	260 786	327 992	167 296	266 583	199 633	103 362	**101 611**	**38 857**				
6	290 887	387 624	266 646	261 402	74 815	**72 981**	**290 384**					
7	269 677	574 012	319 171	118 667	**215 074**	**55 194**						
8	310 502	581 810	260 756	**124 006**	**103 965**							
9	453 875	503 656	**519 766**	**358 531**								
10	559 148	**735 001**	**412 373**									

Table 4.2 Cumulative run-off triangle.

i	1	2	3	4	5	6	7	8	9	10	11	12
1	224 029	650 524	841 570	1 204 125	1 249 843	1 298 953	1 341 502	1 418 478	1 424 349	1 432 703	**1 434 332**	**1 446 184**
2	233 083	595 221	845 925	1 015 667	1 263 291	1 299 628	1 434 913	1 489 444	1 617 043	**1 639 466**	**1 642 367**	
3	272 653	653 428	884 259	1 036 665	1 162 959	1 345 251	1 502 704	1 605 577	**1 701 839**	**1 714 306**		
4	270 892	692 542	951 690	1 218 008	1 455 608	1 696 844	1 776 020	**1 803 513**	**1 847 816**			
5	260 786	588 778	756 075	1 022 658	1 222 291	1 325 652	**1 427 264**	**1 466 120**				
6	290 887	678 512	945 159	1 206 561	1 281 377	**1 354 358**	**1 644 742**					
7	269 677	843 689	1 162 860	1 281 528	**1 496 602**	**1 551 796**						
8	310 502	892 312	1 153 068	**1 277 074**	**1 381 039**							
9	453 875	957 532	**1 477 297**	**1 835 829**								
10	559 148	**1 294 150**	**1 706 523**									

4.1.1 Classic Generalized Linear Model

The expected values of future claims obtained with the GLM for loss reserving based on the original run-off triangle (see Table 4.1) are given in Table 4.3. Development years 2009 and 2010 has not been considered in this calculation. The estimations are achieved by maximum likelihood method. All calculations has been done in R software. By summing up all expected future claims we get an overall reserve of 6 982 482$ and the mean square error is 6 893.

Table 4.3 Predicted incremental run-off triangle.

i	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	9 484
3	0	0	0	0	0	0	0	0	73 694	9 849
4	0	0	0	0	0	0	0	97 277	85 982	11 491
5	0	0	0	0	0	0	97 405	77 945	68 894	9 207
6	0	0	0	0	0	123 486	103 225	82 602	73 011	9 757
7	0	0	0	0	178 108	140 665	117 586	94 093	83 167	11 115
8	0	0	0	288 409	200 338	158 222	132 262	105 837	93 548	12 502
9	0	0	332 970	322 785	224 216	177 081	148 026	118 452	104 698	13 992
10	0	857 438	492 602	477 533	331 709	261 976	218 992	175 240	154 892	20 700

4.1.2 Robust Generalized Linear Model

In order to illustrate the interest of the robust version of the generalized linear model, we introduce an outlier into the original run-off triangle (see Table 4.1) by multiplying **one** claim amount by 10. Then, we apply both, the classic and the robust generalized linear model on this modified run-off triangle.

All calculations have been done in R software using the "robustbase" package. Parameters are estimated with a maximum likelihood aproach. We present the results in Table 4.4. The position of an outlier has a direct impact on the amount of reserve, i.e., in presence of an outlier, the reserve amount becomes larger or smaller than the original amount of reserve.

When there is no outlier in the run-off triangle, the reserve amount is 7 004 980$ and the mean square error is 7 057. As we see in the Table 4.4, reserve amounts in

Table 4.4 Reserve amounts using the Classic and the Robust GLM in presence of outlier.

Outlier	Classic GLM	Robust GLM
No	6 982 482	7 004 980
(1, 3)	6 650 656	6 194 158
(2, 1)	4 197 880	6 753 206
(2, 3)	6 747 584	6 863 685
(3, 1)	4 337 635	6 555 842
(4, 3)	7 614 270	6 347 523
(6, 5)	8 070 429	7 822 721
(10, 1)	33 902 260	33 238 412

the Robust GLM are very close to the amount of reserve in absence of an outlier (7 004 980$). We conclude that the robust version of the generalized linear model for reserves is an improvement and the reserve obtained with this approach is more accurate than the one obtained with the Classic GLM. For the last row, as expected, the robust method has no impact on the amount of reserve.

4.2 Stochastic Chain-Ladder Model

As in the previous section, we consider only years 1999 to 2008 in our run-off triangle in order to adjust the stochastic Chain-ladder model (see the incremental dataset in Table 4.1).

4.2.1 Classic Stochastic Chain-Ladder Model

Development factors $\widehat{\lambda}_j$ calculated based on Equation (2.4.2). The total amount of reserve is based on Equation (2.4.1) and the mean square errors are calculated based on Definition 2.3. All calculations been done using the R software. Package "ChainLadder" and function "MackChainLadder" are used for Classic Mack Chain-Ladder calculations. Robust Mack Chain-Ladder calculated by the algorithm in subsection 3.2.2 and the simulation carried out with the "BootChain-Ladder" function. Detailed results for the classic model are presented in Table 4.5.

Table 4.5 Detailed results with the classic stochastic Chain-Ladder model.

i	Reserve	Std. error	coef. of var.
1	0	0	NaN
2	9 484	34 618	3.652
3	83 543	115 961	1.388
4	194 751	132 782	0.681
5	253 452	133 401	0.526
6	392 084	175 550	0.448
7	624 737	236 979	0.379
8	991 121	316 454	0.319
9	1 442 224	349 665	0.242
10	2 991 086	592 948	0.198

The evaluated development factor are in Table 4.6. As expected, the development factors getting smaller as we get closer to the end of the development period. The predicted cumulative run-off triangle with the classic Chain-Ladder method is given in Table 4.7. The evaluated total amount of reserve with the classic stochastic Chain-Ladder model is 6 982 482$ and the mean square error is 1 190 662. Here we are able to compare observed payments in years 2009 and 2010 for claims occurred between years 1999 to 2008. It enables us to evaluate how accurate the Classic stochastic Chain-Ladder model is. In Table 4.8, we give the

Table 4.6 Development factors with the classic stochastic Chain-Ladder model.

i	1	2	3	4	5	6	7	8	9	10
1	2.53	1.35	1.25	1.14	1.10	1.07	1.05	1.05	1.01	1.00

Table 4.7 Run-off triangle with Classic stochastic Chain-Ladder.

i	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	1 626 527
3	0	0	0	0	0	0	0	0	1 679 271	1 689 120
4	0	0	0	0	0	0	0	1 873 298	1 959 280	1 970 771
5	0	0	0	0	0	0	1 423 058	1 501 003	1 569 897	1 579 105
6	0	0	0	0	0	1 404 864	1 508 090	1 590 692	1 663 703	1 673 461
7	0	0	0	0	1 459 636	1 600 302	1 717 888	1 811 982	1 895 150	1 906 265
8	0	0	0	1 441 477	1 641 815	1 800 038	1 932 300	2 038 138	2 131 686	2 144 189
9	0	0	1 290 502	1 613 288	1 837 504	2 014 585	2 162 612	2 281 064	2 385 763	2 399 755
10	0	1 416 587	1 909 189	2 386 723	2 718 432	2 980 408	3 199 401	3 374 642	3 529 534	3 550 235

results. As one can see, differences between observed and predicted payments are not very large.

4.2.2 Robust Stochastic Chain-Ladder Model

An outlying value introduce to the run-off triangle Table 4.1 by multiplying one claim amount by 10, one at the time. Applying the robust Chain-Ladder method on this adjusted run-off triangle results in the estimated reserve. In Table 4.9, the evaluated reserve amounts by both robust and classic Mack Chain-Ladder are given. When there is no outlier, both models give the same amount of reserve. In the presence of outliers, as we see, the evaluated reserve in the robust Chain-

Table 4.8 Predicted and observed cumulative payments.

Case	Predicted payment	Observed payment
(2, 10)	1 626 527	1 639 466
(3, 9)	1 679 271	1 701 839
(3, 10)	1 689 120	1 714 306
(4, 8)	1 873 298	1 803 513
(4, 9)	1 959 280	1 847 816
(5, 7)	1 423 058	1 427 264
(5, 8)	1 501 003	1 466 120
(6, 6)	1 404 864	1 354 358
(6, 7)	1 508 090	1 644 742
(7, 5)	1 459 636	1 496 602
(7, 6)	1 600 302	1 551 796
(8, 4)	1 441 477	1 277 074
(8, 5)	1 641 815	1 381 039
(9, 3)	1 290 502	1 477 297
(9, 4)	1 613 288	1 835 829
(10, 2)	1 416 587	1 294 149
(10, 3)	1 909 189	1 706 522

Ladder model is very close to the reserve with no outlier. But in the classic Chain-Ladder approach, evaluated reserves are much more different than the reserve with no outlier. Therefore the robust Chain-Ladder model is less sensitive against outlier(s) and provides a more reliable evaluation for the total reserve amount.

Table 4.9 Reserve amounts with the classic and robust Chain-Ladder model in presence of outlier.

Outlier	Classic C-L	Robust C-L
NO	6 982 482	6 982 482
(1, 3)	6 650 656	6 650 656
(2, 1)	4 197 880	6 872 534
(2, 3)	6 747 584	7 000 834
(3, 1)	4 337 635	6 966 091
(4, 3)	7 614 270	6 979 768
(6, 5)	8 070 429	6 905 629
(10, 1)	33 902 260	33 902 260

4.3 Percentile match

In this subsection, we will simulate the run-off triangle based on the original triangle with and without the presence of outlier, then we calculate different percentiles in order to have a more complete image of the risk associated with both models (classic/robust stochastic Chain-Ladder/generalized linear model).

In the first step outcomes without any outlier will be considered. Can be seen in Table 4.10, total reserve is the same as we have calculated before (see Table 4.4 and Table 4.9). Using the robust method has no significant impact on the reserve and percentile outcomes. Now we put one outlier by multiplying the $x_{2,1}$ by 10 in the run-off triangle. As we see in Table 4.11 using robust method in the stochastic chain ladder model has decreased the total reserve amount but decreases the percentiles significantly. In the generalized linear model (Poisson), using a robust method has a large effect on both the total reserve and percentiles. It makes the

Table 4.10 Reserve percentile various models without outlier.

Outlier	Total reserve	s.e	95^{th}	99^{th}	99.5^{th}
Classic C-L	6 982 482	1 190 662	9 756 725	10 089 200	11 293 562
Robust Mack C-L	6 422 943	748 512	10 524 593	11 537 933	11 664 600
GLM-Poisson	6 982 482	130 740	6 986 835	6 988 932	6 989 133
Robust GLM-Poisson	7 004 980	126 688	7 009 452	7 011 160	7 011 943

Table 4.11 Reserve percentile in various with an outlier in position $(2, 1)$.

Outlier	Total reserve	s.e	95^{th}	99^{th}	99.5^{th}
Classic C-L	4 197 880	1 264 304	7 151 971	7 867 591	8 219 502
Robust Mack C-L	4 116 383	1 083 685	4 172 744	4 197 023	4 200 057
GLM-Poisson	4 197 880	75 701	4 201 326	4 202 331	4 203 018
Robust GLM-Poisson	6 753 206	122 227	6 757 544	6 759 198	6 760 024

evaluation very close to the real reserve without the presence of an outlier. We conclude that the Robust GLM model gives us the most accurate evaluation of the reserve, since it has the smallest percentile interval and also the mean square error has decreased with presence of outlier.

CONCLUSION

In this study it has been showed that the outliers strongly influence the outstanding amount of reserve. Outliers usually cause an overvaluation of the amount of reserve which causes companies to keep more reserve than needed. Even outliers can cause short estimation of reserve, in which case the result of short estimation could be catastrophic and cause bankruptcy. Therefore in order to evaluate a more accurate value of reserve it is important to use robust methods. We studied two methods of Mack Chain-Ladder and GLM in this research for which a real run-off triangle of a non-life insurance company was used.

In Mack Chain-Ladder model, we first detect the outlier and then we modify the outlier data. As the examples result showed, the robust Mack Chain-Ladder method is more reliable than the classic Mack Chain-Ladder method.

In GLM model, by comparing GLM and Robust GLM evaluated reserve, we conclude that robustifying improves evaluations.

Nowadays, for each individual claim case, all of the details of progress can be recorded, which can help us to develop the individual approach. To expand this study and for further developing this subject, we recommend studying the individual approach.

APPENDIX A

MATHEMATICAL DEFINITIONS

Definition A.1 (Probability Space). *The mathematical triplet $(\Omega, \mathcal{A}, \mathbb{P})$ defines a probability space which presents a model for a particular class of real-world situations.*

- *Ω stands for the fundamental space (or sample space) is a set of outcomes. More precisely, we have*

  $$\Omega = \{\omega_i : i \in \mathcal{I}\},$$

  where $\mathcal{I}$ represents a set of indices such as $\mathcal{I} = \{0, 1, \ldots\}$ or $\mathcal{I} = \{0, 1, 2, \ldots, N\}$. It is clear that by implementing the model we will have outcomes. Outcomes may be of different nature such as possibilities or experimental results. For example consider tossing a six-sided cube or die that has numbers 1 to 6. When the die comes to rest, it will always show one number which is the outcome.

  Every run of the experiment must produce exactly one outcome. If outcomes of different runs of an experiment differ in any way that matters, they are distinct outcomes.

- *$\mathcal{A}$ stands for the σ-algebra. It is a collection of events that could happen. An "event" is a set of zero or more outcomes and it is a subset of the sample* *space. Properties of a σ-algebra defined on the sample space Ω are*

- *$\mathcal{A}$ contains the empty subset, i.e., $\{\emptyset\} \in \mathcal{A}$;*
- *$\mathcal{A}$ is closed under complementation, i.e., $\forall A \in \mathcal{A}$, we have $A^c \in \mathcal{A}$ where A^c is the complement of A in Ω; and*
- *$\mathcal{A}$ is closed under countable unions, i.e., if $A_1, A_2, A_3, \ldots$ are in $\mathcal{A}$, then $A_1 \bigcup A_2 \bigcup A_3 \bigcup \ldots \in \mathcal{A}$.*

- *$\mathbb{P}$ stands for the probability measure. It is a function returning an event's probability (between 0 and 1)*

$$\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$$

and having the following properties:

- *the probability of a countable union of mutually exclusive events must be equal to the countable sum of the probabilities of each of these events*

$$\forall A_i \neq A_j, \qquad \mathbb{P}\left(A_i \bigcup A_j\right) = \mathbb{P}(A_i) + \mathbb{P}(A_j),$$

- *the probability of the sample space Ω must be equal to 1*

$$\mathbb{P}(\Omega) = 1.$$

Example A.1. *We consider an arbitrary experiment which consists in observing car insurance claims. Thus, the fundamental space can be*

$$\Omega = \mathbb{R}^+$$

or

$$\Omega = \mathbb{R}^+ \setminus \{0\}.$$

Some examples of events are:

- *observation of a claim of* 800$; *and*

- *observation of a claim of more than* 100$.

Example A.2. *Consider a fundamental space* $\Omega = \{a, b, c, d\}$. *Then an example of* σ*-algebra on this space is* $\mathcal{A} = \{\emptyset, \Omega, \{a\}, \{b, c, d\}\}$.

Definition A.2 (Measure Space). *A triplet* (χ, Σ, μ), *where* χ *is the fundamental space,* Σ *is a* σ*-algebra over* χ *and* μ *is a measure (but not necessarily a probability measure). A measure on a set is a systematic way to assign a number to each subset of that set intuitively interpreted as its size.*

A function μ *from* Σ *to the extended real number line* ($\mathbb{R} \cup +\infty \cup -\infty$) *is called a measure if it satisfies the following properties:*

- *non-negativity: for all* E *in* Σ, *we have:* $\mu(E) \geq 0$;

- *null empty set:* $\mu(\emptyset) = 0$; *and*

- *countable additivity (or* σ*-additivity): for all countable collections* $\{E_i\}_{i=1}^{\infty}$ *of pairwise disjoint sets in* Σ, *we have*
$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k).$$

One should note that a probability space is a special case of measure space.

Definition A.3 (Functional). *A functional is an operator or a function from a vector space to a scalar. Consider the mapping*

$$x_0 \to f(x_0)$$

which is a function where x_0 *is the argument of the function* f. *In parallel, it is also possible to consider the mapping*

$$f \to f(x_0),$$

which is a functional.

Example A.3. *We consider a sample* $\{2, 4, 6, 8, 10\}$ *and its sample mean. In a function format, we have*

$$\begin{aligned} T_5\left(X_1, \ldots, X_5\right) &= 1/5 \sum_{i=1}^{5} X_i \\ &= \frac{2+4+6+8+10}{5} \\ &= 6 \end{aligned}$$

and in a functional format, we have

$$T(F) = \int x \, dF(x),$$

where we put

$$F(x) = F_5(x) = \begin{cases} 0, & x < 2 \\ 0.2, & 2 \leq x < 4 \\ 0.4, & 4 \leq x < 6 \\ 0.6, & 6 \leq x < 8 \\ 0.8, & 8 \leq x < 10 \\ 1, & x \geq 10. \end{cases}$$

Then we have

$$\begin{aligned} T\left(F_5\right) &= \int x \, dF_5(x) \\ &= \frac{2+4+6+8+10}{5} \\ &= 6. \end{aligned}$$

In the function form, we can analyze the effect of a change in one (or more) value *of the sample*

$$
\begin{aligned}
G(\epsilon) &= T_5(X_1 + \ldots + X_5 + \epsilon) \\
&= \frac{1}{5}\left(\epsilon + \sum_{i=1}^{5} X_i\right) \\
&= \frac{2+4+6+8+10+\epsilon}{5} \\
&= 6 + \frac{\epsilon}{5},
\end{aligned}
$$

while in the functional form, we can analyze the effect of a change in the distribution

$$
\begin{aligned}
G(F^*) &= T_5(F_5 + F^*) \\
&= \int x\, d(F_5 + F^*)(x) \\
&= \int x\, dF_5(x) + \int x\, dF^*(x) \\
&= \frac{2+4+6+8+10}{5} + \int x, dF^*(x) \\
&= 6 + \int x, dF^*(x).
\end{aligned}
$$

If we choose

$$
F^*(x) = \begin{cases} 0, & x < \epsilon \\ 1, & x \geq \epsilon, \end{cases}
$$

we will have the same result which we had in function format

$$
G(F^*) = 6 + \frac{\epsilon}{5}.
$$

APPENDIX B

GENERALIZED LINEAR MODELS

B.1 Exponential family

The exponential family (EF) is a set of probability density functions defined by

$$f_Y(y) = c(y;\phi)\exp\left(\frac{y\theta - a(\theta)}{\phi}\right), \tag{B.1.1}$$

where $a()$ and $c()$ are known functions, θ is the canonical parameter and ϕ is the dispersion parameter. Most of the commonly used distributions in actuarial science such as the Normal distribution, the Gamma distribution and the Poisson distribution are in the exponential family.

This family provides an alternative parameterization framework for distributions which is useful to define sample statistics.

Theorem B.1.1. *Let Y be a random variable following a distribution from the exponential family. We have*

$$\begin{aligned} \mathrm{E}[Y] &= \mu = a'(\theta) \\ \mathrm{Var}[Y] &= \phi a''(\theta), \end{aligned}$$

where $a'(\theta)$ and $a''(\theta)$ are the first derivative and the second derivative of the function $a()$ with respect to θ, respectively.

Proof. An interested reader may consult Wüthrich et Merz (2008) for the proof.

□

Example B.1. *Let Y be a random variable following a Poisson distribution with parameter* λ. *The probability mass function is*

$$\begin{aligned}\Pr[Y = y] &= \frac{\exp^{-\lambda}\lambda^{y}}{y!} \\ &= \frac{\exp^{y\ln(\lambda)-\lambda}}{y!}, \qquad y = 0, 1, \ldots\end{aligned}$$

and 0 *elsewhere. Therefore,* $\theta = \ln(\lambda)$, $\phi = 1$, $a(\theta) = \lambda = \exp^{\theta}$ *and*

$$c(y;\phi) = \begin{cases} 1/y!, & y = 0, 1, \ldots \\ 0, & elsewhere. \end{cases}$$

This shows that the Poisson distribution is a member of the exponential family. Moreover, we have

$$\begin{aligned}\mathrm{E}[Y] = \mu = a'(\theta) &= \frac{d}{d\theta}\exp^{\theta} = \exp^{\theta} = \lambda \\ \mathrm{Var}[Y] = \phi a''(\theta) &= 1 \times \frac{d^2}{d\theta^2}\exp^{\theta} = \exp^{\theta} = \lambda.\end{aligned}$$

B.2 Caracteristics

Let Y_i $(i = 1, \ldots, n)$ be independent random variables following a distribution in the exponential family with a probability density function given by Equation (B.1.1). Define

$$g(\mu_i) = g(\mathrm{E}[Y_i]) = X_i'\boldsymbol{\beta} = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_k X_{ik}, \quad i = 1, \ldots, n.$$

The main characteristics of this model are

- $g()$ is a link function which determines how the mean is related to independent (explanatory) variables X; and

- the distribution of the response variable, Y, which depends on the function $a(\theta)$.

Example B.2. *Let* Y *be a random variable following a Poisson distribution with parameter* λ *and* $\lambda = exp\{X'\beta\}$ *. We want to determine the link function.*
The probability mass function is

$$\begin{aligned}\Pr[Y = y] &= \frac{\exp^{-\lambda}\lambda^{y}}{y!} \\ &= \frac{\exp^{y\ln(\lambda)-\lambda}}{y!}, \qquad y = 0, 1, \dots\end{aligned}$$

and 0 *elsewhere.*
As we showed in the previous example

$$c(y;\phi) = \begin{cases} 1/y!, & y = 0, 1, \dots \\ 0, & elsewhere. \end{cases}$$

$$\theta = \ln(\lambda)$$

$$a(\theta) = \lambda = \exp^{\theta}$$

$$\phi = 1.$$

Hence $\theta = \ln(\lambda) = \ln(\mu)$*, and therefore* $g(\mu) = ln(\mu)$*. So the link function is a logarithmic function.*

B.3 Parameters Estimation

Here we want to estimate parameters β and ϕ by the maximum likelihood approach for a sample y_i, $i = 1, ..., n$ where variables are independent. The log-likelihood function is

$$
\begin{aligned}
l(y;\beta,\phi) &= \sum_{i=1}^{n} \ln(f(y_i;\beta,\phi)) \\
&= \sum_{i=1}^{n} \left(\ln \left(c(y_i,\phi) + \exp\left(\frac{y_i\theta_i - a(\theta_i)}{\phi} \right) \right) \right) \\
&= \frac{1}{\phi} \sum_{i=1}^{n} (y_i\theta_i - a(\theta_i)) + \sum_{i=1}^{n} \ln\left(c(y_i,\phi)\right).
\end{aligned}
$$

Now by partial derivative we will calculate the maximum likelihood estimator for β_j as below:

$$
\begin{aligned}
\frac{\partial l(y;\beta,\phi)}{\partial \beta_t} &= \sum_{i=1}^{n} \left(\frac{\partial \ln\left(f(y_i;\beta,\phi)\right)}{\partial \theta_i} \right) \left(\frac{\partial \theta_i}{\partial \beta_t} \right) \\
&= \frac{1}{\phi} \left(y_i\theta_i - a'(\theta_i) \right) \left(\frac{\partial \theta_i}{\partial \beta_t} \right).
\end{aligned}
$$

By definition

$$
\mu_i = a'(\theta_i) \rightarrow \theta_i = (a')^{-1}(\mu_i)
$$

$$
g(\mu_i) = X_i'\beta \rightarrow \mu_i = g^{-1}(X_i'\beta).
$$

Also,

$$
\begin{aligned}
\frac{\partial \theta_i}{\partial \beta_t} &= \left(\frac{\partial \theta_i}{\partial \mu_i} \right) \left(\frac{\partial \mu_i}{\partial \beta_t} \right) \\
\frac{\partial \theta_i}{\partial \mu_i} &= \frac{\partial (a')^{-1}(\mu_i)}{\partial \mu_i} \\
&= \frac{1}{a''((a')^{-1}(\mu_i))} = \frac{1}{a''(\theta)} \\
&= \frac{1}{\mathrm{Var}[Y_i]/\phi} = \frac{\phi}{\mathrm{Var}[Y_i]} \\
\frac{\partial \mu_i}{\partial \beta_t} &= \frac{\partial g^{-1} X_i'\beta}{\partial \beta_t} \\
&= \left(\frac{1}{g' g^{-1} X_i'\beta} \right) \left(\frac{\partial X_i'\beta}{\partial \beta_t} \right) \\
&= \frac{X_{it}}{g'(\mu_i)} \\
\rightarrow \frac{\partial \theta_i}{\partial \beta_t} &= \left(\frac{\phi}{\mathrm{Var}[Y_i]} \right) \left(\frac{X_{it}}{g'(\mu_i)} \right).
\end{aligned}
$$

Therefore

$$\frac{\partial l(y;\beta,\phi)}{\partial \beta_t} = \frac{1}{\phi}\sum_{i=1}^{n}\frac{(y_i - a'(\theta_i))\phi}{\text{Var}[Y_i]\, g'(\mu_i)}X_{it}$$
$$= \sum_{i=1}^{n}\frac{(y_i - \mu_i)}{\phi g'(\mu_i)a''(\theta_i)}X_{it} = 0$$
$$\rightarrow \sum_{i=1}^{n}\frac{(y_i - \mu_i)}{g'(\mu_i)a''(\theta_i)}X_{it} = 0, \qquad t = 1, 2, \ldots, I.$$

We solve the above equations to estimate β.

Example B.3. *Y is a random variable following a Poisson distribution with parameter λ and $\lambda = exp\{X'\beta\}$. What is the condition of the first derivation for the maximum likelihood estimation of β.*
The probability mass function of Y is

$$\Pr[Y = y] = \frac{\exp^{-\lambda}\lambda^{y}}{y!}$$
$$= \frac{\exp^{y\ln(\lambda)-\lambda}}{y!}, \qquad y = 0, 1, \ldots$$

and 0 elsewhere.
We know in the poisson distribution family the link function is logarithm function. Also

$$X_i'\beta = log(\mu_i) = log(\lambda_i)$$

therefore

$$\mu_i = \lambda_i = exp(X_i'\beta)$$

We have showed

$$g(\mu_i) = ln(\mu_i) \rightarrow g'(\mu_i) = \frac{1}{\mu_i}$$

and

$$a''(\theta_i) = e^{(\theta_i)} = e^{log(\lambda_i)} = \lambda_i = \mu_i$$

Therefore the condition of the first derivation is

$$\begin{aligned}\frac{\partial l(y;\beta,\phi)}{\partial \beta_t} &= \sum_{i=1}^{n}(y_i - \mu_i)x_{it} \\ &= \sum_{i=1}^{n}(y_i - \lambda_i)X_{it} = 0, \quad t = 1, 2, ..., k\end{aligned}$$

BIBLIOGRAPHY

Antonio, K. et Plat, R. (2014). Micro–level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, *2014*(7), 649–669.

Arjas, E. (1989). The claims reserving problem in non–life insurance: some structural ideas. *ASTIN Bulletin*, *19*(2), 139–152.

Arjas, E. et Haastrup, S. (1996). Claims reserving in continuous time: a nonparametric Bayesian approach. *ASTIN Bulletin*, *26*(2), 139–164.

Barnett, G. et Zehnwirth, B. (2000). *Best estimates for reserves*. Casualty Actuarial Society.

Buchwalder, M., Buhlmann, H., Merz, M. et Wüthrich, M. V. (2006). The mean square error of prediction in the chain ladder reserving method. *ASTIN Bulletin*, *2*(36), 521–542.

Cantoni, E. et Ronchetti, E. (1999). Robust inference for generalized linear models. *Journal of the American Statistical Association*, *96*(455), 1022–1030.

Carrato, A., McGuire, G. et Scarth, R. (1999). A practitioner's introduction to stochastic reserving.

Denuit, M. et Charpentier, A. (2004). *Mathématiques de l'assurance non-vie*. Economica.

England, P. D. et Verrall, R. J. (2002). Stochastic claim reserving in general insurance. *British Actuarial Journal*, *8*(3), 443–518.

Gather, U. et Schultze, V. (1999). Robust estimation of scale of an exponential distribution. *Statistica neerlandica, 53*, 327–341.

Hampel, F. (1971a). A general qualitative definition of robustness. *Annals of mathematical statistics, 42*, 1887–1896.

Hampel, F. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Assosiation, 69*, 383–393.

Hampel, F. R. (1971b). A general qualitative definition of robustness. *Ann. Math. Statist., 42*(6), 1887–1896.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. et Stahel, W. A. (1986a). *Robust Statistics-The Approach Based on Influence Functions*. John Wiley & Sons, Inc.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. et Stahel, W. A. (1986b). *Robust Statistics-The Approach Based on Influence Functions*. John Wiley & Sons, Inc.

Hertig, J. (1985). A statistical approach to ibnr-reserves in marine reinsurance. *ASTIN Bulletin, 15*(2), 171–183.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist., 35*(1), 73–101.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proc. 5th Berkeley Symp*, 221–233.

Huber, P. J. (1981). *Robust Statistics*. John Wiley & sons.

Kaas, R., Goovaerts, M., Dhaene, J. et Denuit, M. (2009). *Modern actuarial risk theory*. Springer.

Larsen, C. R. (2007). An individual claims reserving model. *ASTIN Bulletin*, *37*, 113–132.

Maechler, M. et al. (2016). *Package robustbase.*

Maronna, R. A., Martin, R. D. et Yohai, V. J. (2006). *Robust Statistics Theory and Methods.* John Wiley & Sons, Inc.

McCullagh, P. et Nelder, J. (1989). *Generalized Linear Models. Monographs on statistics and applied probability.* Chapman and Hall.

Merz, M. et Wüthrich, M. (2010). Paid-inccured chain claims reserving method. *Insurance mathematics and economics*, *3*(46), 568–579.

Norberg, R. (1999). Prediction of outstanding liabilities II. Model variations and extensions. *ASTIN Bulletin*, *29*(1), 5–25.

Pigeon, M., Antonio, K. et Denuit, M. (2013). Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin*, *43*(3), 399–428.

Pigeon, M., Antonio, K. et Denuit, M. (2014). Individual loss reserving using paid-incurred data. *Insurance: Mathematics and Economics*, *58*, 121–131.

Renshaw, A. E. et Verrall, R. J. (1998a). A stochastic model underlying the chain ladder technique. *British Actuarial Journal*, *4*(4), 903–923.

Renshaw, A. E. et Verrall, R. J. (1998b). A stochastic model underlying the chain ladder technique. *British Actuarial Journal*, *4*(4), 903–923.

Rousseeuw, P. J. et Leroy, A. M. (1986). *Robust Regression and Outlier Detection.* John Wiley & Sons, Inc.

Staudte, R. G. et Sheather, S. J. (1990). *Robust Estimation and Testing.* John Wiley & Sons, Inc.

Stigler, S. M. (1972). Simon newcomb, percy danielle, and the history of robust estimation 1885-1920. *ASTIN Bulletin.*

Taylor, G. (2000). *Loss Reserving: an Actuarial Perspective.* Springer.

Taylor, G. et Ashe, F. R. (1983). Second moments of estimates of outstanding claims. *Journal of Econometrics*, *23*(1), 37–61.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. in contributions to probability and statistics, eds i. olkin, s. ghurye, w. hoeffding,w. madow and h. mann. *Stanford University Press*, p. 448–485.

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, *33*(1), 1–67.

Verdonck, T. et Debruyne, M. (2011). The influence of individual claims on the chain-ladder estimates: Analysis and diagnostic tool. *Insurance: Mathematics and Economics*, *48*(1), 85–98.

Verdonck, T., Wouwe, M. V. et Dhaene, J. (2009). A robustification of the chain-ladder method. *North American Actuarial Journal*, *13*(2), 280–298.

Wright, T. (1990). A stochastic method for claims reserving in general insurance. *J.Inst. Actuar.*, *117*(1), 677–731.

Wüthrich, M. V. et Merz, M. (2008). *Stochastic claims reserving methods in insurance.* John Wiley & Sons Ltd.

Zhao, X. B. et Zhou, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics*, *46*(2), 290–299.

Zhao, X. B., Zhou, X. et Wang, J. L. (2009). Semiparametric model for prediction of individual claim loss reserving. *Insurance: Mathematics and Economics*, *45*(1), 1–8.